



UNIVERSITE MOHAMMED V
FACULTE DES SCIENCES
DEPARTEMENT DE BIOLOGIE

MASTER : Biotechnologie Végétale Pour L'Amélioration Des Plantes

Mémoire de Fin d'étude

Intitulé

**ETUDE DE LA TOLÉRANCE, L'ADAPTATION ET
L'ACCUMULATION DU PLOMB CHEZ *Hirschfeldia*
incana : APPROCHE GENETIQUE DES
POPULATIONS ET TRANSCRIPTOMIQUE**

Présenté par :

ASSAKREM Mehdi

Soutenu le 20 Octobre 2022

Devant le jury :

M. Zine El Abidine TRIQUI, Professeur à la Faculté des Sciences de Rabat.....Modérateur

M. Abdelaziz SMOUNI, Professeur à la Faculté des Sciences de Rabat.....Examineur

Mme. Mouna FAHR, Professeur à la Faculté des Sciences de Rabat.....Encadrante

M. Bouabid Badaoui, Professeur à la Faculté des Sciences de Rabat.....Co-Encadrant



UNIVERSITE MOHAMMED V
FACULTE DES SCIENCES
DEPARTEMENT DE BIOLOGIE

MASTER : Biotechnologie Végétale Pour L'amélioration Des Plantes

Mémoire de Fin d'étude

Intitulé

**ETUDE DE LA TOLÉRANCE, L'ADAPTATION ET
L'ACCUMULATION DU PLOMB CHEZ *Hirschfeldia
incana* : APPROCHE GENETIQUE DES
POPULATIONS ET TRANSCRIPTOMIQUE**

Présenté par :

ASSAKREM Mehdi

Soutenu le 20 Octobre 2022

Devant le jury :

M. Zine El Abidine TRIQUI, Professeur à la Faculté des Sciences de Rabat.....**Modérateur**

M. Abdelaziz SMOUNI, Professeur à la Faculté des Sciences de Rabat.....**Examineur**

Mme. Mouna FAHR, Professeur à la Faculté des Sciences de Rabat.....**Encadrante**

M. Bouabid Badaoui, Professeur à la Faculté des Sciences de Rabat.....**Co-Encadrant**

Dédicace

Avec l'expression de ma reconnaissance, je dédie ce modeste travail à ceux qui, quels que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

A l'homme, mon précieux offre du dieu, qui doit ma vie, ma réussite et tout mon respect : mon cher père Mohamed.

A la femme qui a souffert sans me laisser souffrir, qui n'a jamais dit non à mes exigences et qui n'a épargné aucun effort pour me rendre heureux : mon adorable mère Chowda.

A mes chères sœurs Karima et Salma qui n'ont pas cessée de me conseiller, encourager et soutenir tout au long de mes études. Que Dieu les protège et leurs offre la chance et le bonheur.

Et à tous ceux que ma réussite leur tient à cœur.

Remerciements

En préambule à ce projet, je remercie ALLAH qui m'aide et m'a donné la patience et le courage durant ces longues années d'étude.

Les travaux présentés dans ce manuscrit de Master ont été réalisés au laboratoire de Physiologie et Biotechnologie végétale de la Faculté des Sciences de Rabat sous la direction de Pr Abdelaziz SMOUNI

J'adresse mes sincères remerciements à mon encadrante Madame Mouna FAHR, Professeur à la FSR-UM5, à qui j'exprime toute ma reconnaissance, de m'avoir encadré et encouragé tout le long de mon stage. Merci pour sa bonne volonté, sa patience et ses précieux conseils ainsi que ses remarques.

Je remercie Professeur Bouabid BADAoui qui m'a Co-encadré, pour le temps qu'il a consacré et pour les précieuses informations qu'il m'a prodiguées avec intérêt et compréhension tout au long de la période du projet.

J'exprime aussi mes vifs remerciements aux membres du jury, Professeurs Zine El Abidine TRIQUI et Abdelaziz SMOUNI à la Faculté des Sciences de Rabat, pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner mon travail et de l'enrichir par leurs remarques.

Je n'oublie pas de remercier Said EL HASNAoui étudiant en Doctorat au sein du laboratoire Biotechnologie et Physiologie végétales à la Faculté des Sciences de Rabat, qui m'a consacré beaucoup de temps afin de m'aider et de me diriger avec des conseils qui m'ont illuminé dans mes choix méthodologiques.

Je ne saurais jamais remercier la lumière de ma vie, mes parents, qui par leurs prières et leurs encouragements, j'ai pu surmonter tous les obstacles.

Pour conclure, je souhaite adresser mes remerciements à tout l'ensemble du personnel du département de biologie de la faculté des sciences, aussi tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Résumé

La pollution par le plomb (Pb) est un problème environnemental très répandu en raison de sa toxicité pour les organismes vivants. *Hirschfeldia incana*, une espèce de la famille des *Brassicaceae*, communément trouvée dans les régions méditerranéennes, est caractérisée par sa capacité à tolérer et à accumuler de fortes concentrations du Pb. C'est une pseudométallophyte qui se développe aussi bien sur des sols pollués que sur des sites naturels non contaminés. Cette plante a été proposée comme un excellent modèle pour évaluer la réponse et l'adaptation des plantes au Pb.

L'objectif de cette étude est d'identifier et de comprendre les mécanismes moléculaires de la tolérance, l'adaptation et l'accumulation du Pb chez *H. incana*. Pour cela, une analyse génétique a été réalisée d'une part en vue d'apprécier l'effet du stress métallique sur la spéciation de cette espèce via une étude comparative des séquences introniques de 4 gènes candidats entre une population métallicole (PM) et une population non métallicole (PNM) d'*H. incana*. D'autre part, le profil transcriptomique de cette espèce a été également assemblé et annoté, et les gènes différentiellement exprimés en réponse au Pb ont été identifiés.

Les gènes *MT2a*, *CCH*, *ABCG22* et *GCN2* ont été choisis pour une étude génétique. Suivant le test AMOVA, l'analyse de séquences a montré une variation significative au niveau intrapopulation. Tandis que la variation interpopulation n'est pas importante pour ces gènes. Ces gènes ont montré une évolution suivant un modèle neutre avec une valeur D de Tajima et F_s de F_u non significative. L'analyse phylogénique a montré que pour les 4 gènes qu'il y a une distinction des groupes avec des similarités génétiques entre les individus des deux populations PM et PNM. L'analyse généalogique a révélé que pour les gènes *ABCG22* et *GCN2*, la PNM est issue de la PM (origine ancestrale) alors que l'étude du gène *Mt2a* a montré l'inverse. A ce niveau d'analyse, nous n'avons pas détecté d'effet de sélection sur les gènes étudiés chez les deux populations. Et l'ambiguïté de la relation ancestrale entre les populations nous incite à utiliser d'autres marqueurs moléculaires pour étudier les origines ancestrales de la population métallicole. La différence en termes d'adaptation au Pb des deux populations manifesterait probablement niveau de la régulation des gènes étudiés.

Le séquençage du transcriptome (RNA-seq) d'*H. incana* dans deux conditions, contrôle (sans Pb) et traité avec 100 μM de $\text{Pb}(\text{NO}_3)_2$ en hydroponie pendant 15 jours a été réalisé. Un total de 842 millions de reads a été généré. 77491 séquences de transcrits ont été assemblées. Les analyses de similarité de séquence et l'annotation de ces transcrits ont été effectuées par rapport

à la base de données des protéines d'*Arabidopsis thaliana* et de Gene Ontology (GO). En réponse au Pb, 577 et 270 gènes ont été exprimés de manière différentielle dans les racines et les parties aériennes, respectivement.

Mot clé : *Hirschfeldia incana*, plomb (Pb), tolérance, accumulation, phytoremediation, population, métallicole, RNAseq.

Abstract

Lead (Pb) pollution is a very severe environmental problem due to its toxicity to living organisms. *Hirschfeldia incana*, a species of the *Brassicaceae* family, is found in Mediterranean regions, is characterized by its ability to tolerate and accumulate high concentrations of Pb. It is a pseudometallophyte that grows both on polluted soils and on uncontaminated natural sites. This plant has been proposed as an excellent model to assess plant response and adaptation to Pb.

The objective of this study is to identify and understand the molecular mechanisms of tolerance, adaptation and accumulation of Pb in *H. incana*. For this, a genetic analysis was carried out on the one hand in order to assess the effect of metal stress on the speciation of this species via a comparative study of the intronic sequences of 4 candidate genes between a metal-dwelling population (PM) and a nonmetallic population (PNM) of *H. incana*. On the other hand, the transcriptomic profile of this species was also assembled and annotated, and the genes differentially expressed in response to Pb were identified.

