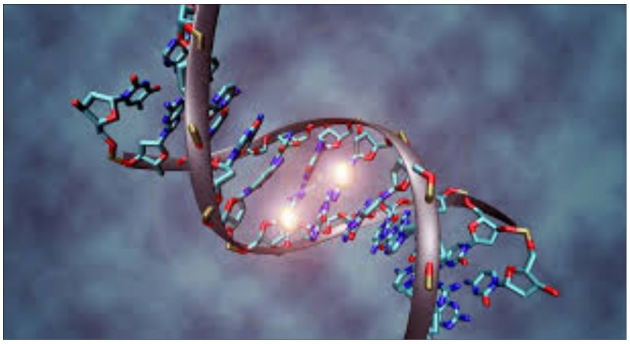
2018-2019

Mention Biologie Végétale



Contrôle Epigénétique de L'adaptation du pétunia au Stress Hydrique

KOUKI Yosra

Sous la direction de Mmes
Leduc Nathalie et Landes Claudine

Membres du jury
Montrichard Françoise | Présidente
Peltier Didier | Auditeur
Jérémy Clotault | Tuteur

Soutenu publiquement
le 03/ 07/ 2019

ENGAGEMENT DE NON PLAGIAT

Je, soussigné(e) Yosra KOUKI.....
déclare être pleinement conscient(e) que le plagiat de documents ou d'une partie d'un document publiée sur toutes formes de support, y compris l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.
En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce rapport ou mémoire.

signé par l'étudiant(e) le **24**/ **06** / **2019**

REMERCIEMENT

Je tiens à remercier mes encadrantes, Nathalie Leduc et Claudine Landès, pour les connaissances qu'elles m'ont transmises, pour leurs patiences, leur disponibilité, et le suivi de mon mémoire.
Un merci tout particulier à Sylvain Gaillard, Ingénieur dans l'équipe BIDEFi, pour le temps qu'il m'a consacré et le partage de ses connaissances du début jusqu'à la fin de mon stage.
Je remercie Jérémy Clotault , pour son implication en tant qu'auditeur et tous les membres du jury.
Je remercie aussi l'ensemble des équipes BioInformatique et ARCH-E pour l'accueil et leur sympathie.

Table des matières

Table des figures

Table des tableaux

Liste de vocabulaires :

Reads: Les séquences des fragments obtenues la fin du séquençage

Mapping =alignement : Trouver la position des lectures dans le génome de référence

Séquençage: déterminer la succession linéaire des bases A, C, G, T de l'ADN, la lecture de cette séquence permet d'étudier l'information biologique contenue par celle-ci

Scaffold : ensemble de contigs orientés et ordonnés.

Contig : ensemble des reads (lectures)

Liste des abréviations :

A : Adénine

ADN : Acide DésoxyriboNucléique

ARN : Acide RiboNucléique

ARNi : Acide RiboNucléique interférent

At : Arabidopsis thaliana

BSMAP : Bisulfite Sequence Mapping Program

C : Cytosine

DMC : Differentially Methylated Cytosine

DMR : Région différentiellement méthylée ou « Differentially

DNMT : DNA Methyltransferase, Méthyltransférase d'ADN Methylated Region »

Gb: giga

kb: kilo base

Mb: mega base

pb : paire de base

PCR: Polymerase Chain Reaction

qRTPCR : «Reverse Transcription Quantitative Polymerase Chain Reaction »

SAM: S-Adenosyl Méthionine

SNPs: single-nucleotide polymorphism

WGBS: Whole Genome Bisulfite Sequencing

Contrôle épigénétique de l’adaptation de pétunia au stress hydrique

1 . Introduction

1.1. Structure d’accueil

Créé en 2012, l’IRHS (Institut de Recherche en Horticulture et Semences) est sous tutelle de l’Université d’Angers, d’Agrocampus Ouest et de l’INRA (Institut National de la Recherche Agronomique). Cette structure regroupe plus de 200 personnes y compris des généticiens, sélectionneurs, phytopathologistes, physiologistes, biochimistes, modélisateurs, bioinformaticiens et statisticiens qui coordonnent leurs expertises pour mettre en œuvre des approches de biologie intégrative. L’institut comprend treize équipes qui sont réparties dans trois pôles: « architecture et floraison sur rosiers et autres ornementales », « qualité et santé des fruits et légumes » et « semences, stress et pathogènes ». Les équipes peuvent être transversales entre les différents domaines, comme l’équipe de bioinformatique ou d’épigénétique. Durant mon stage, j’ai travaillé en collaboration avec deux équipes: l’équipe BIDEFI (Bioinformatics for plant Defense Investigations) sous la direction de Claudine Landes et l’équipe ARCH-E (Biologie Intégrative de l’Interaction Architecture et Environnement) dirigée par Alain Vian **(fig1).** L’équipe BiDEFI, en se basant sur l’informatique, la mathématique et la biologie, développe des méthodes et des outils afin d'analyser, modéliser ou prédire des informations biologiques à partir de résultats expérimentaux obtenus dans les domaines d'expertise de l'IRHS. L’équipe ARCH-E vise à étudier les effets des facteurs environnementaux sur la qualité et la quantité de la production des plantes cultivées, en particulier la qualité visuelle des plantes ornementales. Mon stage s’est déroulé à l’interface entre ces deux équipes en utilisant un outil bioinformatique pour m’aider à étudier l’effet du priming sur la méthylation de l’ADN et sur l’adaptation de pétunia à la sécheresse.

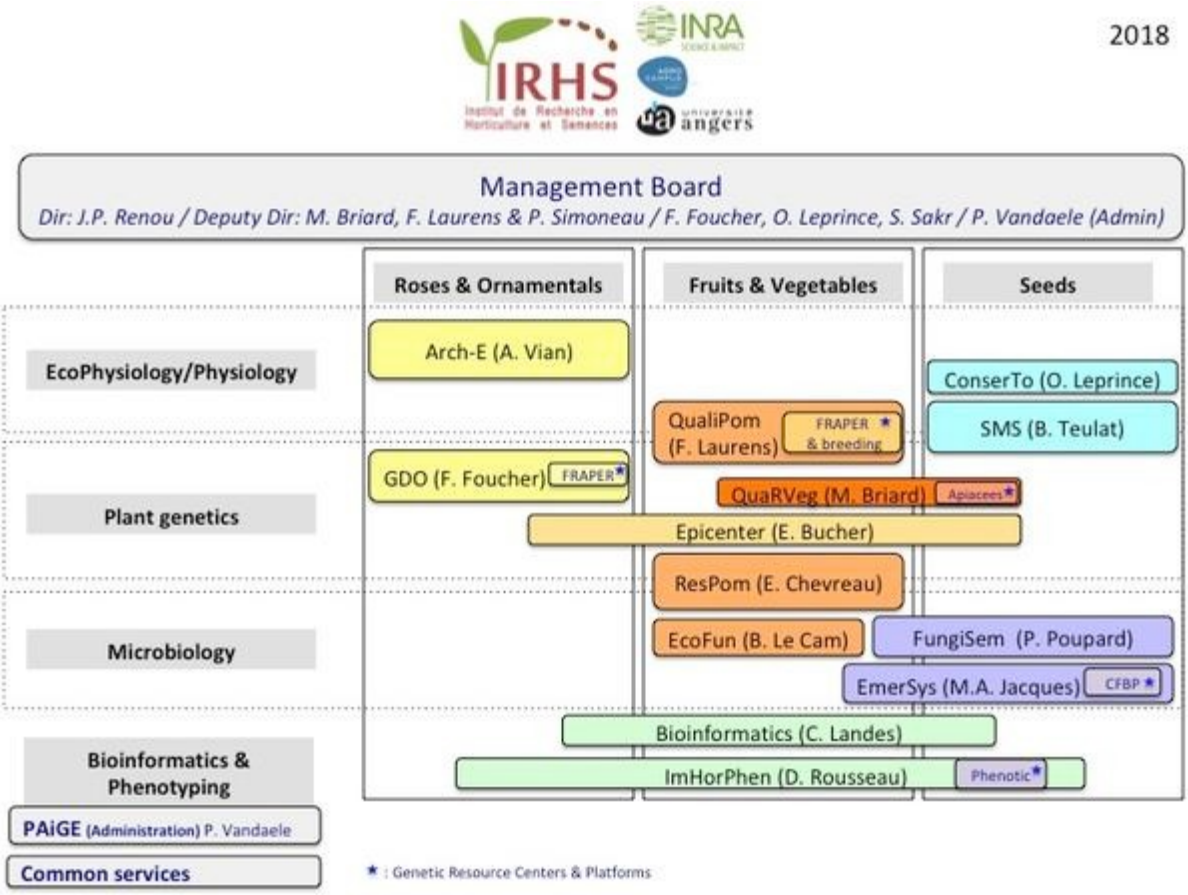


Figure 1: Les différents équipes de l'IRHS

1.2. Bibliographie

1.2.1. Introduction

En général, les animaux réagissent en fonction des contraintes environnementales et peuvent s'échapper. À l'inverse, les plantes sont fixées au sol par leurs racines et sont donc constamment soumises à des perturbations environnementales (Kinoshita et Seki 2014).

A cause du réchauffement climatique et de la forte demande pour des plantes adaptées aux conditions de croissance urbaine, de nombreuses recherches portent sur les mécanismes qui permettent aux plantes de s'adapter à la sécheresse. Le stress hydrique est la principale cause qui limite la croissance des plantes dans les régions arides et semi-arides et déclenche une série de réactions moléculaires, cellulaires et physiologiques (Tafaghodi *et al.,* 2018). Comme les perturbations environnementales peuvent se répéter, les plantes peuvent mémoriser de ces événements passés et utiliser cette mémoire pour mieux réagir lorsque ces événements se reproduisent. Les mécanismes épigénétiques pourraient jouer un rôle dans le contrôle de l'expression des gènes par de petits ARNs, la modification des histones et la méthylation de l'ADN. Ces mécanismes sont hérités par les divisions cellulaires mitotiques et, dans certains cas, peuvent être transmises à la génération suivante au travers la méiose et la reproduction sexuée (Kinoshita et Seki 2014).

1.2.2. L'épigénétique et l'adaptation au stress

L'épigénétique se définie comme l'étude de mécanismes qui modifient l'expression des gènes sans altérer la séquence de l'ADN et qui peuvent engendrer des changements potentiellement héritables dans le phénotype (Bossdorf *et al.,* 2008).

Plusieurs marques épigénétiques ont été étudiées, parmi lesquelles des modifications des bases de l'ADN telles que la méthylation de l'ADN et des modifications post-traductionnelles des histones (Dupont *et al.,* 2009). Notre travail étudie en particulier les changements de méthylation de l'ADN chez le pétunia comme réponse adaptative au stress hydrique. On a choisi d'étudier la méthylation de l'ADN en raison de sa fréquence de transmission par division cellulaire (mitose et / ou méiose), de ses rôles vis-à-vis de l'expression des gènes et de la facilité d'étude par des méthodes récentes en génomique « Whole Genome Bisulfite Sequencing » (WGBS).

