

Génomique de la légumineuse modèle *Medicago truncatula* : état des lieux et perspectives

The model legume *Medicago truncatula*: recent advances and perspectives in genomics

Oléagineux, Corps Gras, Lipides. Volume 8, Numéro 5, 478-84, Septembre - Octobre 2001, La filière

Auteur(s) : Etienne-Pascal JOURNET, Véronique CARREAU, Jérôme GOUZY, Philippe THOQUET, Charles ROSENBERG, David BARKER, Thierry HUGUET, Jean DENARIE, Pascal GAMAS, Laboratoire de biologie moléculaire des relations plantes-micro-organismes (LBMRPM), CNRS-Inra, BP 27, 31326 Castanet-Tolosan Cedex, France.

Résumé : Depuis une vingtaine d'années, les nombreuses recherches focalisées sur la plante modèle *Arabidopsis thaliana* ont permis des progrès considérables dans notre connaissance des bases moléculaires de la biologie des plantes. Avec l'achèvement du séquençage du génome d'*Arabidopsis* [1], la palette complète des outils de génétique et de génomique maintenant disponibles devrait encore accélérer le rythme des découvertes. De son côté, le riz fait l'objet de rapides développements génomiques en tant que modèle pour les monocotylédones. Cependant, il est clair que ces deux espèces ne suffisent pas pour représenter le monde végétal dans toute sa diversité biologique [2]. Ce fait est particulièrement bien illustré par le cas des Légumineuses (Fabacées) qui représentent l'un des taxons végétaux les plus importants, tant du point de vue de la biologie et de l'écologie fondamentales que du point de vue agronomique et environnemental.

Summary : The legume species *Medicago truncatula* is now a model plant used world-wide, and in particular for molecular genetic studies of the endomycorrhizal and nitrogen-fixing root symbioses, associations which do not occur with *Arabidopsis*. Large-scale projects for *M. truncatula* genomics have been initiated within the international community and essential tools are being currently developed for structural genomics (genome mapping, BAC libraries, genome sequencing) and functional genomics (ESTs, microarrays, mutant collections), along with the development of bioinformatics resources. Comparative genomics studies suggest a relatively high level of synteny between legume genomes. It can therefore be anticipated that *M. truncatula* will be used as a nodal legume species to help in the identification of agronomically important genes in crop legumes, e.g. genes for root symbioses, disease/pest resistance, plant architecture, seed quality, and production of specific secondary metabolites.

Keywords : *Medicago truncatula*, root symbioses, crop legumes, plant genomics, bioinformatics.

ARTICLE

La particularité biologique des Légumineuses la plus connue est leur aptitude à s'associer à des bactéries du sol (Rhizobiacées), pour former des organes symbiotiques racinaires (« nodosités » ou « nodules ») au sein desquels ces bactéries transforment l'azote atmosphérique en une forme assimilable par la plante [3]. Les Légumineuses sont également capables, comme la plupart des plantes sauvages ou cultivées, de s'associer à des champignons symbiotiques pour former des endomycorhizes [4, 5]. Ces deux types d'interactions symbiotiques, qui n'existent pas chez *Arabidopsis* ni chez les Crucifères en général, jouent un rôle très important dans la nutrition minérale, le programme de développement et la dynamique des populations des espèces végétales concernées. De plus, les Légumineuses présentent d'autres spécificités biologiques, par exemple concernant leurs interactions avec des pathogènes et des ravageurs (insectes, nématodes), l'architecture de la plante, la formation de graines protéagineuses, la production de métabolites secondaires spécifiques (isoflavonoïdes impliqués dans la signalisation symbiotique, dans les réactions de défense, et présentant aussi un grand intérêt pharmaceutique [6]). Enfin, il n'est pas inutile de rappeler que de nombreuses espèces cultivées appartiennent à la famille des Légumineuses. Elles constituent une source très importante de protéines et de lipides dans l'alimentation humaine et animale : protéagineux tels que le pois, la féverole, le haricot, le pois chiche, les lentilles ; oléoprotéagineux comme le soja et l'arachide ; et fourrages tels que la luzerne et le trèfle. Leur intérêt agronomique provient en premier lieu de leur aptitude à la fixation symbiotique de l'azote, qui leur permet de produire en abondance des protéines végétales même en l'absence de fertilisation azotée, d'où leur intérêt également dans le cadre d'une agriculture « durable » (réduction des intrants, préservation et enrichissement des sols en azote).

Pour ces différentes raisons, la nécessité d'utiliser une espèce modèle pour la génétique moléculaire des Légumineuses a commencé à s'imposer à la communauté scientifique voilà une quinzaine d'années, en particulier dans le domaine de la symbiose fixatrice d'azote. Plus récemment, il est apparu qu'une légumineuse modèle présentait aussi d'autres avantages sur les plans fondamental et appliqué. D'une part, même si les mécanismes de résistance aux maladies sont déjà étudiés de façon extensive chez *Arabidopsis*, il est très intéressant de pouvoir comparer, dans un même végétal, les symbioses avec les interactions pathogènes, dont les processus respectifs présentent des différences mais aussi certains parallèles intrigants [7]. D'autre part, les connaissances acquises sur le génome d'une espèce modèle ne peuvent être efficacement exploitées pour la génétique et l'amélioration des espèces cultivées que si la synténie (ordre d'alignement des marqueurs ou des séquences le long des chromosomes) est suffisamment conservée entre les génomes respectifs. Or la distance phylogénétique entre les familles botaniques représentées parmi les plantes cultivées est trop importante et rend donc nécessaire de disposer dans chaque famille d'une espèce modèle, comme par exemple *Arabidopsis* pour les Crucifères et le riz pour les Céréales.

