Etude des séquences d'ADN répétées spécifiques des différents génomes de riz

Michel DELSENY *, S.A. REDDY **, F. CORDESSE *, M.C. KIEFER-MEYER *, Alexandre DE KOCHKO ***, Gérard SECOND ****

Résumé : Les riz cultivés et sauvages constituent un complexe d'espèces au sein duquel sont représentés divers génomes. Nous avons isolé et caractérisé divers types de séquences d'ADN répétées qui sont spécifiques soit des génomes AA, soit des génomes CC. L'organisation de ces séquences, en tandem ou dispersées dans le génome, le nombre de copies, leur représentation et leur variabilité dans un éventail d'espèces et d'accessions ont été étudiés. Les principaux résultats obtenus dans notre groupe seront présentés.

Mots-clés : *Oryza*, séquences répétées d'ADN, gènes ribosomiques, évolution, polymorphisme.

Abstract : Cultivated and wild rice can be classified as a complex group of species in which various genome types have been recognized. We isolated and characterized several types of repeated DNA sequences which are specific either for the AA genome or the CC genome. Their organization as tandemly or dispersed repeated sequences, their copy number, their distribution and polymorphism in a number of distinct accessions have been studied. Main results obtained in our group will be presented.

Introduction

Les riz cultivés et sauvages sont répandus sur pratiquement tous les continents et ont fait l'objet d'études systématiques intensives ayant fait appel à l'observation morphogénétique et écologique, à la cytogénétique, ainsi qu'à l'analyse des profils isoenzymatiques (Oka, 1974 ; Second, 1982, 1985 ; Glaszmann, 1987). A partir de ces travaux, il a été possible de distinguer divers types de génomes et de typer des sous-groupes pour chaque

* Laboratoire de Physiologie et Biologie Moléculaire Végétales, URA 565 du CNRS, Université de Perpignan, 66860 Perpignan, France.
** Biology Dept., Texas A and M University, College Station, Tx 77843-3258, USA.
*** Division of Plant Biology, MRC7, The Scripps Research Institute, La Jolla, Ca 92037, USA.
**** Plant Breeding and Molecular Genetics, IRRI, PO Box 933, Manila, Philippines.

génome. Ces études sont permis de proposer des scénarios évolutifs de ce complexe d'espèces et de retracer l'histoire de sa domestication par l'homme (Second, 1982, 1985). Ainsi sur ces bases deux grands complexes ont été reconnus.

Le complexe *sativa* regroupe toutes les espèces du génome AA, avec les riz cultivés asiatiques et africains *O. sativa* et *O. glaberrima* et des espèces sauvages qui dérivent probablement des ancêtres des riz cultivés, telles que *O. rufipogon, O. longistaminata* et *O. breviligulata*.

Les riz du complexe *latifolia* comprennent uniquement des espèces sauvages. Elles appartiennent au génome BB (*O. minuta*), au génome CC (*O. officinalis, O. eichingeri*), au génome EE (*O. australiensis*) et au génome FF (*O. brachyantha*). Certaines espèces correspondent à des allotétraploïdes constituées par les génomes BBCC (*O. minuta*) ou CCDD (*O. latifolia, O. alta*). Le génome DD demeure encore inconnu à l'état diploïde. Certains de ces génomes ont une localisation géographique bien définie : ainsi le génome EE est-il endémique d'Australie et les génomes CCDD ne sont présents qu'en Amérique. *O. glaberrima* et *O. breviligulata* sont des espèces africaines (Tsunoda et Takahashi, 1984).

Les études utilisant l'analyse des profils isoenzymatiques ont montré l'existence de plusieurs grands écotypes au sein des riz cultivés qui se regroupent en type *indica* et type *japonica* avec d'autres groupes considérés comme intermédiaires (Second, 1982, 1985 ; Glaszmann, 1987). Ces études ont en fait largement confirmé la distinction établie entre ces deux grands types sur la base de la fertilité ou de la stérilité des produits de croisements avec quelques variétés testeurs. Ce type d'analyse met en évidence des barrières à la reproduction entre espèces relativement proches dont la nature demeure encore inconnue (Oka, 1974).