1.2.3. Le stress hydrique et la mémoire du stress

L'eau possède des propriétés permettant le maintien de la vie et constitue une molécule fondamentale dans la plupart des processus cellulaires. En effet, sa disponibilité a un impact direct sur le développement et la survie des organismes vivants, en particulier les plantes. L'importance de l'eau a été démontrée par exemple chez *Arabidopsis thaliana* chez laquelle les signaux de sécheresse se traduisent par des effets sur l'expression génétique et un contrôle épigénétique (Yamaguchi-Shinozaki et Shinozaki 2005). Chez cette plante, les transcrits des gènes *RD29A (*Responsive to Dessication), *RD20* (préoxygénasse impliqué dans le métabolisme d'oxylipine en cas de stress biotique et abiotique) et *AtGOLS2* (galactitol synthase 2 qui catalyse la formation du galactitol à partir de UDP-galactose de myo-inositol) s'accumulent en réponse au stress hydrique et diminuent en quantité

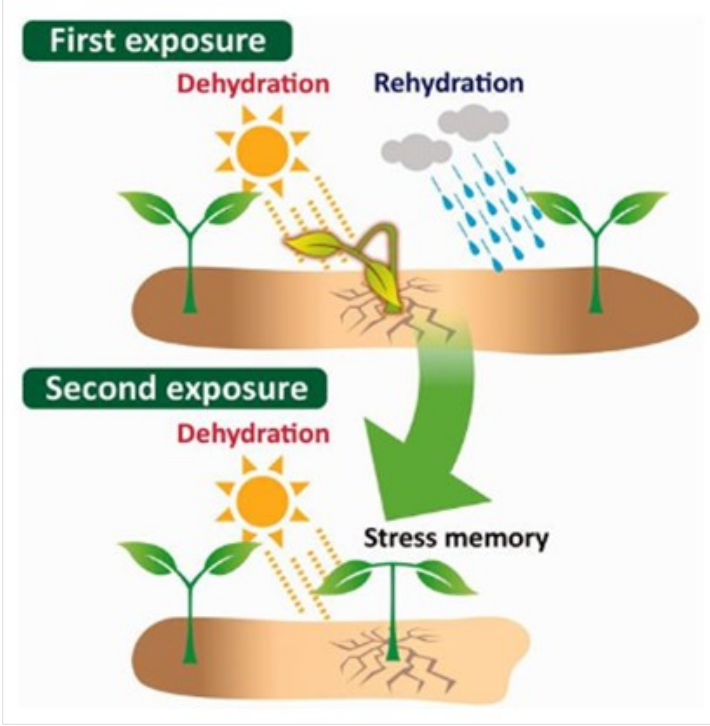


Figure 2 : Un exemple de mémoire du stress chez les plantes.

Les plantes disposent de mécanismes permettant la mémorisation du stress. Par exemple, une plante qui subit une période de sécheresse se flétrisse sous le stress de la déshydratation puis se rétablit après une réhydratation (panneau supérieur); au cours d'un second stress hydrique, la plante mémorise sa première exposition au stress, ce qui permet de mieux résister à la déshydratation et d'améliorer ses chances de survie (panneau inférieur) (Ding et al., 2012).

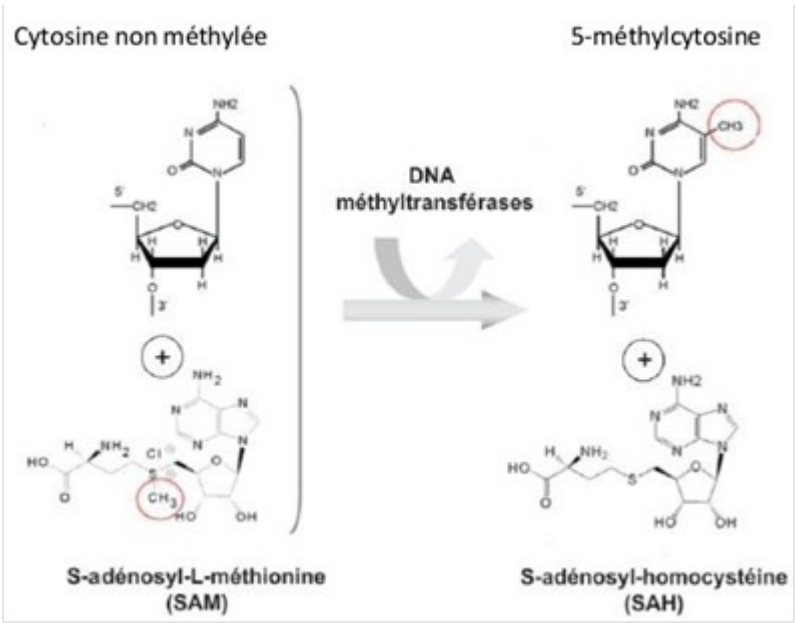


Figure 3: Processus de conversion d'une cytosine non méthylée en 5-méthylcytosine.

SAM correspond à « S-adénosyl méthionine » groupement donneur de méthyle (−CH 3). Une enzyme ADN méthyltransférase est nécessaire pour la réaction de méthylation au niveau d'une cytosine (Jammes et Renard, 2010).

après une réhydratation. En condition de stress hydrique, l'évolution des niveaux de transcrits est corrélée avec l'acétylation de la lysine 9 de l'histone H3 (H3K9ac) et l'abondance d'ARN polymérase II de manière transitoire. Après la réhydratation les niveaux de transcrits diminuent. Les modifications de l'histone triméthylée H3 corrélées à la transcription active, suggèrent que cette marque de la chromatine pourrait jouer un rôle dans la mémorisation du stress hydrique et agirait par une contrôle épigénétique sur la transcription de ces gènes. En effet, cette marque est induite par le stress hydrique et maintenue à certain niveau pendant la réhydratation (Kim *et al.,* 2012).

L'exposition multiple à la sécheresse permet aux plantes de mieux réagir à un nouveau stress en modulant plus rapidement l'expression de leurs gènes que des plantes non exposées auparavant à un stress hydrique **(fig2)**. Ding *et al.,* 2012 ont montré que, chez *A.thaliana*, les transcrits RD29A et COR15A s'accumulent progressivement: c'est-à-dire qu'ils sont plus abondants lors d'un second stress hydrique que lors du premier. Le changement progressif dans l'expression génique et l'accumulation des transcrits peuvent être le résultat d'une augmentation progressive de H3K4me3 et de la phosphorylation de la sérine 5 (Ser5P) de l'ARN polymérase II pendant le processus de récupération après déshydratation. Bien que la transcription des gènes diminue jusqu'au niveau basal chez les plantes d'*A.Thaliana* non soumises à un stress, les niveaux relativement élevés de H3K4me3 et l'abondance de l'ARN polymérase II (Ser5P) ont été maintenus et peuvent contribuer à la mémorisation du stress (Ding *et al.,* 2012).

1.2.4. La méthylation de l'ADN et mémoire du stress

La méthylation de l'ADN est la modification épigénétique la plus analysée. Elle consiste en une addition d'un groupe méthyle sur une cytosine, une adénine ou une guanine. Cependant, la méthylation de la cytosine est la plus fréquente, en particulier la méthylation 5C. Chez les plantes, la méthylation de l'ADN est une modification covalente de type CHG et CHH (H = A, G ou T) où le groupe CH_3 est apporté par le S-Adenosyl Méthionine (SAM) (**fig3**). Elle est médiée par une famille d'enzymes : les ADN méthyl transférases, (DNMT ; Bestor 2000) dont certaines sont responsables de nouvelles méthylations tandis que d'autres maintiennent les méthylations à travers la mitose et participent donc à la conservation de la mémoire épigénétique. La méthylation n'affecte que 3 à 8% des cytosines chez les mammifères contre 50% chez les plantes. En effet, la méthylation de novo de l'ADN est considérée comme une réponse adaptative de la plante à son environnement avec la possibilité de changements rapides et stables de l'expression des gènes en fonction des conditions environnementales (Zhang *et al.,* 2013). Plusieurs études ont montré une corrélation négative entre la méthylation de l'ADN, plus précisément la méthylation du promoteur, et l'expression des gènes tels que chez *A.thaliana* où une perte de méthylation (déméthylation) de l'ADN entraîne une induction de la transcription **(fig4**, Zilberman *et al.,* 2007).

La méthylation de nombreux gènes en réponse à un stress ou lors de processus développementaux indique le rôle important de cette modification épigénétique chez la plante (Kawakatsu *et al.,* 2016, Wang *et al.,* 2014; Meyer, 2015). Chez le soja, par exemple lors d'un stress salin, une réduction de la méthylation de l'ADN et une activation transcriptionnelle des gènes qui codent pour des facteurs de transcription liés à la réponse au stress ont été mis en évidence (Kim *et al.,* 2015).

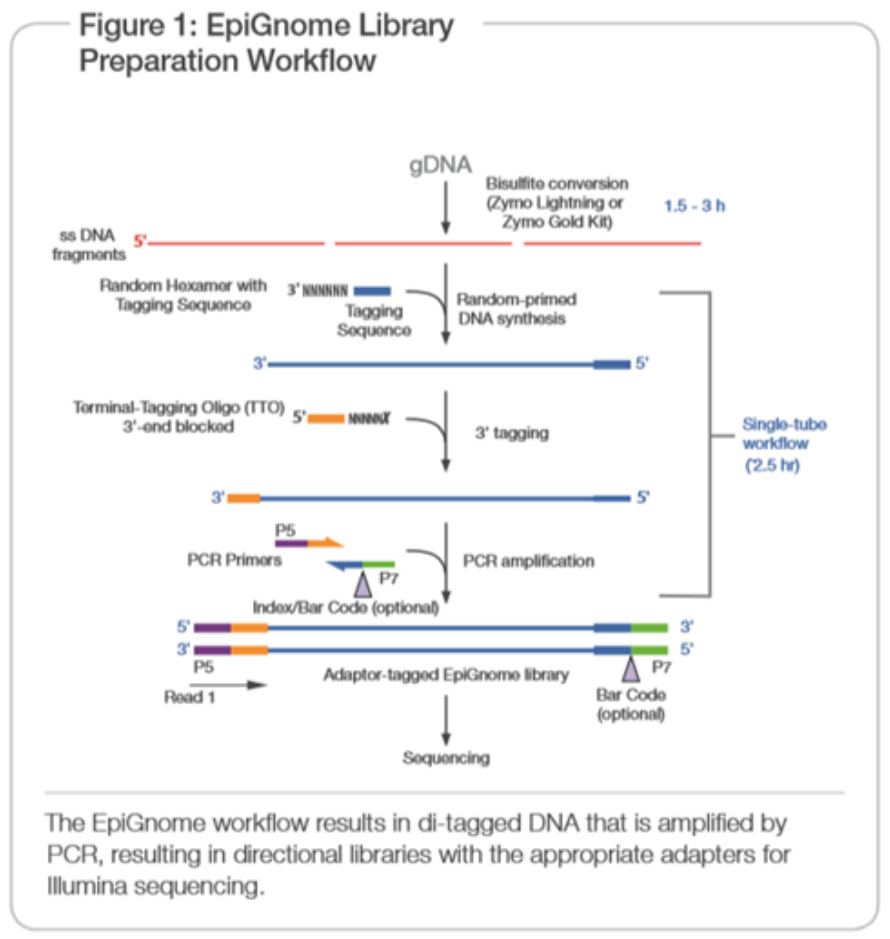


Figure 4: WGBS general workflow

(https://www.illumina.com/content/dam/illumina marketing/documents/products/appnotes/appnote-methylseq-wgbs.pdf) L'ADN traité au bisulfite puis purifié sur une colonne de centrifugation et est utilisé pour préparer la bibliothèque de séquençage à l'aide du kit EpiGnome ™. Dans cette procédure, l'ADN simple brin traité au bisulfite est amorcé de manière aléatoire en utilisant une polymérase capable de lire les nucléotides de l'uracile afin de synthétiser l'ADN contenant un marqueur de séquence spécifique. Les extrémités 3' des brins d'ADN nouvellement synthétisés sont ensuite sélectivement marquées avec une seconde séquence spécifique, ce qui donne des molécules d'ADN bi-marquées avec des étiquettes de séquence connues à leurs extrémités 5' et 3'. Ces étiquettes sont ensuite utilisées pour ajouter des adaptateurs Illumina P7 et P5 par PCR aux extrémités 5 'et 3', respectivement, du brin d'ADN original.