Chez les Légumineuses, plusieurs espèces cultivées (luzerne, pois, soja) figurent parmi les plantes les mieux caractérisées génétiquement, avec de nombreux marqueurs génétiques classiques, des cartes génétiques bien développées fondées sur des marqueurs moléculaires, et un certain nombre d'outils génomiques. Toutefois, ces espèces ne se prêtent pas à des études poussées de génétique moléculaire en raison de traits défavorables tels que ploïdie complexe, allogamie, génomes de grande taille ou difficultés de transformation. C'est ainsi que deux légumineuses peu connues à l'origine, *Medicago truncatula* et *Lotus japonicus*, ont été proposées comme plantes modèles dans

les quinze dernières années et sont maintenant utilisées par de nombreux laboratoires dans le monde [8-14]. Ces deux espèces présentent des caractéristiques génétiques similaires mais développent des nodules de type différent (croissance indéterminée/déterminée). Par ailleurs, *M. truncatula* est plus proche sur le plan phylogénétique de la plupart des Légumineuses cultivées en Europe : elle fait partie du groupe des Galégoïdes qui contient les tribus des Trifoliées (luzernes, trèfles), Viciées (pois, féveroles, lentilles, vesces) et Cicérées (pois chiches).

Propriétés du modèle *Medicago truncatula*

Historiquement, le choix de *M. truncatula* résulte d'un programme Inra (1985-1986) dont l'objectif était de définir une espèce modèle à l'intérieur du genre *Medicago*, afin de l'associer à *Sinorhizobium meliloti*, le microsymbiote bien étudié de la luzerne, et constituer ainsi un système symbiotique modèle plante-bactérie [8]. *M. truncatula* est une légumineuse annuelle originaire du pourtour méditerranéen et proche de la luzerne cultivée *M. sativa*. Elle est diploïde ($n = 8$) et autogame, produit des graines en abondance avec un temps de génération de 10 à 12 semaines (de graine à graine). La taille de son génome est d'environ 500 Mb/1C [15], c'est-à-dire trois à quatre fois supérieure à celle d'*A. thaliana* et équivalente à celle du riz. La biodiversité de l'espèce *M. truncatula* est caractérisée par une forte variabilité morphologique et génétique intra- et interpopulations et par une importante homozygotie au niveau individuel [16]. *M. truncatula* est nodulée par *S. meliloti*, espèce la mieux caractérisée parmi les rhizobia et dont le génome a été récemment séquencé (<http://sequence.toulouse.inra.fr/meliloti.html> [17]). En particulier, le nombre très élevé de mutants symbiotiques disponibles chez *S. meliloti* représente un excellent atout pour disséquer les mécanismes de la symbiose fixatrice d'azote.

Les deux principaux génotypes de *M. truncatula* actuellement utilisés en laboratoire, l'un (A17) issu du cultivar Jemalong et l'autre (R-108-1) issu de l'écotype 108-1, ont été initialement sélectionnés pour leurs capacités symbiotiques et de régénération *in vitro*. Leur manipulation au laboratoire est maintenant bien maîtrisée : cycle vital et amplification des lignées, croisements, cultures dans différentes conditions, établissement des symbioses racinaires, etc. Leur transformation par *Agrobacterium tumefaciens* et leur régénération *in vitro* ont été progressivement optimisées [18-22]. Une méthode rapide et efficace de transformation par *A. rhizogenes* a été récemment établie [23], permettant d'étudier et d'utiliser divers transgènes dans les racines transgéniques ainsi obtenues (principe du *hairy root*), en particulier dans le contexte des deux symbioses.

M. truncatula a déjà fait l'objet de nombreux travaux de biologie moléculaire, qui ont permis d'identifier et d'étudier plusieurs dizaines de gènes impliqués dans l'interaction symbiotique avec *Sinorhizobium* (gènes de nodulines [24, 25]) ou dans la symbiose endomycorhizienne [26]. Nous présentons ci-dessous les faits principaux concernant le développement des approches et des outils et les avancées des connaissances dans le domaine de la génétique et de la génomique de *M. truncatula*.

Coordination internationale des initiatives de génomique sur *M. truncatula*

L'expérience du modèle *Arabidopsis* a démontré l'efficacité de la concentration et de l'organisation des efforts de recherche au niveau international sur un système végétal modèle. Un réseau similaire s'est organisé en quelques années autour de *M. truncatula* et a bénéficié du contexte scientifique favorable au développement de la génomique végétale pour atteindre aujourd'hui une dimension

conséquence. En particulier, plusieurs grands programmes de génomique de *M. truncatula* multinationaux et impliquant plusieurs instituts ont vu le jour. Les deux principaux programmes en cours sont soutenus, l'un par le Plant Genome Program de la NSF - National Science Foundation - (3,4 M \$; 1999-2002 ; <http://chrysie.tamu.edu/medicago>), et l'autre par le 5^e PCRD (Programme Cadre de Recherche et Développement) de l'Union européenne (2,2 M euros ; 2000-2002 ; <http://medicago.toulouse.inra.fr/EU>). L'institution privée Noble Foundation (Ardmore, Oklahoma) finance également un important programme de génomique fonctionnelle, protéomique, métabolomique et bioinformatique (Center for *Medicago* Genomics Research ; www.noble.org/medicago/index.htm). En France, un premier programme « Génome de *Medicago truncatula* » a été financé conjointement par le CNRS et l'Inra et avec le soutien du Génoscope (Évry), sur la période 1998-2000, pour fédérer les travaux de huit laboratoires cherchant à identifier et étudier la part du génome végétal impliquée dans les symbioses racinaires. Une « Action transversale structurante » Inra (2000-2002) reprend cet effort et l'étend à d'autres aspects importants pour les Légumineuses cultivées. Enfin, les interactions entre les laboratoires qui travaillent avec *M. truncatula* sont ouvertes et très actives et des colloques dévolus à cette plante modèle ont lieu au moins une fois par an (un 4^e « Workshop on *M. truncatula* » s'est tenu en juillet 2001, en marge du « 10th International Congress on Molecular Plant-Microbe Interactions » ; www.plantpath.wisc.edu/mpmi/satellite.html). Un cours pratique EMBO (novembre 2001 ; Institut des sciences du végétal du CNRS, Gif-sur-Yvette ; www.isv.cnrs-gif.fr) permettra de former de nouveaux chercheurs à ce système modèle.