Avec le développement de la biologie moléculaire, il devient possible de disposer de marqueurs moléculaires qui permettent de typer les génomes de façon beaucoup plus précise. Notre travail antérieur sur les crucifères (Grellet *et al.*, 1986 ; Delcasso-Tremousaygue *et al.*, 1988) nous avait montré que certaines séquences d'ADN répétées étaient très spécifiques d'une espèce ou d'un genre donné. Nous avons donc entrepris une étude systématique des séquences répétées des riz cultivés et sauvages, en nous focalisant sur les génomes AA et CC. La raison du choix du génome CC est que ce génome constitue une bonne source de gènes pour introduire des caractères intéressants tels que des résistances à des pathogènes ou des résistances à la sécheresse dans les riz cultivés. Par ailleurs, il existait des lignées d'addition et d'introgression dérivées d'*O. officinalis* dans un fond génétique *O. sativa* (Jena et Khush, 1989). Les buts du travail réalisé étaient 1) d'isoler et de caractériser des séquences qui soient spécifiques d'un génome, 2) de les utiliser pour décrire leur polymorphisme et de suivre leur introgression au cours de croisements interspécifiques, 3) de mieux comprendre le rôle des séquences répétées dans l'évolution des génomes.

Matériel et méthodes

La stratégie suivie pour isoler les séquences répétées a consisté à isoler des fragments de restriction visibles à l'état de bandes discrètes superposées à un fond continu et diffus après coloration de l'ADN digéré, et résolu par électrophorèse, en fonction de sa taille. Ces fragments sont marqués radioac-

tivement et hybridés à des membranes portant de l'ADN digéré par différentes enzymes de restriction. Après autoradiographie, le profil d'hybridation permet généralement de dire si le fragment contient une séquence répétée en tandem, dispersée ou autre. Les fragments donnant des profils typiques de séquence en tandem ou de séquences dispersées ont été clonés et séquencés, puis utilisés comme sondes pour analyser la distribution de la séquence dans différentes accessions. La plupart des séquences isolées étant de relativement petite taille, nous avons construit une banque génomique dans un phage lambda à partir d'une accession d'*O. officinalis* (W 1278). Les fragments clonés dans ce vecteur sont de l'ordre de 15 kbp et permettent donc de mieux étudier l'arrangement des séquences étudiées avec les séquences voisines. Toutes les méthodes utilisées sont classiques (Maniatis *et al.*, 1982). Nous avons aussi obtenu de nos collègues japonais (Takaiwa *et al.*, 1984) un clone contenant un gène ribosomique complet avec la séquence espaceur intergénique. En effet, nos travaux antérieurs nous avaient montré que chez les crucifères, la région espaceur contenait également des séquences répétées extrêmement spécifiques du genre (Delcasso-Tremousaygue *et al.*, 1988).

Résultats

Une séquence répétée en tandem de 350-360 bp est spécifique du génome A

Une première séquence répétée a été caractérisée à partir de la variété « Cigalon » du riz *O. sativa*. Un fragment de 352 bp a été cloné et séquencé (De Kochko *et al.*, 1991). Cette séquence permet de révéler, après hybridation à des ADN de riz cultivé, des profils de restriction caractéristiques en échelle, indiquant une organisation en tandem. Des blocs contigus de près d'une centaine d'unités ont été observés. Cette séquence s'hybride avec une intensité plus ou moins forte à toutes les accessions appartenant au génome AA, mais par contre aucun signal n'est détecté avec les espèces appartenant aux autres génomes. Le nombre de copies est modéré et variable. Il est généralement de l'ordre du millier chez les *indica*, de l'ordre de quelques centaines chez les *japonica*, chez *O. rufipogon* et *O. longistaminata*. Des observations quasiment identiques avaient été précédemment rapportées (Wu et Wu, 1987). Toutefois, il semble y avoir chez les *O. rufipogon* deux classes distinctes qui pourraient correspondre respectivement aux ancêtres du type *indica* et à ceux du type *japonica*.

