

Décoder la complexité de la variabilité naturelle pour la croissance et la réponse à l'environnement chez *Arabidopsis thaliana*

Par

Charlotte TRONTIN

Contexte scientifique

Des génotypes adaptés à des environnements contrastés ont de grandes chances de se comporter différemment lorsqu'ils sont placés dans des conditions similaires et contrôlées, notamment si leur sensibilité aux signaux environnementaux et/ou leur croissance intrinsèque sont limitées à différents niveaux. De ce fait, la variabilité observée dans les populations naturelles peut être utilisée comme une source illimitée de nouveaux allèles ou gènes pour l'étude des bases génétiques de la variation des traits quantitatifs. Le but des approches de génétique quantitative est de comprendre comment la diversité génétique et épigénétique contrôle la variabilité phénotypique observée dans les populations à différentes échelles, au cours du développement et sous différentes contraintes environnementales. De plus, ces analyses ont pour objectif de comprendre comment les processus adaptatifs et démographiques influencent la fréquence de ces variants dans les populations en fonction de leur environnement local.

Arabidopsis thaliana est une des plantes modèles les plus étudiées en raison de son cycle de vie court, de sa petite taille, de sa production importante de graines et de l'efficacité de la transformation de cette espèce. De plus la séquence complète de son petit génome est disponible ainsi que de nombreuses lignées mutantes. En ce qui concerne la variabilité naturelle, c'est également un très bon modèle puisque cette petite herbacée est retrouvée dans des habitats relativement contrastés, dans le monde entier. Son potentiel adaptatif particulièrement élevé est aussi suggéré par l'importante diversité génétique, épigénétique et phénotypique observée chez cette espèce. Le terme d' 'accession' est utilisé pour faire référence à des individus collectés sur un site donné.

L'étude de la variabilité naturelle peut être appréhendée en utilisant diverses approches. Tout d'abord, afin d'identifier les loci (on parle de QTL pour *Quantitative Trait Loci*), les gènes (QTG) ou même les polymorphismes (QTP) responsable de la variabilité phénotypique complexe observée entre différentes accessions dans des conditions environnementales données, des approches de génétique quantitative sont utilisées. Ces approches

sont basées sur l'association entre des données génétiques et des données phénotypiques issues du génotypage et du phénotypage de populations de cartographies telles que des F2 ou des RILs ou directement des accessions. Dans le premier cas on parle en général de cartographie de QTL (*QTL mapping*) et dans le deuxième cas on parle de génétique d'association (*Genome-Wide Association mapping* ou *GWA*). L'identification des bases génétiques à l'origine des variations des traits quantitatifs observées dans les populations naturelles peut être plus ou moins difficile en fonction de l'architecture génétique des traits en question et de leur interaction avec l'environnement. Suite à l'identification et la confirmation des QTGs, il est possible d'utiliser des approches plus classiques de biologie moléculaire et cellulaire et de physiologie pour comprendre la fonction du QTG en relation avec la variation phénotypique observée et l'impact du/des polymorphisme(s) en cause sur la fonction du gène. Par ailleurs, l'identification des QTGs et QTPs permet de rechercher des traces sélections au niveau du loci d'intérêt en utilisant des approches plus évolutives. Enfin des corrélations entre la fréquence des variants génétiques, leurs effets sur la fitness et les facteurs environnementaux caractérisant les régions dans lesquelles on retrouve ces variants peuvent aussi être très utiles pour comprendre comment les populations s'adaptent localement.

Mon travail de doctorat a consisté en l'analyse de la variabilité naturelle pour la croissance et la réponse à l'environnement chez *A. thaliana*. J'ai eu la chance de participer à trois projets indépendants qui exploitent tous la variabilité naturelle d'*A. thaliana* et qui m'ont permis d'aborder les différentes approches disponibles pour l'étude des variations des traits quantitatifs dans les populations naturelles.

