

La génomique du riz : un outil pour l'amélioration du riz et des autres céréales

Michel Delseny¹
Jérôme Salse²
Richard Cooke¹

¹ Université de Perpignan UPVD
CNRS UMR 5096 CNRS
Laboratoire Génome et développement
des plantes
52, avenue Paul-Alduy
66860 Perpignan
France
<delseny@univ-perp.fr>
<cooke@univ-perp.fr>

² Inra Site de Crouël
UMR 1095 Génétique, diversité et
écophysiologie des céréales
5, chemin de Beaulieu
63100 Clermont-Ferrand
France
<Jerome.salse@clermont.inra.fr>

Résumé

Le séquençage du génome du riz a permis d'établir un catalogue de quelque 35 000 gènes et des éléments transposables qu'il contient. Ce génome constitue à la fois une référence et un outil pour découvrir la fonction des gènes du riz, comprendre l'organisation générale de ce génome et pour progresser dans la compréhension et la connaissance des génomes des autres céréales. L'introduction des nouvelles technologies à haut débit révolutionne l'analyse de la biodiversité des espèces sauvages et des innombrables variétés cultivées et accélère l'étude de l'évolution du genre *Oryza* et des céréales. Elles offrent un outil de génotypage d'une résolution exceptionnelle et ouvrent la voie à l'isolement et la caractérisation d'un nombre de gènes d'intérêt toujours plus grand tant pour le riz que pour les autres céréales. Certains d'entre eux sont déjà utilisés par les sélectionneurs.

Mots clés : céréales ; gènes ; génomes ; riz ; sciences végétales.

Thèmes : amélioration génétique ; productions végétales.

Abstract

Rice genomics: A tool for breeding rice and other cereals

Rice genome sequencing has made it possible to establish a catalogue of its ~35,000 genes and of its transposable elements. This genome is both a reference and a tool for discovering gene functions, understanding its general organisation and for progressing in our knowledge of other cereal genomes. Development of next generation sequencing technologies has revolutionized analysis of the biodiversity of both cultivated varieties and related wild species. It has considerably boosted the analysis of the evolution of the *Oryza* genus, as well as of other cereals. It provides a fantastic genotyping tool and opens avenues for isolating and characterizing an ever-increasing number of important genes, some of which are already manipulated in breeding programmes.

Key words: cereals; genes; genomes; plant sciences; rice.

Subjects: genetic improvement; vegetal productions.

Le riz est une espèce majeure pour l'alimentation humaine puisqu'il est consommé par la moitié de l'humanité. C'est aussi la céréale pour laquelle les études de génomique sont les plus avancées. En effet, parmi les céréales, son génome est l'un des plus petits (~ 400 Mbp) ; ce dernier est largement colinéaire avec celui des autres céréales (c'est-à-dire que l'ordre des

gènes le long des chromosomes est sensiblement le même que celui qui est observé chez les autres céréales ; la conservation de l'ordre des gènes entre chromosomes ou fragments de chromosomes au sein d'une espèce ou d'espèces distinctes est aussi appelée synténie). Elle se prête plus facilement à la transgénèse, donc à l'obtention de plantes génétiquement modifiées. C'est la première espèce cultivée dont

Tirés à part : M. Delseny

doi: 10.1684/agr.2013.0626

Pour citer cet article : Delseny M, Salse J, Cooke R, 2013. La génomique du riz : un outil pour l'amélioration du riz et des autres céréales. *Cah Agric* 22 : 466-74. doi : 10.1684/agr.2013.0626

le génome a été séquençé avec une haute précision. Sa variabilité génétique est importante et répertoriée dans des collections publiques (plus de 110 000 accessions à l'*International Rice Research Institute* (<http://www.irri.org/our-science/genetic-diversity-conservation>)). Des programmes de recherche et d'amélioration existent dans pratiquement chaque pays producteur.

La génomique du riz commence au début des années 1990 avec le programme japonais de détermination des *expressed sequenced tags* (EST) ou séquences partielles d'ADNc) et la mise en place d'outils, tels que les cartes génétiques (<http://www.gramene.org/>) et les cartes physiques (<http://www.omap.org/>), qui ont permis de déterminer, de façon ordonnée, la séquence « complète » du génome de la variété Nipponbare selon la méthode classique de Sanger (International Rice Genome Sequencing Project, 2005). Simultanément, d'autres outils se sont mis en place pour élucider la fonction des gènes. L'achèvement de la séquence du génome du riz, après celui de l'espèce modèle *Arabidopsis thaliana*, a permis de montrer qu'il était possible d'avoir accès à la séquence d'un génome de plante et d'en tirer des informations importantes pour élaborer des stratégies d'amélioration des plantes. Dans le même temps des technologies nouvelles de séquençage à très haut débit (*next generation sequencing techniques* [NGST]) ont été développées, divisant les coûts par 10 000 en l'espace de dix ans et réduisant les délais d'obtention d'une ébauche de séquence de plusieurs années à quelques semaines (Delseny *et al.*, 2010 ; Schneeberger et Weigel, 2011). Ces avancées permettent maintenant d'obtenir une séquence de qualité raisonnable pour un nombre toujours plus grand d'espèces (plus d'une trentaine de génomes de plantes supérieures ont déjà été publiés et bien d'autres seront prochainement disponibles). Ces données nouvelles permettent la comparaison entre génomes d'espèces voisines, comme les céréales, à une résolution encore jamais atteinte. L'apport le plus important des nouvelles technologies de séquençage est qu'elles donnent accès à la quasi-totalité de la variabilité génétique d'une espèce.

L'objectif de cet article est de retracer les grandes étapes de la génomique du

riz, de montrer ce que l'on sait aujourd'hui de l'organisation de son génome et de celui des autres graminées, et d'illustrer l'utilisation de ces informations en amélioration des plantes.

