Le séquençage des génomes de plantes : vers une nouvelle révolution en biologie végétale

Michel Delseny

Laboratoire Génome et développement des plantes (LGDP)
UMR 5096
CNRS, IRD
Université de Perpignan
66860 Perpignan
France
<delseny@univ–perp.fr>

Résumé

L'accélération récente des programmes de séquençage des génomes végétaux permet maintenant l'essor de la génomique comparative, dont les buts sont de faciliter l'isolement de gènes d'intérêt chez les différentes espèces et de comprendre les processus d'évolution et de domestication des génomes. Le séquençage révèle ainsi des phases d'envahissement des génomes par des éléments transposables qui contribuent à modifier les profils d'expression des gènes et quelques fois à les inactiver. Quelques aspects des apports de la génomique comparative sont illustrés. Finalement les perspectives ouvertes par les nouvelles méthodologies de séquençage à très haut débit sont discutées.

Mots clés : génomique végétale ; séquençage.

Thèmes : métabolisme ; productions végétales ; ressources naturelles et environnement.

Abstract

Sequencing plant genomes: towards a revolution in plant biology

Recent speeding up in the sequencing of plant genomes has resulted in the development of comparative genomics. Applying this strategy makes it possible to both facilitate and speed up the cloning and characterisation of genes of interest in many crop species and to better understand evolution and domestication processes. Comparative sequencing has also revealed that genomes have been repeatedly invaded by transposable elements which contribute to modifying gene expression and genome size. The power of comparative genomics in isolating new genes and in revealing genetic diversity is illustrated. Finally the perspectives and challenges opened up by new very high throughput sequencing methods are discussed.

Key words: plant genomics; sequencing.

Themes: métabolism; natural resources and environment; vegetal productions.

La détermination de la séquence complète des génomes *d'Arabidopsis thaliana* et de riz, de celles encore partielles des génomes de peuplier, de vigne, du papayer et du sorgho, ainsi que de quelques autres espèces, et la disponibilité de larges collections d'EST[1] marquent une étape importante dans notre connaissance des génomes de plantes. Ces données, sur quelques espèces considérées comme modèles, ont déjà considérablement accru nos connaissances de la biologie de ces plantes et ont déjà des implications importantes en sélection. L'auteur a résumé dans un article quelques-uns des acquis les plus récents (Delseny, 2009).

Cette première révolution est en passe d'être achevée, et l'on assiste maintenant à une accélération considérable des recherches sur d'autres espèces, d'intérêt agronomique ou évolutif, avec toutes les conséquences qui en découlent au plan des outils et de la connaissance de la biologie de ces espèces ou de leur amélioration. Enfin des nouvelles méthodes de

[1] *Expressed Sequenced Tag*, séquence partielle de gènes exprimés à partir d'ADN copies des ARN messagers.

Tirés à part : M. Delseny

doi: 10.1684/agr.2009.0342

séquençage, beaucoup plus efficaces en termes de débit et beaucoup plus abordables en termes de coût, commencent à se vulgariser et permettent d'envisager des projets qui auraient paru insensés il y a seulement 3-4 ans. Le *tableau 1* donne un aperçu des génomes en cours de séquençage et il en existe une cinquantaine d'autres pour lesquels des projets sont en attente d'un financement pour démarrer.

Dans cette présentation, nous abordons les perspectives ouvertes par l'amélioration des technologies de séquençage et leur application croissante à l'ensemble des plantes cultivées.

La génomique comparative

La génomique comparative constitue une nouvelle branche de la génomique en plein essor depuis que l'on peut comparer des génomes à l'aide de marqueurs communs, et surtout directement au niveau des séquences nucléotidiques. Cette évolution va encore s'accélérer avec l'accumulation de séquences provenant des organismes les plus divers.

Les premiers apports de la génomique comparative, décrits dans l'article précédent, concernent l'aide à l'annotation, l'observation de régions synténiques[2] entre chromosomes d'espèces différentes, ou l'observation de duplications globales de larges segments chromosomiques. Plusieurs autres aspects de la génomique comparative méritent d'être présentés.

