



Variabilité naturelle, évolution et adaptation : Les apports de la plante modèle *Arabidopsis*.

Olivier Loudet

INRA, Institut Jean-Pierre Bourgin, Versailles, France

Manuscrit révisé le 07 décembre 2012 - Publié le 28 octobre 2013

Résumé : La croissance et le développement de la plante sont des processus complexes intégrant des signaux internes et externes, qui optimisent sa biologie en fonction de l'environnement. La capacité d'adaptation rapide aux fluctuations environnementales est essentielle aux organismes sessiles comme les plantes. Les fluctuations trop importantes sont causes de stress, qui induisent des stratégies adaptatives. Au cours de l'évolution, les plantes ont ainsi développé une myriade de mécanismes pour minimiser les effets négatifs des conditions défavorables, constituant un réservoir d'adaptations génétiques encore largement inconnu et inexploité. Il devient maintenant courant d'exploiter la variabilité naturelle contenue par exemple dans les accessions d'*Arabidopsis thaliana* (isolées des populations naturelles génétiquement distantes les unes des autres) comme source d'approches de génomique quantitative. Dans ces approches, les marqueurs moléculaires permettent de cartographier et d'identifier les régions contrôlant la variation d'un caractère quantitatif (QTL), révélant ainsi des types de variation allélique qui sont plus facilement exploitables que les variations obtenues par mutagenèse, surtout parce qu'ils sont observés dans des fonds génétiques variables. L'approche génomique permet aussi d'identifier des régions contrôlant la variabilité épigénétique. Quelques exemples d'identification de la source moléculaire de variations naturelles issus de nos travaux sont présentés. Ces approches seront décuplés par l'utilisation de cribles phénotypiques finement quantitatifs. Un nouvel outil récemment développé à l'INRA de Versailles pour répondre à ce besoin est présenté, qui pourrait contribuer à renforcer l'intérêt d'un modèle de petite taille pour comprendre la réponse d'une plante à l'environnement.

Mots-Clés : Diversité génétique naturelle, cartographie QTL, réponse au stress, incompatibilité, épigénétique.

Introduction

L'objectif des travaux décrits ici est d'identifier de nouveaux gènes ou allèles responsables de la variation naturelle chez *Arabidopsis*, une petite crucifère sauvage à rosette qui a été choisie comme plante modèle par les généticiens. Cette plante présente plusieurs avantages pour la génétique (en particulier la faible taille physique de son génome, un nombre élevé de gènes, sa diploïdie et son autogamie), mais aussi pour son côté pratique (son cycle court, sa petite taille, qui permet de cultiver et d'observer facilement un grand nombre d'individus dans des conditions contrôlées et avec une surface de culture réduite, sa plasticité phénotypique, sa forte prolificité). Nous pouvons par exemple classiquement utiliser les outils de génétique quantitative sur différents types de caractères : depuis un caractère classique et très intégrateur (la croissance) jusqu'à un caractère dont la variation est plus directement reliée à son contrôle moléculaire (le niveau d'expression des gènes). En combinant l'utilisation d'outils

de phénotypage spécifiques, de la cartographie fine poussée jusqu'à l'échelle du gène et d'approches de complémentation ou de génétique d'association, l'objectif est de mettre en évidence un nombre significatif de QTL ou d'eQTL (QTL contrôlant l'expression d'un gène) au niveau moléculaire, de façon à comprendre l'architecture moléculaire de ces caractères (1). L'analyse de la diversité moléculaire et fonctionnelle conduisant à ces variations de croissance

Glossaire

Accession : Lignée fixée dérivée d'un individu issu d'une population sauvage d'*Arabidopsis thaliana* collectée sur le terrain.

Fitness : Valeur reproductive d'un individu, en d'autres termes le nombre de descendants viables produits par celui-ci.

Haplotype : Combinaison particulière d'allèles sur plusieurs loci (sites génomiques) déterminés.