Validation de l'évolution non-neutre du gène *MOT1* codant pour un transporteur de molybdate

Ce projet a été initié par Olivier Loudet en 1999 suite à l'observation que l'accession Shahdara (ou Sha) (collectée dans la Vallée Shakh dara au Tajikistan) montrait une réduction de croissance bien plus importante que l'accession Bay-0 (collectée à Bayreuth en Allemagne) quand ces accessions étaient cultivées sur un milieu riche en tourbe et donc relativement acide. En utilisant une population de RILs et des lignées de HIFs, Claire Le Metté et Kian Poormohammad ont montré qu'un locus majeur de 80kb, en haut du chromosome 2 expliquait cette différence de croissance entre les deux accessions. Cet interval génétique contenait parmi 18 autres, le gène *MOT1* codant pour un transporteur de molybdate (la forme du molybdène (Mo) assimilable par la plante). Ce gène était un très bon candidat pour le QTL de croissance observé sur tourbe car la disponibilité en Mo pour la plante varie en fonction du pH du sol et une déficience en Mo, cofacteur essentiel

à de nombreuses enzymes dont la nitrate reductase, peut conduire à des phénotypes similaires à ceux que présente l'accession Sha sur tourbe. Au cours de son post-doc, Kian a montré qu'une mutation dans l'allèle Sha du gène *MOT1* (D104Y) rend ce transporteur hypofonctionnel et que le phénotype 'tourbe' est bien le résultat d'une carence en Mo. Dans deux études de 2007 et de 2008, une délétion de 53 paires de bases (Del-53-pb) dans le promoteur du gène *MOT1* a été observée chez plusieurs accessions (dont Ler-0 et Van-0) et a été associée à une réduction du niveau d'expression du gène *MOT1* ainsi qu'à une diminution du contenu en Mo dans les feuilles et les racines de ces accessions. Partant de ce constat, Kian a montré que toutes les accessions de type 'Ler' (Del-53-pb) présentaient aussi le phénotype 'tourbe' et les données de contenu en Mo disponibles ont par ailleurs montré que toutes les accessions de type 'Sha' (D104Y) présentent un contenu en Mo réduit en condition standard. Ainsi au moins deux allèles différents du gène *MOT1* ségrègent dans les populations naturelles d'*A. thaliana* et sont responsables d'un phénotype similaire à savoir une réduction de la quantité de Mo dans la plante ainsi qu'une réduction de croissance sur des sols pauvres en Mo. Cette forme d'hétérogénéité allélique est probablement assez fréquente bien que peu décrite.

Quand je suis arrivée dans l'équipe, l'importance évolutive de l'hétérogénéité allélique observée à *MOT1* était mal connue. Au cours de ma thèse, j'ai réalisé avec l'aide de Matthieu Simon et de Thierry Robert des analyses de génétique des populations dans le but de détecter de possibles traces de sélection au niveau du locus *MOT1*. Le séquençage de 102 accessions d'*A. thaliana* réparties dans tout l'hémisphère nord, dont 28 Sha-like, 19 Ler-like et 55 Col-like a montré que l'allèle Sha présente un faible niveau de polymorphismes comparé à *MOT1[Ler]* et *MOT1[Col]*. Ce faible taux de polymorphismes a aussi été observé à une échelle régionale en 'Asie de l'ouest' où sont confinées les accessions de type Sha et peut refléter une expansion récente et rapide de l'allèle Sha dans cette région. Cette expansion peut être le résultat de processus évolutifs neutres au cours de la recolonisation postglaciaire de 'Asie de l'ouest' tel que du '*gene surfing*' ou peut refléter des événements de sélection locale en faveur de l'allèle Sha. Cette dernière hypothèse est en accord avec la significativité de plusieurs tests de sélection réalisés sur un échantillon mondial et régional d'accessions, dont le test de Tajima ($D < 0$), le test de Mc Donald-Kreitman ($NI > 1$) et le test HKA, suggérant que la diversité génétique observée à *MOT1* est le résultat de la sélection de différents haplotypes. De plus, l'analyse élémentaire de plusieurs sols sur lesquels des accessions d'Asie de l'ouest ont été collectées a montré que les accessions de type Sha poussaient sur des sols relativement riches en Mo. Enfin, des analyses de toxicité effectuées sur *A. thaliana* en serre et *in vitro* ont montré que de fortes concentrations en Mo diminuaient la croissance racinaire et le nombre de graines