L'analyse de ces séquences par électrophorèse en gel d'agarose révèle peu ou pas de polymorphisme sauf chez les *O. rufipogon* et les *O. longistaminata* chez lesquels des événements d'amplification et de délétion ont donné lieu à l'apparition de fragments atypiques. Un polymorphisme intéressant permettant habituellement de discriminer entre types *japonica* et *indica* n'est observé qu'à condition d'utiliser des enzymes reconnaissant quatre paires de bases et de réaliser les électrophorèses, de façon résolutive, en gel d'acrylamide dénaturant (Gebhardt *et al.*, 1989). Le même polymorphisme est retrouvé au sein des *O. rufipogon*, ce qui confirme l'hypothèse initialement émise d'une domestication indépendante des *O. sativa* de type *indica* et

japonica à partir de deux populations distinctes et géographiquement éloignées d'*O. rufipogon*. Récemment, des expériences d'hybridation *in situ* (Wu *et al.*, 1991) ont montré que cette séquence était localisée dans l'une des deux régions subtélomériques sur 8 des 12 paires de chromosomes.

L'espaceur des gènes ribosomiques contient des répétitions partiellement spécifiques des divers génomes

A partir de l'unité clonée, nous avons dérivé toute une série de sondes qui nous ont permis de tester leur spécificité vis-à-vis des différents génomes.

Des travaux préliminaires (Olmedilla *et al.*, 1984 ; Cordesse *et al.*, 1990) avaient permis d'étudier la variabilité des gènes ribosomiques et de montrer que pour l'essentiel, celle-ci est localisée dans la région espaceur. Nous avons établi que chez les espèces cultivées asiatiques, il existait 8 classes d'espaceur dont les tailles diffèrent chacune d'un incrément régulier de 250-260 bp. Les espèces du groupe *japonica* sont parmi les classes courtes, alors que les *indica* ont des espaceurs plus longs. Les riz africains, *O. glaberrima* et *O. brevigulata,* ne varient pas du tout. La variabilité est beaucoup plus grande et moins régulière chez les espèces sauvages. Ces résultats ont été confirmés par des chercheurs japonais (Sano et Sano, 1990).

La région espaceur a été entièrement séquencée (Takaiwa *et al.*, 1990 ; Cordesse, 1990) et a révélé l'existence de trois répétitions de 250-260 bp. Par ailleurs, chacune de ces répétitions contient un site potentiel d'initiation de la transcription.

La dissection de l'espaceur a révélé que cette région des répétitions a une spécificité large puisqu'une séquence homologue existe dans les génomes BB, CC et FF. Dans le génome EE, cette séquence est absente ou a divergé suffisamment pour ne plus être reconnue en conditions d'hybridation moyennement sévères. Nous avons trouvé, en aval des répétitions, deux séquences adjacentes, l'une de 93 bp, strictement spécifique du génome AA, et l'autre de 61 bp spécifique des génomes AA et FF (Cordesse *et al.*, 1992). La spécificité vis-à-vis du génome AA porte sur toutes les accessions analysées, alors que celle vis-à-vis du génome FF est restreinte à la seule accession dont nous disposions.

Isolement et caractérisation de deux séquences répétées en tandem de 360 bp spécifiques du génome C

Plusieurs enzymes de restriction ont permis de révéler des bandes pouvant correspondre à de l'ADN répété dans l'espèce *O. officinalis*. Eco RV et Sph I révèlent une bande de 360-370 bp, Hind II une bande de 195 bp et Bgl II une bande de 520 bp. L'ADN de chacune de ces bandes a été cloné et s'est révélé correspondre à des séquences spécifiques du génome CC en conditions rigoureuses d'hybridation. Les séquences Eco RV, Sph I et Hind II correspondent à des répétitions en tandem, alors que la séquence Bgl II correspond à une séquence dispersée. Dans les 3 premiers cas des profils en échelle sont observés, alors que dans le dernier, ce sont des fragments de taille variable qui sont observés.

La séquence de 3 clones indépendants Eco RV et de 3 clones Sph I a été déterminée (Reddy *et al.*, 1992). Elle a montré que les deux types de fragments appartiennent à la même famille puisque l'homologie à l'intérieur de chaque type est supérieure à 95 % et que l'homologie entre chaque classe est supérieure à 90 %. Ces deux séquences sont homologues à environ 85 % à une séquence décrite récemment dans une autre accession (Zhao *et al.*, 1989).