Le séquençage du génome du riz et son organisation générale

Le séquençage du génome du riz a été réalisé en plusieurs étapes qui ont permis de mettre en place toutes les ressources nécessaires à la réalisation d'un séquençage de très haute qualité et à l'annotation structurale et fonctionnelle de la séquence (l'annotation est la démarche qui consiste à délimiter, dans la séquence, les gènes, les éléments transposables [séquences d'ADN capables de se déplacer et de se multiplier de manière autonome dans un génome], les séquences régulatrices et à élucider la fonction biochimique ou biologique). Le séquençage génomique proprement dit a été le fait d'un consortium, l'*International Rice Genome Sequencing Project* (International Rice Genome Sequencing Project, 2005), qui a travaillé sur la variété japonaise Nipponbare (*Japonica*). Dans le même temps, des initiatives chinoises ou industrielles ont produit des brouillons du génome de la même variété ainsi que de la variété élite chinoise 93-11 de type *Indica*. Au total, ~ 390 Mbp, représentant 12 chromosomes, ont été séquencées. Bien que la séquence soit réputée complète, il reste encore une trentaine de trous, dont 9 régions centromériques, et 10 des 24 régions télomériques.

Depuis 2005, l'essentiel du travail consiste à augmenter les ressources biologiques utiles aux études génomiques, à annoter la séquence de référence, à cloner et caractériser de façon fonctionnelle les gènes d'intérêt et à explorer la biodiversité génétique du riz de façon à améliorer les variétés existantes.

Actuellement, environ 1 253 000 EST sont répertoriées dans la base de données dbEST ([http://www.ncbi.](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)

http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). L'étiquetage des gènes par des séquences partielles d'ADNc fournit à la fois des informations sur l'expression des gènes, mais représente également un apport essentiel pour l'annotation du génome. Des collections d'ADNc complets existent, aussi bien pour Nipponbare que pour des variétés *Indica* (Jiang *et al.*, 2012). Cette ressource a été complétée récemment par les stratégies de séquençage à très haut débit RNA-Seq (Zhang *et al.*, 2010 ; Lu *et al.*, 2010). La qualité de l'annotation du génome, basée dans un premier temps sur des prédictions bio-informatiques, a été largement améliorée par l'apport de ces informations, par les données de protéomique (Helmy *et al.*, 2011), ainsi que par la comparaison avec des séquences déterminées sur d'autres céréales. La séquence a été annotée indépendamment par le consortium international (<http://www.rapdb.dna.affrc.go.jp/>) et une équipe américaine (<http://www.rice.plantbiology.msu.edu/>). L'annotation est régulièrement mise à jour sur ces sites. Le premier groupe estime qu'il y a 31 439 gènes, validés par l'existence d'un ADNc, alors que le second considère qu'il y en a 42 653, essentiellement sur la base de prédictions bio-informatiques. Cette différence est due au fait que, jusqu'à récemment, l'assemblage des 12 pseudo-molécules représentatives des 12 chromosomes ne faisait pas consensus, que différentes méthodes de validation des gènes sont utilisées et que de nombreuses ambiguïtés sont créées par la présence de rétrotransposons (RT). Depuis octobre 2011, des séquences de pseudo-molécules communes ont été définies, mais les processus d'annotation restent distincts. Malgré ces divergences, qui illustrent la difficulté à annoter correctement un génome, le nombre de gènes est estimé à ~ 35 000 et est voisin de celui d'*Arabidopsis*. Ces gènes se répartissent en ~ 13 000 familles. Les éléments transposables, complets ou partiels, dont plus de 250 familles de RT, (<http://www.retroryza.fr/>), représentent 35 % du génome. L'annotation fonctionnelle repose, elle, sur la comparaison de séquences avec des gènes déjà connus chez d'autres espèces, sur l'exploitation des collections de mutants et de lignées transgéniques qui permettent

d'associer un gène à un phénotype (Krishnan *et al.*, 2009 ; Sakurai *et al.*, 2011), sur le clonage positionnel de gènes d'intérêt, sur les profils de coexpression des gènes déduits de l'analyse de puces à ADN (Childs *et al.*, 2011 ; Jiang *et al.*, 2012), sur l'analyse des interactions protéines-protéines ou interactome (Sapkota *et al.*, 2011), et sur des validations expérimentales.

La génomique comparative du riz et des céréales

Les premières études de cartographie du riz à l'aide des marqueurs molé-

culaires ont suggéré que les chromosomes 1 et 5, ainsi que les chromosomes 11 et 12, portaient des gènes dupliqués. La détermination de la séquence a permis la comparaison, deux à deux, de tous les chromosomes et a abouti à une caractérisation de l'ensemble des duplications inter- et intrachromosomiques (International Rice Genome Sequencing Project, 2005 ; Yu *et al.*, 2005). L'analyse la plus récente (Salse *et al.*, 2008) conclut à une duplication du génome entier (WGD, *Whole Genome Duplication*) qui se serait produite il y a ~ 70 millions d'années (c'est-à-dire peu avant la divergence des génomes de céréales) impliquant les chromosomes r11-r12, r5-r1, r10-r3, r7-r3, r4-r2, r9-r8, r2-r6. Ces sept duplications majeures couvrent ~ 70 % du génome (*figure 1*). Ces travaux suggèrent que le riz moderne diploïde

est un ancien polyploïde, à l'instar de toutes les plantes supérieures (Salse, 2012).

Depuis la publication de la séquence, de nouvelles technologies (NGST), plus efficaces en termes de production, ont permis d'obtenir les séquences plus ou moins élaborées des génomes d'autres céréales ou graminées (Delseny *et al.*, 2010). Celles du sorgho, du maïs et du millet des oiseaux (*Setaria italica*) ont été publiées, permettant des comparaisons entre ces génomes et celui du riz. Il en est de même pour le *Brachypodium*, une graminée de plus en plus utilisée comme plante modèle pour l'étude du génome graminée, du fait de la petite taille de son génome et de la facilité de sa culture en conditions contrôlées. Ainsi, parmi les 13 000 familles géniques du riz, 10 900 sont représentées chez le maïs et 11 500 chez le sorgho alors que 1 200 semblent spécifiques du riz. Environ 75 % des gènes sont conservés entre le riz et les autres graminées étudiées, c'est-à-dire 67 % riz-maïs, 79 % riz-sorgho, 73 % riz-blé, et 82 % riz-*Brachypodium* (Schnable *et al.*, 2009 ; Salse, 2012).

Les techniques de cartographie génétique suggéraient que le génome du riz était colinéaire avec celui des autres céréales et avaient permis de représenter la synténie chez les céréales sous la forme de cercles concentriques (Moore *et al.*, 1995). Ces observations ont été largement confirmées (*figure 2*) au fur et à mesure que la résolution des cartes s'est affinée, puis que les séquences génomiques sont devenues disponibles (Salse et Feuillet, 2011). Elles ont permis de proposer un scénario évolutif à partir d'un ancêtre commun (*figure 3*) possédant 5 chromosomes (Salse, 2012).