Tableau 1. Génomes séquencés ou en cours de séquençage d'après *Entrez Genome Project* (NCBI).

Table 1. Sequenced genes and sequencing underway accroding to the Entrez Genome Project (NCBI).

Espèce	Taille	État
Arabidopsis thaliana Col 0	120 Mbp	Achevé, publié
Arabidopsis thaliana Landsberg		Brouillon, publié
Oryza sativa (Nipponbare)	390 Mbp	Achevé, publié
Oryza sativa (93-11)	400 Mbp	Achevé, publié
Populus trichocarpa	485 Mbp	Brouillon assemblé, publié
Vitis vinifera	500 Mbp	Brouillon assemblé, publié
Carica papaya	370 Mbp	Brouillon assemblé, publié
Zea mays	2 300 Mbp	Brouillon assemblé
Sorghum bicolor	730 Mbp	Brouillon assemblé, publié
Brachypodium distachyon	320 Mbp	Brouillon assemblé
Lotus japonicus	470 Mbp	Brouillon assemblé
Ricinus communis	400 Mbp	Brouillon assemblé
Aquilegia formosa	350 Mb	En cours
Arabidopsis lyrata	230 Mbp	En cours
Brassica napus	1 100 Mbp	En cours
Brassica oleracea	600 Mbp	En cours
Brassica rapa	500 Mbp	En cours
Capsella rubella		En cours
Cassava manihot		En cours
Citrus sinensis	380 Mbp	En cours
Eucalyptus globulus	600 Mbp	En cours
Glycine max	1 200 Mbp	En cours
Medicago truncatula		En cours
Nicotiana tabacum	4 500 Mbp	En cours
Panicum virgatum		En cours
Pinus taeda		En cours
Poncirus trifoliata	380 Mbp	En cours
Solanum bulbocastanum		En cours
Solanum demissum		En cours
Solanum lycopersicum	950 Mbp	En cours
Solanum tuberosum	840 Mbp	En cours
Triticum aestivum	16 000 Mbp	En cours

Mbp : million de paires de base.

Variabilité et fluidité des génomes

La taille des génomes d'espèces voisines appartenant à un même genre peut varier dans des proportions importantes. La disponibilité de séquences, même partielles, permet de déterminer les causes de cette variation. Les premiers loci séquencés de façon comparative entre riz, maïs et sorgho suggèrent que l'ordre des gènes est à peu près conservé – avec des exceptions correspondant à des ruptures de synténie –, mais que les distances intergéniques peuvent varier de façon importante du fait de l'insertion d'éléments transposables (Cooke *et al.*, 2007).

Une illustration spectaculaire de ce phénomène au niveau non plus d'un locus, mais d'un génome entier, vient de l'analyse du génome des riz sauvage et cultivé. Ainsi le génome de l'espèce cultivée *Oryza sativa* est deux fois plus petit que celui de l'espèce sauvage *Oryza australiensis*. Dans le cadre du projet OMAP[3], des banques BAC[4] ont été réalisées pour une douzaine d'espèces de riz, ordonnées en cartes physiques, et les clones ont été séquencés à chacune des extrémités (Ammiraju *et al.*, 2006). Leur analyse a permis de reconstituer chez *O. australiensis* les séquences de trois rétroéléments, *RIRE1*, *Wallaby* et *Kangourou*. Ces trois éléments rendent compte de près de 60 % du génome d'*O. australiensis* (Piegu *et al.*, 2006). Par ailleurs, l'alignement des deux cartes physiques d'*O.sativa* et d'*O. australiensis* à l'aide de séquences non répétées montre globalement une très bonne conservation de l'ordre des gènes. La plus grande taille du génome *d'O. australiensis* résulte donc essentiellement de l'envahissement du génome par ces trois rétroéléments. Des résultats similaires ont été obtenus sur *O. granulata* (Ammiraju *et al.*, 2007).

[2] Organisations identiques des gènes sur une portion de chromosome pour des espèces apparentées.

[3] OMAP : *Oryza Map Alignment Project*. Voir http://www.omap.org/

[4] BAC : *Bacterial Artificial Chromosome*.