The *MT2a*, *CCH*, *ABCG22* and *GCN2* genes were chosen for genetic study. Following the AMOVA test, sequence analysis showed significant variation at the intrapopulation level. While the interpopulation variation is not important for these genes. These genes showed evolution following a neutral pattern with insignificant D of Tajima and Fs of Fu. The phylogenetic analysis showed that for the 4 genes that there is a distinction of groups with genetic similarities between the individuals of the two populations PM and PNM. The genealogical analysis revealed that for the *ABCG22* and *GCN2* genes, the PNM comes from the PM (ancestral origin) while the study of the *Mt2a* gene showed the opposite. At this level of analysis, we did not detect any selection effect on the species studied in the two populations. And the ambiguity of the ancestral relationship between the populations encourages us to use other molecular markers to study the ancestral origins of the metal-dwelling population. The difference in terms of adaptation to Pb of the two populations would probably manifest the level of regulation of complex genes.

Transcriptome sequencing (RNA-seq) of *H. incana* under two conditions, control (without Pb) and treated with 100 μ M of Pb(NO₃)₂ in hydroponics for 15 days was carried out. A total of 842 million reads were generated. 77491 transcript sequences were assembled. Sequence similarity analyzes and annotation of these transcripts were performed against the *Arabidopsis*

thaliana transcriptome and Gene Ontology (GO) protein database. In response to Pb, 577 and 270 genes were differentially expressed in roots and aerial parts, respectively.

Key word: *Hirschfeldia incana*, lead (Pb), tolerance, accumulation, phytoremediation, population, metalicolous, RNAseq.

Table des matières

Dédicace	ii
Remerciements	iii
Résumé	iv
Abstract	vi
Liste des figures	xi
Listes des tableaux	xiii
Liste des abréviations	xiv
Introduction	1
Revue bibliographique	4
1. Le plomb : propriétés physico-chimiques, effets sur la santé et l'environnement.....	5
1. 1. Propriétés physico-chimiques du Pb	5
1. 2. Origine naturelles et anthropiques	5
2. Effet/toxicité sur l'environnement et l'homme	7
3. Effet sur la plante.....	8
3.1 Biodisponibilité	8
3.2 Effet du pH	9
3.3 Potentiel redox.....	9
4. Activité biologique et biodisponibilité du Pb.....	10
5. Toxicité pour les plantes.....	10
Les symptômes de la toxicité chez les plantes	11
6. Les plantes hyperaccumulatrices	11
7. Mécanismes moléculaires de tolérance.....	12
7.1. L'absorption	14
7.2. Détoxification : la chélation et la séquestration.....	15
7.2.1- Chélation	15
a. Phytochélatines (PC)	15
b. Métallothionéines (MTs)	15
c. Acides organiques	16
7.2.2 Translocation	16
7.3 Transport et séquestration du plomb.....	17
A. Étude des populations : outil pour la compréhension de l'adaptation des plantes aux métaux lourds	18
1. Adaptation des plantes et réponse au stress métallique.....	18
2. Variabilité de la tolérance et l'accumulation des éléments-traces métalliques (ETMs)	20
B. Etude transcriptomique.....	22

1. Technologies pour l'analyse à large échelle du transcriptome	22
1.1. Les puces à ADN.....	23
1.2. Le RNA-seq.....	23
2. Méthodes d'assemblage (sans génome de référence).....	27
3. Quelques études transcriptomiques en réponse au métaux lourds.....	28
C. <i>Hirschfeldia incana</i> comme model d'étude de l'étude de la tolérance et l'accumulation du plomb	30
1. Description botanique de la plante « <i>Hirschfeldia incana</i> »	30
2. Cycle de vie et habitat	30
3. <i>H.incana</i> : plante modèle.....	30
Matériels et méthode	32
I- Étude de la diversité intra et inter- populations d' <i>Hirschfeldia incana</i>	33
1. Matériel végétal.....	33
2. Analyse des produits de séquençage.....	33
3. Analyse bio-informatique.....	34
4. Test de la diversité génétique des populations d' <i>H.incana</i>	34
5. Les tests statistiques	34
II- Étude transcriptomique	35
1. Matériel végétal.....	35
2. Culture des plantes	36
3. Préparation des cDNA et séquençage (RNAseq).....	36
4. Assemblage a novo (<i>De Novo</i> assembly).....	36
5. Annotation et profil fonctionnel	37
6. Estimation de l'abondance des transcrits et analyse de l'expression des gènes.....	37
7. Analyse des gènes différentiellement exprimés (DEGs)	37
Résultats et discussion	39
I- L'analyse de la diversité intra et inter population d' <i>Hirschfeldia incana</i>	40
1. Choix des gènes candidats	40
2. Équilibre de Hardy-Weinberg.....	42
3. Tests de neutralité.....	43
4. Analyse de la structure génétique	45
5. Analyse multivariée de variance des 4 gènes.....	46
6. Analyse phylogénique et généalogique des gènes	49
II- Étude transcriptomique d' <i>Hirschfeldia incana</i>	56
1. Contrôle qualité des reads.....	56
2. Contrôle de la qualité du transcriptome d' <i>H.incana</i>	57

3. Annotation du transcriptome d' <i>H.incana</i>	58
4. Analyse de l'expression des gènes en réponse au plomb	60
5. Exemple de quelques gènes différentiellement exprimés en réponse au Pb.....	61
Conclusion générale	67
Références bibliographiques	70
Annexes.....	78
Annexe 1	78
Annexe 2	81
Annexe 3	82
Annexe 4	83
Annexe 5	84
Annexe 6	84
Annexe 7	86

Liste des figures

Figure 1 : Pays comptant les plus grandes réserves de plomb dans le monde en 2019 (en millions de tonnes) (https://fr.statista.com/statistiques/571219/reserves-mondiales-de-plomb-par-pays/ : visiter le 10/09/2022).....	6
Figure 2 : Facteurs influençant les interactions de la rhizosphère (Lynch et al., 2001).	9
Figure 3 : Effet du potentiel redox et du pH sur l'absorption du Pb par les racines (Reddy and Patrick Jr, 1977).	10
Figure 4 : Un modèle des mécanismes qui se produisent dans les plantes d'hyperaccumulation lors de l'exposition au zinc (Zn) : absorption, chélation, transport et séquestration des ions Zn. La biodisponibilité du zinc peut être influencée par plusieurs facteurs, tels que les micro-organismes, les racines ramifiées, le pH et les exsudats. Une fois adsorbé par les racines, le Zn peut être absorbé par une voie apoplastique : une diffusion passive à travers les cellules, ou par une voie symplastique via des transporteurs. Dans cette dernière voie, l'absorption du Zn par les cellules de l'épiderme est principalement favorisée. sous forme de micro-organismes, de pH de ramification et d'exsudats. Une fois adsorbé par les racines, le Zn peut être absorbé par une voie apoplastique : une diffusion passive à travers les cellules, ou par une voie symplastique via des transporteurs. Dans cette dernière voie, l'absorption de Zn par les cellules de l'épiderme est principalement favorisée par IRT1, ZIP19 et ZIP23. Pour atteindre le cortex, le Zn peut être diffusé directement ou par l'intermédiaire de ZNT2 et ZNT5. Ensuite, le Zn peut être soit stocké dans des vacuoles (promues par ZIF2) soit transporté vers l'endoderme via les transporteurs suivants : ZIP23, ZIP19, ZIP5 et IRT3. Le Zn suivant la voie apoplastique est stoppé par la bande casparienne, puis pénètre dans l'endoderme via ZNT1/ZIP4. A ce niveau, le Zn peut être chélaté par la nicotianamine (NA) ou diffusé directement aux cellules du péricycle où une partie peut également être associée à l'histidine (His). Le Zn non chélaté peut atteindre le xylème par diffusion directe ou via YSL, ferric réductase défectueuse 3 (FRD3,) et HMA4. Le Zn traverse ensuite le xylème sous une forme sans Zn ou couplé à His, citrate ou malate. Pour pénétrer dans les cellules foliaires, le Zn peut pénétrer passivement sous forme chélatée ou sous forme sans Zn via les protéines ZIP4 et ZIP6. Il est ensuite séquestré à l'intérieur de la vacuole par les transporteurs MTP1 (protéines de tolérance aux métaux 1), MTP8, MTP11, NRAMP3, NRAMP4, HMA3 et HMA4, ou bloqué dans la paroi cellulaire (Balafrej et al., 2020).	13
Figure 5 : Les différentes technologies de séquençage. Graphique représentant le nombre de lectures (reads) obtenues par run en fonction de la longueur des lectures pour les différentes technologies de séquençage de première, deuxième et troisième génération (Lex Nederbragt , 2012-2016).....	24
Figure 6 : Principes de base des techniques NGS. (a) pyroséquençage : l'incorporation d'un nouveau nucléotide génère une lumière détectable. (b) Séquençage 454 : l'incorporation de nucléotides est associée à la libération de pyrophosphate résultant en un signal lumineux. (c) Solexa : les fragments d'ADN forment des ponts double brin et après l'ajout des terminateurs marqués, le cycle de séquençage commence. (d) SOLiD : si les adaptateurs sont liés, une PCR en émulsion est effectuée pour générer ce que l'on appelle des clones de billes (Mutz et al., 2013).	25
Figure 7 : Séquençage illumina. A) L'ADN à séquencer est fragmenté et des adaptateurs sont liées aux extrémités de chaque fragment obtenu. B) Les fragments à séquencer vont être amplifiés par PCR, après avoir été fixés sur la "flow cell" (lame de verre). On obtient des clusters correspondants chacun à un fragment initial et contenant environ 1000 copies de celui-ci. C) Dans chaque cluster, une base est incorporée dans chaque fragment, celle-ci émet un signal fluorescent qui est enregistré. Ces étapes sont répétées jusqu'à séquençage du fragment de taille souhaitée (Illumina 2016).	26
Figure 8 : Etapes de Novo assembly jusqu'à l'annotation des transcrits. (A) Contrôle de la qualité des lectures brutes en filtrant les lectures erronées. (B) Assemblage de séquences, y compris le regroupement en groupes d'isoformes et la suppression des séquences redondantes (les isoformes sont des variants de transcription résultant d'un épissage alternatif). (C) Cartographier les lectures brutes sur les séquences assemblées pour le contrôle de la qualité de l'assemblage ou pour l'analyse de l'expression différentielle. (D) Application de tests statistiques pour l'identification des changements	