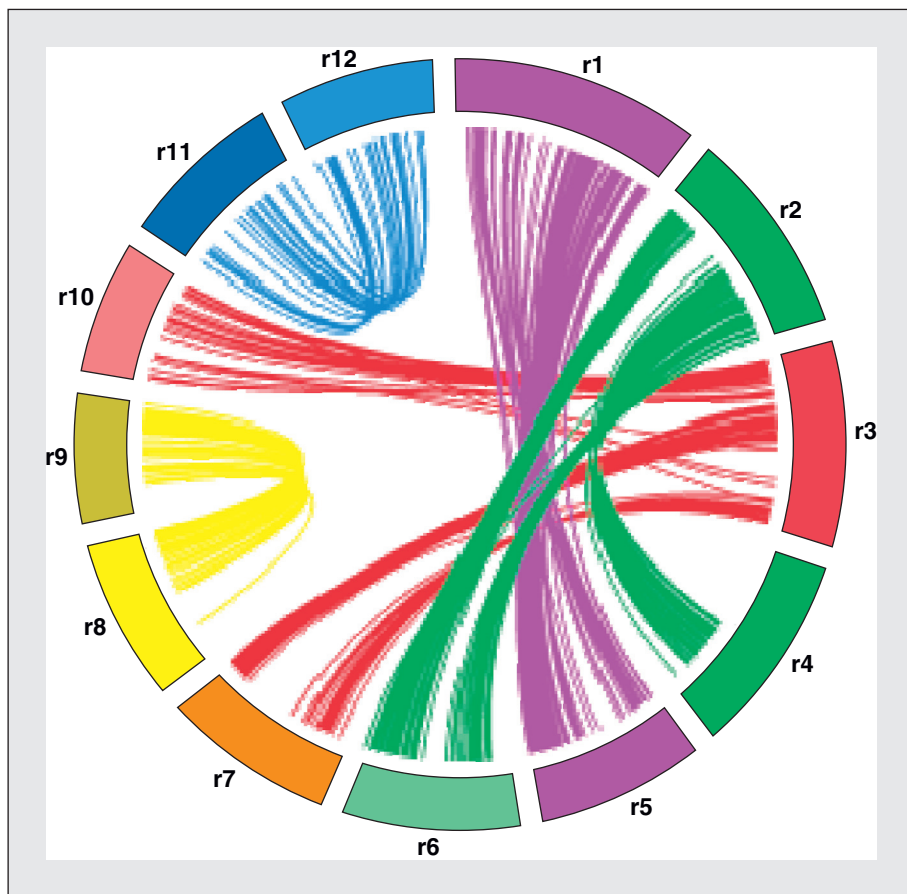


Figure 1. Les duplications du génome du riz.

Figure 1. Duplications in the rice genome.

Les 12 chromosomes du riz (r1-r12) sont représentés selon un modèle circulaire. Les gènes paralogues (homologues au sein de l'espèce et dérivant d'un ancêtre commun par duplication) qui définissent les 7 duplications majeures sont reliés entre eux par une ligne colorée.

Adapté de Salse *et al.* (2008).

Séquençage des génomes du genre *Oryza*

Outre les deux espèces cultivées, *Oryza sativa* et *Oryza glaberrima*, le genre *Oryza* comprend 22 espèces sauvages représentant dix types de génomes, qui ont divergé entre un et 25 millions d'années, avant la



Figure 2. Représentation de la synténie entre les chromosomes de céréales selon le modèle des cercles concentriques.

Figure 2. Synteny between cereal chromosomes, according to the concentric circular model.

Les génomes de blé, maïs, sorgho, *Brachypodium* et riz sont représentés par leurs chromosomes, disposés sur des cercles concentriques. Le génome du riz est au centre avec ses 12 chromosomes caractérisés chacun par une couleur et celui du blé à l'extérieur. Le code couleur permet de repérer les segments chromosomiques homologues d'un chromosome de riz chez chacune des autres graminées. Les flèches indiquent les fusions et réarrangements de chromosomes qui se sont produites chez les graminées et qui expliquent des nombres de chromosomes distincts : ainsi, les chromosomes 1 et 2 du sorgho (troisième cercle à partir du centre) résultent respectivement de la fusion des chromosomes 3 et 10 et 7 et 9 du riz. Ces fusions sont matérialisées par des flèches reliant les fragments correspondants. Seul le génome haploïde du blé est représenté et celui du maïs, tétraploïde cryptique, est représenté par deux cercles rapprochés. Le long d'un rayon du cercle, les gènes sont très largement conservés d'une espèce à l'autre. Adapté de Salse et Feuillet (2011).

domestication. Un important programme américain, the *Oryza Mapping Alignment Project* ([OMAP], <http://www.omap.org/>) s'est fixé pour objectif d'analyser en détail les différents génomes des riz sauvages et cultivés, d'abord au niveau des cartes physiques, puis, au travers d'un consortium international, au niveau de la séquence (Goicoechea *et al.*, 2010). Cette ressource permet d'avoir accès à l'évolution des gènes et des génomes au sein du genre *Oryza* et au processus de domestication. Les séquences produi-

tes par les NGST, qui consistent à découper le génome en très petits fragments (centaines de nucléotides), puis à les lire et les assembler, peuvent être de qualité très inégale et dépendent en fait de l'objectif recherché. Si l'on veut obtenir une séquence de référence, précise et correctement assemblée, un taux de couverture élevé (lecture de plusieurs copies de chaque fragment) est nécessaire et une combinaison des différentes technologies est indispensable. Si, l'objectif est d'obtenir du polymorphisme ou un génotypage,

une couverture beaucoup plus faible est suffisante, mais il est indispensable de disposer d'une séquence de référence de bonne qualité, comme celle de Nipponbare.