Modifications post-traductionnelles : Ensemble des modifications des protéines qui ont lieu après le stade de la traduction de l'ARN messager en protéine.

Phénotypage : Mesure de caractères (quantitatifs ou qualitatifs) exprimés par les individus. Cela peut par exemple être la surface foliaire d'une plante, le temps de floraison, ou bien le niveau d'expression d'un gène...

QTL : Acronyme anglais de '*Quantitative Trait Locus*' (locus de caractère quantitatif); il s'agit d'une région génomique au niveau de laquelle on a pu déterminer statistiquement l'effet du génotype (allèle ou haplotype) sur la variation d'un caractère (phénotype).

Régulation en cis : Contrôle du niveau d'expression d'un gène correspondant à des séquences régulatrices placées sur le même brin d'ADN que le gène et généralement à proximité (régions promotrices...).

Régulation Transcriptionnelle ou post-transcriptionnelle : Ensemble des mécanismes de régulation du niveau d'expression des gènes (dont les régulations en cis) qui agissent au moment de la transcription ou bien après la transcription de l'ARN messager à partir de l'ADN.

Transcriptome : Ensemble des niveaux d'accumulation des ARN messagers d'un individu, issus de l'expression des gènes.

et/ou d'accumulation de transcrit en interaction avec l'environnement, donne des pistes pour comprendre comment l'adaptation façonne la variabilité naturelle et aide à décoder les origines moléculaires de la régulation du transcriptome. De plus, les gènes et les processus physiologiques identifiés représentent potentiellement des cibles pour les programmes de sélection en amélioration des plantes, et participent à une meilleure compréhension intégrée de la biologie de l'espèce et de son évolution (2). Parmi les divers mécanismes que le clonage de ces QTL commence à révéler, la régulation de l'expression des gènes (cis ou trans) pourrait jouer un rôle important (3). D'ailleurs certains postulent que les séquences régulatrices pourraient être moins contraintes au cours de l'évolution que les séquences codantes. Cependant, les polymorphismes ayant un rôle en cis restent plus difficiles à reconnaître et à confirmer, tandis qu'un niveau de diversité supplémentaire, les régulations d'origine épigénétique (4), reste mal évalué.

Variabilité naturelle de la plante modèle *Arabidopsis thaliana*

Une des grandes questions en biologie réside dans la caractérisation des supports moléculaires de la variation des caractères, que ce soit chez les plantes cultivées ou chez les espèces sauvages. Les facteurs de variation des caractères complexes sont traditionnellement décomposés en effets de loci individuels et interactions entre loci, associés à des marqueurs moléculaires correspondant à des variations de séquence, et permettant de cartographier le génome. Si la génétique

quantitative (la génétique des caractères à variation continue) fut initialement appliquée aux plantes cultivées, il est devenu intéressant également d'exploiter la variabilité naturelle contenue par exemple dans les accessions d'*Arabidopsis thaliana* (isolées de populations naturelles génétiquement distantes les unes des autres) comme source d'approches de génomique quantitative. *Arabidopsis* est une plante qui pousse à peu près partout sur l'ensemble du continent euro-asiatique, à l'état sauvage mais aussi dans des milieux qui peuvent être très perturbés par l'homme. Sont présentées sur la carte ci-dessous (Figure 1), un certain nombre de populations d'*Arabidopsis*, dans lesquelles on a pu collecter des individus. Les lignées consanguines expérimentales dérivées de ces individus sont appelées des accessions, qui représentent donc des génotypes fixés (homozygotes) issus de ces populations sauvages. Ces accessions permettent de constater l'extrême diversité à la fois phénotypique et génotypique de cette espèce.