dans les niveaux d'expression. (E) Classification des séquences par espèces d'ARN et traduction en séquences protéiques avant annotation. (F) Annotation des séquences sur la base de la similarité des séquences, identification des caractéristiques des séquences (telles que les domaines fonctionnels) et annotation des termes d'ontologie génétique (Raghavan et al., 2022).....	27
Figure 9 : Plant d' <i>Hirschfeldia incana</i> poussant sur les scories d'Oued EL Heimer	35
Figure 10 : Les différentes étapes d'analyses des données RNA-seq	38
Figure 11 : Analyse discriminante des composantes principales des 4 gènes (a : gène ABCG22, .b : gène CCH, .c : gène GCN2, .d : gène Mt2a)	47
Figure 12 : Diagramme de la composition génotypique de chaque population pour les quatre gènes (A : gène ACG22, B : gène CCH, .C : gène GCN2, D : gène Mt2a) (1. : Population métallicoles, .2 : population non métallicoles).	48
Figure 13 : la position des SNP indicative de la différenciation entre les deux populations (a : gène ABCG22, b : gène CCH, .c : gène GCN2, d : gène Mt2a).....	49
Figure 14 : Arbre phylogénétique d'alignement de séquences multiples du gène ABCG22 avec Neighbor Joining algorithme.	50
Figure 15 : Graphique de généalogie des haplotypes du gène ABCG22.	51
Figure 16 : Arbre phylogénétique d'alignement de séquences multiples du gène CCH avec Neighbor Joining algorithme.....	52
Figure 17 : Arbre phylogénétique d'alignement de séquences multiples du gène GCN2 avec Neighbor Joining algorithme.....	53
Figure 18 : Graphique de généalogie des haplotypes du gène GCN2.....	53
Figure 19 : Arbre phylogénétique d'alignement de séquences multiples de gène Mt2a avec Neighbor Joining algorithme.....	54
Figure 20 : Graphique de généalogie des haplotypes du gène Mt2a.....	55
Figure 21 : Exemple d'un résultat de contrôle de la qualité de nos reads.	56
Figure 22 : Résultat de l'évaluation de la qualité de transcriptome par BUSCO tool.....	57
Figure 23 : Les dix GO les plus dominants en fonction du nombre de gènes selon les trois domaines fonctionnels : Processus Biologique (BP), Composants cellulaires (CC) et Fonction moléculaire (MF).	59
Figure 24 : Diagramme de venn des gènes communs entre les plantes traitées au plomb et non traitées (a. transcrit de la partie racinaire, b. transcrit de la partie aérienne) et c : l'histogramme de nombre des gènes différentiellement exprimés	61

Listes des tableaux

Tableau 1 : Consommation de plomb par industries, 1962 (en tonnes courtes) (Brännvall et al., 1999)	7
Tableau 2 : Gène Candidats avec les amorces spécifiques d' <i>H. incana</i> correspondants	33
Tableau 3 : Liste des gènes candidats	40
Tableau 4 : Les valeurs P d'équilibre de Hardy Weinberg	43
Tableau 5 : Les résultats de test de D de Tajima et Fs de Fu	44
Tableau 6 : La structure génétique des populations et l'indice de fixation et l'indice de tri généalogique.	46
Tableau 7 : Résultats d'alignement avec Bowtie2 tool	58
Tableau 8 : Les 5 premier gènes (surexprimés et sous exprimés) en réponse au Pb dans la partie racinaire chez <i>H. incana</i> .	62
Tableau 9 : Les 5 premier gènes sur-exprimés en réponse au stress par le Pb dans la partie aérienne d' <i>H. incana</i> .	63
Tableau 10 : Les 5 premier gènes sous-exprimés en réponse au stress par le Pb dans la partie aérienne d' <i>H. incana</i> .	64

Liste des abréviations

ABC : ATP Binding Cassette
ADN : Acide désoxyribonucleique
ADNc : ADN complémentaire
AFLP : Amplified Fragment Length Polymorphism
Ag : Alluminium
AMOVA : Analysis of Molecular variance
ARN : Acide ribonucleique
ARNm : Acide ribonucleique messenger
ARP : Arlequin Project
ATM : ABC Transporter of the Mitochondrion
ATP : Adenosine Triphosphate
BD : Broughton et Dilworth
BP : Biological Process
BUSCO : Benchmarking Universal Single-Copy Orthologs
Ca : Calcium
CAPS : Cleaved Amplified Polymorphic Sequence
CBP : Calmodulin-binding protein
CC : Cellular Component
Cd : Cadmium
CNGC : Cyclic nucleotide-gated ion channel
Co : Cobalt
Cu : Cuivre
DACC : Depolarization-activated calcium channels
DAPC : Discriminant Analysis of Principal Component