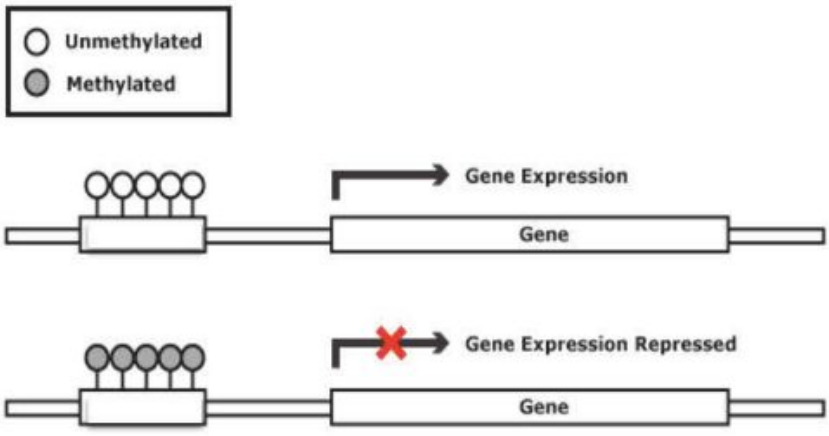


Figure 5: La méthylation de l'ADN et l'expression des gènes

En général, la méthylation de l'ADN est réversible. Chez le riz, par exemple, les modifications de la méthylation de l'ADN induites par le stress hydrique reviennent à leurs états initiaux après ré-arrosage (Wang *et al.,* 2011). Toutefois, certaines études ont montré que ces variations de méthylation pouvaient être mémorisées. On parle alors de « priming» ou de « mémoire du stress » (Crisp *et al.,* 2016). Ceci a été prouvé chez le riz cultivé avec une carence en phosphate suivie d'un apport en phosphate, où les éléments transposables (ETs) proches des gènes induits par l'environnement sont réprimés par hyperméthylation. Une corrélation temporelle a été montrée entre les changements transcriptionnels et les régulations épigénétiques en réponse à un stress (Secco *et al.,* 2015). Chez *A. thaliana,* en réponse à des contraintes osmotiques répétées, une mémoire du stress héritable sur au moins une génération a été observée. Cette mémoire est progressivement perdue en absence d'un stress répété. Il s'agit donc d'une mémoire transitoire qui repose sur la méthylation de l'ADN et transmise par la lignée germinale femelle grâce à l'activité d'ADN glycosylase dans la lignée germinale mâle (Wibowo *et al.,* 2016).

1.2.5. Séquençage bisulfite

De nombreuses méthodes d'analyse de la méthylation de l'ADN sont disponibles de nos jours (Plomion *et al.,* 2016; *Yong et al.,* 2016). Ces méthodes utilisent soit des anticorps anti-5-méthycytosine soit un traitement au bisulfite couplé à des analyses de type microarray ou de séquençage à haut débit (Yong *et al.,* 2016). La méthode de la plus utilisée actuellement est le « Whole Genome Bisulfite Sequencing » (WGBS ; **fig4**). Cette technique consiste en un séquençage Illumina classique après un traitement bisulfite de l'ADN suivi d'une amplification par PCR. Le traitement bisulfite conduit à une conversion chimique des cytosines (Cs) non méthylés en thymines (Ts) sans affecter les autres bases (**fig6**). Les thymines détectées dans les reads (lectures) après séquençage sont soit de la thymine originelle, soit des cytosines non méthylées du génome après traitement bisulfite. Le taux de conversion des Cs en Ts est généralement supérieur à 95 %. Des programmes d'alignement des lectures comme Bismark ou BSMAP, sont ensuite utilisés pour le mapping des reads sur le génome de référence. Une fois que le mapping des reads est effectué, le taux de méthylation de chaque cytosine le long du génome est déterminé. Théoriquement, pour une seule cellule haploïde et avec un taux de conversion de 100% en bisulfite, le taux de méthylation devrait toujours être égal à zéro ou à 1. Chez les cellules diploïdes (deux allèles de méthylation), le taux de méthylation peut avoir des valeurs entre 0 et 1 à cause des cellules ayant différents états de méthylation, de la variation biologique entre les cellules et les tissus, du taux imparfait de la conversion de C en T et du bruit généré par le mapping en raison des régions répétitives du génome. Toutefois, les valeurs comprises entre zéro et un sont minoritaires (Daccord, 2018).

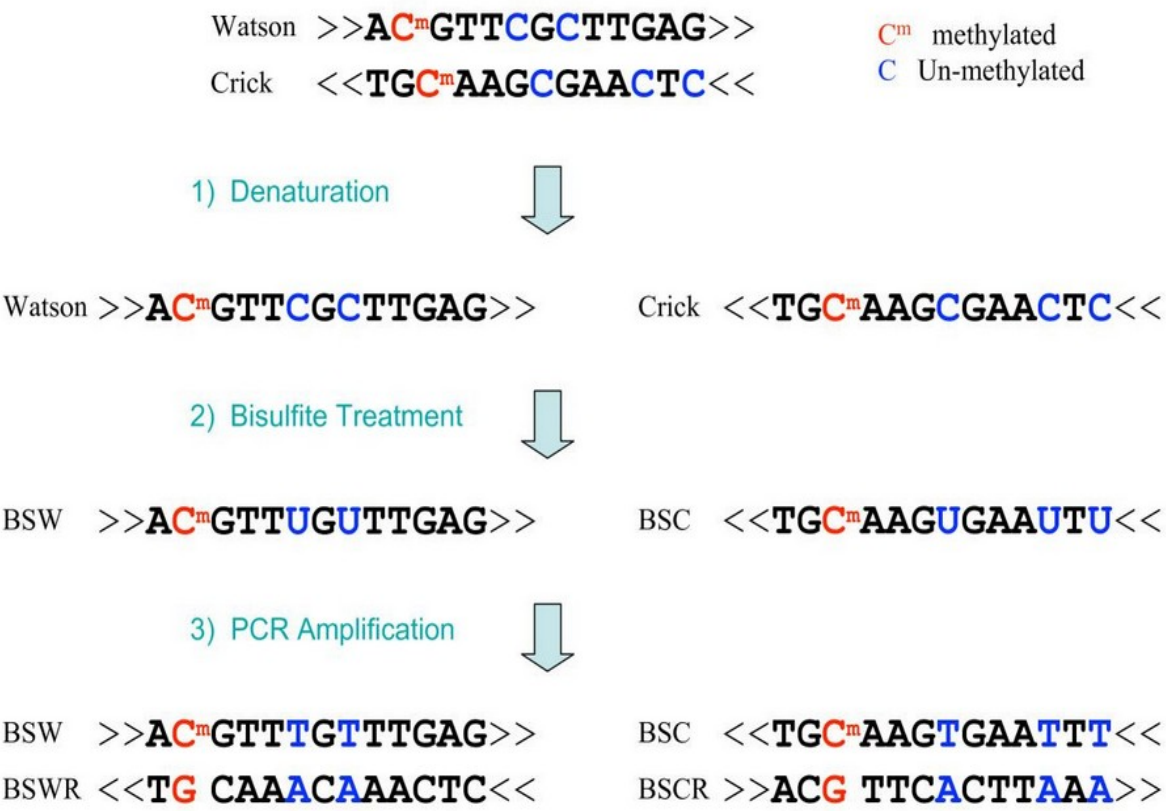


Figure 6: Pipeline de séquençage bisulfite.

1) dénaturation: séparer les brins de Watson et de Crick; 2) traitement au bisulfite: conversion de cytosines non méthylées (bleu) en uraciles; les cytosines méthylées (rouge) restent inchangées; 3) amplification par PCR des séquences traitées au bisulfite donnant lieu à quatre brins distincts: bisulfite Watson (BSW), bisulfite Crick (BSC), complément inverse de BSWR (BSWR) et complément inverse de BSC (BSCR) (Xi et Li 2009)

Tableau 1: Résumé des statistiques des assemblages des génomes de *P.axillaris N* et *P. inflata S6.*

Species	Category	Number	L50 (kb)	N50 (seqs)	Longest (Mb)	Size (Gb)
P. axillaris N	Total contigs	109,892	95.17	3,943	0.57	1.22
	Total scaffolds	83,639	1,236.73	309	8.56	1.26
P. inflata S6	Total contigs	213,254	34.99	9,813	0.57	1.20
	Total scaffolds	136,283	884.43	406	5.81	1.29

The assemblies are version Peaxi162 for *P. axillaris* N and version Peinf101 for *P. inflata* S6.

1.2.6. Analyse de la méthylation différentielle:

L'approche la plus utilisée pour détecter les différences de méthylation entre deux échantillons consiste à rechercher des régions différentiellement méthylées (DMR) à partir des données de séquençage au bisulfite du génome entier (WGBS). Pour chaque condition et par rapport au génome de référence, les taux de méthylation différentiels sont calculés en comparant le nombre de reads méthylés et non méthylées de deux échantillons. Chaque locus du génome où les niveaux de méthylation sont significativement différents entre les deux échantillons est appelé DMR. Il existe plusieurs outils de recherche de DMRs qui sont basés sur des méthodes statistiques. Deux types de méthodes de recherche de DMR ont été utilisés: les méthodes qui recherchent d'abord les cytosines différentiellement méthylées (DMC) et fusionnent les fragments denses contre les DMRs, et les méthodes qui calculent directement les différences de méthylation sur des régions (Daccord, 2018).

1.2.7. Pétunia : un modèle biologique

Le pétunia à fleur blanche (*Petunia axillaris*) est de la famille des solanacées originaire d'Amérique du Sud. C'est un parent sauvage de l'une des fleurs les plus populaires de l'horticulture, le Pétunia de jardin (*Petunia hybrida*). Son génome diploïde de 1,4 Gb donne un aperçu de l'évolution du pétunia cultivé. Le Pétunia, avec un nombre de chromosomes de base de x = 7, est le seul genre génétiquement accessible qui est caractérisé par une période de génération relativement courte, la facilité de culture et en particulier la disponibilité du marquage par transposon et de la transformation génétique. Pour cela Petunia est considéré comme un excellent modèle pour étudier l'ARNi (Napoli *et al.,* 1990), le développement, l'activité des transposons, l'auto-incompatibilité génétique et les interactions avec les microbes, les herbivores et les pollinisateurs (Bombarely *et al.,* 2016). De plus, des travaux précurseurs sur les régulations épigénétiques s'appuyant sur les variations de couleur de la fleur ont été conduits sur cette espèce (Gerats et Vandenbussche, 2005).

Le *P. hybrida* du commerce est issu de croisements entre *P. axillaris* pollinisé par le papillon et des *P. inflata* pollinisé par des abeilles **(fig7)** (Segatto *et al.,* 2014). Les premiers hybrides ont été produits par des horticulteurs européens au début du XIXe siècle. De ce fait, la grande diversité phénotypique actuelle des pétunias de jardin est le résultat de deux siècles de culture commerciale intense et de croisements successifs des différentes accessions des deux parents. La figure 4 montre *P. axillaris* N et *P. inflata* S6, deux accessions de laboratoire représentant les parents de *P. hybrida* (Bombarely *et al.,* 2016).

Génomes parentaux :

Les génomes parentaux ont été séquencés récemment (Bombarely *et al.,* 2016).