L'analyse de plusieurs accessions a révélé des profils très semblables les uns aux autres. Quelques échantillons ne présentent pas cette séquence. Le nombre de copies est variable, mais élevé dans la plupart des échantillons, de l'ordre de 200 000. L'analyse des lignées d'addition (Jena et Khush, 1989) a permis d'observer la séquence, avec le même profil en échelle pour toutes les plantes ayant le chromosome 8 en surnombre (MAAL 8), ainsi que sur une des quatre plantes analysées appartenant à la famille MAAL 6.

La répétition en tandem de 195 bp est apparentée aux répétitions de l'espaceur des gènes ribosomiques

Le fragment cloné Hind II de 195 bp a été utilisé comme sonde de la même façon que les fragments précédents. Avec la plupart des échantillons *O. officinalis,* il donne un profil caractéristique en échelle régulière après une digestion Hind II. Avec quelques autres, il ne donne pas de signal indiquant que dans ces accessions, la séquence n'est pas ou peu représentée. Lorsque l'hybridation est réalisée en conditions douces, un signal est généralement observé sur tous les types génomiques testés. La séquence amplifiée chez *O. officinalis* est donc apparentée à une séquence présente sans doute en plus faible nombre de copies et avec un autre arrangement. Lorsque d'autres espèces du génome CC telles que *O. collina* et *O. eichingeri* sont analysées, des profils polymorphes très différents des profils en échelle son observés, suggérant que cette séquence est organisée de façon différente dans ces espèces.

La comparaison de la séquence de 3 clones indépendants de cette répétition a révélé une homologie inattendue avec les répétitions de 250 bp décrites dans l'espaceur des gènes ribosomiques de type AA (Takaiwa *et al.*, 1990 ; Cordesse *et al.*, 1992). La séquence répétée spécifique d'*O. officinalis* est plus courte que celle des gènes ribosomiques et la région délétée correspond en fait à la séquence encadrant le site présumé d'initiation de la transcription. Ainsi, il semble que la séquence des répétitions de l'espaceur des gènes ribosomiques ait été amplifiée ailleurs dans le génome après avoir été altérée de telle sorte que les éléments fonctionnels aient disparu. Le nombre de copies, estimé à 200 000, est très largement supérieur à celui attendu si cette séquence résultait entièrement de l'espaceur des gènes ribosomiques. En effet, il y a au maximum une dizaine de répétitions dans l'espaceur et un ou deux milliers d'espaceurs, soit un maximum de 20 000 copies. Chez *O. officinalis* nous en trouvons en fait 10 fois plus. Nous avons observé la présence de la séquence amplifiée sur les plantes de la lignée d'addition MAAL 11.

Une séquence de 520 bp est dispersée dans le génome des riz sauvages de type CC

La séquence de 520 bp qui a été isolée sous la forme d'un fragment Bgl II conduit à des profils irréguliers superposés à un bruit de fond lorsque l'ADN est digéré par divers types d'enzymes. Ceci indique que cette séquence, dont le nombre de copies a été estimé à 30 000, est dispersée dans le génome. La séquence ne présente aucune homologie significative avec quoi que ce soit de connu. Lorsque différents échantillons sont analysés par digestion Bgl II, pratiquement aucune variabilité n'est observée suggérant que les sites Bgl II sont à l'intérieur de l'élément répété plutôt qu'à l'extérieur. Une fraction des fragments de haut poids moléculaire est retrouvé sur les plantes de la lignée MAAL 8.

En vue d'étudier plus en détail l'organisation de cette famille d'éléments, nous avons isolé 14 clones génomiques de 14-18 kbp hybridant avec le fragment Bgl II initial. Des fragments Bgl II de taille différente ont été observés. Ces fragments sont présents au nombre d'un, quelquefois deux ou trois par phage et, dans tous les cas, il est clair que l'insertion dans le phage contient une proportion importante de séquences qui n'hybrident pas à la sonde. Ces observations constituent donc une preuve directe de la dispersion des séquences Bgl II au milieu d'autres séquences sans relation apparente. Par contre, l'origine de ces séquences dispersées et leur agencement dans le génome sont encore complètement inconnus.

Discussion

Le travail présenté a atteint le premier objectif recherché, à savoir disposer de sondes spécifiques du génome AA et du génome CC. Par ailleurs, comme aucune des sondes spécifiques du génome CC ne s'hybridait à l'ADN des échantillons appartenant aux allotétraploïdes CCDD, nous avons isolé et caractérisé une séquence spécifique du génome CCDD.