Actuellement, les séquences des bras courts du chromosome 3 de 14 espèces, ainsi que les séquences complètes d'*O. glaberrima*, *O. rufipogon*, *O. nivara*, *O. brachyantha* et *O. longistaminata* ont été déterminées et plusieurs autres sont en progrès. L'impact des éléments transposables, et notamment des RT, sur la taille des génomes a été mis en évidence. L'amplification récente de seulement trois familles de RT a conduit au doublement de la taille du génome chez *O. australiensis*. De même, deux familles de RT sont responsables d'une part importante de la variabilité de taille entre 12 des génomes *Oryza* (Zuccolo *et al.*, 2007). Au niveau génique, une première étude portant sur neuf génomes, ciblée sur les gènes *Adh1-Adh2* codant l'alcool-déshydrogénase, a mis en évidence une grande plasticité du génome, avec une rupture de la synténie provoquée par des pertes ou des acquisitions de gènes. Une étude plus récente (Lu *et al.*, 2009) autour de *Monoculm*, un gène impliqué dans le tallage, a confirmé cette observation et révélé une variabilité des éléments transposables, ceux-ci étant conservés seulement entre espèces ayant le même génome (par ex. AA ou BB). Si une grande conservation de la présence et de l'ordre des gènes est généralement observée, chez certaines espèces, certains gènes auraient été créés *de novo*, maintenus préférentiellement, ou insérés. L'exploitation des ressources d'OMAP et l'acquisition de séquences d'autres génomes a également révélé l'importance de l'évolution concertée dans l'évolution des génomes des céréales (ce terme traduit le fait que des mécanismes, encore mal connus, homogénéisent certaines familles multigéniques, de sorte qu'elles paraissent avoir évolué moins vite que d'autres séquences et de façon coordonnée). L'analyse de la séquence du génome du sorgho, puis de *Brachypodium* a démontré qu'une duplication du riz dans la région subtélomérique des chromosomes 11 et 12, d'abord considérée comme une duplication segmentaire récente (Yu *et al.*, 2005), avait en fait pour origine la duplication génomique commune à l'ensemble

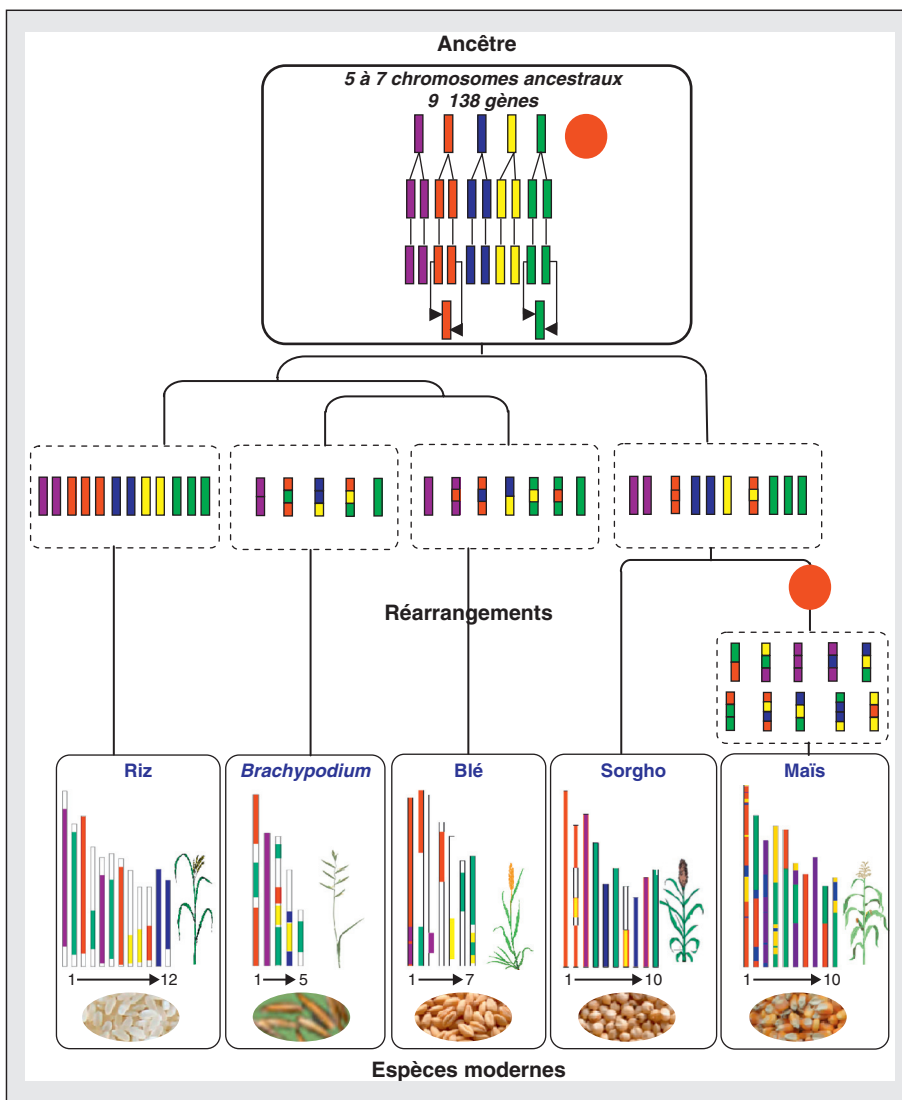


Figure 3. Modèle d'évolution des génomes de graminées à partir d'un ancêtre à 5 chromosomes

Figure 3. Evolution model of cereal genomes from 5 chromosome ancestors.

Les chromosomes sont représentés avec des codes couleur pour illustrer l'évolution des segments à partir d'un ancêtre commun avec 5 protochromosomes. Les événements de duplication majeurs sont représentés par un cercle rouge. Les autres remaniements supposés dans ce scénario sont des duplications et des fusions de chromosomes. La structure des génomes actuels est représentée au bas de la figure, avec les mêmes codes couleurs.

Adapté de Salse (2012).

des céréales, datée à ~ 70 millions d'années (Wang *et al.*, 2011). La comparaison des séquences dans cette région dupliquée chez trois espèces de riz, *O. sativa* (AA), *O. glaberrima* (AA) et *O. brachyantha* (FF), a permis de conclure à l'homogénéisation de la séquence, à la fois des gènes et des séquences non codantes (Jacquemin *et al.*, 2011). L'évolution concertée de cette région semble avoir été un phénomène récurrent dans l'histoire du genre *Oryza* et probablement chez l'ensemble des céréales.