Le locus *Bronze* (*Bz*) du maïs, localisé sur le chromosome 9 a été séquencé sur deux variétés, McC et B73, ainsi que chez le riz. Ces séquences révèlent, d'une part, une rupture de la synténie entre les deux espèces, les gènes voisins de *Bz* étant différents, mais aussi des différences importantes entre les deux variétés, certains gènes étant présents dans l'une, d'autres non (Fu et Dooner, 2002). Ce type d'observation a été étendu en séquençant 2,8 Mbp sur deux autres lignées de maïs, B 73 et Mol 17 : cela confirme que plus de 50 % de la séquence n'est pas colinéaire entre les deux variétés et que les deux tiers des gènes inventoriés sont absents de cette région chez l'une ou l'autre des variétés (Brunner *et al.*, 2005). Le séquençage récent des génomes du sorgho (Paterson *et al.*, 2009), du maïs et de *Brachypodium* (*tableau 1*) devrait permettre de généraliser ces observations.

La généralisation des méthodes à haut débit et coût réduit permet maintenant le re-séquençage de différentes variétés. Ces méthodes sont déjà mises en oeuvre sur différents écotypes d'*Arabidopsis* et différentes variétés du riz : elles permettent d'avoir une vision très précise des mutations ponctuelles (SNP, *Single Nucleotide Polymorphism*), ainsi que des insertions/délétions et des remaniements chromosomiques.

Aide à la cartographie physique et au clonage de gènes d'intérêt

Du fait de la relativement bonne conservation de l'ordre des gènes le long des chromosomes des espèces d'une même famille botanique, il est en principe possible de s'appuyer sur l'espèce modèle dont le génome est disponible pour construire des cartes génétiques et physiques d'espèces apparentées et éventuellement de réaliser plus facilement des clonages positionnels. Cette stratégie se révèle d'autant plus efficace que les espèces sont plus proches, mais devient problématique dès que l'on change de famille botanique. Dès que les séquences d'*Arabidopsis* ont été disponibles, elles ont été utilisées pour dresser des cartes génétiques d'espèces voisines comme le chou ou le colza (Lan *et al.*, 2000 ; Babula *et al.*, 2003 ; Parkin *et al.*, 2005 ; Qiu *et al.*, 2006).

L'aide au clonage positionnel, en fournissant des marqueurs et une trame de carte, est bien réelle, comme l'illustre le *tableau 2* (Cooke *et al.*, 2007) : sur les 25 premiers gènes isolés par clonage positionnel sur des graminées autres que le riz, 17 ont pu l'être grâce à l'utilisation de la synténie avec le riz. Dans 4 cas, une rupture de la synténie a été observée, principalement dans le cas de gènes de résistance à des pathogènes dont l'évolution est plus rapide que celle des autres gènes. Les 4 gènes restants, pour lesquels la synténie n'a pas été utilisée, correspondent à des gènes isolés par étiquetage par un transposon.

Tableau 2. Utilisation de la synténie et de l'orthologie avec le riz pour le clonage de gènes importants chez d'autres graminées.

Table 2. Using synteny and orthology with rice to clone genes important to other graminates.