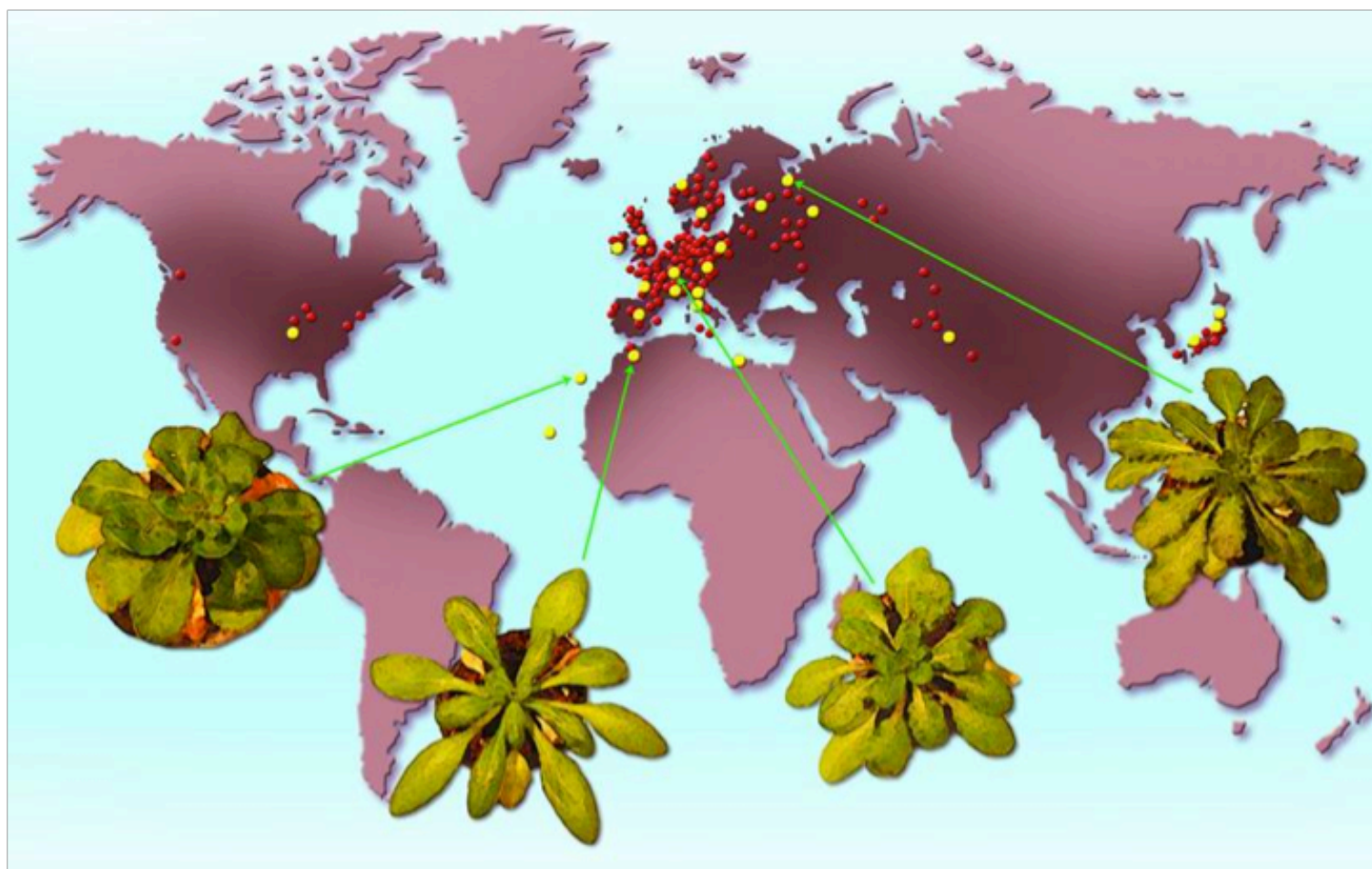


Figure 1 Carte de répartition mondiale des lieux d'origine des accessions d'*Arabidopsis* à la disposition de la collectivité scientifique.

