



A la découverte des gènes et des génomes des plantes

par **Michel DELSENY**, Membre élu dans la discipline « Biologie intégrative »

Mesdames, Mesdemoiselles, Messieurs, chères Consoeurs et chers Confrères, les plantes supérieures sont indispensables à notre survie : elles sont l'une des sources de l'oxygène que nous respirons et elles assurent une fraction significative de la fixation du gaz carbonique. Nous les utilisons pour nous nourrir et nous soigner, pour nos industries et pour la production d'énergie. Ce sont de fabuleuses usines chimiques miniatures fonctionnant à l'énergie solaire, qui produisent des milliers de molécules différentes. La production de ces molécules est régie par les gènes, mais ce n'est que récemment que nous avons commencé à décrypter cette machinerie.

Le début de ma carrière a coïncidé avec les premiers balbutiements de la biologie moléculaire et du génie génétique végétal. Il m'a conduit à caractériser les ARNm des protéines de réserve du radis et ceux de protéines énigmatiques, présentes dans presque toutes les graines, les protéines « Late Embryogenesis Abundant » ou LEA.

J'ai alors réalisé l'intérêt qu'il y avait à travailler sur une plante modèle, *Arabidopsis thaliana* ou Arabette des dames, l'équivalent de la drosophile pour les animaux. Cette plante se prête remarquablement à l'analyse génétique et possède l'un des plus petits génomes connus avec « seulement » 130 millions de paires de bases, à comparer avec les 16 milliards du génome du blé. Un consortium CNRS-INRA a entrepris de séquencer systématiquement des copies d'ARNm, ou ADNc, préparés à partir de différents organes, une stratégie qui permet d'« étiqueter » chaque gène exprimé. Nous avons pris en charge les ADNc de graines. A cette époque, moins de 500 gènes et ADNc avaient été caractérisés chez les plantes, dont 200 chez *Arabidopsis*. En quelques mois, les efforts du consortium ont porté ce chiffre à plus de 10 000.

Nous étions alors prêts à participer à un projet pilote européen de séquençage d'un fragment de 2 millions de paires de bases du chromosome IV. Ce projet a permis l'émergence du consortium mondial qui a publié, fin 2000, le génome complet d'*Arabidopsis*, le premier d'une plante, avec environ 25000 gènes identifiés. Avec une contribution de l'ordre de 1% du génome, notre équipe ne pouvait entrer en compétition avec les grands centres de séquençage qui se mettaient en place. Au travers de cette expérience, nous avons compris que le séquençage n'était pas une fin en soi, mais qu'il fallait, une fois la séquence acquise, l'interpréter, l'annoter pour délimiter les gènes, en élucider les fonctions, et, au-delà, les intégrer dans le fonctionnement global de la plante entière.

Notre équipe s'est alors consacrée à l'analyse fonctionnelle. Elle a répertorié et annoté la soixantaine de gènes codant pour des protéines LEA et a analysé leur régulation. Ces protéines permettent la survie de la graine en conditions de déshydratation extrême mais nombre d'entre elles protègent aussi les stades végétatifs de la sécheresse. Nous avons participé à l'analyse d'une collection de mutants d'insertion OGM, créée par nos collègues de l'INRA. Ces insertions créent des mutations qui nous informent sur la fonction du gène muté et elles permettent d'isoler le gène affecté. Une dizaine de gènes dont la fonction est essentielle à la formation de la graine ont ainsi été identifiés et analysés. Nous avons aussi caractérisé plusieurs gènes contrôlant la biosynthèse des acides gras, la tolérance aux stress hydriques et oxydatifs, et la mort cellulaire.



Le séquençage du génome de la plante modèle avait montré qu'il était possible de séquencer le génome d'une plante et que la moisson d'informations était tout à fait considérable. Il est vite apparu que le génome suivant serait celui du riz. Dans le cadre de Génoplante nous avons entraîné nos collègues du CIRAD, de l'IRD et du secteur privé, dans l'analyse fonctionnelle de ce génome. La stratégie utilisée a été la même que pour Arabidopsis: construction d'une collection de mutants d'insertion avec détermination des séquences flanquant les insertions. Trente mille lignées OGM indépendantes ont été produites, multipliées et partiellement évaluées au champ. Ces informations et les graines correspondantes ont été mises à la disposition de la communauté scientifique grâce à un site web. Notre équipe a également identifié quelques gènes importants comme le premier gène de résistance au virus de la panachure jaune, une maladie endémique du riz en Afrique, et celui qui contrôle l'arôme du riz basmati.

Cependant notre découverte la plus originale a été de réaliser que la plupart des gènes étaient présents en double exemplaire, voire davantage. Nous avons déterminé comment ces différentes copies étaient distribuées dans le génome et avons ainsi montré que ce génome, que nous croyions simple, était en fait passablement compliqué et se présentait comme un patchwork de segments dupliqués. Cette observation était totalement inattendue et s'est révélée être l'une des clefs de la compréhension de l'évolution des génomes de plantes. Il a cependant fallu attendre le séquençage du génome de la vigne par Genoscope pour que ces observations prennent un sens. La vigne a un génome plus proche de celui de l'ancêtre des dicotylédones et il permet de retrouver l'origine des blocs dupliqués. Entre le génome ancestral et celui d'Arabidopsis, deux duplications globales suivies de réarrangements se sont produites, dont la plus récente il y a environ 25 millions d'années. Nous avons ensuite montré que le génome du riz était lui aussi dupliqué, ainsi que ceux du maïs et du blé, avant même que leurs séquences ne soient connues. Ces duplications sont conservées chez les graminées et résultent d'un évènement qui s'est produit il y a environ 70 millions d'années chez leur ancêtre commun. Ces analyses ont conduit à produire un modèle d'évolution du génome des graminées, modèle qui est toujours d'actualité et que les séquences de sorgho, de maïs, et de millet sont venues confirmer.

Vous avez ainsi un bref aperçu des travaux qui m'ont conduit à être devant vous cet après-midi. J'ai le sentiment d'avoir participé à une aventure extraordinaire, où presque chaque jour, avec mes collaborateurs, nous apprenions quelque chose de nouveau. Sans eux, rien de tout cela n'aurait été possible et je profite de l'occasion pour leur adresser tous mes remerciements. L'aventure ne fait que commencer, car la génomique révolutionne aujourd'hui la biologie. Le coût du séquençage a été réduit d'un facteur X10000 en 15 ans. Il en résulte une accumulation de données tout à fait spectaculaire : plus de soixante génomes de plantes, avec chacun 25 à 40 000 gènes, sont disponibles. Nous avons aujourd'hui accès à la variabilité moléculaire de pratiquement chaque grande espèce cultivée et nous savons modifier leur génome. Il nous faut maintenant comprendre comment ces différents gènes sont régulés et comment leur expression peut être modulée par l'environnement. Il reste à tirer parti de cette diversité génétique, au bénéfice de l'amélioration des plantes, pour faire face à l'immense défi d'assurer de façon durable la sécurité alimentaire d'une population mondiale en forte croissance, alors même que le climat change, que les surfaces cultivables se réduisent, et qu'il est impératif de préserver notre environnement.

Je vous remercie pour votre attention.