EST, clustering et analyses *in silico*

Le résultat le plus visible du développement rapide du modèle *M. truncatula* est la quantité d'EST (*expressed sequence tags*) déposées en moins de deux ans dans les bases de données publiques. Les deux premiers lots d'EST ont été produits à partir d'une banque d'ADNc issus de pointes racinaires enrichies en poils absorbants (environ 900 EST [27] ; <http://bio-SRL8.stanford.edu/>) et d'une banque issue de jeunes nodules fixateurs d'azote (environ 500 EST [28]). Depuis, les programmes de séquençage systématique du consortium NSF, de la Noble Foundation et du Génoscope ont permis d'atteindre, en septembre 2001, le nombre de 145 000 EST déposées, correspondant principalement à des extrémités 5'. Elles sont issues de plus de 30 banques d'ADNc qui représentent différents organes, stades de développement et traitements par des stimuli biotiques (symbiotes, pathogènes, ravageurs, éliciteurs) ou abiotiques (c'est-à-dire carences minérales, stress hydrique, etc.). Le séquençage systématique se poursuit aux États-Unis avec un objectif de 200 000 EST pour le début 2002.

Le traitement et l'exploitation de cette masse énorme de données de séquences partielles nécessitent l'utilisation d'outils bio-informatiques performants. Il s'agit d'identifier les gènes représentés et les polypeptides correspondants, de les annoter par une prédiction de leur fonction, et d'obtenir des informations sur leurs profils d'expression. Une base de données regroupant toutes les EST de *M. truncatula* en *clusters* et régulièrement mise à jour a été construite au TIGR (The Institute for Genomic Research, Rockville, Maryland) selon une procédure déjà utilisée pour plus d'une dizaine d'organismes animaux ou végétaux [29] (www.tigr.org/tdb/tgi.shtml).

Ce « *Medicago truncatula* Gene Index » fournit des séquences consensus provisoires assorties d'une annotation automatique, donnant ainsi un premier aperçu des différents gènes de *M. truncatula* représentés dans les bases de données. À partir de 141 000 EST, environ 29 000 séquences distinctes ont été obtenues par le TIGR (14 700 séquences consensus et 14 500 singletons), chiffre à comparer aux 42 000 séquences distinctes de l'index d'*Arabidopsis* (septembre 2001).

Au LBMRPM de Toulouse (CNRS-Inra), une stratégie sensiblement différente a été adoptée pour générer et analyser des EST de racines de *M. truncatula* (collab. V. Gianinazzi-Pearson et D. van Tuinen, Inra-Univ. Bourgogne, Dijon). Le Génoscope (Évry) a réalisé en 1999 le séquençage d'EST en 5' et 3' à partir de 3 banques d'ADNc (racines témoins, jeunes nodules, racines mycorhizées ; 5 000 clones chacune). Environ 24 000 EST exploitables ont été obtenues, correspondant à environ 14 000 clones (<http://sequence.toulouse.inra.fr/Mtruncatula.html>). La connaissance de la séquence 3' pour la plupart de ces clones, en sus de la séquence 5', a permis de créer un noyau de *clusters* très fiable sur lequel peuvent venir se greffer les autres EST. Selon cette approche, les 14 000 clones analysés représentent environ 6 400 gènes distincts. La comparaison avec le « Gene Index du TIGR » a révélé des différences dans le groupement de mêmes EST et souligne la nécessité de perfectionner les procédures de *clustering*. En outre, comme les procédés d'annotation automatique introduisent ou transfèrent des erreurs qui polluent les bases de données, nous avons fait le choix d'une annotation semi-automatique de nos EST. Une interface d'annotation (iANT), permettant une analyse en profondeur des homologies de séquences par l'annotateur grâce à divers outils bio-informatiques disponibles, a été développée et utilisée pour associer aux *clusters* une information validée « manuellement », plus sûre et d'une plus grande valeur sur le plan biologique. L'ensemble de ces données devrait être disponible sur un site Web fin 2001.

Une fois un *cluster* défini de manière fiable, la distribution des clones ADNc qui le composent en fonction de leurs différentes banques d'origine permet d'esquisser le profil d'expression du gène correspondant. Par cette approche *in silico*, nous avons pu identifier de nouveaux gènes candidats à expression différentielle, spécifique d'un tissu ou d'une condition particulière, par exemple dans les jeunes nodules. Néanmoins, ces résultats ne sont statistiquement valables que pour les gènes fortement exprimés et doivent être confirmés directement par des quantifications au niveau moléculaire (voir ci-dessous).

« Unigene Set », *microarrays* et analyse du transcriptome

L'une des principales applications du séquençage massif d'EST est le développement des microréseaux d'ADN (*microarrays*) pour l'analyse du transcriptome. En effet, suite au *clustering* des EST, on connaît l'ensemble des gènes distincts représentés et on peut choisir, pour représenter physiquement chaque gène, le meilleur clone ADNc. Le consortium NSF travaille à la construction d'une collection non redondante et universelle d'ADNc (*universal unigene set*), comportant au départ 6 000 gènes et qui représentera à terme tous les gènes transcrits de *M. truncatula* (estimés entre 20 000 et 30 000) et sera accessible à la communauté sous la forme de clones individuels et de *microarrays*.