Tout le continent euro-asiatique, y compris l'ensemble du bassin méditerranéen, est donc la zone la plus importante pour la diversification de l'espèce, et l'on peut constater que la distribution des populations auxquelles on a accès dans les centres de ressources n'est pas parfaitement représentative de la distribution de l'espèce sur ce continent. Pour étudier ces populations au laboratoire on procède assez classiquement à des croisements entre les accessions 2 à 2 pour obtenir des ensembles de lignées recombinantes. Ces lignées génotypées avec un grand nombre de marqueurs servent à faire des cartographies QTL (5), mais les accessions peuvent être aussi utilisées directement en génétique d'association, ce qui permet dans les 2 cas de relier les variations phénotypiques aux variations de l'ADN.

Exemple : réponse à un micro-élément du sol, le molybdate

Un exemple de réponse à un stress abiotique, permet d'illustrer ces approches. Sur des substrats acides à base de tourbe, les individus de l'accession Shahdara (Tadjikistan) développent des nécroses et expriment des réductions de croissance et de fitness très importantes, ce qui n'est pas le cas des individus de l'accession Bay-0 (Allemagne). Dans le croisement entre ces deux accessions, la tolérance à cette forme d'acidité du sol est liée à un gène (QTL) majeur. Par cartographie fine et clonage positionnel, nous avons pu réduire l'intervalle candidat qui contrôle ce phénotype à une zone dans laquelle il n'y a qu'un petit nombre de gènes. Surtout, nous avons dans cet intervalle un gène candidat particulièrement intéressant (*MOT1*), connu pour être un transporteur de molybdate (Mo) selon Baxter *et al.* (6). Il s'agit d'un bon gène candidat fonctionnel, car, en condition de sol acide, le molybdate, comme d'autres éléments, est moins accessible. Le phénotype observé sur la plante pouvait donc être une conséquence secondaire de l'acidité, le défaut de transport dans la plante renforçant la carence en Mo. Pour confirmer cette hypothèse nous avons effectué des tests de complémentation chimiques (augmentation du pH ou de la teneur en Mo du sol) et génétiques (croisement avec un mutant artificiel dans *MOT1*) qui ont tous deux confirmé cette hypothèse : le défaut de croissance de Shahdara est bien lié à un défaut du gène *MOT1* (en l'occurrence un changement d'acide aminé dans la protéine).

Ce qui peut être particulièrement intéressant lorsqu'on identifie dans un croisement spécifique un phénotype associé à un génotype particulier, c'est de revenir ensuite à l'échelle de l'espèce et de voir quelle est l'importance de cette variation à cette échelle. Cela peut permettre d'identifier des variations ayant répondu positivement à la sélection naturelle. Dans le cas de *MOT1*, la mutation naturelle de Shahdara est présente uniquement en Asie Centrale. Nous pouvons donc essayer, à l'échelle de cette région, de relier la présence de cet allèle à une caractéristique particulière de l'environnement de ces populations, pour tenter de mettre en évidence une pression de sélection qui affecterait cette variation, c'est-à-dire de lui donner un éventuel sens adaptatif. La Figure 2 montre clairement que les populations possédant l'allèle déficient au locus *MOT1* (carrés rouges) poussent sur des sols particulièrement riches en Mo, mais qu'il n'y a pas de corrélation avec l'acidité du sol.