ddNTP : dideoxynucleotide triphosphate

DEG : Differentially Expressed Gene

dNTP : désoxynucleotide triphosphates

DTPA : Diethylenetriamine pentaacetate

DW : Dry Weight

EDTA : Ethylenediaminetetraacetic

eIF2 α : the α -subunit of eukaryotic initiation factor 2

ETM : Elements de Traces Metaliques

FastQC : Fast Quality Control

FC : Fold Change

Fe : Fer

FRD : Ferric reductase defective

FST : Fixation index

GO : Gene Ontology

GS : Glutathione Synthase

GSH : Glutathion

gsi : genealogical sorting index

H₂O₂ : peroxyde d'hydrogene

HA : H⁺ ATPase

HACC : Hyperpolarization activated calcium channels

HEDTA : Hydroxyethylenediaminetetraacetic

His : Histidine

HMA : Heavy Metal ATPase

IRT : Iron-regulated Transporter

JC : Jesus Christ

K : Potassium

K80 : Kimura 80

LRR-RLK : Leucine-rich repeat-receptor like kinase

MANOVA : Multivariate Analysis of Variance

MAPK : Mitogen-activated protein kinases

MF : Molecular Function

Mg : Magnesium

MGDG : Monogalactosyldiacylglycerol

MHX : Magnesium/proton Exchanger

Mn : Manganese

MRP : Multidrug Resistance-Associated protein

MT : Metallothioneines

MTP : Metal Tolerance Protein

NGS : New Generation Sequencing

Ni : Nickel

NRAMP : Natural Resistance-Associated Macrophage Protein

NRT : Nitrate Transporter

NTA : Nitrilotriacetic acid

O₂⁻ : radical superoxide

OH⁻ : radical hydroxyl

p : probability

Pb : Lead

Pbs : Galena

PC : Phytochelatins

PCR : Polymerase Chain Reaction

PCS : Phytochelatin synthase

pH : potential d'hydrogene

PIB : P-type ATPase of the class IB

PM : Population Metallicole

PNM : Population Non Metallicole

ppm : point per million

PR : Pathogene related

QPCR : Quantitative PCR

RAPD : Random Amplified Polymorphic Sequence

ROS : Espèces réactives à l'oxygène

RSEM : RNA-seq by Expectation-Maximization

RT-PCR : Reverse Transcription Polymerase Chain Reaction

SNP : Single Nucleotide Polymorphism

SOD : superoxide dismutase

VICC : Voltage-intensive cation channels

YSL : Yellow Strip Like

ZIP : Zinc and Iron rgulated trasporters protein

Zn : Zinc

Introduction

Le plomb (Pb) est l'un des principaux métaux lourds présents dans l'environnement, à la fois à partir de sources naturelles et principalement à partir d'activités anthropiques. Le Pb est connu pour être toxique pour les plantes, les animaux et les micro-organismes. En tant qu'élément non essentiel, le Pb n'a aucune fonction biologique et il est toxique pour les organismes vivants, même à de faibles concentrations. La croissance des plantes est inhibée à une concentration du Pb dans le sol de 30mg/kg (Usman et al., 2020). Ainsi, il provoque divers effets morphologiques, physiologiques et biochimiques. Cependant, plusieurs espèces végétales ont développé diverses stratégies de tolérance et d'hyperaccumulation qui sont appelées « hyperaccumulatrices ». Les plantes hyperaccumulatrices sont capables de stocker le Pb à une concentration qui dépasse 1000 ppm dans la matière sèche des feuilles (Baker and Brooks, 1989) et ont ainsi des potentialités d'utilisation dans des procédés de phytoremédiation (phytoextraction). Cette capacité à ne pas avoir d'effet néfaste sur la croissance implique l'existence de mécanismes de détoxification des métaux au sein de la plante. Les principaux mécanismes sont la chélation du cation métallique par des ligands ou la séquestration des métaux hors des sites du métabolisme dans le cytoplasme, notamment dans la vacuole et la paroi cellulaire. Cependant, les mécanismes moléculaires élucidant ces processus de tolérance et d'accumulation du Pb sont très fragmentaires.

Une des approches, permettant d'étudier ces mécanismes et mieux comprendre les réponses des plantes vis-à-vis des métaux, est l'étude des populations des espèces pseudométallophytes. Ces dernières dites également métallophytes facultatives, se développant aussi bien sur des sites métallifères (population métallicole) que sur des sites non métallifères (population non métallicole). Ces populations présentent généralement des différences au niveau physiologique et moléculaire qui leur permettent de s'adapter chacune à son environnement. En particulier, la génétique des populations permet d'étudier les facteurs qui déterminent, maintiennent et modifient la variation des populations. Elle étudie la distribution et les changements de la fréquence des versions d'un gène (allèles) dans les populations, sous l'influence des « pressions évolutives » du fait que les changements de fréquence des allèles sont un aspect majeur de l'évolution.

Récemment avec le développement des techniques omiques, l'étude et la compréhension des mécanismes moléculaires de la réponse aux métaux lourds à grande échelle sont rendues possibles. En effet, la transcriptomique permet d'explorer le profil de transcription de la plante dans de multiples conditions environnementales. La transcriptomique aide aussi à comprendre les réponses physiologiques et moléculaires en termes de régulation des gènes, de

différenciation des gènes, de modifications post-transcriptionnelles et d'épissage des gènes (Imadi et al., 2015). Ces différents profils permettent de retrouver les gènes clés différentiellement exprimés sous stress métallique pour une recherche ciblée de ces gènes. Ainsi, la transcriptomique, associée à d'autres approches biotechnologiques, contribue à la compréhension des mécanismes mis en place par ces plantes hyperaccumulatrices en réponse au Pb.

Hirschfeldia incana, pseudométallophyte de la famille des *Brassicaceae*, est caractérisée par sa capacité à tolérer et à accumuler de fortes concentrations du Pb dans ses parties aériennes (Becerra-Castro et al., 2012). L'accumulation du Pb chez *H. incana* a été également confirmée au laboratoire en conditions contrôlées (Fahr et al., 2015 ; Auguy et al., 2013). Cette plante est un modèle idéal pour l'étude moléculaire et physiologique de la tolérance au Pb chez les plantes. Il est à noter qu'à nos jours, aucune plante-modèle pour l'étude de la tolérance et l'accumulation du Pb n'a été décrite.

L'objectif de notre étude est d'une part (i) faire une analyse intra et inter population des séquences de gènes candidats (*ABCG22*, *CCH*, *GCN2*, *MT2a*) afin de détecter de signature de sélection permettront ainsi d'identifier les gènes associés aux phénomènes d'adaptation au Pb. D'autre part (ii) réaliser un assemblage de novo du transcriptome par RNA-seq, annoter les gènes du transcriptome d'*H. incana* et caractériser les changements transcriptionnels qui se produisent dans cette espèce en réponse au Pb.

Revue bibliographique

1. Le plomb : propriétés physico-chimiques, effets sur la santé et l'environnement

1. 1. Propriétés physico-chimiques du Pb

Le mot français plomb est lié au latin « plumbum », qui signifie liquide argenté et qui a donné à l'élément son symbole chimique Pb (Boldyrev, 2018). Parmi les métaux lourds dont le numéro atomique est supérieur à 60, le Pb est le plus abondant dans la croûte terrestre (Adriano, 2001). C'est un métal bleuté brillant, mou, très malléable et ductile. En raison de son caractère chalcophile, le Pb natif est rare. Il est associé au gisement de sulfures ; on l'extrait de sa source minérale principale, la galène (PbS) qui en contient 86.6% en poids (Cheng and Hu, 2010), mais aussi des minerais associés au zinc (Zn) (la sphalérite), à l'argent et le plus abondamment au cuivre (Cu) (Hettiarachchi and Pierzynski, 2004). La plupart des autres minéraux de plomb sont liés à la galène d'une manière ou d'une autre : (i) la boulangerite, $Pb_5Sb_4S_{11}$, est un sulfure mixte dérivé de la galène ; (ii) l'anglesite, $PbSO_4$, est un produit de l'oxydation de la galène, et la cérusite ou (iii) le minerai de plomb blanc, $PbCO_3$, est un produit de décomposition de la galène (Boldyrev, 2018).

1. 2. Origine naturelles et anthropiques

Le Pb est présent dans tous les compartiments de la biosphère. Les gisements de Pb les plus importants sont ceux hébergés dans les bassins de rift continental et d'affaissement du Protérozoïque entre environ 1,65 et 1,4 million d'années (Leach et al., 2010). Les ressources du Pb sont réparties dans le monde entier avec une réserve de 226.1 million de tonnes (Mudd et al., 2017) et la production annuelle du Pb est de 3,4 millions de tonnes avec une réserve exploitable estimée de 70 millions de tonnes (Wood and Nimmo, 1994). Mais d'autres processus naturels contribuent de façon significative à la libération du plomb comme la dégradation et l'érosion des roches mères, les feux de forêt, et les éruptions volcaniques (Patterson and Settle, 1987). Les réserves dans le monde étaient de 95.6 millions de tonnes en 2019 (figure 1).

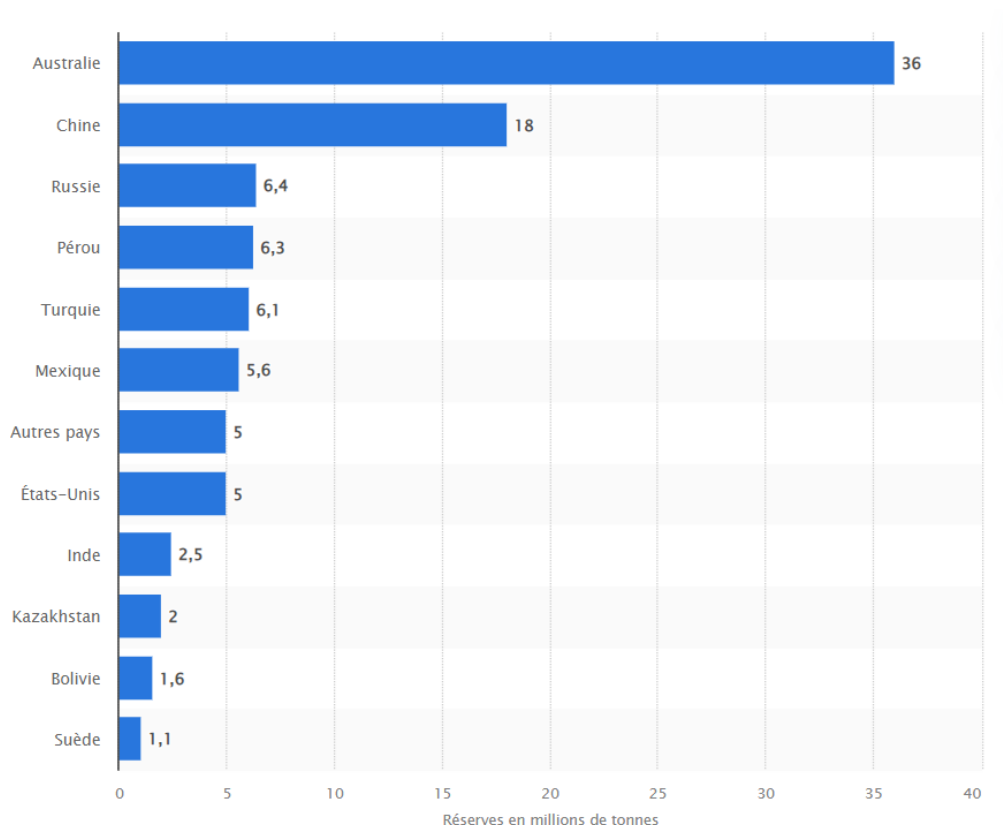


Figure 1 : Pays comptant les plus grandes réserves de plomb dans le monde en 2019 (en millions de tonnes) (<https://fr.statista.com/statistiques/571219/reserves-mondiales-de-plomb-par-pays/>: visiter le 10/09/2022)