produites par plante. Ces tests ont aussi montré que les plantes portant un allèle défectueux à *MOT1* produisaient des graines en moyenne légèrement plus grosses que les plantes de type sauvage et ce spécifiquement sur des sols très riches en Mo. Ainsi, l'allèle Sha pourrait avoir à la fois un effet protecteur en réponse à de fortes accumulations de Mo dans l'environnement et être délétère pour la plante sur des sols pauvres en Mo.

La découverte de variants génétiques associés à des phénomènes de trade-off environnementaux reste assez rare malgré leur importance dans l'évolution et la répartition des haplotypes au sein des populations et des espèces. De plus notre travail met en évidence l'importance des paramètres environnementaux hétérogènes, tel que les propriétés des sols, comme agents de sélection à l'origine de l'adaptation. Ce travail a été publié dans la revue [PLoS Genetics](#) le 12 Juillet 2012.

Deux récepteurs kinases dupliqués en tandem sont responsables de la variation de croissance observée en réponse à un traitement au mannitol dans des populations naturelles d'*A. thaliana*

Depuis ses débuts, le groupe VAST s'emploie à décoder les bases génétiques de la tolérance aux contraintes abiotiques dans des populations naturelles d'*A. thaliana* en utilisant des approches de génétique quantitative. Pour ce faire, des protocoles de phénotypage haut débit *in vitro* ont été développés. Le mannitol, un polyol connu pour ne pas être synthétisé par *A. thaliana* ni perturber son métabolisme, est généralement utilisé dans les milieux gélosés pour induire des stress osmotiques. Dans cette étude, les bases génétiques de la variabilité pour la croissance en réponse au stress induit par un milieu contenant 60mM de mannitol ont été recherchées dans une population de RILs issus du croisement des accessions Col-0 et Cvi-0. Un QTL majeur et spécifique du milieu mannitol a été identifié en haut du chromosome 1 et sa ségrégation a été confirmée dans une famille de HIFs (pour '*Heterogeneous Inbred Family*'). Ce QTL a été nommé EGM pour '*Enhanced Growth under Mannitol stress*' (augmentation de croissance sous stress induit par le mannitol) à cause de l'effet positif de l'allèle Cvi du QTL sur la croissance en condition de stress induit par le mannitol. Kian Poormohammad, a cartographié à l'échelle du gène le QTL EGM et l'intervalle candidat de 10kb identifié a été validé en utilisant une famille de lignées appelées arHIFs pour '*advanced recombined HIF*' qui ségègent uniquement pour la région d'intérêt. L'analyse de plusieurs mutants T-DNA pour chacun des 3 gènes présents dans l'intervalle candidat a finalement suggéré qu'une mutation dans le gène *At1g11300* devait être la cause du QTL EGM (i.e le QTG).

En ce qui concerne ce projet, le but de ma thèse a été de confirmer le QTG, d'identifier le ou les polymorphismes responsable(s) du QTL et de comprendre la fonction du gène vis à vis du stress induit par le mannitol.

Tout d'abord, j'ai montré que le gène annoté *At1g11300* dans TAIR10 recouvre en réalité deux gènes paralogues indépendants apparus par duplication en tandem et codant pour des récepteurs kinases putatifs de la famille SD1 et caractérisés par un peptide signal, une région extracellulaire et un domaine kinase intracellulaire. Étant donné que les mutants T-DNA disponibles dans ces deux gènes sont plus grands que leur fond génétique sauvage en condition de stress induit par du mannitol, ces deux gènes ont été renommés *EGM1* (*At1g11300*) et *EGM2* (*At1g11305*).