Gène	Espèce	Fonction	Orthologie riz
Lr10	*T. aestivum*	Résistance maladie	Non
Pm3	*T. aestivum*	Résistance maladie	Non
Vrn1	*T. monococcum*	Vernalisation	Oui, # 3
Vrn2	*T. monococcum*	Vernalisation	Oui, # 3
Q	*T. aestivum*	Egrenage	Non utilisée
Lr21	*Aegilops tauschii*	Résistance maladie	Non
Ph1	*T. aestivum*	Appariement chromosomique	Oui, # 9 (plus Brachypodium)
Ph2	*T. aestivum*	Appariement chromosomique	Oui, # 1
Ppd-H1	*H. vulgare*	Réponse photopériode	Oui, # 7
Ror1	*H. vulgare*	Résistance maladie	Non utilisée
Ror2	*H. vulgare*	Résistance maladie	Oui, # 3
Rpg1	*H. vulgare*	Résistance maladie	Oui, # 6
Mlle	*H. vulgare*	Résistance maladie	Non utilisée
Rym4/Rym5	*H. vulgare*	Résistance virus	Non utilisée
Rar1	*H. vulgare*	Résistance maladie	Non
Rht1	*T. aestivum*	Nanisme	Oui, # 3
Tga1	*Zea mays*	Architecture épi	Oui, # 8
Vgt1	*Zea mays*	Transition phase vegetative vers reproductrice	Oui, # 8
Ra2	*Zea mays*	Architecture épi	Oui, # 1
Ra3	*Zea mays*	Architecture épi	Oui, # 2
Ba1	*Zea mays*	Architecture épi	Oui, # 11
Th	*Zea mays*	Architecture épi	Oui, # 6
Ts4	*Zea mays*	Architecture épi	Oui, # 12
Bru	*Saccharum officinarum*	Résistance maladie	Oui, # 2
ASGR	*Pennisetum squamulatum*	Aposporie	Oui, # 11

« Oui » indique que la synténie a été utilisée et que l'on a isolé le gène orthologue à celui du riz (le chromosome du riz est indiqué) ; « non » signifie que l'on a essayé d'utiliser la synténie, mais que le gène orthologue n'existe pas ou n'est pas à la position attendue ; « non utilisée » indique que la synténie n'a pas été utilisée. Les références concernant ces informations se trouvent dans Cooke *et al.*, 2007.

La synténie a été utilisée massivement pour aligner les cartes physiques des différents génomes du riz (Ammiraju *et al.*, 2006 ; Ammiraju *et al.*, 2007) et, plus récemment, pour construire la carte physique du chromosome 3B du blé (Paux *et al.*, 2008).

Évolution de grandes familles de gènes

Les EST et le séquençage génomique fournissent des catalogues de gènes qui se ressemblent dans les différentes familles botaniques. La comparaison de deux génomes révèle en principe ce qui est spécifique à chacun. On est maintenant convaincu que, bien que très largement partagé, sans doute à plus de 80 % entre les différentes espèces, le répertoire des

gènes présente des spécificités. Certains gènes présents chez *Arabidopsis* sont absents chez le riz, comme par exemple plusieurs gènes qui fonctionnent dans le déterminisme de la mise à fleur. Il existe aussi des gènes présents chez le maïs, mais absents chez le riz, comme *Ramosa 1* qui contrôle l'architecture de l'inflorescence et de l'épi chez le maïs. Il en va de même du locus *Ha* qui contrôle la dureté du grain chez l'orge et le blé, mais qui n'existe pas chez le riz (Chantret *et al.*, 2005). L'analyse récente du génome du sorgho (Paterson *et al.*, 2009) révèle que 24 % des gènes sont spécifiques des graminées et 7 % sont spécifiques du sorgho. Ces spécificités sont sans doute importantes, car c'est en partie sur elles que reposent les capacités d'adaptation et les qualités propres à chaque espèce.