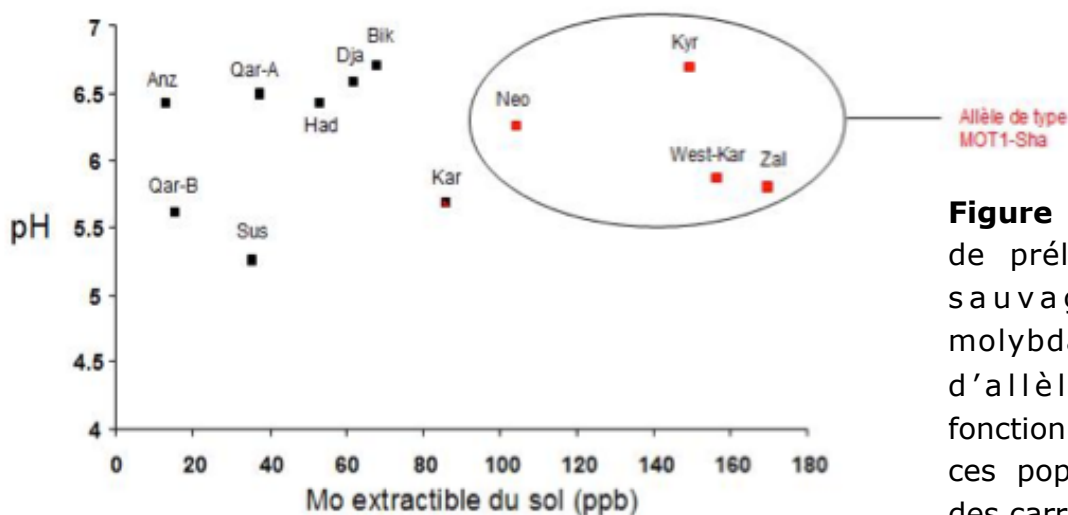


Figure 2 Caractères du sol au lieu de prélèvement des populations sauvages - pH, teneur en molybdate (Mo) - et présence d'allèles fonctionnels/non-fonctionnels au locus *MOT1* dans ces populations (représentées par des carrés noirs/ rouges)

Bien qu'ayant identifié cette variation sur la base d'un défaut de croissance en réponse à une carence en Mo, il semble que le sens biologique de cette variation chez *Arabidopsis* soit plutôt à rechercher dans la réponse à la toxicité des teneurs élevées de Mo dans le sol, toxicité que nous avons décrite et qui conduit effectivement à des réductions de fitness (production de graines) pouvant atteindre 50% chez *Arabidopsis*. *MOT1* représente donc un bon candidat pour jouer sur ces contraintes d'intérêt agronomique.

Incompatibilité allélique et épigénétique

Aujourd'hui, nous savons que des variations de niveau épigénétique - contrôlant potentiellement l'activité du génome - apparaissent spontanément au cours des générations dans un fond génétique donné ne subissant pas de pression de sélection, comme le montrent les analyses de méthylation de l'ADN à l'échelle du génome dans des lignées d'accumulation de variations conduites sur plusieurs dizaines de générations (7, 8). Il est d'ailleurs remarquable que, si l'on travaille non pas à l'échelle de la base méthylée (cytosine), mais à l'échelle de la région méthylée (quelques centaines de paires de bases contenant plusieurs cytosines sujettes à variation de méthylation de façon coordonnée), le nombre de variations (épimutations) observées est du même ordre (7) que celui des mutations classiquement observé au niveau nucléotidique (9), soit environ un par génome et par génération. Par ailleurs, nous savons aussi que des variations de méthylation de l'ADN existent naturellement, même si seulement une partie de ces variations affecte le niveau d'expression de gènes (10). *Arabidopsis thaliana* offre ainsi l'opportunité de prendre en compte ce niveau de régulation dans nos approches.