Pour *P. axillaris,* un assemblage de novo hybride a été réalisé en utilisant une combinaison de technologies à lecture courte (Illumina; couverture 137X) et à lecture longue (PacBio; couverture 21X), tandis que pour *P. inflata S6*, seulement des lectures courtes ont été utilisées (Illumina; Illumina; couverture 135X). Les séquences d'assemblage ont une taille de 1,26 Gb pour *P. axillaris* et de *1,29* Gb pour *P. inflata* (Tableau 1). La taille estimée des deux génomes est de 1,4 Go, en utilisant des k-mer de taille 31. Les statistiques d'assemblages des génomes montrent que pour *P. axillaris*, on a 83,639 scaffolds avec un L50 de 1236.73 kb (L50 est la taille du scaffold médian) et pour *P. inflata,* on a 136,283 scaffolds avec un L50 de 884,43 Kb. La fraction non assemblée estimée du génome comprend environ 140 Mo pour *P. axillaris* et environ 110 Mo pour P. *inflata* qui est sans doute dûe au grand nombre de séquences répétitives.

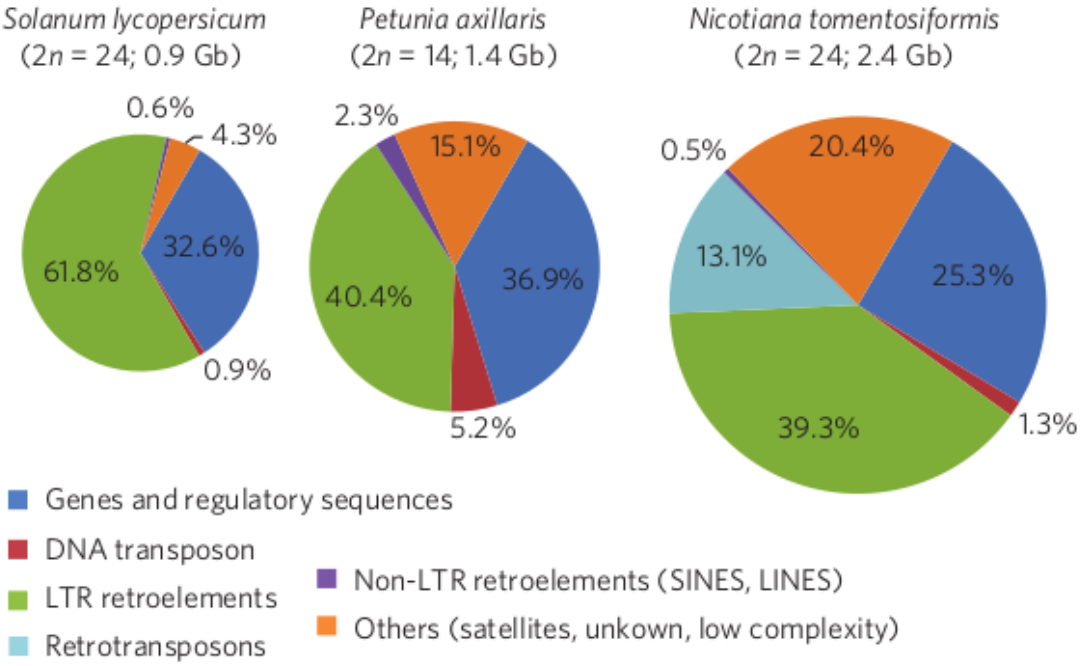


Figure 7: Organisation comparée du génome de *Solanum lycopersicum, P. axillaris et Nicotiana tomentosiformis.*

Les cercles sont proportionnels à la taille du génome; les séquences régulatrices et les classes de répétition sont indiquées dans les segments

L'annotation du génome a identifié 32 928 gènes codant pour des protéines de *P. axillaris* et 36 697 gènes codant pour des protéines de *P. inflata* avec une moyenne de 5,2 et 5,1 exons par gène codant pour une protéine et une taille de protéine prédite moyenne de 393 et 386 acides aminés, respectivement. Le génome de *P.axillaris* est moins fragmenté que celui de *P.inflata* (**tableau 1,** Bombarely *et al.,* 2016)..

La figure montre que la proportion des gènes dans le génome de *P. axillaris* est plus importante que dans ceux de *N. tomentosiformis* et de *Solanum lycopersicum*. Le génome de pétunia contient plus de DNA transposon (5,2%) par rapport aux deux autres génomes avec une proportion de LRT retroelements proche de celle de *N. tomentosiformis*. Par ce qu'il s'agit d'une première annotation du génome de *P. axillaris,* on remarque une proportion non négligeable des « others » contenant des satellites est des éléments inconnus (15,1%) (Bombarely *et al.,* 2016)..

Origine des génomes de *P. hybrida* : Les comparaisons des deux séquences du génome avec les données transcriptomiques de trois lignées non apparentées de *P. hybrida,* nommées Mitchell, R27 et R143 ont révélé une histoire complexe du pétunia de jardin. La plupart des gènes analysés (environ 20 000) pourraient être attribués à *P. axillaris* (environ 15 000) avec seulement 600 gènes environ qui seraient attribués à *P. inflata*. Cela indique que le parent *P. inflata* n'apporte qu'une contribution mineure dans le génome de *P. hybrida*. La dominance du génome du parent sauvage pourrait être le résultat de la sélection des couleurs qui nécessite un fond génétique avec des mutations récessives responsable de la pigmentation. Environ 2 000 gènes de *P. hybrida* contiennent un pourcentage élevé de SNPs (single-nucleotide polymorphism) qui proviennent probablement d'un ancêtre inconnu (Bombarely *et al.,* 2016).

1.3. Objectifs du stage

La capacité de fixer la « mémoire du stress » au travers des générations et l'identification de gènes affectés lors de la réponse au stress représentent des outils pertinents pour les sélectionneurs (Rodríguez López & Wilkinson, 2015). Le sujet de ce stage s'inscrit dans le cadre de l'amélioration des plantes cultivées. L'hypothèse du projet de recherche dans lequel s'inscrit ce stage est que le stress hydrique sur des pieds-mères pourrait imprimer des changements épigénétiques stables dans le génome de ces plantes qui pourraient être transmis par multiplication végétative à des clones et ainsi conférer à ces plantes une meilleure adaptation à des stress hydriques ultérieurs. Les plantes étudiées sont issues d'une multiplication végétative par bouturage à partir d'une seule plante-mère. L'objectif du mon stage est d'analyser grâce à des outils bio-informatiques le méthylome de clones issus de cette plante-mère de pétunia cultivés soit en condition de confort hydrique, soit en condition de stress hydrique afin d'identifier les régions puis les gènes différentiellement méthylés par la comparaison des séquences génomiques.

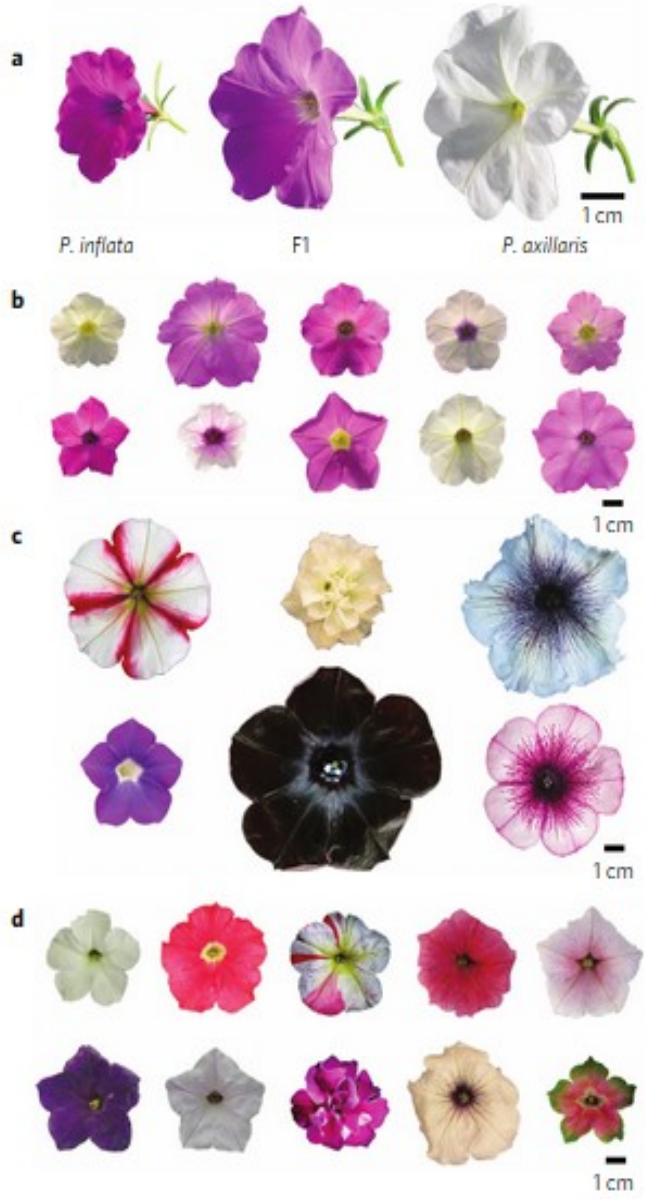


Figure 8: Origine et diversité des fleurs de *P. hybrida.*
a: *P. inflata* S6, *P. axillaris* N et leur F1. **b,** individus sélectionnés parmi la population de *P. inflata* S6 × *P. axillaris* N F2. **c,** accessions commerciales de *P. hybrida*. **d,** p. les hybrides et les mutants. Rangée 1, de gauche à droite: Mitchell (W115); R27; ligne de transposon W138; R143; ph3 vacuolaire mutant de couleur pâle comparé à l'isogénique R143. Mitchell, R27 et R143 ont été utilisés pour l'analyse transcriptomique. Rangée 2, de gauche à droite: V26; V26 avec transgène ARNi CHS (images fournies par J. Kooter, VU Amsterdam); mutant homéotique pMADS3RNAi / fbp6; mutant an2; aveugle mutante homéotique (Bombarely *et al.,* 2016).

2. Matériel et Méthodes

2.1. Matériel Végétal

Une plante de *Petunia hybrida* var Carnival Red Star, issue de semis a été sélectionnée comme unique pied-mère pour l'expérimentation. Cette plante a été multipliée par bouturage et les clones produits ont été soumis à un stress hydrique contrôlé pendant 7 générations successives de multiplication végétative par bouturage. D'autres plantes issues du même pied-mère ont, elles, été cultivées en confort hydrique pendant les 7 générations clonales successives. Les réponses phénotypiques et physiologiques des plantes au stress ont été analysées en parallèle de ce stage. L'ADN des plantes du clone de première génération cultivé en confort hydrique ainsi que l'ADN des plantes de 7eme génération cultivées en stress hydrique a été isolé et adressé en amont du stage pour WGBS à la plateforme BGI en Chine. L'analyse bio-informatique des méthylomes des 2 échantillons transmis par la plateforme BGI fait l'objet de ce stage. La comparaison du niveau de méthylation global de l'ADN du méthylome des plantes de pétunia exposés au stress hydrique par rapport aux plantes qui n'ont pas subi le stress va permettre de déterminer des corrélations entre les changements épigénétiques et l'acclimatation des plantes au stress hydrique. La stabilité des changements épigénétiques au fil des générations sera étudiées ultérieurement par l'étude des niveaux de méthylation des DMR identifiées lors de ce stage.