Deux gènes homologues dans deux espèces qui dérivent par spéciation à partir d'un ancêtre commun sont nommés orthologues. Lorsque plusieurs gènes dans une même espèce sont similaires et ont dérivé à partir d'un gène ancestral par duplication (WGD[5] ou tandem) au sein de l'espèce, on dit qu'ils sont paralogues. Il est donc possible de comparer les gènes qui, dans une même espèce, se ressemblent, avec ceux qui leur ressemblent dans une autre espèce et tenter de comprendre quelle est leur généalogie, et comment leurs fonctions ont été conservées ou se sont différenciées au cours de l'évolution. Le *tableau 3* illustre une comparaison de quelques grandes familles entre *Arabidopsis* et riz. Elle montre que chez chacune des deux espèces, des familles de gènes ont connu une amplification importante, mais que ce ne sont pas toujours les mêmes familles qui ont connu une expansion. Ainsi, on observe beaucoup plus de gènes chez le riz pour des familles telles que les RLK (*Receptor-like kinases*), ou les NBS-LRR[6] (protéines à répétitions riches en leucine et à domaine de liaison aux nucléotides), classiquement associées à la perception de signaux et à la défense contre les pathogènes, ou encore les cytochromes P 450 qui interviennent dans de nombreuses réactions du métabolisme secondaire. À l'opposé, les gènes pour les facteurs de transcription à *MADS box*[7], impliqués dans les processus de développement ou les facteurs de transcription de type MYB[8], sont plus abondants chez *Arabidopsis*. Un autre exemple est constitué par les facteurs CBF[9] considérés comme des régulateurs clés de la réponse à la sécheresse et au froid : il existe 6 gènes chez *Arabidopsis*, 10 chez le riz et au moins 20 chez l'orge. La correspondance d'orthologie n'est donc pas toujours évidente à établir et il est clair que, chez chacune des espèces, certains gènes se sont spécialisés. Ainsi, on observe deux grandes classes de gènes de résistance aux pathogènes à domaine LRR chez *Arabidopsis* mais une seule chez le riz et les graminées. Ces dernières ne présentent que la classe des protéines NBS-LRR et sont dépourvues de LRR[10] à motif TIR (*Toll Interleukine Receptor*).

L'analyse phylogénétique a maintenant été réalisée dans la plupart des grandes familles et l'on observe différents clades. Par exemple, l'analyse des facteurs de transcription de type GRAS[11], impliqués dans une variété de processus développementaux montre l'existence de 6 sous-familles. La classe I comprend 6 gènes clairement orthologues dans chaque espèce. La classe II présente 2 gènes distincts chez le riz, mais 7 chez *Arabidopsis* : 2 correspondent au premier gène du riz et les 5 autres au second gène. La classe III présente la situation réciproque avec 5 gènes distincts chez *Arabidopsis* auxquels correspondent pour chacun 2 gènes chez le riz. La classe IV comprend deux sous-familles de paralogues chez chaque espèce sans que l'on puisse clairement déterminer leur filiation. Les classes V et VI sont constituées respectivement de 26 gènes spécifiques au riz et de 4 gènes spécifiques à *Arabidopsis* (Tian *et al.*, 2004). Le même type d'observation prévaut également pour les autres familles de facteurs de transcription analysées (Cooke *et al.*, 2007), illustrant la diversité propre à chaque espèce. Ce type d'analyse va s'étoffer au fur et à mesure de l'annotation des génomes disponibles. L'analyse des taux de substitutions synonymes ou non synonymes permettra de dater les événements de duplication et de déterminer si ces groupes de gènes ont fait l'objet d'une sélection positive ou neutre.

Tableau 3. Quelques exemples de familles multigéniques.

Table 3. Examples of multigenetic families.

Famille protéique	*Arabidopsis*	Riz
RLK	> 600	> 1 100
NBS-LRR	128	> 800
Cyt P 450	272	455
AP2/ERF TF	146	161
CBF/DREB TF	6	10
WRKY	72	100
NAC TF	105	149
GRAS TF	32	57
bZIP TF	76	94
ARF TF	23	29
MAPKK	10	16
Cyclines	50	44
MAPK	20	15
MADS TF	106	77
DOF TF	36	30
MYB TF	130	85

TF indique que la famille protéique correspond à des facteurs de transcription. Les références concernant ces informations se trouvent dans Cooke *et al.*, 2007.

Séquençage massivement parallèle et nouvelle révolution du séquençage

De nouvelles méthodes

Pratiquement tout ce qui a été exposé précédemment a été acquis avec les méthodes classiques de séquençage à partir de fragments clonés, utilisant l'électrophorèse capillaire et analysant simultanément un nombre limité d'échantillons. Depuis environ trois ans, des nouvelles méthodes, qui permettent d'éviter le clonage, apparaissent et constituent une deuxième révolution.

[5] WDG : *Whole Genome Duplication.*
[6] *Nuclear Binding Site-Leucine Rich Repeat.*
[7] Acronyme pour des facteurs de transcription initialement découverts chez la levure mais présents chez tous les organismes, dont les séquences sont assez conservées, impliquées dans les processus de développement notamment la floraison pour les plantes.
[8] Présents chez tous les organismes, impliqués dans le contrôle du cycle cellulaire.
[9] *Cold Binding Factors*
[10] LRR : *Leucine Rich Repeat.*
[11] GRAS : acronyme constitué des initiales des quatre gènes fondateur de la famille chez Arabidopsis.