Le réseau européen *M. truncatula* travaille sur une approche similaire, en concertation et en réciprocité avec les équipes américaines, pour générer des *microarrays* et, également, des *macroarrays* pour des objectifs plus ciblés et d'utilisation plus accessible. Ainsi, grâce aux infrastructures de génomique récemment mises en place en France et en Europe, les études du transcriptome de *M. truncatula*, actuellement en phase pilote, vont pouvoir pleinement se développer dans les laboratoires de ce réseau européen.

Collections de variants naturels et de mutants

La forte biodiversité qui existe chez *M. truncatula* a été rendue accessible à l'expérimentation grâce à une collection de plusieurs centaines d'écotypes constituée à l'Inra de Montpellier [30]. Le criblage de ces ressources génétiques a déjà mis en évidence des variants naturels dans la reconnaissance symbiotique plante-*Sinorhizobium* (équipes de J. Dénarié et T. Huguet, CNRS-Inra Toulouse [31]) ainsi que des écotypes présentant divers degrés de résistance ou de sensibilité à des champignons pathogènes (*Phytophthora medicaginis*, *Colletotrichum trifolii*) et nématodes (*Meloidogyne incognita*) [32]. La recherche de sources de résistance à des insectes ravageurs (pucerons) est en cours en France et en Australie. Après une caractérisation fine des pathosystèmes chez *M. truncatula*, le clonage positionnel de gènes de résistance à ces divers bio-agresseurs devrait ouvrir de nouvelles perspectives pour l'amélioration des Légumineuses cultivées.

Plusieurs programmes de mutagenèse par des agents physico-chimiques classiques ont été réalisés : par rayons gamma [33], par l'EMS [34] et par neutrons rapides (équipe de S. Long, Stanford). Ils ont fourni le matériel nécessaire pour des criblages à grande échelle sur divers phénotypes et, en premier lieu, pour des mutants affectés dans les interactions symbiotiques fixatrices d'azote et/ou mycorhizienne. On peut s'attendre à trouver ainsi des gènes impliqués dans divers processus biologiques, par exemple la perception/ transduction de signaux biotiques, le développement (morphogénèse et organogénèse), la régulation de la nutrition minérale, etc. En outre, il devrait être possible d'identifier par cette approche la plupart des gènes essentiels et spécifiques à ces deux symbioses, car celles-ci ne sont pas indispensables à la vie de la plante lorsqu'elle est cultivée sur milieu minéral complet (*a priori*, pas de mutants létaux). De nombreux mutants de nodulation ont déjà été identifiés chez *M. truncatula* et leur caractérisation phénotypique et génétique a permis de commencer à disséquer les mécanismes de reconnaissance entre bactérie et plante, d'infection et de formation des nodules [35-40]. Comme chez d'autres Légumineuses, plusieurs gènes de *M. truncatula* contrôlant les étapes précoces de la nodulation possèdent une fonction symbiotique plus générale, car ils sont également nécessaires à l'établissement de la mycorhization (mutants Nod⁻/Myc⁻) [33, 38]. Le clonage positionnel de plusieurs gènes de nodulation est en cours, à l'aide des cartes du génome et d'une banque génomique de BAC (voir plus loin). Enfin, le criblage des populations mutagénisées sur d'autres phénotypes a permis d'isoler les premiers mutants de développement [34, 41].

Projets de mutagenèse insertionnelle

Une étape cruciale pour le développement de la génomique fonctionnelle de *M. truncatula* est de réaliser une mutagenèse à haute densité, voire à saturation, du génome par insertion d'étiquettes (ADN-T ou transposons). Cette stratégie a déjà fait preuve de son efficacité chez *A. thaliana* pour la génétique directe ou inverse [42]. En particulier, la possibilité de transformer efficacement *Arabidopsis in planta* a permis de construire plusieurs collections de plus de 50 000 lignées

d'insertion d'ADN-T [43-45]. Concernant *M. truncatula*, le nombre de lignées d'insertion estimé nécessaire pour atteindre la saturation du génome est de 300 000 à terme. Une équipe de la Noble Foundation a récemment publié un protocole d'infiltration sous vide qui permettrait d'obtenir des fréquences de transformation supérieures à celles observées chez *Arabidopsis* [46]. Toutefois, aucun des laboratoires ayant cherché à reproduire cette méthode n'a encore réussi (en septembre 2001) à obtenir de transformant. La reproductibilité de cette méthode reste donc à montrer, et des recherches supplémentaires sont nécessaires avant de pouvoir construire des collections de mutants ADN-T à grande échelle. En attendant, la transgénèse classique par embryogenèse somatique *in vitro* reste d'actualité et, constitue l'approche utilisée par P. Ratet (Institut des sciences du végétal [ISV], Gif-sur-Yvette) et M. Schultze (Université de York, Royaume-Uni) pour générer environ 2 500 lignées avant fin 2002. Par ailleurs, des systèmes de mutagenèse par transposons sont en cours d'évaluation sur *M. truncatula* à l'ISV (Gif-sur-Yvette, P. Ratet) et chez Plant Research International (Wageningen, Pays-Bas, A. Pereira).