Les émissions totales du cycle anthropique du plomb en 2010 ont été calculées comme étant de 3,56 millions tonnes. Ils proviennent de différentes ressources selon l'étape du processus de raffinage atteint par la mine, depuis la production, fabrication vers la gestion et le recyclage des déchets. Les déchets de batteries plomb-acide est la première sources qui cause un grand taux de contamination mais le processus de production de gaines de câbles est parmi les sources majeures d'émission de Pb (Liang and Mao, 2015).

Les premiers objets en Pb étaient datés de 7000 avant JC. À cette époque, le Pb avait peu d'applications en raison de sa douceur et de son aspect terne (Boldyrev, 2018). En temps moderne, il existe une variété d'industries et de produits dans lesquels des tonnages de plomb sont utilisés. Plus de 40% de tout le Pb utilisé est sous forme de composés chimiques comme le Pb blanc, le Pb rouge, la litharge... Le reste (60%) est largement consommé sous forme de Pb pur ou d'alliages. L'utilisation la plus importante était la batterie de stockage avec une consommation de 387 000 tonnes (Tableau 1) (Brännvall et al., 1999).

Tableau 1: Consommation de plomb par industries, 1962 (en tonnes courtes) (Brännvall et al., 1999)

Storage batteries	387,000
Oil refining and gasoline	175,000
Cable	55,500
Construction	123,000
Paint and varnish	29,000
Ammunition	47,000
Brass manufacturing	21,000
Printing	27,000
Ceramics	17,000
Colors	20,000
Can manufacturing	7,000
Automobiles (other than batteries)	12,000
Railroads	7,000
Insecticides	2,500
Steel and wire	4,000
Foil	4,000
Collapsible tubes	12,000
Rubber and hose	3,500
Coatings	1,500
Unclassified	122,000
Total	1,077,000

La pollution actuelle de l'environnement par le Pb était due à l'industrie métallurgique médiévale qui était parallèle à l'augmentation de la population et à la croissance économique puis à l'expansion de l'exploitation minière avec la découverte de nouveaux gisements de minerai et la réouverture d'anciennes mines (Brännvall et al., 1999). Environ 357 à 857 millions de kg/an de Pb sont rejetés dans l'environnement par les activités d'extraction et de fusion. À l'échelle mondiale, on estime que le traitement du minerai de Pb a rejeté environ 300 millions de tonnes de déchets dans l'environnement au cours des cinq derniers millénaires, principalement au cours des 500 dernières années (Zhang et al., 2012). Au fur et à mesure que les débris provenant des mines arrivent au contact de l'eau, les propriétés physiques des particules de débris provoquent une pollution et cette toxicité est renforcée par la réaction avec l'eau et la formation de sel du Pb (Carpenter, 1924).

2. Effet/toxicité sur l'environnement et l'homme

Le Pb est présent dans tout l'environnement à des taux différents. Les concentrations de fond typiques de Pb ne dépassent pas 0,1 $\mu\text{g}/\text{m}^3$ dans l'atmosphère ; 100 mg/kg dans le sol ; et 5 $\mu\text{g}/\text{L}$ en eau douce et en eau de mer (Boldyrev, 2018). Lorsque sa concentration dépasse les seuils, le Pb induit des effets toxiques pour l'environnement et pour l'homme puisqu'il n'a pas de rôle biologique. Sa prévalence dans le corps humain - à une moyenne de 120 mg- n'est cependant dépassée que par le zinc (2500 mg) et le fer (4000 mg) parmi les métaux lourds (Kumar et al., 2020). Les principaux symptômes de l'empoisonnement au Pb chez les adultes sont un

dysfonctionnement des organes vitaux, des troubles du système nerveux central, une diminution de la fonction cognitive, des troubles cardiovasculaires et une toxicité rénale accrue ainsi que d'autres symptômes (Prendergast, 1910).

À partir des années 1900, les pays ont commencé à réglementer l'émission de Pb et à limiter son utilisation dans les industries spécialisées en limitant sa présence dans la peinture. Car il y a eu une diminution notable des émissions de Pb de 96% entre 1970 et 1987 (Farber, 1992).

3. Effet sur la plante

Le plomb contenu dans les plantes provient principalement du processus d'absorption à partir de la racine, mais les feuilles des plantes peuvent également capturer la suspension de Pb dans l'air grâce à leur surface pelucheuse. Après son absorption, le Pb est soit stocké dans la racine, soit transporté vers d'autres zones de la plante.

3.1 Biodisponibilité

Le terme biodisponibilité désigne la portion d'une substance ou d'un élément dans un sol qui est disponible pour être absorbée par des organismes vivants, tels que les humains, les animaux ou les plantes (Hettiarachchi and Pierzynski, 2004).

La concentration du Pb dans le sol est le premier facteur qui contribue à son absorption par les plantes. Plus la concentration du Pb augmente dans le sol plus le taux d'absorption augmente suivant une courbe linéaire (De Silva et al., 2019). Mais nous ne négligeons pas les autres critères comme l'état du sol ou l'âge de la plante ou de l'espèce comme facteurs importants (figure 2).

La rhizosphère, le champ d'action ou d'influence d'une racine, est généralement considérée comme une zone étroite du sol soumise à l'influence des racines vivantes, où les exsudats racinaires stimulent ou inhibent les populations microbiennes et leurs activités (Lynch et al., 2001).

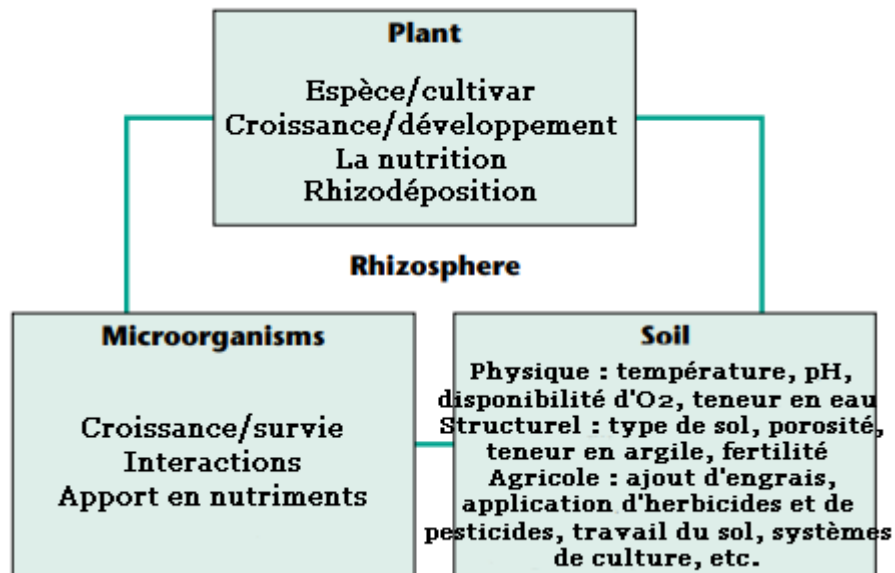


Figure 2 : Facteurs influençant les interactions de la rhizosphère (Lynch et al., 2001).

Le principal facteur qui contribue à la stabilité du Pb dans le sol est le type de composé moléculaire qu'il forme. Les phosphates de Pb, et en particulier les pyromorphites, sont l'une des formes les plus stables de Pb dans les sols dans un large éventail de conditions environnementales (Sauvé et al., 1998). Mais on peut citer d'autres minéraux comme le fer et l'aluminium dont l'adsorption avec le Pb peut être irréversible (Hettiarachchi and Pierzynski, 2004).