Tableau 4. Les différentes méthodes de séquençage.

Table 4. Methods of sequencing.

	Séquençage capillaire (ABI)	Roche 454	*Illumina*	SOLiD
Stratégie	ADN polymérase	Pyroséquençage	ADN polymérase	Ligation
Amplification PCR	aucune	Émulsion-PCR	Amplification de ponts	Émulsion PCR
Mbp/lecture	0,08 Mbp	100 Mbp	1300 Mbp	3000 Mbp
Longueur lue	800 bp	250 bp	32-40 bp	35 bp
Coût/Mbp (dollars US)	20 000	85	6	5,8

Certaines, s'appuyant sur la connaissance préalable d'un génome de référence, utilisent des techniques d'hybridation sur des puces à ADN pour reséquencer le génome d'une autre variété ou le génome d'une espèce très voisine. Cette méthode, développée en particulier par la société Perlgen, a été récemment utilisée pour reséquencer plusieurs variétés de riz (McNally *et al.*, 2006). Les séquences ainsi obtenues sont incomplètes, mais sont suffisantes pour repérer les mutations (principalement SNP[12]) qui distinguent les différentes variétés.

Les autres méthodes sont les méthodes dites de séquençage massivement parallèle dans lesquelles plusieurs dizaines de milliers de molécules sont immobilisées sur une lame de verre, amplifiées par PCR[13] et séquencées en parallèle en utilisant différentes réactions chimiques (*tableau 4*). Ces méthodes ont le double avantage d'augmenter considérablement le débit du séquençage et d'en abaisser le coût (Mardis, 2007). Elles ont le défaut de ne permettre la lecture que d'une courte séquence, de l'ordre de 250 paires de base (pb) pour la technologie Roche 454, et de l'ordre de 30 à 50 pb pour les autres approches. Il en résulte des difficultés d'assemblage, mais comme la profondeur du séquençage (10 fois ou davantage) est nettement supérieure et moins coûteuse que les technologies classiques, ces techniques gagnent du terrain et sont de plus en plus utilisées pour des génomes aussi complexes que ceux des plantes. Ces technologies ont été testées dans le cadre de projets de reséquençage d'autres écotypes ou variétés d'*Arabidopsis* ou du riz, et sont maintenant utilisées en routine pour « dégrossir » le travail sur les nouveaux génomes en cours de séquençage, en combinaison avec les stratégies classiques (*tableau 2*). Plusieurs clones BAC ordonnés peuvent être séquencés en vrac et faire l'objet d'un assemblage partiel qui fournit le plus souvent plus de 90 % de la séquence. Il devient donc possible de séquencer des nouveaux génomes, au moins au stade de brouillon assemblé, très rapidement et de reséquencer les génomes des variétés d'une espèce pour laquelle on dispose d'une séquence de référence. Il est même possible, en utilisant des systèmes de code-barres oligonucléotidiques, de séquencer simultanément des échantillons correspondant à différents génomes et ainsi de réaliser un génotypage particulièrement précis.

Des applications multiples

Au-delà du fait de faciliter l'obtention d'une nouvelle séquence, l'intérêt majeur est de détecter facilement SNP et In/Del (insertion/délétion) qui sont indispensables pour réaliser les clonages positionnels ou identifier les gènes responsables de traits phénotypiques par génétique d'association en évaluant le déséquilibre de liaison à un locus donné.

L'utilisation de ces méthodes change radicalement le paysage dans le domaine de l'analyse de la biodiversité : Il est désormais possible de séquencer des fragments amplifiés d'un gène donné sur plusieurs centaines d'accessions pour un coût réduit et d'obtenir une vision très précise des haplotypes[14]. Mieux, on peut en avoir une vision globale en séquençant plusieurs génomes proches pour quelques dizaines de milliers de dollars.