Le génome de *P.axillaris* est moins fragmenté que celui de *P. inflata* (tableau 1), nous l'avons choisi, alors, comme génome de référence pour les analyses informatiques de méthylation de l'ADN. Deux autres génomes qui correspondent au lot 1 de *P. hybrida* (1ère génération clonale sans stress hydrique) **(fig8)** et au lot 2 de *P. hybrida* (7-ème génération clonale soumise au stress hydrique) ont été utilisés pour mettre en évidence la méthylation différentielle en réponse au stress hydrique **(fig9).**

2.2. Séquençage bisulfite

Le séquençage bisulfite génome entier (WGBS) a été réalisé à la plateforme BGI en Chine. Le séquençage a été réalisé en « paired end » (2 x 100 bases). Les résultats ont été fournis sous la forme de fichiers au format fastqz avec une couverture théorique de « 30 X » minimum. Ces fichiers contiennent les séquences et leurs scores de qualité (score Phred). Ce score évalue la confiance du séquençage. Les lectures sont des courtes séquences de 150 pbs obtenus par le WGBS. Pour le lot1, les reads correspondent au fichier FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_1_fastqc (R1) de 4096 octets et au fichier FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_1_fastqc.zip (R2) de 727374 octets. Pour le lot2 le fichier les reads correspondent au fichier FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_2_fastqc (R1) de 4096 octets au fichier FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_2_fastqc.zip (R1) de 698310 octets.

Figure 9: Plantes du lot 1 de pétunia cultivé sous confort hydrique

Figure 10: Plantes du lot 2 de pétunia cultivé sous stress hydrique

2.3. Mapping des reads bisulfite sur le génome de référence

La première étape de la recherche de DMRs avec les données WGBS est l'alignement des reads sur le génome de référence de P. Axillaris à l'aide du logiciel Bisulfite Sequence Mapping Program BSMAP (Xi et Li 2009) avec les paramètres suivants:

```
bsmap -a lot1/FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_1_fastqc -b lot1/FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_1_fastqc.zip -d Petunia_axillaris_v1.6.2_genome.fasta -o lot1_out.bam -q 2
bsmap -a lot2/FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_2_fastqc -b lot2/FCHC3JFDSXX_L1_PETcabMAAAAAAA-30_2_fastqc.zip -d Petunia_axillaris_v1.6.2_genome.fasta -o lot2_out.bam -q 2
```

Chaque échantillon a été traité séparément avec les éléments suivants : 2 fichiers d'entrée (qui sont les résultats du bisulfite sequencing) R1 et R2 lus dans les 2 sens), 1 fichier correspondant au génome de référence, le programme bsmap, et le fichier de sortie. L'option « -a »sert à indiquer le nom du fichier contenant les reads correspondant à un extrait de la paire (séquençage paire end). l'option « -b » sert à indiquer le nom du fichier de reads de l'autre extrémité de la séquence. l'ption « -d » sert à indiquer le nom du fichier contenat le génome de référence (*P.axillaris*). L'option « -o » sert à désigner le nom du fichier de sortie. l'option « -q 2 » est un paramètre de qualité pour le nettoyage des extrémités 3' des reads. Que les bases ayant un score PHRED superieur ou égal à 2 ont été retenus. Avec ce programme, un fichier au format bam (version de sam compressée) a été obtenu pour chaque lot. Ce logiciel construit un tableau pré-compilé divisé en plusieurs sous tableaux pour minimiser le temps de latence perdu pendant la recherche de la position potentielle des k-mers sur le génome.

2.4. Annotation des gènes de stress hydrique

Sur le site de NCBI et en utilisant des mots clé, aucun gènes annotés est impliqué dans la réponse au stress hydrique n'a été trouvé dans le génome de *P. axillaris* de référence.

On a donc procédé à une annotation semi-automatique des gènes de stress hydrique chez le pétunia qui se fait en deux étapes :

-La récupération automatique de tous les identifiants de gènes de plantes annotés comme liés au stress hydrique chez quelques solanacées

-Le téléchargement les gènes des solanacées et blaster un à un tous les gènes candidats contre le génome de *P. axillaris* pour trouver leur homologue chez pétunia.

Une recherche dans la banque « gene » du site NCBI à l'aide du mot clé « drought » en limitant la recherche aux « greens plants » dans le champ organisme a été effectuée. Un tableau de 1270 gènes de réponse au stress hydrique chez les plantes a été obtenu. Une commande a été faite unix pour extraire les colonnes 6 et 7 du fichier gene_result1.txt en considérant la tabulation comme séparateur de colonnes qui contiennent les identifiants et la description des gènes. Puis, la redirection Unix a été faite pour stocker ce résultat dans un fichier appelé « select_gene.txt » qui contient seulement deux colonnes.

```
cut -f6,7 gene_result1.txt > select_gene.txt
```

```
def listegene(filename, fileout):
        fpi=open(filename)
        fpo=open(fileout, 'w')
        identifiant=''
        for i in fpi.readlines():
                colonnes=i.strip().split('\t')
                try :
                        if 'AT' in colonnes[1]:
                                identifiant=colonnes[1]
                        elif 'AT' in colonnes[0]:
                                identifiant=colonnes[0]
                        elif 'LOC' in colonnes[0]:
                                identifiant=colonnes[0]
                        elif 'LOC' in colonnes[1]:
                                identifiant=colonnes[1]
                        else:
                                print(colonnes[0], colonnes[1])
                except IndexError:
                        identifiant= colonnes[0]
                fpo.write(identifiant + '\n')
        fpi.close()
        fpo.close()
        return

##############################

filename=input('nom du fichier: ')
fileout=input('nom du fichier de sortie: ')
listegene(filename,fileout)
```

Figure 11: « Script « select_gene_ID.py » de sélection des gènes.

La première condition est de voir si les gènes qui commencent par AT sont dans les colonnes 1 ou 0, la deuxième condition vérifie si les gènes qui commencent par LOC sont dans la colonne 0 ou 1 et la dernière condition permet d'afficher les gènes qui ne commencent ni par LOC ni par AT dans le terminal.

```
LOC4328874      OSNPB_0202
COP1    AT2G32950
MPK6    AT2G43790
FT      AT1G65480
GI      AT1G22770
DREB1A  AT4G25480
PHOT1   AT3G45780
AGB1    AT4G34460
ABF3    AT4G34000
ABI4    AT2G40220
MYB96   AT5G62470
SIZ1    AT5G60410
COR15A  AT2G42540
YUC6    AT5G25620
DREB2A  AT5G05410
ABI5    AT2G36270
ABA3    AT1G16540
KIN11   AT3G29160
TOC1    AT5G61380
MAX2    AT2G42620
PLDDELTA        AT4G35790
```

```
AT1G56280
AT5G49230
AT1G76080
AT3G05700
AT4G02200
AT3G06760
AT1G02750
AT4G15910
AT1G80710
AT5G26990
AT1G73330
LOC110922489
LOC113734562
LOC4337576
LOC100272898
LOC100193039
LOC100282272
LOC100282084
LOC110915766
LOC113736428
LOC110796218
```

Figure 12: Exemple d'un fichier d'entrée (à gauche) et un fichier de sortie (à droite)

Un script python appelé **« select_gene_ID.py** » a été écrit pour mettre dans une seule colonne les gènes qui commencent par LOC ou AT et d'afficher le reste des gènes dans le terminal **(fig11)**. Le programme « select_gene_ID.py » crée le fichier de sortie `select_gene.txt` qui contient la liste des identifiants de gène liés au stress hydrique chez les plantes. Le fichier `select_gene.txt` contient des répétitions, la commande suivante sur unix a été utilisée pour les éliminer:

```
uniq select_gene.txt > select_genes.txt
```

A l'issue de cette étape, un tableau avec une seule colonne contenant les gènes d'intérêt a été obtenu. Afin d'obtenir les séquences des gènes de manière automatique, un script biopython appelé « **extract_genes.py** » a été écrit et qui permet le téléchargement des séquences au format fasta depuis le site du NCBI **(fig13).**

Dans la liste de gènes obtenue précédemment, aucun gène de réponse au stress hydrique de pétunia a été identifié. Aussi, en partant du même tableau de la première approche qui contient une liste de 1270 gènes candidats, des blastn des séquences fasta de ces gènes ont été effectués sur le génome de *P. axillaris* en utilisant le site de Sol Genomics Network (https://solgenomics.net/tools/blast/?db_id=272). Un seuil de p-value inférieur ou égal à 10^{-100} a été choisi. Ce travail n'a pas donné de résultats intéressants. Une nouvelle recherche a été effectuée en ciblant les gènes de réponses au stress hydrique chez des plantes de la famille des solanacées. On a identifié la position de début et fin pour chaque gène (début et fin) sur les scaffolds de *P. axillaris* pour pouvoir voir après, sur le génome browser déployé à l'IRHS pour ce projet, si ces gènes sont différentiellement méthylés ou pas.

2.5. Recherche des régions différentiellement méthylées

Pour rechercher les zones différentiellement méthylées entre les méthylomes de pétunia cultivé en confort hydrique et de pétunia cultivés sous stress hydrique, nous avons utilisés deux approches en parallèle :

- Une approche par gènes candidats (DMC par région)

- Une approche sans *a priori* (DMR)

2.5.1. L'approche *a priori* :

Le génome de *P. axillaris* est disponible sur le génome browser de l'IRHS à l'adresse: http://irhs-001/jbrowse/?data=petunia&loc=Peaxi162Scf00000%3A1..286800&tracks=DNA%2CGene%20models&highlight=. Grâce aux positions des gènes qu'on a annoté comme étant lié au stress, les reads bissulfite mappés sur le génome de *P. axillaris* par BSMAP peuvent être visualisés. Une inspection visuelle permettra d'apprécier si oui ou non les gènes sont différentiellement méthylés (étape de screening). Pour les gènes d'intérêt, un test statistique peut être effectué pour valider ou non la zone comme différentiellement méthylée.

2.5.2. L'approche sans a priori : le programme DMR tools

Pour identifier les DMRs entre les 2 lots de pétunia, le génome est divisé en petites régions chevauchantes. Un test Wilcoxon entre les deux conditions va être appliqué. En plus de la p-value du test Wilcoxon, autres valeurs seront calculées : la moyenne de la couverture des reads par région, la moyenne de la différence de méthylation par région. Après, un seuil de p-value doit être défini pour filtrer les DMRs. Les DMRs seront filtrés selon aussi un seuil minimum de différence du niveau de méthylation.

```
from Bio import Entrez
from Bio import SeqIO

Entrez.email = "toto@gmail.com"

# ====== EXTRACTION D'UNE SEQUENCE =====
# La fonction Entrez.efetch( ) se cdonnecte sur le NCBI et recupere la sequence seqID
handle = Entrez.efetch(db="nucleotide", rettype="fasta", retmode="text",id="6273291")
# la fonction SeqIO.read() lit la sequence nucleotidique au format fasta
seq_record = SeqIO.read(handle, "fasta")
handle.close()
#ecriture de la taille de la sequence
print "%s with %i nt" % (seq_record.id, len(seq_record.seq))
#ecriture de fichier des resultats
fpo=open("toto.fasta", 'w')
SeqIO.write(seq_record, fpo, "fasta")
fpo.close()
#=======================================
```

Figure 13: script « extract_genes.py » d'extraction de la séquence fasta du gène 6273291 toto fasta est un exemple du nom de fichier fasta de sortie.

3. Résultats

3.1. Résultats du BSMAP

L'exécution du programme bsmap a duré environ 1310000 secondes soit 15 jours de calculs pour chaque échantillon répartis sur 6 processeurs. La taille des fichiers de sortie de format bam (fichier binaire) est respectivement 36 661 988 404 octets (environ 36 Go) pour le lot 1 et 35646608798 octets (35 Go) pour le lot 2. La figure 10 montre que 52,8 % des reads du lot1 qui s'alignent sur le génome de référence et 47,4% pour le lot 2 **(fig14)**. Les fichiers de sortie de cette étape ont été mis sur génome browser.