Cartes génétique, physique et cytogénétique

Deux cartes génétiques relativement denses ont été construites. La première a été réalisée par l'équipe de T. Huguët (CNRS-Inra, Toulouse) à partir d'un croisement entre les lignées Jemalong J6 et DZA 315-16. Les 290 marqueurs (surtout de type AFLP et RAPD) qui ont pu être placés sur cette carte se répartissent sur huit groupes de liaison ($2n = 16$) pour un total de 1 225 cM, ce qui correspond à une valeur moyenne de 470 kb/cM. La densité de répartition des marqueurs est très homogène et les distorsions de ségrégation sont modérées (concernant trois groupes de liaison ; taux global de 27 %). D'autres marqueurs y sont ajoutés, en particulier 250 de type microsatellite qui ont été identifiés *in silico* à partir de la collection d'EST, ce qui permet au passage de localiser les gènes correspondants sur la carte. De plus, une collection de 195 lignées recombinantes (RIL), toutes génotypées et la plupart en génération F7, a été réalisée en collaboration avec J.M. Prosperi de l'Inra de Montpellier. Ainsi, la disponibilité pour la communauté scientifique d'une collection de marqueurs microsatellites et de lignées recombinantes devrait grandement faciliter les travaux de cartographie dans des conditions standard et reproductibles. La seconde carte génétique a été réalisée par l'équipe de D. Cook (Univ. Californie, Davis) à partir d'un croisement entre la lignée Jemalong A17 et l'écotype A20. Deux cent trente marqueurs PCR codominants ont été cartographiés. Des liens entre ces deux cartes vont être établis par le biais de marqueurs microsatellites montrant du polymorphisme entre les trois écotypes parentaux.

L'équipe de D. Cook s'est également investie dans la construction de deux banques génomiques dans un vecteur BAC [47], qui atteignent maintenant l'équivalent de 6 et 23 fois le génome. Ces banques sont utilisées actuellement pour réaliser une carte physique complète par *fingerprinting* des BAC et par séquençage des extrémités de BAC, carte qui devrait être terminée à la fin de l'année 2002 (D. Cook ; N. Young, Univ. Minnesota, Saint-Paul). Cette carte devrait constituer un outil de première importance pour le clonage positionnel de gènes, pour la génomique comparative des Légumineuses et pour le séquençage massif du génome de *M. truncatula*.

Enfin, il était important de développer chez *M. truncatula* des outils de cytogénétique moléculaire. En effet, l'élaboration d'une carte cytogénétique fournira des informations précieuses sur l'organisation du génome, la répartition des régions d'euchromatine et d'hétérochromatine, et la position des régions centromériques. L'intérêt d'une cartographie par hybridation d'ADN (FISH) est

qu'elle n'est pas fondée sur la recombinaison et ne nécessite pas l'existence d'un polymorphisme dans la région considérée : cette méthode devrait donc permettre de cartographier des gènes dans des régions présentant des fréquences de recombinaison faibles ou peu polymorphes. Un caryotype a été réalisé récemment pour deux génotypes en utilisant la technique de FISH sur des plaques métaphasiques [48]. D'autres travaux conduits par l'équipe de Ton Bisseling (Univ. Wageningen, Pays-Bas) ont montré la faisabilité chez *M. truncatula* d'une carte cytogénétique moléculaire fine en positionnant et en ordonnant des BAC sur des chromosomes au stade pachytène, qui offrent une bien meilleure résolution (environ 60 kb) qu'au stade métaphasique [49] (pour revue, voir [50]). L'intégration des cartes génétique, physique et cytogénétique a ainsi commencé, à l'aide de divers marqueurs moléculaires et des BAC. Par exemple, 150 marqueurs de la seconde carte génétique ont été localisés sur des clones BAC, ce qui a permis, en retour, de placer sur cette carte plus de 100 EST. L'existence d'une telle carte intégrée chez *M. truncatula* va faciliter et accélérer la localisation précise et le clonage positionnel de gènes d'intérêt.

***M. truncatula*, espèce de référence pour la génomique comparative des Légumineuses**

Plusieurs travaux de génomique comparative en cours semblent montrer que le degré de conservation de la synténie entre *M. truncatula* et *Arabidopsis* est faible. L'étude des distributions respectives de marqueurs et la comparaison des séquences de clones BAC de *M. truncatula* avec celle du génome d'*Arabidopsis* donnent une estimation moyenne de la conservation microsyténique entre les deux génomes d'environ 10 %. En outre, l'étude de la phylogénie de gènes de résistance chez plusieurs légumineuses et chez *Arabidopsis* a permis d'identifier des gènes de résistance présents chez les Légumineuses qui sont mal représentés, voire absents, chez *Arabidopsis* (équipes de D. Cook et de N. Young, États-Unis). Ce dernier résultat suggère qu'*Arabidopsis* n'est sans doute pas un bon modèle pour étudier les gènes de résistance aux maladies des Légumineuses cultivées.

Cette constatation souligne l'intérêt que présente l'étude de la conservation de la synténie entre *M. truncatula* et les principales légumineuses cultivées (luzerne, pois et soja). Les premiers résultats montrent un degré élevé de conservation macrosyténique avec la luzerne *M. sativa* (à partir de 80 marqueurs environ) et de microsyténie avec le soja (40 % en moyenne sur 20 Mb de BAC du soja). De plus, une comparaison détaillée des génomes du pois et de *Medicago* est en cours dans le cadre d'un programme européen (collab. N. Ellis, John Innes Institute, Angleterre, G. Kiss, Szeged, Hongrie, T. Huguet, Toulouse). Le pois et *Medicago* sont très proches d'un point de vue phylogénétique. Or, le pois possède 2 x 7 chromosomes alors que *M. truncatula* en possède 2 x 8. Il s'avère que, pour cinq chromosomes, la répartition des marqueurs est essentiellement colinéaire, tandis que des marqueurs situés sur les deux autres chromosomes du pois se retrouvent répartis sur trois chromosomes de *Medicago*. Ainsi, le développement de tels marqueurs-ancres devrait permettre d'établir des liens entre les cartes génomiques de plusieurs légumineuses (*M. truncatula*, luzerne, pois, soja, *Lotus japonicus*). L'objectif à terme est de construire une carte composite des génomes de Légumineuses, qui devrait permettre de passer aisément d'un génome à l'autre et faciliter ainsi le clonage des gènes d'intérêt agronomique chez les Légumineuses cultivées (tétraploïdes et/ou avec de trop grands génomes) en clonant d'abord leurs orthologues chez *M. truncatula* [13].