3.2 Effet du pH

Le facteur déterminant de la mobilité du Pb dans le sol est le pH. Il affecte le nombre de charge électrique des ions métalliques dans la solution ce qui influence la mobilité de ces ions. Dans un sol alcalin, la mobilité du Pb diminue et inversement sa biodisponibilité augmente en diminuant le pH (De Silva et al., 2019) (figure 3).

La plante contribue à la biodisponibilité du Pb par l'excrétion des exsudats racinaires (Alford et al., 2010). Pour prélever l'eau et les éléments nutritifs nécessaires à sa croissance, la plante libère diverses substances minérales et organiques : exsudats, sécrétions et excréments racinaires. Les racines peuvent affecter le pH en libérant des ions H^+ (pour prélever des cations comme NH_4^+) ou OH^- (pour prélever des anions comme NO_3^-) dans le milieu (Fahr, 2014).

3.3 Potentiel redox

Le potentiel redox est une mesure quantitative de la capacité d'une substance ou d'un système biogéochimique à recevoir ou à donner des électrons. Dans les études environnementales, il est

généralement utilisé pour caractériser la capacité réductrice ou oxydante d'un système, généralement dans un environnement anoxique ou proche de l'anoxie comme, par exemple, les sédiments de surface des systèmes d'eau douce, marins ou estuariens (Likens, 2010).

A un pH donné, la translocation du Pb de la racine vers les parties aériennes augmente avec l'augmentation du potentiel redox (Reddy and Patrick Jr, 1977) (figure 3).

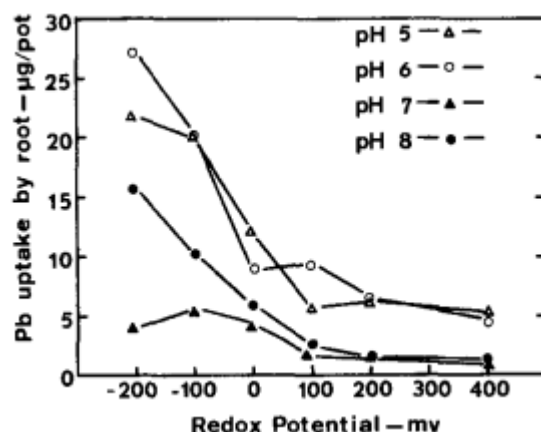


Figure 3 : Effet du potentiel redox et du pH sur l'absorption du Pb par les racines (Reddy and Patrick Jr, 1977).

4. Activité biologique et biodisponibilité du Pb

L'activité biologique joue un rôle très important sur la solubilité des métaux lourds dans les sols. Les végétaux supérieurs et les microorganismes ont une influence sur l'ensemble du milieu à l'échelle du temps et de l'espace. Parmi les microorganismes, il y a les bactéries et les champignons dont les activités métaboliques influencent la mobilité des métaux lourds comme le Pb. Epelde *et al* (Epelde et al., 2008) et Jusselme *et al.*, (Jusselme et al., 2012) ont montré que l'accumulation du Pb au niveau racinaire augmente en présence des vers de terre comme le cas du *Pontoscolex corethrurus* qui ont un impact sur la distribution des microorganismes en stimulant l'activité microbienne.

5. Toxicité pour les plantes

Les symptômes généraux de la toxicité du Pb sont une inhibition rapide de la croissance des racines, croissance lente de la plante, noircissement du système racinaire et chlorose. Le Pb inhibe la photosynthèse, la nutrition minérale et l'équilibre hydrique et les activités enzymatiques (Burton et al., 1984; Sharma and Dubey, 2005). Ces troubles perturbent les activités physiologiques normales de la plante. À des concentrations élevées, le Pb peut entraîner la mort cellulaire (Seregin and Ivanov, 2001).

Les symptômes de la toxicité chez les plantes

Le Pb inhibe la germination des graines et retarde la croissance des semis, diminue le pourcentage de germination, la longueur des racines, l'architecture racinaire (Mishra and Choudhuri, 1998). Le développement, la biomasse et la longueur des racines et des feuilles ont été négativement affectés par la présence de différentes concentrations du Pb dans les plantes (Zhou et al., 2018). Des résultats similaires ont été obtenus chez *Pisum Sativum* (Kevresan et al., 2001), chez *Zea mays* (Çimrin et al., 2007; Małkowski et al., 2002) et chez *Elsholtzia argyi* (Islam et al., 2007). Une concentration extrême de Pb provoque des effets phytotoxiques sur les ultrastructures des organites cellulaires tels que la membrane cellulaire (Gupta et al., 2009) et chloroplastique (Cunha Neto et al., 2020), le noyau et les mitochondries (Sengar et al., 2008). Le symptôme le plus évident est le changement de coloration naturelle des feuilles qui est souvent un signe visible de toxicité du plomb pour les plantes. Le plomb inhibe la photosynthèse jusqu'à environ 60 %, inhibe la synthèse de la chlorophylle en provoquant une mauvaise absorption des éléments essentiels tels que le Mg et le Fe et modifie la composition lipidique des membranes thylakoïdes (Sengar et al., 2008). La présence de forte concentration du Pb, induit de façon directe ou indirecte la production d'espèces réactives à l'oxygène (ROS) : peroxyde d'hydrogène (H_2O_2), radical superoxyde ($O_2^{\cdot -}$) et radical hydroxyle (OH^{\cdot}) (Malecka et al., 2001; Verma and Dubey, 2003). Ces ROS, si elles ne sont pas neutralisées, vont causer des dommages irréversibles à différentes molécules comme les lipides membranaires, les acides nucléiques ou encore les pigments chloroplastiques. Des enzymes, comme la catalase, les superoxyde dismutases (SODs) et les peroxydases vont participer à la dégradation des ROS. Il existe également une voie non enzymatique de neutralisation des ROS par le glutathion, la cystéine, l'acide ascorbique et le tocophérol (Verma and Dubey, 2003). Le Pb n'est pas un métal oxydo-réducteur, comme le fer par exemple, et son effet sera donc indirect (Sharma and Dubey, 2005).

6. Les plantes hyperaccumulatrices

L'hyperaccumulation de métaux est associée à une capacité fortement accrue à détoxifier l'accumulation de métaux dans les tissus au-dessus du sol. Une plante est qualifiée d'hyperaccumulatrice de Pb lorsqu'elle est capable d'absorber plus de 1000 ppm de Pb dans ses parties aériennes (Baker and Brooks, 1989). Ces plantes hyperaccumulatrices sont très présentes dans la famille des *brassicaceae* principalement *Arabidopsis halleri*, *Noccaea*

caerulescens et *Allysum* sp. qui sont considérées comme des plantes modèles pour l'étude de l'hypertolérance aux métaux Zn, Cd et Ni (Krämer, 2010).

Aujourd'hui, 500 espèces de plantes présentant des individus tolérants et hyperaccumulateurs sont connues, aussi bien dans les zones tropicales que tempérées. La majorité d'entre elles (317 espèces) concerne l'hyperaccumulation du nickel (Reeves and Baker, 2000). Les espèces hyperaccumulatrices de zinc arrivent en deuxième position, bien qu'elles soient en nombre beaucoup plus réduit (Balafrej et al., 2020). Par contre, le nombre de plantes décrites comme hyperaccumulatrices du Pb est restreint. Selon (Bhargava et al., 2012), seules 14 espèces auraient été identifiées comme hyperaccumulatrices du Pb et, parmi celles-ci, *Sesbania drummondii*, *Arabis paniculata* et *Plantago orbignyana*.

7. Mécanismes moléculaires de tolérance

Comme déjà mentionné, le plomb peut provoquer des effets délétères chez les plantes sensibles aux métaux, il se manifeste dans le fonctionnement des organes et des cellules : altérations de l'absorption des cations (K^+ , Ca^{2+} , Mg^{2+} , Mn^{2+} , Zn^{2+} , Cu^{2+} , Fe^{3+}) et des anions (NO_3^-) et diminution des composés maintenant la turgescence cellulaire et la paroi cellulaire, la taille des cellules de garde, l'ouverture des stomates, le niveau d'acide abscisique et la surface foliaire et affecte le fonctionnement sous-cellulaire comme le chloroplaste-photosynthèse, les irrégularités noyau-mitose et les mitochondries-respiration (Arshad, 2009).

Cependant, les plantes hyperaccumulatrices, contrairement aux plantes sensibles aux métaux, pourraient immobiliser, exclure, chélater et compartimenter les ions métalliques (Peng et al., 2021).