A titre illustratif, voici un exemple d'une variation due à une modification de la méthylation de l'ADN et dont les conséquences sont potentiellement importantes sur le plan évolutif et en amélioration des plantes (11). Certaines plantes issues de croisement entre les accessions d'*Arabidopsis* Columbia (Col) et Shahdara (Sha) produisent des quantités réduites de pollen et de graines, ce qui affecte fortement leur capacité de reproduction. Nous avons découvert que cela correspond à une incompatibilité entre deux régions chromosomiques de Col et de Sha. Cette incompatibilité concerne des gènes nommés *AtFOLT*, impliqués dans le transport des folates, dont l'absence est connue pour affecter la formation des cellules reproductrices ainsi que les étapes précoces du développement embryonnaire chez de multiples organismes. Chez Col, un seul gène *AtFOLT* est présent ; il est nommé *AtFOLT1* et est localisé sur le chromosome 5. Chez Shahdara, en plus du gène *AtFOLT1* sur le chromosome 5, il existe un gène similaire, nommé *AtFOLT2*, situé sur le chromosome 4. Ce gène, qui remplit le même rôle qu'*AtFOLT1*, se trouve au sein d'une structure très complexe qui comprend notamment des séquences d'ADN répétées. Ces dernières sont responsables de la production de petits ARN, entraînant des modifications épigénétiques (et héréditaires) du gène *AtFOLT1* et donc son inactivation. Ainsi, dans la descendance du croisement entre Col et Sha, les plantes héritant du chromosome 4 de Col (sans le gène *AtFOLT2*) et du chromosome 5 de Sha (avec un gène *AtFOLT1* inactivé) ne sont pas fertiles. Le gène *AtFOLT1* se maintient dans la descendance sous sa forme inactivée même si le gène *AtFOLT2* n'est plus présent. De plus, l'introduction dans une plante de la structure particulière du chromosome 4 peut provoquer en quelques générations l'inactivation du gène *AtFOLT1*. La nouveauté de ce mécanisme réside dans sa nature épigénétique qui a la capacité d'accélérer l'apparition de conséquences délétères dans la descendance d'un croisement. Ces

résultats suggèrent que des modifications épigénétiques naturelles provenant de variations structurales peuvent jouer un rôle important dans la mise en place rapide d'incompatibilités génétiques entre individus d'une même espèce. Ils contribuent probablement à expliquer les mortalités observées lors de certaines hybridations entre variétés ou entre espèces et à comprendre la mise en place de la séparation des espèces. En outre, ces phénomènes pourraient avoir des conséquences importantes dans le domaine de l'amélioration des plantes en limitant les fragments chromosomiques qui peuvent être introduits par croisement d'une variété dans une autre.

Phénotypage haut-débit pour la réponse au stress

Dans toutes ces approches, et en particulier celles visant à une analyse quantitative fine des caractères, il apparaît clairement que le phénotypage est le facteur limitant, car il faut réduire au maximum la variance environnementale (qui perturbe fortement l'évaluation de l'effet phénotypique de loci individuels) tout en évaluant des populations de très grande taille (que ce soit pour la cartographie initiale ou en cartographie fine). Ceci est particulièrement complexe lorsqu'on s'intéresse à des stress en interaction avec tous les facteurs de l'environnement d'une chambre de culture (température, humidité, flux d'air, lumière, ...) comme le stress hydrique. A titre d'exemple, notre réponse sur ce point a demandé plusieurs années de mise au point grâce à l'effort de nombreuses personnes au sein de l'Institut Jean-Pierre Bourgin (IJPB) à Versailles. Nous avons développé un robot de phénotypage original (appelé le *Phenoscope*) qui permet de cultiver simultanément sur une table 735 plantes dans des pots individuels qui suivent un mouvement de déplacement le long de rails leur permettant de prendre successivement toutes les positions possibles sur la table à chaque cycle (environ 4 heures), ce qui homogénéise les conditions de culture perçues par chaque individu (Figure 3).



Figure 3 Un des 2 robots '*Phenoscope*' de l'IJPB, prenant en charge 735 plantes individuelles, âgées d'environ 25 jours sur cette photo.

De plus, à chaque passage le robot ajuste par pesée/arrosage la teneur en eau dans le sol qui porte la plante (permettant d'obtenir, pour chaque plante individuellement, le niveau de stress hydrique désiré par l'utilisateur) et prend une photo verticale qui est automatiquement analysée pour en extraire la surface foliaire projetée de chaque plante ainsi que d'autres paramètres phénotypiques. Nous savons d'ores et déjà que ce robot permet non seulement de réaliser des expériences qui ne seraient pas faisables à la main, mais il les réalise dans des conditions de répétabilité et d'homogénéité qui sont bien meilleures que tout ce que nous avons pu faire auparavant. Nous disposons maintenant de 2 Phenoscopes installés dans une chambre de culture dédiée. Nous pensons que cet outil nous permettra d'étudier des QTL de croissance et de réponse au stress hydrique avec une précision inégalée.