3.2. Résultats de démarche *a priori*

Grâce au positionnement de quelques gènes candidats sur le génome (tableau), les scaffolds correspondant sur les génomes du lot1 et du lot2 ont pu être identifiés **(fig15).** Les positions des SNPs permettant de voir si ces gènes sont différentiellement méthylés en réponse au stress hydrique ne sont pas encore présents dans génome browser. Ce problème est en cours de résolution par les ingénieurs de l'équipe.

3.3. Résultats de l'approche sans *a priori*

Pour l'instant, les résultats résument à l'annotation semi-automatique des gènes de la résistance au stress chez Pétunia. Un script nommé « extract_UID.py » a été créé pour permettre l'extraction automatique d'une liste de gènes au format fasta à partir des identifiants GI dans la banque « nucleotide » du NCBI. Ce programme génère un fichier fasta par séquence dont le nom est construit en ajoutant l'extension « .fasta » après le nom de l'identifiant GI **(fig16)**. Dans l'étape suivante, on va tourner DMRs Tools actuellement en cours de débogage.

3.4. Résultats de l'approche gènes candidats

Le tableau montre quelques gènes impliqués dans la réponse au stress hydrique chez des solanacées, *A. thaliana* et *Glycine max.* Ces gènes ont des tailles différents et positionnés sur des scaffolds différents. Parmi ces gènes, il y a ceux qui sont liés à la voie de signalisation de l'ABA tels que UGT75C1, SlbZIP1, AREB1 chez *Solanum lycopersicum* **(tableau2).** Ces gènes sont probablement parmi les 12, 398 gènes « en commun » qui existe chez chez *Solanum lycopersicum, Solanum tuberosum P. hybrida, P. inflata* et *Nicotiana benthamian* **(fig17).**

a

```
ykouki@node2:~$ bsmap -a lot1.bam -b lot1.bam -d  Petunia_axillaris_v1.6.2_genome.fasta -o lot1_out.bam -q 2 -p 6
[bsmap] @Thu May 23 10:58:55 2019       loading reference file: Petunia_axillaris_v1.6.2_genome.fasta   (format: FASTA)
[bsmap] @Thu May 23 10:59:12 2019       83639 reference seqs loaded, total size 1259220201 bp. 17 secs passed
[bsmap] @Thu May 23 11:04:36 2019       create seed table. 341 secs passed
[bsmap] @Thu May 23 11:04:36 2019       Pair-end alignment(6 threads),
        Input read file #1: lot1.bam    (format: BAM)
        Input read file #2: lot1.bam    (format: BAM)
        Output file: lot1_out.bam        (format: SAM, automatically convert to BAM)

[bsmap] @Fri Jun  7 15:11:13 2019       total read pairs: 300602919     total time consumed:  1311138 secs
        aligned pairs: 158579504 (52.8%), unique pairs: 129008808 (42.9%), non-unique pairs: 29570696 (9.8%)
        unpaired read #1: 66046547 (22.8%), unique reads: 48478063 (16.1%), non-unique reads: 17568484 (5.8%)
        unpaired read #2: 62341090 (20.7%), unique reads: 45234925 (15.0%), non-unique reads: 17106165 (5.7%)
```

b

```
ykouki@node2:~$ bsmap -a lot2.bam -b lot2.bam -d  Petunia_axillaris_v1.6.2_genome.fasta -o lot2_out.bam -q 2 -p 6
[bsmap] @Thu May 23 11:32:28 2019       loading reference file: Petunia_axillaris_v1.6.2_genome.fasta   (format: FASTA
[bsmap] @Thu May 23 11:32:46 2019       83639 reference seqs loaded, total size 1259220201 bp. 18 secs passed
[bsmap] @Thu May 23 11:39:24 2019       create seed table. 416 secs passed
[bsmap] @Thu May 23 11:39:24 2019       Pair-end alignment(6 threads),
        Input read file #1: lot2.bam    (format: BAM)
        Input read file #2: lot2.bam    (format: BAM)
        Output file: lot2_out.bam        (format: SAM, automatically convert to BAM)
[bsmap] @Fri Jun  7 16:34:13 2019       total read pairs: 300314151     total time consumed:  1314105 secs
        aligned pairs: 142437266 (47.4%), unique pairs: 116368918 (38.7%), non-unique pairs: 26068348 (8.7%)
        unpaired read #1: 87886920 (29.3%), unique reads: 62991958 (21.0%), non-unique reads: 24894962 (8.3%)
        unpaired read #2: 63417556 (21.1%), unique reads: 45077191 (15.0%), non-unique reads: 18340365 (6.1%)
ykouki@node2:~$
```

Figure 14: résultats d'alignement avec bsmap (a) lot1 et (b) lot2.

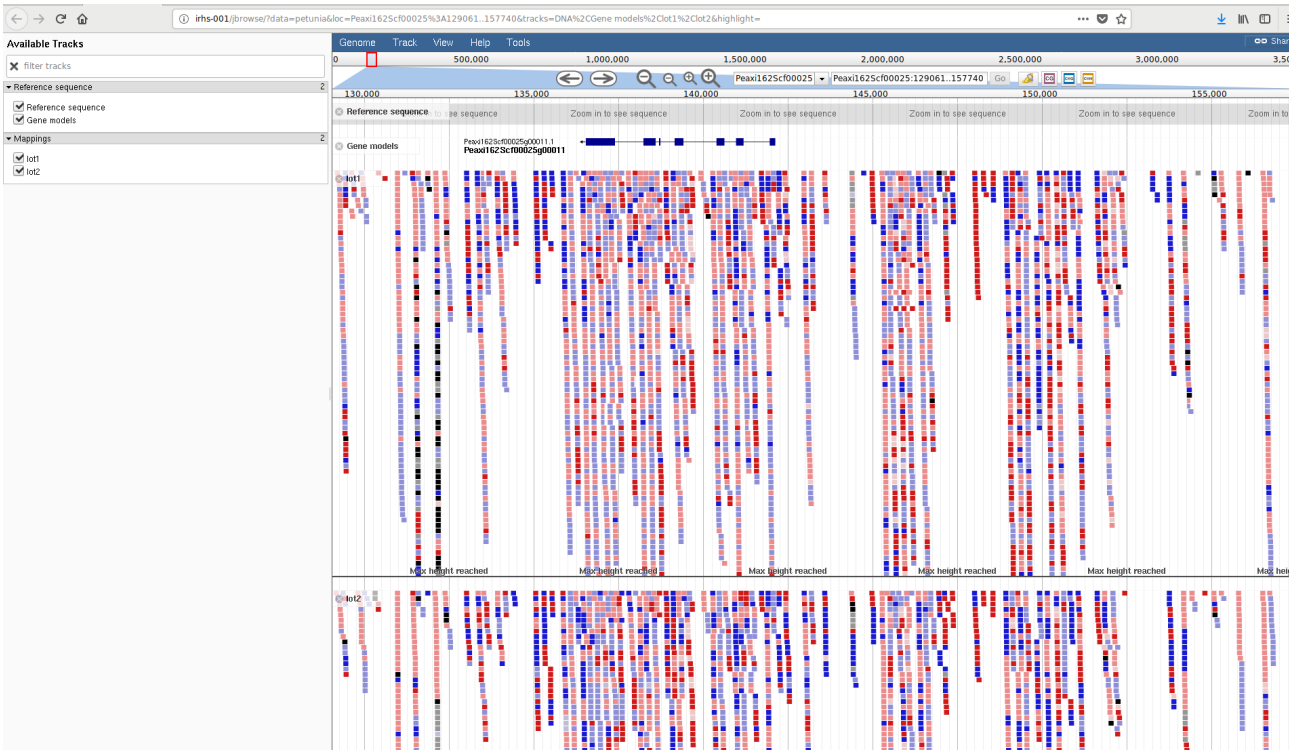


Figure 15 : Exemple de positionnement d'un gène sur les scaffolds de lot1 et 2 sur génome browser.

Tableau 2: Positionnement de quelques gènes impliqués dans la réponse au stress hydrique sur les scaffolds de lot1 et lot 2.			
Gènes	**scaffold**	**Position**	**e-value**
sfr2 beta-glycosidase-like [*Solanum lycopersicum* (tomato)]: La découverte de la fonction SFR2 de la tomate dans la sécheresse et la résistance au sel fournit des informations supplémentaires sur les mécanismes généraux de tolérance au stress basés sur le remodelage des lipides membranaires (Wang K, *et al*. 2016)	Peaxi162Scf00228	773727-772971 780490-780144	0.0 e-104
ftsH6 FtsH protease [*Solanum lycopersicum* (tomato)]: Le gène LeftsH6 est un gène ftsH typique induit par la chaleur, dont l'expression n'est pas induite par d'autres stress tels que le froid, la sécheresse, la salinité et une lumière intense Sun AQ, *et al*., (2006)	Peaxi162Scf00344	1693768-1693195 1692904-1692518 1691961-1691516	0.0 e-148 e-127
NPR1 regulatory protein NPR1 [*Solanum lycopersicum* (tomato)]: Les données suggèrent que SlNPR1 est impliqué dans la régulation de la réponse à la sécheresse chez les tomates.	Peaxi162Scf00221	23627-24678	0.0
UGT75C1 UDP-glycosyltransferase 75C1 [*Solanum lycopersicum* (tomato)] : Le SlUGT75C1 joue un rôle crucial dans la maturation des fruits médiée par l'acide abscissique (ABA), la germination des graines et la réponse à la sécheresse chez la tomate. Il est de type glycosyltransférase_GTB; Les glycosyltransférases catalysent le transfert de fragments de sucre de molécules donneuses activées à des molécules accepteuses spécifiques, en formant des liaisons glycosidiques. La molécule accepteuse peut être un lipide, une protéine, un composé hétérocyclique ou un autre glucide (Sun Y, *et al*., 2017)	Peaxi162Scf00286	1030995-1029584	0.0
mpka1;1 Mitogen-activated protein kinase [*Solanum lycopersicum* (tomato)]: La désactivation des gènes SpMPK1, SpMPK2 et SpMPK3 chez la tomate réduit la tolérance à la sécheresse induite par l'acide abscissique (Li C, *et al*. 2013)	Peaxi162Scf00483	130602-131004	e-178

CURL3 brassinosteroid LRR receptor kinase [*Solanum lycopersicum* (tomato)]: L'amélioration de la tolérance à la sécheresse en réponse aux BR indique une voie de signalisation en aval des BR et qui diffère de celle de BRI1 (Nie *et al.*, 2019).	Peaxi162Scf00025	3449330-3452379	0,0
LOC107760279 protein DEHYDRATION-INDUCED 19 homolog 3-like [Nicotiana tabacum (common tobacco)]	Peaxi162Scf00358 (>200) Peaxi162Scf00224 (>200) Peaxi162Scf00073	97373-97641 620173-619905 83041-82656	e-139 e-139 e-127
LOC107793787 thioredoxin-like protein CDSP32, chloroplastic [*Nicotiana tabacum* (common tobacco)]	Peaxi162Scf00174	837485-837017	e-168
LOC107771551 protein DEHYDRATION-INDUCED 19-like [Nicotiana tabacum (common tobacco)]	Peaxi162Scf00638	496203-496602	e-103
CYP97C11 cytochrome P450-type monooxygenase 97C11 [*Solanum lycopersicum* (tomato)]: Expression de SlLUT1 [LeLUT1] est induite par le stress hydrique. Les plants de tabac transgéniques présentent une teneur en lutéine plus élevée que le tabac de type sauvage (WT). Sous l'effet de la sécheresse, les plantes transgéniques surexprimant SlLUT1 ont montré de meilleures performances de croissance, des teneurs en chlorophylle et en eau relatives plus élevées et des structures complexes de chloroplaste et de PSII plus intactes que le tabac WT. La surexpression du gène ε-hydroxylase de la tomate caroténoïde (SlLUT1) a amélioré la tolérance à la sécheresse du tabac transgénique Wang S, *et al* 2018).	Peaxi162Scf00110	790446-789752	0.0
SIZ1 E3 SUMO-protein ligase SIZ1 [*Solanum lycopersicum* (tomato)]:	Peaxi162Scf00351	315335-314120	0.0