Les plantes ont mis en place un réseau complexe de mécanismes pour maintenir l'homéostasie des éléments trace métallique. Le Pb reste le métal le moins étudié dans les interactions plantes-métaux lourds. Ceci est principalement dû à sa faible biodisponibilité dans le sol rendant le travail expérimental plus compliqué. Il existe néanmoins chez les plantes des mécanismes d'absorption et d'accumulation du Pb dont certains ont été caractérisés. Les mécanismes moléculaires de la tolérance et de l'accumulation les plus étudiés et décrits concernent en particulier le Zn (Balafrej et al., 2020). Les principales stratégies de la tolérance et l'accumulation des métaux lourds sont : (a) Le transport comprenant l'absorption, l'exclusion ou la translocation vers les zones de stockage ; (b) La chélation ; (c) La séquestration (Figure 4)

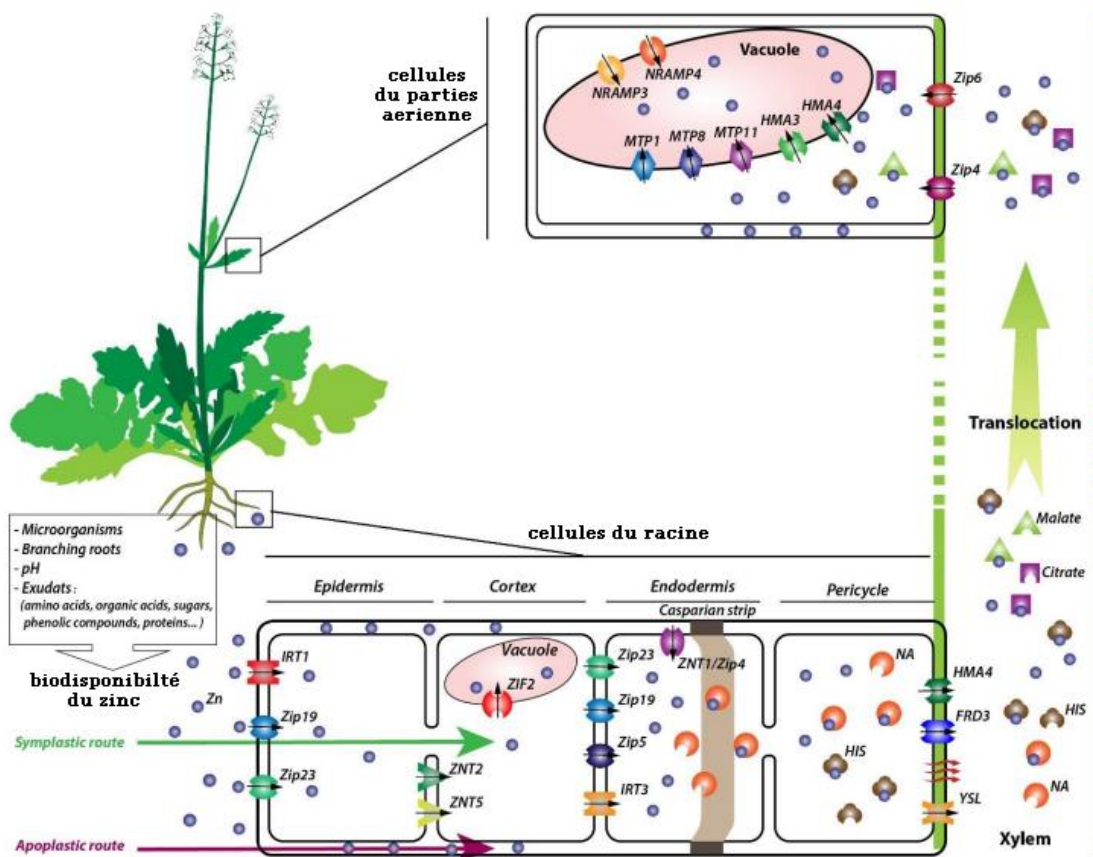


Figure 4 : Un modèle des mécanismes qui se produisent dans les plantes d'hyperaccumulation lors de l'exposition au zinc (Zn) : absorption, chélation, transport et séquestration des ions Zn. La biodisponibilité du zinc peut être influencée par plusieurs facteurs, tels que les micro-organismes, les racines ramifiées, le pH et les exsudats. Une fois adsorbé par les racines, le Zn peut être absorbé par une voie apoplastique : une diffusion passive à travers les cellules, ou par une voie symplastique via des transporteurs. Dans cette dernière voie, l'absorption du Zn par les cellules de l'épiderme est principalement favorisée sous forme de micro-organismes, de pH de ramification et d'exsudats. Une fois adsorbé par les racines, le Zn peut être absorbé par une voie apoplastique : une diffusion passive à travers les cellules, ou par une voie symplastique via des transporteurs. Dans cette dernière voie, l'absorption de Zn par les cellules de l'épiderme est principalement favorisée par IRT1, ZIP19 et ZIP23. Pour atteindre le cortex, le Zn peut être diffusé directement ou par l'intermédiaire de ZNT2 et ZNT5. Ensuite, le Zn peut être soit stocké dans des vacuoles (promues par ZIF2) soit transporté vers l'endoderme via les transporteurs suivants : ZIP23, ZIP19, ZIP5 et IRT3. Le Zn suivant la voie apoplastique est stoppé par la bande casparienne, puis pénètre dans l'endoderme via ZNT1/ZIP4. A ce niveau, le Zn peut être chélaté par la nicotianamine (NA) ou diffusé directement aux cellules du péricycle où une partie peut également être associée à l'histidine (His). Le Zn non chélaté peut atteindre le xylème par diffusion directe ou via YSL, ferric réductase déficiente 3 (FRD3,) et HMA4. Le Zn traverse ensuite le xylème sous une forme sans Zn ou couplé à His, citrate ou malate. Pour pénétrer dans les cellules foliaires, le Zn peut pénétrer passivement sous forme chélatée ou sous forme sans Zn via les protéines ZIP4 et ZIP6. Il est ensuite séquestré à l'intérieur de la vacuole par les transporteurs MTP1 (protéines de tolérance aux métaux 1), MTP8, MTP11, NRAMP3, NRAMP4, HMA3 et HMA4, ou bloqué dans la paroi cellulaire (Balafrej et al., 2020).

7.1. L'absorption

L'absorption racinaire du Pb a fait l'objet de nombreuses recherches sur diverses espèces végétales. Ces travaux ont montré que le Pb, après qu'il se fixe au rhizoderme, pénètre dans le système racinaire de façon passive et il suit le système de conduction de l'eau (Seregin et al., 2004). Cette pénétration se fait en deux voies : voie apoplastique et voire symplastique.

Le Pb passe par la voie apoplastique ce qui explique qu'il se trouve principalement dans la paroi cellulaire et précipité dans l'espace intercellulaire. Il a une forte affinité pour les groupes carboxyle et les pectines et dans une moindre mesure pour les molécules d'hémicellulose et de cellulose et la lignine des parois cellulaires. Cette forte affinité pour les composants de la paroi cellulaire peut expliquer la faible mobilité du Pb (Pourrut et al., 2013).

L'utilisation de la voie symplastique se produit dans les zones où les cellules se divisent activement, comme dans la zone apicale ou dans le protoderme, mais elle se trouve confinée à certains compartiments cellulaires, tels que les vacuoles, les vésicules dictyosomales et les vésicules du réticulum endoplasmique (Pourrut et al., 2013).

De point de vue moléculaire, le Pb profite de la non-spécificité de certains canaux/transporteurs et de la très forte différence de potentiel membranaire, pouvant dépasser -200 mV dans les cellules du rhizoderme pour se diffuser à l'intérieur de la racine. Parmi les transporteurs membranaires impliqués dans l'entrée des ETMs dans les cellules de l'épiderme de la racine, on trouve la famille des transporteurs régulés par le Zn et le Fe « Zinc and Iron regulated transporters Protein » (*ZIP*), qui sont, en général, fortement exprimés chez les plantes hyperaccumulatrices en conditions normales (Krämer et al., 2007).

D'autre part, il y'a les canaux cationiques non sélectifs. Parmi ces canaux, on trouve : les canaux calciques activés par la dépolarisation (DACC), les canaux calciques activés par l'hyperpolarisation (HACC) et les canaux cationiques insensibles à la tension (VICC) sont considérés comme l'une des principales voies d'entrée du Pb dans les cellules racinaires. L'entrée de Pb dans la cellule nécessite que la cellule dépense de l'énergie par l'excrétion de protons dans l'environnement externe via des pompes $H^+ / ATPase$ qui provoque un pH bas qui favorise la solubilité des métaux et une concentration localement élevée de Pb dans la solution du sol (Pourrut et al., 2013).