Dans la région candidate, neuf polymorphismes différencient l'allèle Col de l'allèle Cvi, quatre d'entre eux étant non-synonymes. Afin d'identifier parmi ces SNPs celui responsable du QTL EGM, j'ai testé la ségrégation d'EGM dans plusieurs populations de F2 spécialement créées pour ségréger pour diverses combinaisons de polymorphismes de type Col et Cvi, une approche que l'on a appelé *génétique d'association spécifique*. Cette approche m'a permis de réduire le nombre de SNPs candidats à 3 : une mutation de T→C dans le promoteur d'*EGM1*, une mutation non-synonyme d'une sérine conservée dans la famille SD1 en une glycine dans le domaine lectine d'*EGM2* et une mutation non-synonyme d'une cystéine très conservée et impliquée dans la formation d'un pont disulfure important pour le repliement du domaine PAN-APPLE d'*EGM2*. Des complémentations transgéniques ont montré que l'allèle Cvi du gène *EGM2* est au moins hypofonctionnel probablement du aux deux mutations non-synonymes indiquées ci-dessus. De plus des analyses d'expression des gènes EGMs dans les arHIFs, les mutants *egm1* et *egm2* et dans différentes accessions suggèrent que le polymorphisme dans le promoteur d'*EGM1* pourrait aussi contribuer au phénotype EGM. Finalement, certaines de nos données montrent que ces deux récepteurs kinase putatifs ne sont pas complètement redondants bien qu'impliqués dans un pathway commun. Le rôle exacte de chacun reste à déterminer.

Au début de notre étude, le mannitol a été utilisé pour induire un stress osmotique. Cependant, comme les deux gènes responsables du QTL présentent des domaines de liaison au mannose (domaine B-lectine), nous nous sommes demandés si le phénotype EGM que nous observions sur mannitol était bien le résultat d'un stress osmotique. Différentes contraintes osmotiques ont été testées mais la ségrégation d'EGM n'a été observée que sur mannitol suggérant que le mannitol en lui même pouvait avoir un effet sur la croissance indépendamment de la contrainte osmotique qu'il impose. Une analyse transcriptomique a montré qu'environ 200 gènes sont différentiellement exprimés entre les 2 arHIFs sur mannitol. Parmi eux, un enrichissement en gènes répondant à des contraintes biotiques a été observé, enrichissement qui a par ailleurs été montré dans une analyse comparant le transcriptome de plantes Col-0 traitées ou non au mannitol. De plus, l'induction de plusieurs

de ces gènes de stress biotiques est spécifique du milieu mannitol et n'est pas observé sur un milieu contenant du sorbitol. L'ensemble de ces observations nous ont amenés à penser que les gènes *EGM1* et *EGM2* pouvaient jouer un rôle dans la réponse aux contraintes biotiques. En effet, on sait que certains pathogènes, dont les champignons, utilisent le mannitol comme source de carbone. De plus, quelques papiers suggèrent qu'au cours de l'infection, les pathogènes sécrètent du mannitol pour faire face à la production de ROS par la plante. En accord avec cette hypothèse, nous avons montré que les mutants *egm1* et *egm2* étaient plus sensibles à *Botrytis cinerea* que les fonds génétiques sauvages respectifs.

En conclusion, ce projet nous a permis d'identifier deux récepteurs kinase putatifs potentiellement impliqués dans les réponses de défense contre certains pathogènes via la perception directe ou indirecte du mannitol produit par ces derniers. Ce travail sera bientôt soumis dans une revue scientifique (PNAS ou Plant Cell).