Reséquençage et génotypage de nouvelles variétés

Le terme reséquençage désigne le séquençage complet ou partiel du génome d'individus, de variétés ou d'accessions d'espèces pour lesquelles on dispose déjà d'une séquence de référence. Il vise à détecter la variabilité génétique qui existe chez ces

échantillons sous forme d'insertions/délétions (InDels) de segments chromosomiques ou de différences de nature d'un nucléotide (*single nucleotide polymorphism* ou SNP). Pour le riz, deux stratégies de reséquençage ont été développées. La première a consisté en un reséquençage partiel d'environ 100 Mbp non répétées, par hybridation à l'aide de puces à ADN, l'autre en un reséquençage global par NGST. L'hybridation ne permet de détecter qu'une partie des SNP et ne donne pas accès à la variabilité due aux InDels ou au nombre de copies d'un gène. Elle a été appliquée à une vingtaine de variétés cultivées de type *Indica* ou *Japonica*. Après les étapes de validation, plus de 150 000 SNP fiables ont été découverts et, en moyenne, entre 26 700 et 57 000 SNP permettent de distinguer chaque couple de variétés. Le niveau d'introgresion et les flux de gènes entre ces variétés ont ainsi pu être déterminés avec une haute résolution. À partir de cette étude, des puces de génotypage plus simples, contenant 44 000 oligonucléotides identifiant chacun un SNP, ont été réalisées et ont été utilisées pour de la cartographie génétique par association à l'échelle du génome (Famoso *et al.*, 2011). La combinaison de différentes stratégies de séquençage et un taux de couverture élevé (10-50 \times) révèlent un nombre important de différences autres que les SNP, non seulement entre variétés *Indica* et *Japonica*, mais aussi au sein de chaque groupe (Arai-Kichise *et al.*, 2011). Ainsi, la variété Omachi (*Japonica*) diffère de Nipponbare par $\sim 132\,500$ SNP, 16 500 insertions et 19 300 délétions. La technique de reséquençage direct à faible couverture ($< 2 \times$) a été appliquée au génotypage de lignées recombinantes (RIL) issues du croisement 93-11 \times Nipponbare et du croisement Zhenshan \times Minghui 63 (Xie *et al.*, 2010). Grâce à l'utilisation d'un code barre, les génomes des différentes lignées sont séquencés simultanément en une ou quelques expériences. Environ 1 500 000 SNP fiables ont été détectés et ont permis de cartographier un peu plus de 5 000 points de recombinaison sur le premier croisement. Les 238 lignées du second croisement ont été génotypées en une seule expérience, avec une couverture de 0,055 \times par lignée. À cette résolution, les deux parents

diffèrent par ~ 13 200 SNP et le gène *GW5*, correspondant à un *quantitative trait locus* (QTL) de taille du grain précédemment cloné, a pu être localisé en une seule expérience.

Reséquençage et isolement de gènes d'intérêt

En permettant le génotypage d'un grand nombre d'individus avec une grande résolution, le reséquençage à très haut débit permet de développer une nouvelle approche pour identifier et caractériser des gènes d'intérêt. Il s'agit de la cartographie d'association qui consiste à rechercher quel polymorphisme est associé, dans une population représentative de la variabilité génétique, au caractère recherché. Ainsi, les génomes de 517 variétés représentatives de la variabilité génétique des riz chinois, ont été reséquencés à un taux de couverture de ~ 1 ×. Les variétés ont été analysées en parallèle pour 14 caractères agronomiques (tallage, taille du grain, nombre d'épillets, couleur du grain, date de floraison, teneur en amylose...) et les SNP associés à chaque caractère ont été recherchés. Près de 3,6 millions de SNP ont été identifiés et, pour une majorité de caractères, il a été possible d'en identifier quelques-uns, associés à un ou quelques gènes qui contrôlent le caractère étudié. Cette stratégie est validée par le fait que plusieurs des gènes ainsi identifiés avaient précédemment été isolés par l'approche classique du clonage positionnel. Plus récemment, ce sont les génomes de 950 lignées, représentatives de la diversité mondiale qui ont été reséquencés et analysés pour la date de floraison et la taille du grain (Huang *et al.*, 2011). Trente-deux (32) nouveaux loci pour la date de floraison et 10 pour la taille du grain ont ainsi été identifiés, illustrant l'extraordinaire potentiel de cette approche. Enfin, le reséquençage des génomes de plus de 1 500 accessions de riz cultivé ou sauvage a permis de proposer de nouvelles hypothèses sur la domestication du riz (Huang *et al.*, 2012). De façon générale, le reséquençage devient une approche rapide et d'un

coût raisonnable pour détecter des différences entre lignées proches, ce qui est souvent le cas du matériel utilisé par les sélectionneurs, dans lequel le polymorphisme des fragments de restriction ou des microsatellites est limité. À l'opposé, des régions quasi invariantes ont été recherchées en reséquençant 66 lignées (22 *Indica*, 22 *Japonica* et 22 *O. rufipogon*). Elles correspondent souvent à des régions contenant un ou des gènes liés à la domestication. Treize gènes candidats nouveaux ont été ainsi révélés (He *et al.*, 2011).

Une dernière application a consisté à analyser les causes de la variabilité somatique qui apparaît au cours de la culture de tissus ou de la production de plantes transgéniques. Les génomes de plusieurs lignées des collections de mutants d'insertion (Krishnan *et al.*, 2009) ont été séquencés et un nombre important de SNP ou de nouvelles insertions d'éléments transposables ont été mis en évidence (Sabot *et al.*, 2011 ; Miyao *et al.*, 2012), rendant ainsi compte de la variabilité observée.

Bénéfices pour l'amélioration des plantes

La génomique du riz a déjà produit un grand nombre de connaissances et de ressources biologiques utiles à l'amélioration de cette espèce.