Identification de régions riches en gènes et projets de séquençage du génome

Il est intéressant de constater que des espèces de Légumineuses proches du point de vue évolutif possèdent des génomes de tailles très différentes. Ainsi, le génome du pois est environ 7 fois plus grand que celui de *M. truncatula*. Or, il apparaît que cette différence de taille est due non à la duplication de régions génomiques, mais plutôt à l'accumulation récente de rétrotransposons et de régions d'ADN non codantes (travaux conduits par N. Ellis, John Innes Institute, Angleterre). Ces observations ont soulevé la question de l'existence de régions riches en gènes, une hypothèse nourrie par des résultats de cytogénétique qui montrent une organisation particulièrement simple du génome de *M. truncatula*. En effet, il semble qu'environ 80 % du génome est organisé en blocs compacts d'hétérochromatine péri-centromérique, tandis que les 20 % restants sont situés sur les bras des chromosomes, essentiellement sous forme d'euchromatine supposée contenir la plupart des gènes transcrits et peu de séquences répétées. Ainsi, les régions riches en gènes pourraient être concentrées sur environ 100 Mb seulement chez *M. truncatula* [49].

Des contacts ont été pris aux États-Unis et en Europe pour discuter d'un projet de séquençage massif du génome de *M. truncatula*. Un tel projet est considéré à l'heure actuelle par plusieurs institutions (USDA-ARS - US Department of Agriculture, Agricultural Research Service -, Inra, John Innes Institute) comme le projet de génomique qui pourrait faire le mieux progresser les connaissances sur l'organisation du génome des Légumineuses en général. Deux stratégies sont envisagées : un séquençage restreint aux régions potentiellement riches en gènes ou bien un séquençage *shotgun* complet du génome. Dans un premier temps, l'USDA a décidé de financer la construction de la carte physique complète de *M. truncatula* (voir plus haut). Les *contigs* de BAC correspondant aux régions riches en gènes pourront alors être sélectionnés à l'aide des marqueurs EST déjà placés sur les cartes génétique ou physique. De son côté, la Noble Foundation (Ardmore, Oklahoma) a pris l'initiative de financer le séquençage en masse de l'équivalent d'une fois le génome, en un an (www.genome.ou.edu/). La réalisation de ces projets complémentaires devrait ouvrir la voie au séquençage massif du génome de cette légumineuse-modèle à partir de 2002.

Ressources bio-informatiques pour *M. truncatula*

L'un des défis liés aux approches globales de la génomique concerne le stockage, la restitution intelligible, le partage, l'harmonisation et l'exploitation de quantités énormes de données d'origines multiples. Toutes ces opérations ne peuvent être réalisées qu'avec de solides outils bio-informatiques, qui restent à adapter ou à créer pour *M. truncatula*. Les bio-informaticiens d'un certain nombre de laboratoires étudiant cette espèce développent actuellement des bases de données librement accessibles et interrogeables *via* Internet. Les sites existants ont été présentés plus haut, concernant les EST (bases de données MtDB du programme NSF et MtB du programme français « *M. truncatula* Gene Index » du TIGR) et les séquences génomiques (BAC, séquençage *shotgun*). En outre, le National Center for Genome Resources (NCGR, Santa Fe, Nouveau-Mexique) et la Noble Foundation ont créé une base de données intégrative (« the *Medicago* Genome Initiative » [51] ; www.ncgr.org/research/mgi/) qui a pour ambition de réunir et relier entre elles les données concernant les séquences d'EST, les profils d'expression des gènes, les différentes collections de mutants d'insertion, le protéome et le métabolome.

Enfin, un site fédérateur a été ouvert à l'Université du Minnesota (www.medicago.org), comportant des liens avec les différents programmes en cours sur *M. truncatula* et donnant accès à divers projets propres à cette université (volet bio-informatique du programme NSF, génomique comparative des Légumineuses) ainsi qu'à un forum de discussion.

Note :

¹ Une première version de cette étude a été rédigée pour la 2^e école thématique de biologie végétale Inra-CNRS (« Génomique fonctionnelle chez les végétaux : du gène à la fonction » ; mars 2001).

CONCLUSION

Perspectives générales

Les divers programmes de génomique sur *M. truncatula* ont déjà permis de réaliser des avancées importantes sur des axes essentiels, en particulier concernant les EST, la mutagenèse, l'intégration des cartes génomiques et une banque génomique en vecteur BAC. D'autres développements plus récents devraient également fournir à terme de nouveaux outils permettant d'analyser et d'utiliser plus facilement le génome de cette espèce modèle : les projets de séquençage génomique, une mutagenèse sur le principe du *gene tilling* [52] comme outil pour la génétique inverse, le développement d'une carte composite des Légumineuses. Ces différents outils vont être rendus accessibles par la création de centres de ressources génétiques et génomiques en Europe et aux États-Unis. L'analyse globale de l'expression du génome (transcriptome) en fonction des contextes tissulaires, physiologiques ou génétiques, et de ses conséquences aux niveaux protéique (protéome) et métabolique (métabolome) a commencé. Enfin, la montée en puissance des ressources bio-informatiques dédiées à *M. truncatula* devrait fournir les moyens indispensables pour exploiter et valoriser les masses de données produites. Ainsi, le développement génomique du modèle *M. truncatula* crée une nouvelle opportunité, après *Arabidopsis* et le riz, pour des recherches fondamentales non seulement sur des aspects originaux comme les interactions symbiotiques, mais aussi sur des questions plus générales de biologie et physiologie végétales, et pour l'utilisation privilégiée de *M. truncatula* comme espèce de référence pour les Légumineuses cultivées.