La variabilité épigénétique observée au locus *QQS* dans des populations naturelles d'*A. thaliana*

Ce projet a été initié par Michel Vincentz (CBMEG, Brésil) et son étudiante en thèse Amanda Silveira. Alors qu'ils étudiaient un facteur de transcription potentiellement impliqué dans l'homéostasie cellulaire du carbone, ils ont observé par hasard que le gène *QQS* (pour *Qua-Quine Starch*) était spécifiquement et fortement up-régulé dans un de leur stock de graines Col-0, stock qu'ils ont appelé 'Col-0*'. En utilisant des approches de génétique et d'épigénétique, ils ont montré que différents épialèles stables du gène *QQS* ségrègent dans leur stock de graines Col-0* et dans quelques accessions d'*A. thaliana*. Ces épialèles sont caractérisés par une forte corrélation négative entre le niveau de méthylation du promoteur et du 5'UTR du gène *QQS* et le niveau d'accumulation de son transcrite. De plus ils ont montré que ces variations de méthylations sont indépendantes de la séquence génétique observée au niveau de *QQS* ainsi que de l'état de méthylation des transposons présents autour de ce gène. Partant de ces résultats, nous avons participé à trois aspects différents du projet.

Tout d'abord, plusieurs personnes au laboratoire travaillent sur l'identification et la confirmation des QTLs d'expression (eQTL) dans les populations de RILs issues des croisements Cvi-0 x Col-0 et Bur-0 x Col-0. Un des eQTLs locaux les plus significatifs identifié parmi ces deux populations colocalise avec le gène *QQS* dont il régule l'expression. Avec l'aide de Matthieu Canut et grâce à l'expertise du laboratoire dans ce domaine, nous avons confirmé que *QQS* est bien contrôlé en cis par cet eQTL en utilisant des tests ASE (pour *Allele Specific Expression*) dans des F1 issus des croisements de Col-0 (qui présente

un allèle méthylé) avec les accessions Jea et Kondara (qui présentent des allèles déméthylés comme Cvi-0). De plus nous avons montré qu'il était possible de restaurer l'expression de *QQS* dans des accessions présentant un allèle méthylé en traitant les plantes avec un inhibiteur de méthylation. Ces résultats ont fortement renforcé l'hypothèse selon laquelle les épivariants observés au gène *QQS* sont stables et 'pure' dans le sens où ils ne semblent pas dépendre de variants génétiques.

Ensuite, à partir de graines collectées directement sur le terrain, en Asie Centrale, par Olivier Loudet, nous avons montré que les épiallèles au gène *QQS* ségrégent bien dans les populations naturelles d'*A. thaliana* et ne sont pas juste le résultat de plusieurs générations d'autofécondation en condition expérimentale en serre. Au sein de trois populations différentes, des variants méthylés et non méthylés ont été isolés posant la question du potentiel adaptatif de ces épivariants.

Enfin nous avons essayé de tester ce dernier en cherchant un phénotype associé aux différents épivariants (méthylé et non méthylé). Le gène *QQS* ayant précédemment été caractérisé comme un régulateur négatif de la synthèse d'amidon, nous avons testé deux phénotypes –la croissance et l'accumulation et dégradation de l'amidon au cours de 24h– dans différentes conditions –*in vitro* et *in vivo*, en condition normales ou en condition de contraintes hydrique, osmotique ou froid–. Cependant aucune différence significative n'a été observée entre les différentes lignées suggérant que les épivariants au gène *QQS* n'ont pas de conséquence phénotypique évidente (en tout cas dans nos conditions) et donc n'ont peut être pas de potentiel adaptatif.

En conclusion ce travail a permis la caractérisation précise d'un épiallèle 'pure' ségrégeant à une fréquence relativement élevée dans des populations naturelles d'*A. thaliana* –caractérisation qui chez les plantes reste rare–. Bien que l'importance phénotypique et écologique de ces épiallèles reste à démontrer, ce travail souligne la contribution des variations épigénétiques aux variations stables du niveau d'expression de certains gènes dans les populations. Enfin, ce travail suggère que les gènes nouvellement formés (neo-gènes) pourraient être particulièrement enclins aux variations de niveau de méthylation. Ce travail a été publié dans la revue [PLoS Genetics](#) le 11 Avril 2013.