Elle a permis de dresser un premier catalogue des gènes du riz, mais le plus souvent, il reste encore à en découvrir et valider la fonction. Pour cela, il faut cloner le gène, lui associer un phénotype, ainsi qu'une fonction biochimique. Plus de 600 gènes, dont certains ont une importance majeure, ont ainsi déjà été validés (Jiang *et al.*, 2012) et quelques exemples sont présentés (*tableau 1*) : ce sont, pour l'essentiel, des gènes contrôlant l'architecture de la plante, la date de floraison, le poids des grains, l'égrenage, la fertilité, des qualités culinaires, la tolérance aux pathogènes, la tolérance aux stress abiotiques, et l'utilisation de l'azote. Des allèles favorables ont été repérés, même si l'inventaire de leur diversité ne fait que

commencer, offrant déjà une première base de travail aux sélectionneurs. Deux grandes stratégies s'offrent aux sélectionneurs pour utiliser ces informations : repérer dans les ressources génétiques les variétés présentant les allèles favorables et les introduire par croisement dans les meilleures variétés actuelles, en se servant de la sélection assistée par marqueurs (SAM), ou bien introduire directement l'allèle intéressant par transgénèse. L'objectif est de réaliser un « super riz » qui posséderait des résistances multiples aux insectes et aux maladies, qui serait tolérant à la sécheresse et aurait une bonne efficacité d'utilisation de l'azote et de l'eau. Une deuxième retombée importante est la simplification et l'amélioration de la précision du génotypage. L'utilisation du reséquençage en clonage positionnel et en cartographie d'association accélère considérablement l'isolement et la caractérisation des gènes importants. Il est illustré de façon plus ciblée par la recherche systématique de polymorphismes liés à la production de biomasse (Jahn *et al.*, 2011). Dans un avenir très proche, ces technologies devraient se généraliser aux nombreuses lignées RIL, de substitution de segment de chromosome (CSSL) ou quasi isogéniques (NIL), qui ont été créées pour cartographier des QTL d'intérêt (Miura *et al.*, 2011 ; Bocco *et al.*, 2012) et permettre la caractérisation d'un plus grand nombre de gènes importants pour l'agriculture.

Un troisième produit de la génomique du riz est la fourniture d'informations pour d'autres espèces aux génomes moins bien connus à travers la génomique comparative. Dans ce contexte, aussi intitulé génomique translationnelle, les informations obtenues chez le riz (par exemple la structure ainsi que la fonction des gènes) peuvent être transférées aux autres espèces proches dont le génome complexe n'est pas encore correctement séquencé. Des exemples multiples existent dans la littérature sur l'utilisation de la séquence génomique du riz pour développer des marqueurs moléculaires ou isoler des gènes candidats permettant la dissection de caractères agronomiques quantitatifs chez le sorgho, l'orge ou le maïs avant que leurs génomes ne soient séquencés (Cooke *et al.*, 2007). Aujourd'hui, la génomique translationnelle est

Tableau 1. Quelques gènes importants caractérisés chez le riz.

Table 1. Some important genes characterized in rice.

Fonction contrôlée	Noms des gènes	Références
<i>Architecture</i>	<i>MOC1, DWARF10, DWARF27, D3, OsTB1, HTD1, OsSPL14</i>	Jiang <i>et al.</i> , 2012
<i>Date de floraison</i>	<i>OsG1, Hd1, Hd3a, RFT1, RCN1, Ehd1, Ehd2, SE5, PHYB, ETR2, Hd6, Ghd7, OsMADS50, OsMADS51, RFL, OsMADS56, OsMADS14, OsLFL1</i>	Tsuji <i>et al.</i> , 2011
<i>Taille de la panicule</i>	<i>RCN1, RCN2, LAX1, Gn1a, Ghd7, APO1, LOG, RFL, LRK1, EP3, Ghd8, SPA, FZP, ASP1</i>	Jiang <i>et al.</i> , 2012, Yoshida <i>et al.</i> , 2012
<i>Poids des grains</i>	<i>GS3, GW2, GW5, GIF1, RISBZ1, RPF</i>	Jiang <i>et al.</i> , 2012
<i>Égrenage</i>	<i>Sh4, qSH1</i>	Miura <i>et al.</i> , 2011
<i>Fertilité</i>	<i>S5, SaM, SaF, S1</i>	Ouyang <i>et al.</i> , 2010, Guyot <i>et al.</i> , 2011
<i>Arôme</i>	<i>BADH2</i>	Bourgis <i>et al.</i> , 2008
<i>Stress biotiques</i>		
Résistance à pyriculariose	<i>Pib, Pi-ta, Pi-k, Pi9, Pi21, Pi36, Pi37, Pikm, Pi5-1, Pi5-2, Pid3, Pb1, Pi-d2</i>	Jiang <i>et al.</i> , 2012
Résistance à <i>Xanthomonas</i>	<i>Xa1, Xa3, xa5, xa13, Xa21, Xa27</i>	Jiang <i>et al.</i> , 2012
Résistance à cicadelle brune	<i>OsH1-LOX, OsLox1, Bph14</i>	Jiang <i>et al.</i> , 2012
Résistance à virus RYMV	<i>Rymv1</i>	Albar <i>et al.</i> , 2006
<i>Stress abiotiques</i>		
Sécheresse	<i>OsSKIPa, DSM1, DSM2, OsCIPK12, OsGH3.13</i>	Jiang <i>et al.</i> , 2012
Salinité	<i>SKC1, OsNAC6, OsKAT1, OsCIPK15</i>	Jiang <i>et al.</i> , 2012
Sécheresse et salinité	<i>SNAC1, OsbZIP23, DST, AP59, OsSIK1, OsNAC10</i>	Jiang <i>et al.</i> , 2012
Froid	<i>OsCIPK03, qLTG3-1, Ctb1 OsMYB3R-2, MYBS3</i>	Jiang <i>et al.</i> , 2012
Submersion	<i>Sub1A, SNORKEL1, SNORKEL2</i>	Jiang <i>et al.</i> , 2012
<i>Utilisation de l'azote</i>	<i>GS1.1, GS1.2, GlnA, GOGAT, OsAAT1, OsAAT2</i>	Jiang <i>et al.</i> , 2012

Source : ce tableau est adapté de Jiang *et al.* (2012).

notamment utilisée chez le blé. Ainsi, l'utilisation des informations acquises grâce à la séquence des génomes du riz, du maïs et du sorgho, a-t-elle permis de cloner, chez le blé, un gène de glutamate synthase, responsable d'un QTL majeur d'efficacité de l'utilisation de l'azote (Quraishi *et al.*, 2011).

Conclusions et perspectives

Dans cet article, nous avons essayé de présenter les développements les plus récents de la génomique du riz.