L'autre application majeure de ces méthodes est dans le domaine de l'analyse du transcriptome : plusieurs centaines de milliers d'ADNc sont séquencés en une seule expérience et l'on a ainsi une vision quantitative du transcriptome en dénombrant les transcrits identiques et les différents transcrits. Cette approche a révolutionné, entre autre, l'étude des petits ARN (miRNA et siRNA) en permettant leur dénombrement et leur classification (Lu *et al.*, 2006 ; Nobuta *et al.*, 2007). Elle rend également caduque l'élaboration de puces à ADN pour analyser le transcriptome : il est devenu moins cher et plus précis de réaliser un séquençage massivement parallèle d'un transcriptome que d'élaborer une puce.

On est maintenant arrivé à un niveau de fiabilité et de coût tel que l'on peut aussi se demander si les stratégies de génotypage et de clonage positionnel sont encore réalistes et s'il ne vaut pas mieux directement reséquencer un mutant ou une lignée quasi isogénique. La question mérite d'autant plus d'être posée que ces technologies vont encore évoluer, l'ambition étant de ramener à 1 000 dollars US le séquençage d'un génome humain !

Conclusion

Nous avons essayé d'exposer, dans cette présentation et dans l'article précédent, les progrès fulgurants réalisés en matière de séquençage des génomes végétaux au cours des vingt dernières années.

Les données acquises, notamment sur le repérage ou l'identification des gènes d'intérêt, ont déjà un impact important, car la plupart des firmes semencières et des instituts responsables de l'amélioration des plantes dans le monde ont commencé à intégrer les méthodologies de sélection assistée par marqueurs. Les plus avancés utilisent déjà les marqueurs SNP qui sont directement dérivés du séquençage.

L'objectif prioritaire de la génomique végétale reste de continuer à identifier des nouveaux gènes d'intérêt : cela

[12] *Single nucleotide polymorphism*
[13] *Polymerase chain reaction*

[14] Gènes liés sur un même chromosome.

requiert une accélération à la fois du séquençage et de l'analyse fonctionnelle de nouveaux génomes. Ces méthodes permettent aussi d'aller au-delà de la simple identification d'un gène d'intérêt, en analysant son fonctionnement et sa diversité dans les différentes variétés et espèces. Le plus souvent, les caractères phénotypiques ne sont pas des caractères simples monogéniques, mais le résultat du fonctionnement de plusieurs gènes distincts, repérés par des QTL[15], entre lesquels existent des relations d'épistasie. Il est clairement nécessaire de comprendre toute cette complexité au niveau moléculaire et d'en analyser la variabilité. Cela suppose de préserver les ressources actuelles et de créer les populations qui permettront de les caractériser.

Plusieurs boîtes noires persistent encore malgré tous les développements récents. L'une d'elle, cruciale pour le rendement des différentes productions, est la compréhension de l'hétérosis. Au moins deux pistes de recherche sont en cours d'exploration : l'analyse des différents allèles dans les croisements avec l'étude de leur expression et l'analyse des modifications épigénétiques des génomes, qui, au moins chez *Arabidopsis* et le riz, deviennent accessibles à l'expérimentation moléculaire. La détection des séquences méthylées par séquençage direct, celle des modifications structurelles de la chromatine à l'aide de puces, constituent de ce point de vue des outils performants que le séquençage a permis de mettre au point (Lister *et al.*, 2009). L'autre boîte noire est de comprendre les interactions génome / environnement qui vont faire qu'une variété sera mieux adaptée dans certaines circonstances climatiques ou édaphiques que d'autres. La compréhension de la façon dont les plantes répondent aux variations environnementales constitue un champ thématique important pour la génomique, avec la recherche de gène clés et l'analyse de leur variabilité. Avec le séquençage à haut débit, cette variabilité génétique devient accessible et il devrait être possible de la corréler avec la variabilité phénotypique.