REFERENCES

1. The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408 : 796-815.
2. ADAM D (2000). *Arabidopsis thaliana* genome. Now for the hard ones. *Nature*, 408 : 792-3.
3. MYLONA P, PAWLOWSKI K, BISSELING T (1995). Symbiotic nitrogen fixation. *Plant Cell*, 7 : 869-85.
4. GIANINAZZI-PEARSON V (1996). Plant cell responses to arbuscular mycorrhizal fungi: getting to the roots of the symbiosis. *Plant Cell*, 8 : 1871-83.
5. HARRISON MJ (1997). The arbuscular mycorrhizal symbiosis: an underground association. *Trends Plant Sci*, 2 : 54-6.

6. DIXON RA (1999). Isoflavonoids: biochemistry, molecular biology, and biological functions. In : *Comprehensive natural products chemistry*. Sankawa U, ed. Oxford : Elsevier, vol. 1 : 773-823.
7. PARNISKE M (2000). Intracellular accommodation of microbes by plants: a common developmental program for symbiosis and disease? *Curr Opin Plant Biol*, 3 : 320-8.
8. BARKER DG, BIANCHI S, LONDON F, *et al.* (1990). *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. *Plant Mol Biol Rep*, 8 : 40-9.
9. COOK D, VANDENBOSCH K, DE BRUIJN F, HUGUET T (1997). Model legumes get the Nod. *Plant Cell*, 9 : 275-81.
10. JIANG Q, GRESSHOFF PM (1997). Classical and molecular genetics of the model legume *Lotus japonicus*. *Mol Plant Microbe Interact*, 10 : 59-68.
11. COOK DR (1999). *Medicago truncatula*, a model in the making! *Curr Opin Plant Biol*, 2 : 301-4.
12. HARRISON MJ (2000). Molecular genetics of model legumes. *Trends Plant Sci*, 5 : 414-5.
13. FRUGOLI J, HARRIS J (2001). *Medicago truncatula* on the move! *Plant Cell*, 13 : 458-63.
14. UDVARDI MK (2001). Legume models strut their stuff. *Mol Plant Microbe Interact*, 14 : 6-9.
15. BLONDON F, MARIE D, BROWN S, KONDOROSI A (1994). Genome size and base composition in *Medicago sativa* and *M. truncatula* species. *Genome*, 37 : 264-70.
16. BONNIN I, HUGUET T, GHÉRARDI M, PROSPÉRI JM, OLIVIÉRI I (1996). High level of polymorphism and spatial structure in a selfing plant species *Medicago truncatula* Gaertn. using RAPDs markers. *Am J Bot*, 83 : 843-55.
17. GALIBERT F, *et al.* (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, 293 : 668-72.
18. THOMAS MR, ROSE RJ, NOLAN KE (1992). Genetic transformation of *Medicago truncatula* using *Agrobacterium* with genetically modified Ri and disarmed Ti plasmids. *Plant Cell Rep*, 11 : 113-7.
19. CHABAUD M, LARSONNAUD C, MARMOUGET C, HUGUET T (1996). Transformation of barrel medic (*Medicago truncatula* Gaertn.) by *Agrobacterium tumefaciens* and regeneration via somatic embryogenesis of transgenic plants with the *MtENOD12* nodulin promoter fused to the *gus* reporter gene. *Plant Cell Rep*, 15 : 305-10.
20. TRIEU AT, HARRISON MJ (1996). Rapid transformation of *Medicago truncatula*: regeneration via shoot organogenesis. *Plant Cell Rep*, 16 : 6-11.
21. HOFFMANN B, TRINH TH, LEUNG J, KONDOROSI A, KONDOROSI E (1997). A new *Medicago truncatula* line with superior *in vitro* regeneration, transformation and symbiotic properties isolated through cell culture selection. *Mol Plant Microbe Interact*, 10 : 307-15.

22. TRINH TH, RATET P, KONDOROSI E, DURAND P, KAMATE K, BAUER P, KONDOROSI A (1998). Rapid and efficient transformation of diploid *Medicago truncatula* and *Medicago sativa* ssp. *falcata* germ lines improved for somatic embryogenesis. *Plant Cell Rep*, 17 : 345-55.
23. BOISSON-DERNIER A, CHABAUD M, GARCIA F, BECARD G, ROSENBERG C, BARKER DG (2001). *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol Plant Microbe Interact*, 14 : 693-700.
24. PAWLOWSKI K (1997). Nodule-specific gene expression. *Physiol Plantarum*, 99 : 617-31.
25. SCHULTZE M, KONDOROSI A (1998). Regulation of symbiotic root nodule development. *Annu Rev Genet*, 32 : 33-57.
26. HARRISON MJ (1999). Molecular and cellular aspects of the arbuscular mycorrhizal symbiosis. *Annu Rev Plant Physiol Plant Mol Biol*, 50 : 361-89.
27. COVITZ PA, SMITH LS, LONG SR (1998). Expressed sequence tags from a root-hair-enriched *Medicago truncatula* cDNA library. *Plant Physiol*, 117 : 1325-32.
28. GYORGYEY J, VAUBERT D, JIMENEZ-ZURDO JI, CHARON C, TROUSSARD L, KONDOROSI A, KONDOROSI E (2000). Analysis of *Medicago truncatula* nodule Expressed Sequence Tags. *Mol Plant Microbe Interact*, 13 : 62-71.
29. QUACKENBUSH J, LIANG F, HOLT I, PERTEA G, UPTON J (2000). The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucl Acids Res*, 28 : 141-5.
30. PROSPERI JM (2000). The genus *Medicago* and the genetic resources of *Medicago truncatula*. *Grain Legumes*, 28 : 14-5.
31. TIRICHINE L, DE BILLY F, HUGUET T (2000). *Mtsym6*, a gene conditioning *Sinorhizobium* strain-specific nitrogen fixation in *Medicago truncatula*. *Plant Physiol*, 123 : 845-51.
32. COOK DR, KIM D, ZHU H, URIBE P (2000). Plant-pathogen interactions in *Medicago truncatula*. *Grain Legumes*, 28 : 20.
33. SAGAN M, MORANDI D, TARENGHI E, DUC G (1995). Selection of nodulation and mycorrhizal mutants in the model plant *Medicago truncatula* (Gaertn.) after gamma-ray mutagenesis. *Plant Sci*, 111 : 63-71.
34. PENMETSVA RV, COOK DR (2000). Production and characterization of diverse developmental mutants of *Medicago truncatula*. *Plant Physiol*, 123 : 1387-97.
35. BÉNABEN V, DUC G, LEFEBVRE V, HUGUET T (1995). TE7: an inefficient symbiotic mutant of *Medicago truncatula* Gaertn cv. Jemalong. *Plant Physiol*, 107 : 53-62.