Nous avons observé qu'un certain nombre d'échantillons du génome CC ne donnaient pas de signal avec nos sondes. On peut donc se demander si à l'intérieur des *O. officinalis* il ne faudrait pas distinguer deux groupes, selon la présence ou l'absence de certaines séquences répétées. Une autre hypothèse serait que ces échantillons atypiques correspondraient à des erreurs de classification ou d'étiquetage lors de la propagation des collections. Pour au moins l'un des échantillons, il est probable que ce soit le cas, car ni l'analyse RFLP, ni l'analyse isozymique ne confirme l'appartenance au groupe *O. officinalis*.

L'analyse du polymorphisme de ces séquences répétées s'est jusqu'à présent révélée décevante et n'a pas permis de les localiser sur les cartes RFLP en cours d'élaboration, à l'exception des séquences subtélomériques AA. Pour tenter de remédier à cette difficulté, nous avons recherché des clones génomiques de grande taille (14-16 kbp) susceptibles de contenir des fragments polymorphes, adjacents aux séquences répétées. Dans cette approche, les clones contenant les séquences en tandem se sont révélés instables ou non

viables. Les seuls clones que nous ayons pu maintenir sont ceux qui contiennent la séquence Bgl II dispersée. Il reste à les caractériser et à établir la nature de l'élément dispersé, ce qui donnera peut être des idées sur les mécanismes qui contrôlent cette dispersion.

L'analyse des lignées d'addition, qui malheureusement ne constituent pas un stock complet puisque quatre d'entre elles ont été perdues depuis leur isolement initial, suggère que ces séquences répétées ne sont présentes que sur un nombre limité de chromosomes. Cependant, cela ne correspond pas aux observations d'hybridation *in situ* sur les séquences AA-spécifiques (Wu *et al.*, 1991) ou sur d'autres espèces (Bedbrook *et al.*, 1980) et il est probable que ces lignées d'addition ont déjà subi un certain nombre de réarrangements ayant conduit à la perte des séquences répétées spécifiques. Cette élimination suggère que certaines des séquences répétées pourraient jouer un rôle dans l'établissement des barrières entre espèces.

Au terme de ce travail, plusieurs voies de recherche demeurent ouvertes et mériteraient d'être explorées plus avant : quel est le sort des séquences répétées dans des croisements interspécifiques ? Comment et à quelle vitesse sont-elles éliminées ? Dans quelles régions chromosomiques ces séquences sont-elles localisées ? Il serait particulièrement utile de les encadrer avec divers marqueurs RFLP afin d'analyser leur comportement dans les croisements : sont-elles le siège d'événements de recombinaison plus fréquents, ou bien au contraire inhibent-elles les recombinaisons ? Jouent-elles un rôle dans la mécanique chromosomique ? Est-il en particulier possible d'isoler des protéines qui se lient spécifiquement à ces séquences et qui pourraient interagir avec le fuseau chromatique ? Toutes ces questions peuvent maintenant être abordées avec les outils disponibles.

Remerciements

Ce travail n'a été possible que grâce au soutien de la Fondation Rockefeller et au soutien continu du CNRS et de l'ORSTOM. Il a également bénéficié d'une aide du Centre Français du Riz.

Bibliographie

BEDBROOK J.R., JONES J., O'DELL M., THOMPSON R.D., FLAVELL R.B., 1980 — A molecular description of telomeric heterochromatin in *Secale* species. *Cell*, 19 : 545-560.

CORDESSE F., SECOND G. and DELSENY M., 1990 — Ribosomal gene spacer length variability in cultivated and wild rice species. *Theor. Appl. Genet.*, 79 : 81-88.

CORDESSE F., 1990 — Etude des gènes ribosomiques chez des riz cultivés et sauvages. PhD thesis dissertation. Université de Perpignan.

CORDESSE F., GRELLET F., REDY A.S. and DELSENY M., 1992 — Genome spécificity of rDNA spacer fragments from *Oryza sativa*. *Theor. Appl. Genet.* (in press).

DE KOCHKO A., KIEFER M.C., CORDESSE F., REDDY A.S. and DELSENY M., 1991 — Distribution and organization of a tandemly repeated 352 bp sequence in the *Oryzae* family. *Theor. Appl. Genet.*, 82 : 57-64.