AREB1 ABA-response element binding factor AREB1 [*Solanum lycopersicum* (tomato)] Le SlAREB permet la régulation de l'expression de certains gènes sensibles au stress. Sa surproduction améliore la tolérance des plantes au déficit hydrique et au stress salin. AREB1 est un facteur de transcription bZIP induit par différents stress. Il peut réguler la transcription des gènes associés à la tolérance à la sécheresse (Hsieh TH, et al., 2010).	Peaxi162Scf00252	284875-284098	0.0
bZIP1 bZIP transcription factor6 [*Solanum lycopersicum* (tomato)] SlbZIP1 joue un rôle essentiel dans la tolérance au stress salin et à la sécheresse en modulant une voie médiée par l'ABA. SlbZIP1 pourrait également être utilisé dans l'ingénierie de cultivars de tomates tolérantes au sel et à la sécheresse (Zhu M, *et al*., 2018).	Peaxi162Scf00691	244061-244404	e-101
LOC100301981 N-acetyl-glutamate synthase [Solanum lycopersicum (tomato)] Les plantes transgéniques d'Arabidopsis exprimant le gène SlNAGS1 présentent une accumulation importante d'ornithine dans les feuilles et une plus grande tolérance au sel et à la sécheresse par rapport aux plantes WT (Kalamaki MS, *et al*., 2009).	Peaxi162Scf00118	1883207-1883676	0.0
MPK2 mitogen-activated protein kinase 2 [*Solanum lycopersicum* (tomato)]	Peaxi162Scf00075 Peaxi162Scf00483	1256915-1256564 130632-130975	e-162 e-135
MPK3 mitogen-activated protein kinase 3 [*Solanum lycopersicum* (tomato)] SpMPK1, SpMPK2 et SpMPK3 peuvent jouer un rôle crucial dans l'amélioration de la tolérance à la sécheresse des plants de tomates en influençant l'activité stomatique et la production de H2O2 via la voie ABA-H2O2 Muhammad T, *et al*., 2019).	Peaxi162Scf02118	18299-17938	e-150
SRN1 stress-related NAC1 [*Solanum lycopersicum* (tomato)] SlSRN1 est un régulateur positif de la réponse de défense contre B. cinerea et Pst DC3000 mais est un régulateur négatif de la réponse au stress oxydatif et à la sécheresse chez la tomate (Liu B, *et al*., 2014).	Peaxi162Scf00160	1678967-1678496	e-141
BIP2 ER-binding immunoglobulin protein BIP2 [*Glycine max* (soybean)] La protéine de liaison chaperon moléculaire BiP empêche la perturbation de l'homéostasie cellulaire induite par la déshydratation des feuilles. Le retard dans la sénescence des feuilles dû à la surexpression de BiP pourrait être lié à l'absence de réponse à la sécheresse (Valente MA, *et al*., 2009).	Peaxi162Scf00074	433066-432594	e-101

LOC107793723 PAX3- and PAX7-binding protein 1-like [Nicotiana tabacum (common tobacco)]	Peaxi162Scf00102	886687-886034	e-139
HSP90.1 heat shock-like protein [Arabidopsis thaliana (thale cress)] Le réseau de mémoire du cafier montre que la protéine Hsp90.1 pouvait interagir indirectement et / ou directement avec 15 autres gènes / protéines de mémoire y compris des interactions avec plusieurs protéines de mémoire contenant le domaine LRR [+/−] et chaperons et avec le TF putativeRAD-like (Watanabe E, et al., 2017)	Peaxi162Scf00444 Peaxi162Scf00102	237145-238615 228855-230116 1622906-1624015	0.0 0.0 e-165

```python
# ====================================================
# Ce programme extrait au format fasta une liste de genes
# a partir des identifiants GI dans la banque nucleotide du NCBI
# Il genere un fichier fasta par sequence dont le nom est construit
# en ajoutant l'extension .fasta apres le nom de l'identifiant GI
# ====================================================

from Bio import Entrez
from Bio import SeqIO

Entrez.email = "toto@gmail.com"

def extractGenes(filename):
   # lecture de la liste des identifiants des sequences
   fpi=open(filename)
   mylist=fpi.readlines()
   fpi.close()
   # recuperation des sequences et sauvegarde au format fasta
   for id in mylist:
         #recuperer un identifiant
         seqID=id.strip()
         print(seqID)
         #recuperer la sequence sur le NCBI
         try :
            handle = Entrez.efetch(db="nucleotide", rettype="fasta", retmode="xml",id=seqID)
            # la fonction SeqIO.read() lit la sequence nucleotidique au format fasta
            seq_record = SeqIO.read(handle, "fasta")
            handle.close()
            #sauvegarde au format fasta
            # i) creer un nom de fichier pour la sequence
            outname=seqID+".fasta"
            # ii) creation et ecriture du fichier fasta
            with open(outname, "w") as fpo:
                SeqIO.write(seq_record, fpo, "fasta")
         except ValueError :
            print("Entry",seqID,"not found")


#########################
"""
filename=input('liste:')
extractGenes(filename)
"""

# code test sur la banque Gene avec une sequence
seqID = "843940"
handle = Entrez.efetch(db="gene", id="819739",retmode="xml")
seq_record = SeqIO.read(handle,"fasta")
handle.close()
outname=seqID+".fasta"
with open(outname, "w") as fpo:
    SeqIO.write(seq_record, fpo, "fasta")
```

Figure 16: script « extract_UID.py »

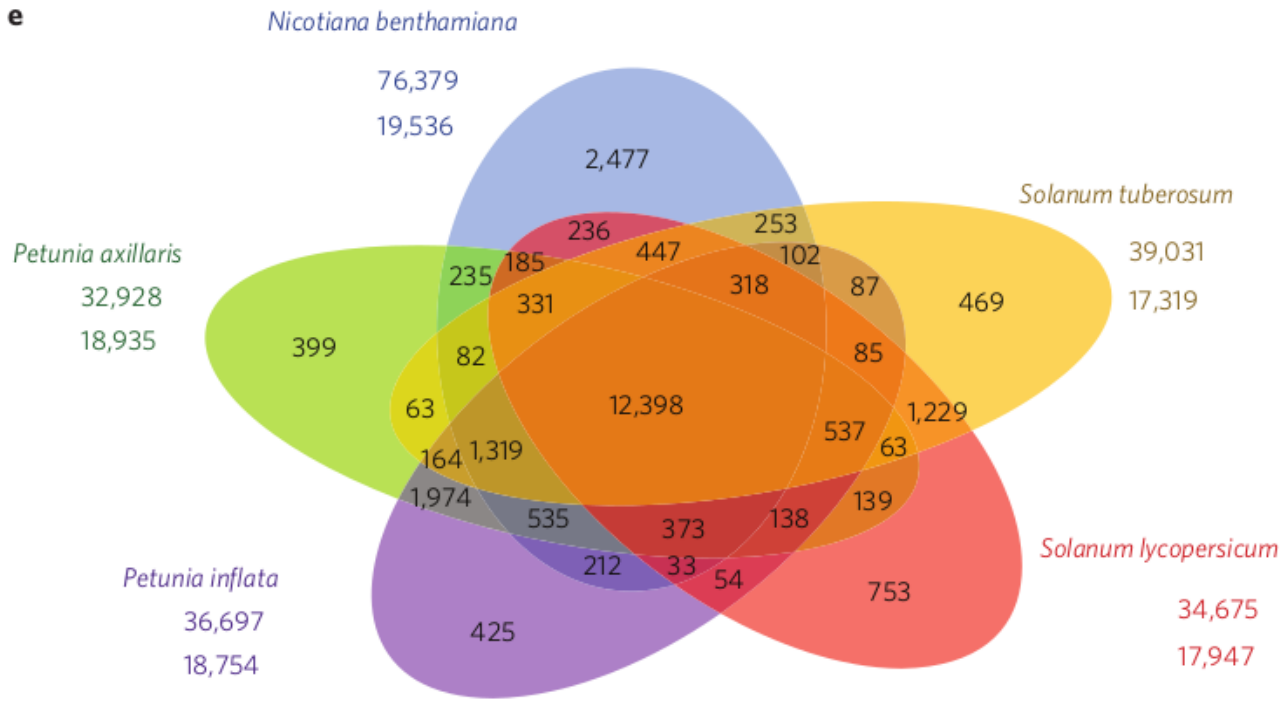


Figure 17: Diagramme de Venn basé sur l'analyse en grappes de la famille de gènes de cinq espèces de solanacées. Les chiffres situés sous le nom de l'espèce indiquent le nombre de gènes codant pour les protéines (en haut) et le nombre de groupes de familles de gènes (en bas).

4. Discussion

L'analyse des données obtenues à l'issue du WGBS nous indique que le nombre de reads et la distribution des nucléotides pour chaque échantillon biologique sont homogènes. Lors de l'alignement des séquences en moyenne 52,8 % des reads pour le lot 1 et 47,4 % pour le lot2 ont pu être alignés en utilisant BSMAP sur le génome de *P. axillaris*. Le nombre de reads mappés avec BSMAP est plus élevé si l'on compare avec d'autre logiciels d'alignement comme Bowtie2 où seulement 20 % des reads ont été mappés sur le génome de référence pour les deux lots.

L'approche sans a priori :

Pour faire l'alignement des reads sur le génome de référence, on a utilisé au début le logiciel bowtie2 qui fait une indexation du génome de référence pour optimiser la lecture et faciliter la recherche des motifs dans la séquence en parcourant le fichier fasta. A l'issue de cette étape, on a obtenu deux fichiers de sortie de format SAM (Sequence Alignment/Map). L'utilisation de samtools a permis de les convertir en format BAM. Malheureusement, on a obtenu un taux d'alignement faible environ 20 % pour les deux lots. Aussi, une nouvelle analyse a été conduite cette fois avec le logiciel bsmap avec les paramètres suivants en exploitant les fichiers bam:

```
bsmap -a lot1.bam -b lot1.bam -d Petunia_axillaris_v1.6.2_genome.fasta -o lot1_out.bam -q 2 -p 6.
bsmap -a lot2.bam -b lot2.bam -d Petunia_axillaris_v1.6.2_genome.fasta -o lot2_out.bam -q 2 -p 6.
```

Les calculs ont été lourds en mémoire et en en temps. En effet, lorsqu'on a lancé cette commande sur la machine locale, les fichiers de sortie ont été trop grands pour l'espace disponible sur le disque dur. Pour résoudre ce problème, nous nous sommes connectés sur node2 avec la commande :

```
ssh ykouki@node2
```

Il s'agit d'un serveur multiprocesseur vers lequel nous avons transféré nos fichiers par la commande :

```
rsync –av lnom_fichier ykouki@node2:~/Petunia
```

Le programme a tourné quelques jours mais on a eu des problèmes de coupure de réseau. Du coup nous avons relancé nos commandes dans un screen pour que le programme tourne indépendamment du réseau. Au bout de 15 jours, les résultats de bsmap ont été obtenus. Pour l'étape suivante, un pipeline d'analyse a été lancé qui a déjà été appliqué sur le génome de pommier. Toutefois, cette pipieline s'est avéré incomplètement compatible avec nos données et a nécessité des modifications pour le relancer. Pour réaliser cette étape, le génome de référence a été divisé en quatre fichiers et une boucle permettant de traiter chaque fichier a été créé :

```
#!/bin/bash
list=$1
for f in $(cat $list)
do
        echo $f
        touch test_$f
        python dmr_pipeline.py ../Petunia_axillaris_v1.6.2_genome.fasta 200 50 100 test_$f  $f ech.txt
done
```

Ce travail a permis de rechercher les DMRs entre le génome de l'echantillon1 et celui de l'échantillon2, chaque échantillon possédant un fichier bam binaire (plus facilement lisible pour une machine), et compressé.
A l'issue de cette étape, un fichier gff3 compatible avec un génome browser sera obtenu. Les DMRs les plus pertinentes seront sélectionnés. Les critères de la sélection varient selon l'organisme et les données de séquençage, et devront être définis. Un tri en ordre croissant sur la colonne 8 du fichier gff qui indique la différence moyenne de méthylation entre les deux échantillons dans une région bien déterminée pourra être fait. Ceci permettra d'afficher les DMRs ayant les plus grandes différences de méthylation en premier.