Ils démontrent la très grande variabilité du génome du riz d'une variété à l'autre, phénomène que l'on qualifie aussi de fluidité ou de plasticité du génome. Ils ont déjà une influence considérable sur notre connaissance de la biologie du riz et sur les mécanismes physiologiques qui contrôlent nombre de caractères agronomiques importants. Cependant, il est clair que nous n'en sommes qu'au tout début, pour au moins quatre raisons : la technologie va encore évoluer, le nombre de gènes à caractériser est considérable, de nouveaux facteurs de régulation de l'expression des gènes, tels que les petits ARN non codants, commencent à émerger, et enfin nous ne commençons qu'à peine à analyser la diversité génétique.

Au plan de la technologie, la faible longueur des séquences lues par les NGST actuelles est un facteur limitant pour l'assemblage des séquences génomiques. Il devrait être supprimé par les séquenceurs de nouvelle génération. De plus, les coûts devraient encore baisser. Mentionnons aussi que le séquençage permet maintenant d'accéder aux modifications non héréditaires du génome comme les méthylations des résidus cytosine. C'est là une méthode supplémentaire pour analyser les processus épigénétiques dont l'importance dans l'adaptation des plantes à leur environnement est de plus en plus reconnue, mais qui n'a encore fait l'objet que d'études limitées chez le riz (Feng *et al.*, 2010).

Au plan scientifique, les objectifs pour les prochaines années sont clairement définis (Zhang *et al.*, 2008) : identifier la fonction de chacun des gènes du riz

d'ici 2020 en utilisant toutes les ressources disponibles ; recenser et analyser la diversité allélique des gènes utiles pour l'agriculture ; transférer et appliquer les découvertes de la génomique à l'amélioration du riz. Il est d'autre part évident que les avancées réalisées sur le riz ne vont pas se limiter à cette espèce, mais vont aussi bénéficier à l'amélioration des autres céréales. Réciproquement, on peut également espérer que les avancées sur les autres céréales bénéficieront également à la recherche sur le riz. Cet objectif de transfert est certainement le plus ambitieux, mais il n'est pas irréaliste au vu des variétés nouvelles actuellement expérimentées, que ce soit pour la tolérance à la submersion, pour la tolérance aux pathogènes, ou encore pour l'utilisation de l'azote. Le succès passera nécessairement par une formation accrue des sélectionneurs aux nouveaux outils et une sensibilisation plus grande des biologistes aux problèmes agronomiques. ■

Références

Albar L, Bangratz-Reyser M, Hébrard E, Ndjioudjop MN, Jones M, Ghesquière A, 2006. Mutations in the *elF(iso)4G* translation initiation factor confer high resistance of rice to Rice yellow mottle virus. *Plant Journal* 47 : 417-26. doi: 10.1111/j.1365-3113x2006.02792.x.

Arai-Kichise Y, Shiwa Y, Nagasaki H, Ebana K, Yoshikawa H, Yano M, *et al.*, 2011. Discovery of genome-wide DNA polymorphisms in a landrace cultivar of *Japonica* by whole genome sequencing. *Plant Cell Physiology* 52 : 274-82. doi: 10.1093/pcp/pcr003.

Bocco R, Lorieux M, Seck PA, Futakuchi K, Manneh B, Baimey H, *et al.*, 2012. Agro-morphological characterization of a population of introgression lines derived from crosses between IR 64 (*Oryza sativa indica*) and TOG 5681 (*Oryza glaberrima*) for drought tolerance. *Plant Science* 183 : 65-76. doi: 10.1016/j.plantsci.2011.09.010.

Bourgis F, Guyot R, Gherbi H, Tailliez E, Amabile I, Salse J, *et al.*, 2008. Characterization of the major fragrance gene from an aromatic *Japonica* rice and analysis of its diversity in Asian cultivated rice. *Theoretical and Applied Genetics* 117 : 353-68. doi: 10.1007/s00122-008-0780-9.

Childs KL, Davidson RM, Buell CR, 2011. Gene coexpression network analysis as a source of functional annotation for rice genes. *PLoS One* 6 : e22196. doi: 10.1371/journal.pone.0022196.

Cooke R, Piegu B, Panaud O, Guyot R, Salse J, Feuillet C, *et al.*, 2007. From rice to other cereals: comparative genomics. In : Upadhyaya N, éd. *Rice functional genomics*. New York : Springer. doi: 10.1007/0-387-48914-2.

Delseny M, Han B, Hsing YI, 2010. High throughput DNA sequencing: the new sequencing revolution. *Plant Science* 179 : 407-22. doi: 10.1016/j.plantsci.2010.07.019.

Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, Bustamante C, *et al.*, 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genetics* 7 : e1002221. doi: 10.1371/journal.pgen.1002221.

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, *et al.*, 2010. Conservation and divergence of methylation patterning in plants and animals. *Proceedings National Academy Sciences USA* 107 : 8689-94. doi: 10.1073/pnas.1002720107.

Goicoechea JL, Ammiraju JSS, Marri PR, Chen M, Jackson S, Yu Y, *et al.*, 2010. The future of rice genomics: sequencing the collective *Oryza* genome. *Rice* 3 : 89-97. doi: 10.1007/s12284-010-9052-9.

Guyot R, Garavito A, Gavory F, Samain S, Tohme J, Ghesquière A, *et al.*, 2011. Patterns of sequence divergence and evolution of the S orthologous regions between Asian and African cultivated rice species. *PLoS One* 10 : e17726. doi: 10.1371/journal.pone.0017726.

He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, *et al.*, 2011. Two evolutionary histories in the genome of rice: the role of domestication genes. *PLoS Genetics* 7 : e1002100. doi: 10.1371/journal.pgen.1002100.

Helmy M, Tomita M, Ishihama Y, 2011. OryzaPG-DB: rice proteome database based on shotgun proteogenomics. *BMC Plant Biology* 11 : 63. doi: 10.1186/1471-2229-11-63.

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, *et al.*, 2011. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44 : 32-9. doi: 10.1038/ng.1018.

Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, *et al.*, 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490 : 497-501. doi: 10.1038/nature11532.

International Rice Genome Sequencing Project, 2005. The map-based sequence of the rice genome. *Nature* 436 : 793-800. doi: 10.1038/nature03895.

Jacquemin J, Chaparro C, Laudé M, Berger A, Gavory F, Goicoechea JL, *et al.*, 2011. Long-range and targeted ectopic recombination between the two homeologous chromosomes 11 and 12 in *Oryza* species. *Molecular Biology and Evolution* 28 : 3139-50. doi: 10.1093/molbev/msr144.