La conséquence incontournable de cette révolution technologique est la mise dans le domaine public d'une masse considérable de données, en croissance exponentielle, que la plupart des laboratoires ne sont pas ou peu préparés à traiter. Si l'on veut tirer le plein bénéfice des investissements réalisés, il est crucial de se doter au plus vite des outils bioinformatiques indispensables, de se former à leur utilisation et de les intégrer dans les cursus de formation des étudiants L'exploitation, à des fins d'amélioration des plantes, des progrès du séquençage, et de façon plus générale de la biologie moléculaire, ne fait donc que commencer et représente un immense espoir pour que l'agriculture puisse faire face aux multiples défis de l'augmentation de la population de la planète, de la réduction de la surface des terres arables, de la nécessité de rendre l'agriculture plus respectueuse de l'environnement et mieux adaptée aux changements climatiques annoncés.

Remerciements

L'auteur remercie ces collègues du laboratoire LGDP et les nombreux autres collègues, qu'il n'a pas été possible de citer, faute de place, mais qui ont contribué par leurs travaux à l'essor de la génomique végétale. ■

Références

Ammiraju JSS, Luo M, Goicoechea JL, *et al.* The Oryza bacterial artificial chromosome library resource : construction and analysis of 12 deep-coverage large insert bac libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res* 2006 ; 16 : 140-7.

Ammiraju JSS, Zuccolo A, Yu Y, *et al.* Evolutionary dynamics of an ancient retrotransposon family provides insight into evolution size in the genus *Oryza*. *Plant J* 2007 ; 52 : 342-51.

Babula B, Kaczmarek M, Barakat A, Delseny M, Quiros CF, Sadowski J. Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. *Mol Gen Genomics* 2003 ; 268 : 656-65.

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A. Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* 2005 ; 17 : 343-60.

Chantret N, Salse J, Sabot F, *et al.* Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* 2005 ; 17 : 1033-45.

Cooke R, Piegu B, Panaud O, *et al.* From rice to other cereals: comparative genomics. In : Uppadyahia N, Dennis E, eds. *Rice functional genomics*. Heidelberg : Springer-Verlag, 2007.

Delseny M. Le séquençage des plantes : les acquis. *Cah Agri* 2009 ; 19 epub. doi: 10.1684/agr.2009.0343.

Fu H, Dooner HK. Intraspecific violation of genetic colinearity and its implications in maize. *Proc Natl Acad Sci USA* 2002 ; 99 : 9573-8.

Lan TH, DelMonte TA, Reischmann KP, *et al.* An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. *Genome Res* 2000 ; 10 : 776-88.

Lister R, Gregory BD, Ecker JR. Next is now: new technologies for sequencing genomes, transcriptomes, and beyond. *Current Opin Plant Biol* 2009 ; 12 : 107-18.

Lu C, Kulkarni K, Muthu Valliappan R, *et al.* Micro RNA s and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase 2 mutant. *Genome Res* 2006 ; 16 : 1276-88.

Mardis ER. The impact of next generation sequencing technology on genetics. *Trends in Genet* 2007 ; 24 : 133-41.

Mc Nally KL, Bruskiewich R, MacKill D, *et al.* Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol* 2006 ; 141 : 26-31.

Nobuta K, Venu RC, Lu C, *et al.* An expression atlas of rice mRNAs and small RNAs. *Nature Biotechnol* 2007 ; 25 : 473-7.

Parkin IAP, Gulden SM, Sharpe AG, *et al.* Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 2005 ; 171 : 765-81.

Paterson AH, Bowers JE, Bruggmann R, *et al.* The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 2009 ; 457 : 551-6.

Paux E, Sourdille P, Salse J, *et al.* A physical map of the 1-Gigabase bread wheat chromosome 3B. *Science* 2008 ; 322 : 101-4.

Piegu B, Guyot R, Picault N, *et al.* Doubling genome size without polyploïdization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 2006 ; 16 : 1262-9.

Qiu D, Morgan C, Shi J, *et al.* A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor Appl Genet* 2006 ; 114 : 67-80.

Tian C, Wan P, Sun S, Li J, Chen M. Genome-wide analysis of the GRAS gene family in rice and Arabidopsis. *Plant Mol Biol* 2004 ; 54 : 519-32.

[15] *Quantitative trait locus*