Conclusions et perspectives

Le stress hydrique est une contrainte environnementale qui peut aboutir à des modifications épigénétiques diverses. Nous avons étudié en particulier la méthylation de l'ADN chez le pétunia qui est une plante de grande valeur ornementale. Le but de ce travail a été d'identifier les régions différentiellement méthylées en réponse au stress hydrique par des outils bio-informatiques. Pour cela nous avons utilisé deux approches complémentaires en parallèle . Les calculs ont été lourds en mémoire et en temps. A ce stade, nous essayons de déboguer les programmes pour pouvoir avancer dans les analyses. L'expression des gènes différentiellement méthylés et identifiés par les analyses bio-informatiques sera ensuite étudiée par RTqPCR en temps réel pour valider ces résultats et ainsi avancer dans la compréhension du rôle de la méthylation de l'ADN dans la réponse au stress hydrique ainsi que dans la mémorisation de ce stress.

Références bibliographiques

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., et al. (2016). Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants.* **2**:16074.

Bossdorf O, Zhang Y. (2011) A truly ecological epigenetics study. *Mol Ecol.* ***20***(8): 1572–1574.

Bossdorf, O., Richards, C.L. & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology Letters*, **11**, 106-115.

cerevisiae. FEMS Yeast Res 13(2):200-18

Crisp PA, Ganguly D, Eichten SR, Borevitz JO, Pogson BJ. (2016). Reconsidering plant memory: Intersections between stress recovery, RNA turnover, and epigenetics. *Science Advances.* **2:** 1501340–1501340.

Ding, Y. Fromm, M, Avramova, Z. (2012). Multiple exposures to drought 'train' transcriptional responses in *Arabidopsis*. *Nat. Commun.* 3:740.

Dupont, Cathérine (Ph.D), D. Randall (Ph.D) Armant, and Carol A. (Ph.D) Brenner. 2009. "Epigenetics: Definition, Mechanisms and Clinical Perspective." *Seminars in Reproductive Medicine* **27**(5): 351–57.

Hsieh TH, Li CW, Su RC, Cheng CP, Sanjaya, Tsai YC, Chan MT. (2010). A tomato bZIP transcription factor, SlAREB, is involved in water deficit and salt stress response. *Planta*,

Jammes H, Renard JP. (2010). Epigénétique et construction du phénotype, un enjeu pour les productions animales. Robustesse, rusticité, flexibilité, plasticité, résilience... les nouveaux critères de qualité des animaux et des systèmes d'élevage. Sauvant D., Perez JM (Eds). *Dossier, INRA Prod. Anim* ***23*****:** 23–42.

Kalamaki MS, Alexandrou D, Lazari D, Merkouropoulos G, Fotopoulos V, Pateraki I, Aggelis A, Carrillo-López A, Rubio-Cabetas MJ, Kanellis AK.(2009). Over-expression of a tomato N-acetyl-L-glutamate synthase gene (SlNAGS1) in Arabidopsis thaliana results in high ornithine levels and increased tolerance in salt and drought stresses. *J Exp Bot,*

Kawakatsu T, Huang S-SC, Jupe F, et al, (2016) Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cel*l **166**: 492–505.

Kim J-M, Sasaki T, Ueda M, Sako K, Seki M. (2015). Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. *Frontiers in Plant Science* 6: 114.

Kim J, Kim I, Han SK, Bowie JU, Kim S. (2012). Network rewiring is an important mechanism of gene essentiality change. *Sci Rep.* 2:900

Kinoshita, T., and Seki, M. (2014). Epigenetic memory for stress response and adaptation in plants. *Plant Cell Physiol.* **55**, 1859–1863. doi: 10.1093/pcp/pcu125

Li C, et al., (2013). Silencing the SpMPK1, SpMPK2, and SpMPK3 genes in tomato reduces abscisic acid-mediated drought tolerance. *Int J Mol Sci,*

Liu B, Ouyang Z1, Zhang Y1, Li X1, Hong Y1, Huang L1, Liu S1, Zhang H1, Li D1, Song F1. (2014). Tomato NAC transcription factor SlSRN1 positively regulates defense response against biotic stress but negatively regulates abiotic stress response. *PLoS One.*

Meyer P. (2015). Epigenetic variation and environmental change. *Journal of Experimental Botany.* **66**: 3541–3548.

Muhammad T, Zhang J3,4, Ma Y5,6, Li Y7,8, Zhang F9, Zhang Y10,11, Liang Y (2019). Overexpression of a Mitogen-Activated Protein Kinase SlMAPK3 Positively Regulates Tomato Tolerance to Cadmium and Drought Stress. *Molecules,.*

N. Daccord. (2018). Thèse : Analyse bioinformatique du génome.

Napoli, C., Lemieux, C. & Jorgensen, R. (1990). Introduction of chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell.* **2,** 279–289.

Nie S, Huang S2, Wang S2, Mao Y2, Liu J2, Ma R2, Wang X (2019). Enhanced brassinosteroid signaling intensity via SlBRI1 overexpression negatively regulates drought resistance in a manner opposite of that via exogenous BR application in tomato. *Plant Physiol Biochem,*

Plomion C, Bastien C, Bogeat-Triboulot M-B, Bouffier L, Déjardin A, Duplessis S, Fady B, Heuertz M, Le Gac A-L, Le Provost G, et al. (2016). Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science,* **73**: 77–103.

Rodríguez López C. M., Wilkinson M. J. (2015). Epi-fingerprinting and epi-interventions for improved crop production and food quality. *Front. Plant Sci*. 6:397.

Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R. (2015). Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *Elife* 4: e09343.

Segatto, A. L. A., Ramos-Fregonezi, A. M. C., Bonatto, S. L. & Freitas, L. B. (2014). Molecular insights into the purple flowered ancestor of garden petunias. *Am. J. Bot*. **101**, 119–127.

Shinozaki K, Yamaguchi-Shinozaki K. (2007). Gene networks involved in drought stress response and tolerance. *J Exp Bot.* 58, 221-227

Sun AQ, Zhi Wu Sheng Li Yu Fen Zi Sheng Wu Xue Xue Bao, (2006) Cloning and molecular characteristic of the metalloprotease gene LeftsH6 from tomato..

Sun Y, 1, Ji K1, Liang B1, Du Y1, Jiang L1, Wang J1, Kai W1, Zhang Y1, Zhai X1, Chen P1, Wang H1, Leng P1., (2017). Suppressing ABA uridine diphosphate glucosyltransferase (SlUGT75C1) alters fruit ripening and the stress response in tomato. *Plant J,*

Tafaghodi R, Marashi H, Moshtaghi N, Zarghami MM. (2018). Expression Patterns of Catalase and Superoxide Dismutase (Cu/Zn-SOD) Genes Under Drought Stress in *Petunia hybrida*. *J Adv Plant Sci* 1: 209

Timothy H. Bestor,(2000). The DNA methyltransferases of mammals, *Human Molecular Genetics*, 9, 2395 -2402

Valente MA, .Faria JA, Soares-Ramos JR, Reis PA, Pinheiro GL, Piovesan ND, Morais AT, Menezes CC, Cano MA, Fietto LG, Loureiro ME, Aragão FJ, Fontes EP., (2009). The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. *J Exp Bot.*

Wang K,Hersh HL1,2, Benning C3 (2016). SENSITIVE TO FREEZING2 Aides in Resilience to Salt and Drought in Freezing-Sensitive Tomato. *Plant Physiol,*

Wang S, Zhuang K1, Zhang S1, Yang M1, Kong F2, Meng Q3.(2018) Overexpression of a tomato carotenoid ε-hydroxylase gene (SlLUT1) improved the drought tolerance of transgenic tobacco. *J Plant Physiol,*

Wang W-S, Pan Y-J, Zhao X-Q, Dwivedi D, Zhu L-H, Ali J, Fu B-Y, Li Z-K. (2011). Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (*Oryza sativa* L.). J*ournal of Experimental Botany.* **62**: 1951–1960.

Watanabe E, Mano S2,3, Hara-Nishimura I4, Nishimura M1, Yamada K1. (2017). HSP90 stabilizes auxin receptor TIR1 and ensures plasticity of auxin responses. *Plant Signal Behav,*

Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J, et al. (2016). Hyperosmotic stress memory in *Arabidopsis* is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *ELife* 5: e13546.

X. Wang, M. Vignjevic, D. Jiang, S. Jacobsen, B. Wollenweber, (2014). Improved tolerance to drought stress after anthesis due to priming before anthesis in wheat (*Triticum aestivum* L.) var. *Vinjett.* **65**, 6441–6456,

Xi Y, Li W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10: 232.

Yong W-S, Hsu F-M, Chen P-Y. 2(016). Profiling genome-wide DNA methylation. Epigenetics & Chromatin 9: 26.

Zhang L, Liu N, Ma X, Jiang L, (2013). The transcriptional control machinery as well as the cell wall integrity and its regulation are involved in the detoxification of the organic solvent dimethyl sulfoxide in Saccharomyces

Zhang S, Zhuang K1, Wang S1, Lv J, Ma N1, Meng Q1. (2017) A novel tomato SUMO E3 ligase, SlSIZ1, confers drought tolerance in transgenic tobacco. *J Integr Plant Biol,*

Zhu M, et al., (2018).Basic leucine zipper transcription factor SlbZIP1 mediates salt and drought stress tolerance in tomato. *BMC Plant Biol.*

Zilberman D., Gehring M., Tran R.K., Ballinger T., Henikoff S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet*. **39:**61–69.

RÉSUMÉ

Petunia hybrida est l'une des espèces de plantes ornementales les plus importantes du point de vue économique et une grande partie de production est faite par multiplication végétative. Le stress hydrique présente une contrainte majeure pour la culture de cette plante. Notre travail est dans le cadre d'améliorer la résistance de pétunia face au stress hydrique avec un priming des générations successives par un déficit hydrique. Ce dernier a créé une marque épigénétique dont la méthylation de l'ADN. Notre étude a été basée sur deux approches bio-informatiques complémentaires pour identifier les gènes différentiellement méthylés en réponse au stress hydrique en utilisant les données de WGBS et un pipeline complet.

Mots-clés: épigénétique, méthylation de l'ADN, stress hydrique, régions différentiellement méthylées (DMR), pétunia.

ABSTRACT

Petunia hybrida is one of the most economically important ornamental plant species and a large part of production is by vegetative propagation. Water stress presents a major constraint for the cultivation of this plant. Our work is in the framework of improving petunia resistance to water stress with a priming of successive generations by a water deficit. The latter has created an epigenetic brand, including methylation of DNA. Our study was based on two complementary bioinformatics approaches to identify differentially methylated genes in response to water stress using WGBS data and a comprehensive pipeline.

keywords: epigenetic, DNA methylation, drought stress, differentially methylated regions (DMRs), petunia.

.