36. SAGAN M, DE LARAMBERGUE H, MORANDI D (1998). Genetic analysis of symbiosis mutants in *Medicago truncatula*. In : *Biological nitrogen fixation for the 21st Century*. Elmerich C, Kondorosi A, Newton WE, eds. Dordrecht : Kluwer Academic Publishers : 317-8.
37. PENMETSА RV, COOK DR (1997). A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. *Science*, 275 : 527-30.
38. CATOIRA R, GALERA C, DE BILLY F, *et al.* (2000). Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *Plant Cell*, 12 : 1647-66.
39. CATOIRA R, TIMMERS AC, MAILLET F, *et al.* (2001). The *HCL* gene of *Medicago truncatula* controls *Rhizobium*-induced root hair curling. *Development*, 128 : 1507-18.
40. WAIS RJ, GALERA C, OLDROYD G, *et al.* (2000). Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula*. *Proc Natl Acad Sci USA*, 97 : 13407-12.
41. NAKATA PA, MCCONN MM (2000). Isolation of *Medicago truncatula* mutants defective in calcium oxalate crystal formation. *Plant Physiol*, 124 : 1097-104.
42. PARINOV S, SUNDARESAN V (2000). Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. *Curr Opin Biotechnol*, 11 : 157-61.
43. BECHTOLD N, PELLETIER G (1998). In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol Biol*, 82 : 259-66.
44. KRYSAN PJ, YOUNG JC, SUSSMAN MR (1999). T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell*, 11 : 2283-90.
45. BENT AF (2000). *Arabidopsis* in planta transformation. Uses, mechanisms, and prospects for transformation of other species. *Plant Physiol*, 124 : 1540-7.
46. TRIEU AT, BURLEIGH SH, KARDAILSKY IV, *et al.* (2000). Transformation of *Medicago truncatula* via infiltration of seedlings or flowering plants with *Agrobacterium*. *Plant J*, 22 : 531-41.
47. NAM YW, PENMETSА RV, ENDRE G, URIBE P, KIM DJ, COOK DR (1999). Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. *Theor Appl Genet*, 98 : 638-46.
48. CERBAH M, KEVEI Z, SILJAK-YAKOVLEV S, KONDOROSI E, KONDOROSI A, TRINH TH (1999). FISH chromosome mapping allowing karyotype analysis in *Medicago truncatula* lines Jemalong J5 and R-108-1. *Mol Plant Microbe Interact*, 12 : 947-50.
49. KULIKOVA O, GUALTIERI G, GEURTS R, *et al.* (2001). Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *Plant J*, 27 : 49-58.
50. DE JONG J, FRAN SZ P, ZABEL P (1999). High resolution FISH in plants - techniques and applications. *Trends Plant Sci*, 4 : 258-63.

51. BELL CJ, DIXON RA, FARMER AD, *et al.* (2001). The *Medicago* Genome Initiative: a model legume database. *Nucleic Acids Res*, 29 : 114-7.

52. MCCALLUM CM, COMAI L, GREENE EA, HENIKOFF S (2000). Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. *Plant Physiol*, 123 : 439-42.

Illustrations



Photo 1. *Fleur de M. truncatula cv Jemalong en cours d'ouverture. À ce stade, l'autopollinisation a déjà eu lieu.*

Crédit photos : Journet et al., CNRS-Inra



Photo 2. *Feuille de M. truncatula cv Jemalong montrant une pigmentation caractéristique chez ce cultivar. Ce nœud porte aussi deux gousses en développement (environ 10 jours après pollinisation).*

Crédit photos : Journet et al., CNRS-Inra

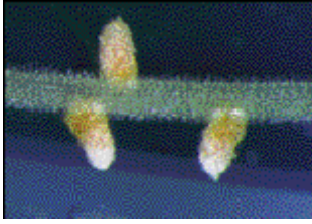


Photo 3. *Nodules fixateurs d'azote formés sur une racine de M. truncatula après inoculation du microsymbiote Sinorhizobium meliloti.*

Crédit photos : Journet et al., CNRS-Inra

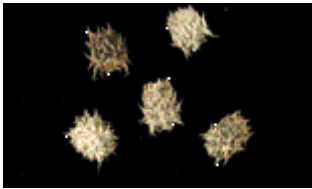


Photo 4. *Les gousses de M. truncatula cv Jemalong, qui peuvent contenir jusqu'à 12 graines, se détachent spontanément de la plante après maturation. Elles présentent une forme compacte (enroulement hélicoïdal) et de fortes épines.*

Crédit photos : Journet et al., CNRS-Inra