DELCASSO-TREMOUSAYGUE D., GRELLET F., PANABIERES F., ANANIEV E.D. and DELSENY M., 1988 — Structural and transcriptional characterization of the external spacer of a ribosomal RNA nuclear gene from a higher plant. *Eur. J. Biochem.*, 172 : 767-776.

GEBHARDT C., BLOMENDAHL C., SCHACHSTABEL U., DEBENER T., SALAMINI F. and RITTER E., 1989 — Identification of 2 n breeding lines and 4 n varieties of potato (*Solanum tuberosum* ssp. *tuberosum*) with RFLP finger prints. *Theor. Appl. Genet.*, 78 : 16-22.

GLASZMANN J.C., 1987 — Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.*, 74 : 21-30.

GRELLET F., DELCASSO D., PANABIERES F. and DELSENY M., 1986 — Organisation and evolution of a higher plant alphoïd-like satellite DNA sequence. *J. Mol. Biol.*, 187 : 495-507.

JENA K.K. and KHUSH G.S., 1989 — Monosomic alien addition lines of rice : production, morphology, cytology and breeding behavior. *Genome*, **32** : 449-455.

MANIATIS T., FRITSCH E.F. and SAMBROOK J., 1982 — Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory Press.

McCOUCH S.R., KOCHERT G., YU Z.H., WANG Z.W., KHUSH G.S., COFFMAN W.R., TANKSLEY S.D., 1988 — Molecular mapping of rice chromosomes. *Theor. Appl. Genet.*, 76 : 815-829.

OKA H.I., 1974 — Experimental studies on the origin of cultivated rice. *Genetics*, 78 : 475-486.

OLMEDILLA A., DELCASSO D. and DELSENY M., 1984 — Methylation pattern of nuclear ribosomal DNA genes from rice (*Oryza sativa*). *Plant Sci. Lett.*, 37 : 123-127.

REDDY A.S., KIEFER-MEYER M.C., CORDESSE F., SECOND G. and DELSENY M., 1992 — Characterization of new variants of a satellite DNA from *Oryza officinalis* specific for the CC genome of rice (en préparation).

SANO Y. and SANO T., 1990 — Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. *Genome*, 33 : 209-218.

SECOND D., 1982 — Origin of the genic diversity of cultivated rice (*Oryza* ssp) : study of the polymorphism scored at 40 isozyme loci. *Jap. J. Genet*,. 57 : 25-57.

SECOND D., 1985 — Evolutionary relationships in the *sativa* group of *Oryza* based on isozyme data. *Genet. Select. Evol.*, 17 : 89-114.

TAKAIWA F., OONO K. and SUGIURA M., 1984 — The complete nucleotide sequence of a rice 17S rDNA gene. *Nucl. Acid Res.*, 12 : 5441-5448.

TAKAIWA F., KIKUCHI S. and OONO K., 1990 — The complete nucleotide sequence of the intergenic spacer between 25S ans 17S rDNAs in rice. *Plant Mol. Biol.*, 15 : 933-935.

TSUNODA S. and TAKAHASHI N., 1984 — *Biology of rice. Development in Crop Science*, 7. Japan Scientific Societies Press, Tokyo.

WU T.Y. and WU R., 1987 — A new rice repetitive DNA shows sequence homology to both 5S RNA and tRNA. *Nucl. Acids Res.*, 15 : 5913-5923.

WU H.K., CHUNG M.C., WU T., NING C-N. and WU R., 1991 — Localization of specific repetitive DNA sequences in individual rice. *Chromosoma*, 100 : 330-338.

ZHAO X., WU T., XIE T. and WU R., 1989 — Genome specific repetitive sequences in the genes *Oryza*. *Theor. Appl. Genet.*, 78 : 201-209.

Delseny M., Reddy S.A., Cordesse F., Kiefermeyer M.C., Kochko Alexandre de, Second Gérard.

Etude des séquences d'ADN répétées spécifiques des différents génomes de riz.

In Complexes d'espèces, flux de gènes et ressources génétiques des plantes.

Paris (FRA) : Bureau des Ressources Génétiques, 1992, p. 225-232. ISBN 2-908447-02-9