Jahn CE, McKay JK, Mauleon R, Stephens J, McNally KL, Bush DR, *et al.*, 2011. Genetic variation in biomass traits among 20 diverse rice varieties. *Plant Physiology* 155 : 157-68. doi: 10.1104/pp.110.165654.

Jiang Y, Cai Z, Xie W, Long T, Yu H, Zhang Q, 2012. Rice functional genomic research: progress and implications for crop genetic improvement. *Biotechnology Advances* 30 : 1059-70. doi: 10.1016/j.biotechadv.2011.08.013.

Krishnan A, Guiderdoni E, An G, Hsing YI, Han CD, Lee MC, *et al.*, 2009. Mutant resources in rice for functional genomics of the grasses. *Plant Physiology* 149 : 165-70. doi: 10.1104/pp.108.128918.

Lu T, Lu G, Fan D, Zhu C, Li W, Zhao Q, *et al.*, 2010. Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Research* 20 : 1238-49. doi: 10.1101/gr.106120.110.

Lu F, Ammiraju JSS, Sanyal A, Zhang S, Song R, Chen J, *et al.*, 2009. Comparative sequence

- analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proceedings National Academy Sciences USA* 106 : 2071-6. doi: 10.1073/pnas.0812798106.
- Miyao A, Nakagome M, Ohnuma T, Yamagata H, Kanamori H, Katayose Y, *et al.*, 2012. Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiology* 53 : 256-64. doi: 10.1093/pcp/pcr172.
- Miura K, Ashikari M, Matsuoka M, 2011. The role of QTL in the breeding of high-yielding rice. *Trends in Plant Science* 16 : 319-26. doi: 10.1016/j.tplants.2011.02.009.
- Moore G, Devos KM, Wang Z, Gale MD, 1995. Cereal genome evolution. Grasses, line up and form a circle. *Current Biology* 5 : 737-9. doi: 10.1016/S0960-9822(95)00148-5.
- Ouyang Y, Liu YG, Zhang Q, 2010. Hybrid sterility in plant: stories from rice. *Current Opinion in Plant Biology* 13 : 186-92. doi: 10.1016/j.pbi.2010.01.002.
- Quraishi UM, Abrouk M, Murat F, Pont C, Foucrier S, Desmaizieres G, *et al.*, 2011. Cross-genome map based dissection of a nitrogene use efficiency ortho-meta QTL in breadwheat unravels concerted cereal genome evolution. *Plant Journal* 65 : 745-56. doi: 10.1111/j.1365-313X.2010.04461.x.
- Sabot F, Picault N, El-Baidouri M, Llauro C, Chaparro C, Piegue B, *et al.*, 2011. Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *Plant Journal* 66 : 241-6. doi: 10.1111/j.1365-313X.2011.04492.x.
- Sakurai T, Kondou Y, Akiyama K, Kurotani A, Higuchi M, Ichikawa T, *et al.*, 2011. RiceFOX: a database of Arabidopsis mutant lines over-expressing rice full-length cDNA that contains a wide range of trait information to facilitate analysis of gene function. *Plant Cell Physiology* 52 : 265-73. doi: 10.1093/pcp/pcq190.
- Salse J, Bolot S, Troude M, Jouffe V, Piegue B, Quraishi UM, *et al.*, 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell* 20 : 11-24. doi: 10.1105/tpc.107.056309.
- Salse J, 2012. *In silico* archeogenomics unveils modern plant genome organisation, regulation and evolution. *Current Opinion in Plant Biology* 15 : 1-9. doi: 10.1016/j.pbi.2012.01.001.
- Salse J, Feuillet C, 2011. Paleogenomics in cereals: modeling of ancestors for modern species improvement. *Comptes Rendus Biologies* 334 : 205-11. doi: 10.1016/j.crv.2010.12.014.
- Sapkota A, Liu X, Zhao XM, Cao Y, Liu J, Liu ZP, *et al.*, 2011. DIPOS: a database of interacting proteins in *Oryza sativa*. *Molecular BioSystems* 7 : 2615-21. doi: 10.1039/c1mb05120b.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, *et al.*, 2009. The B73 maize genome: complexity, diversity and dynamics. *Science* 326 : 1112-5. doi: 10.1126/science.1178534.
- Schneeberger K, Weigel D, 2011. Fast-forward genetics enabled by new sequencing technologies. *Trends in Plant Science* 16 : 282-8. doi: 10.1016/j.tplants.2011.02.006.
- Tsuji H, Taoka K, Shimamoto K, 2011. Regulation of flowering in rice: two florigene genes, a complex gene network and natural variation. *Current Opinion in Plant Biology* 14 : 45-52. doi: 10.1016/j.pbi.2013.01.005.
- Wang X, Tang H, Paterson AH, 2011. Seventy million years of concerted evolution of a homeologous chromosome pair in parallel, in major Poaceae lineage. *The Plant Cell* 23 : 27-37. doi: 10.1105/tpc.110.080622.
- Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, *et al.*, 2010. Parent-independent genotyping for constructing a ultrahigh-density linkage map based on population sequencing. *Proceedings National Academy Sciences USA* 107 : 10578-83. doi: 10.1073/pnas.100593110.
- Yoshida A, Ohmori Y, Kitano H, Taguchi-Shiobara F, Hirano HY, 2012. *ABERRANT SPIKELET AND PANICLE1* encoding a TOPLESS-related transcriptional corepressor is involved in regulation of meristem fate in rice. *Plant Journal* 70 : 327-39. doi: 10.1111/j.1365-313X.2011.04872.x.
- Yu J, Wang J, Lin W, Li S, Li H, Zhou J, *et al.*, 2005. The genomes of *Oryza sativa*: a history of duplications. *PloS Biology* 3 : e38. doi: 10.1371/journal.pbio.0030038.
- Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, *et al.*, 2010. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Research* 20 : 646-54. doi: 10.1101/gr.100677.109.
- Zhang Q, Li J, Xue Y, Han B, Deng XW, 2008. Rice 2020: A call for an international coordinated effort in rice functional genomics. *Molecular Plant* 1 : 715-9. doi: 10.1093/mp/ssn043.
- Zuccolo A, Sebastian A, Talag J, Yu Y, Kim H, Collura K, *et al.*, 2007. Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. *BMC Evolutionary Biology* 7 : 152. doi: 10.1186/1471-2148-7-152.