Conclusion

Savoir quelle fraction de la diversité est contrôlée au niveau du génome (structure, localisation ou nombre de copies du gène), au niveau de l'ARN (de façon transcriptionnelle ou post-transcriptionnelle), au niveau de la protéine (modifications post-traductionnelles, quantité, activité, fonction, interaction...), reste un objectif majeur. Le nombre limité de QTL que nous avons pu étudier peut entraîner un biais dans notre vision de la variabilité naturelle quantitative.

Au niveau mondial, la « communauté Arabidopsis » cherche à identifier les supports moléculaires d'un plus grand nombre de QTL. Cela devrait permettre une vision plus générale des supports de la diversité y compris par l'intermédiaire de loci rarement étudiés (ceux dont les effets alléliques sont faibles -par exemple largement inférieur à la variation environnementale sur la plupart des outils de phénotypage classiques- ou en interaction avec l'environnement ; ceux soumis à des interactions épistatiques ; ceux sans candidat fonctionnel évident malgré une cartographie fine, etc...).

En prenant avantage des ressources uniques disponibles chez cette espèce modèle, nous cherchons à nous donner les moyens de considérer tout type de variation causale, y compris structurale (duplication, translocation de gène, variation du nombre de copies) ou régulatrice (et, particulièrement, épigénétique) qui sont plus rarement considérés car plus difficiles à détecter, à valider et à exploiter. Ceci devrait contribuer à une meilleure compréhension des mécanismes contrôlant la variation quantitative de caractères d'intérêt agronomique et à proposer des pistes en amélioration des plantes

Références bibliographiques

1. TRONTIN C., TISNÉ S., BACH L., LOUDET O., 2011. - What does Arabidopsis natural variation teach us (and does not teach us) about adaptation in plants? *Curr. Opin. Plant Biol.*, **14**, 225-231.
2. ALONSO-BLANCO C., AARTS M.G., BENTSINK L., KEURENTJES J.J., REYMOND M., et al., 2009. - What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell*, **21**, 1877-1896.
3. CUBILLOS F.A., COUSTHAM V., LOUDET O., 2012. - Lessons from eQTL mapping studies: non-coding regions and their role behind natural phenotypic variation in plants. *Curr. Opin. Plant Biol.* **15**, 192-198.

4. SCHMITZ R.J., ECKER J.R., 2012. - Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends Plant Sci.* 17, 149-154.
5. SIMON M., LOUDET O., DURAND S., BÉRARD A., BRUNEL D., et al., 2008. - QTL mapping in five new large RIL populations of *Arabidopsis thaliana* genotyped with consensus SNP markers. *Genetics*, 178, 2253-2264.
6. BAXTER I., MUTHUKUMAR B., PARK H.C., BUCHNER P., LAHNER B., et al., 2008. - Variation in molybdenum content across broadly distributed populations of *Arabidopsis thaliana* is controlled by a mitochondrial molybdenum transporter (MOT1). *PLoS Genet.*, 4, e1000004.
7. BECKER C., HAGMANN J., MÜLLER J., KOENIG D., STEGLE O., et al., 2011. - Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, 480, 245-249.
8. SCHMITZ R.J., SCHULTZ M.D., LEWSEY M.G., O'MALLEY R.C., URICH M.A., et al., 2011. - Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 334, 369-373.
9. OSSOWSKI S., SCHNEEBERGER K., LUCAS-LLEDO J.I., WARTHMAN N., CLARK R.M., et al., 2010. - The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, 327, 92-94.
10. VAUGHN M.W., TANURDZIC M., LIPPMAN Z., JIANG H., CARRASQUILLO R., et al., 2007. - Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.*, 5, e174.
11. DURAND S., BOUCHE N., PEREZ STRAND E., LOUDET O., CAMILLERI C., 2012. - Rapid Establishment of Genetic Incompatibility through Natural Epigenetic Variation. *Curr. Biol.*, 22, 326-331.