Un autre transporteur transmembranaire, *CNGC1* (cyclic nucleotide-gated ion channel) chez *Arabidopsis* et *CBP4* (calmodulin-binding protein) chez le tabac, a été identifié comme

intervenant dans l'assimilation du Pb chez *A. thaliana* et *N. tabaccum* (Sunkar et al., 2000). Ces canaux sont proches de la famille des canaux ioniques mais elles sont spécifiquement modulées par les nucléotides cycliques non sélectifs aux cations.

7.2. Détoxification : la chélation et la séquestration

7.2.1- Chélation

La chélation est une étape importante pour détoxifier les métaux lourds. Ça se fait dans le cytoplasme par des ligands spécifiques tels que le glutathion (GSH), les phytochélatines (PC) et la métallothionéine (MT), avant la compartimentation du complexe ligand-ETM dans les vacuoles (Dubey et al., 2014).

a. Phytochélatines (PC)

La formule générale des PC est $(\gamma\text{-Glu-Cys})_n\text{X}$ où n est un nombre variable de 2 à 11 selon l'organisme, bien que la plupart des formes courantes aient 2-4 peptides, X représente un acide aminé tel que Gly, β -Ala, Ser, Glu ou Gln. En présence des groupements thiol de Cys, les PCs peuvent chélater 1 à 2 ions de Pb à la fois, formant des complexes protégeant le cytosol de ces ions métalliques libres (Arshad, 2009). Le glutathion est le précurseur de la synthèse de ces peptides de liaison aux métaux, qui est catalysée par la phytochélatine synthase (PCS) chez les plantes lors de l'exposition aux métaux. La présence de phytochélatines allant de PC1 à PC3 dans les racines de *Vetiveria zizanioides* L. lors d'une exposition à 1 200 mg Pb L⁻¹. le type et le niveau de PC montrent une relation linéaire avec la teneur en Pb, de sorte que le PC3 n'a été observé que dans les tissus racinaires en raison des niveaux élevés de Pb, mais le PC le plus abondant dans les tissus des racines et des pousses de vétiver était le PC1 (Andra et al., 2010).

b. Métallothionéines (MTs)

Les MT sont des peptides riches en cystéine et de faible poids moléculaire, capables de se lier aux ions métalliques au moyen de liaisons covalente. Les MT sont divisées en trois classes différentes en fonction de leur teneur en cystéine et de leur structure. Les motifs Cys-Cys, Cys-X-Cys et Cys-X-X-Cys (dans lesquels X désigne tout acide aminé) sont caractéristiques et invariants pour les MT (Arshad, 2009). Ces résidus Cys de liaison aux métaux dans les MT peuvent également piéger les espèces réactives de l'oxygène (ROS). Ainsi, au lieu de la remobilisation des métaux, l'induction de la MT pendant la sénescence est corrélée à la signalisation et au piégeage des ROS et peut réguler les phytohormones internes de la plante comme les cytokinines (Joshi et al., 2016). En termes de quantité de transcription, de nombreux

gènes MT végétaux sont exprimés à des niveaux très élevés dans tous les tissus (Arshad, 2009). *MT1* a montré une expression plus élevée dans les racines alors que *MT2* et *MT3* dans les feuilles et *MT4* dans les graines (Joshi et al., 2016). *AtMT1a* et *AtMT2a* semblent s'accumuler dans les trichomes, ce qui permet la séquestration des ions de métaux lourds chez *Arabidopsis thaliana* (Arshad, 2009). Le rôle des MT sont impliqués dans une myriade de processus physiologiques tels que la régulation de la croissance et de la prolifération cellulaire, la protection et l'homéostasie des métaux toxiques, le piégeage des radicaux libres ou la protection contre le stress oxydatif, la réparation des dommages à l'ADN et la défense anti-apoptotique (Joshi et al., 2016).

c. Acides organiques

La chélation extracellulaire par des acides organiques, tels que le citrate et le malate, est importante dans les mécanismes de tolérance aux métaux lourds. Les acides organiques et certaines acides aminés, en particulier l'histidine « His », jouent un rôle dans la chélation des ions métalliques à la fois dans les cellules et dans la sève du xylème (Arshad, 2009). L'acide malique était corrélé à la tolérance au Zn mais pas au Cu ou au Ni chez *S. vulgaris*, mais chez *Deschampsia caespitosa*, l'acide citrique était en excès de 5 à 10 fois par rapport à l'acide malique. Le malate chélatait le Zn dans le cytosol et le complexe était déplacé dans la vacuole (Rauser, 1999).

Les chélateurs métalliques du Pb tels que l'EDTA, l'HEDTA, le DTPA, le NTA et l'acide citrique peuvent faire prospérer la plante sensible dans un environnement à de forte concentration en Pb. Chez *Sesbania drummondii*, il a été reporté que les chélateurs augmentaient le transport du Pb dans l'ordre EDTA > HEDTA > DTPA > NTA > acide citrique (Ruley et al., 2006).

7.2.2 Translocation

La translocation est le transport des ETMs vers les feuilles. C'est la fonction principale qui différencie les plantes tolérantes dites « excluser » de celles qui sont accumulatrices, car les hyperaccumulatrices, présentent une forte capacité de translocation des ETMs vers les parties aériennes (Gupta et al., 2013). Pour ce phénomène, le Pb recrute le système vasculaire en suivant le flux d'eau vers les feuilles où il s'accumule (Verbruggen et al., 2009). Ils ont identifié une ATPase de Type *P1B*, *HMA4* (heavy metal ATPase) chez *A. thaliana* (Mills et al., 2003) puis chez la plante hyperaccumulatrice *Noccaea caerulescens* (Papoyan and Kochian, 2004). Cette protéine membranaire serait impliquée dans le transport longue distance du Cd, Pb, Zn et

du Cu (Wong and Cobbett, 2009). Elle appartient à une famille multigénique qui compte 8 gènes chez *A. thaliana*. En fonction de leurs séquences protéiques, ces transporteurs sont soit des transporteurs de Zn/Cd/Co/Pb, soit des transporteurs de Cu/Ag (Williams and Mills, 2005). Aussi, chez (Mouna et al., 2015) ont montré que l'expression d'*HiHMA4* est stimulée par le Pb chez *H. incana*, ce qui suggère que ce transporteur se lie au Pb et l'amène vers les parties aériennes de la plante ou vers la vacuole pour la séquestration. Lors du passage de Pb dans la sève brute, il peut former des complexes avec des acides aminés tels que l'histidine et des acides organiques tels que l'acide citrique, l'acide fumarique et l'acide malique (Ghnaya et al., 2013; Shahid et al., 2012).

7.3 Transport et séquestration du plomb

Le transport et l'homéostasie des cations métalliques sont essentiels à la nutrition des plantes et à la tolérance aux métaux lourds. Les complexes ligand-ETM formés sont généralement séquestrés dans les vacuoles, au niveau des feuilles et, dans une moindre mesure, au niveau des racines. De nombreux transporteurs situés au niveau du tonoplaste ont été identifiés. Il existe des ATPases de métaux lourds qui sont impliquées dans l'homéostasie et la tolérance globales des ions métalliques chez les plantes. Le gène *HMA3*, qui code pour une ATPase de type P, complémente des mutants de levure hypersensibles au Pb et au Cd. De plus, ce transporteur serait localisé au niveau du tonoplaste (Verret et al., 2004). Des études sur l'ATPase 4 (*HiHMA4*) ont montré leur implication dans la tolérance au Pb chez *H. incana* (Auguy et al., 2013). Pour les transporteurs de type ABC, on peut noter que l'activité du promoteur du gène qui code pour la protéine *MRP3* (Multidrug Resistance-Associated) est augmentée par le Pb (Zientara et al., 2009). Chez *Arabidopsis*, un transporteur mitochondrial de type ABC, *ATM3*, serait impliqué dans la tolérance au Cd et au Pb et dans l'accumulation du Cd dans les feuilles (Jamla et al., 2021). La famille des « Natural Resistance-Associated Macrophage Protein » (*NRAMP*) transporte les cations vers le cytoplasme et leur première fonction biologique est l'homéostasie du Fe (Thomas et al., 2021). Ces transporteurs pourraient en plus, jouer un rôle dans l'accumulation des ETMs. En effet, *NRAMP3* serait impliqué dans la remobilisation du Fe, du Mn, du Zn et du Cd (Thomine et al., 2003). Mais d'autres transporteurs peuvent agir indirectement sur l'absorption du Pb. Par exemple les transporteurs de nitrate NRT1.1 transporte les nitrates NO_3^- vers la cellule mais il régule également l'absorption du Pb en consommant les protons libres dans le sol ce qui augmente le pH et provoque la diminution de biodisponibilité du Pb (Zhu et al., 2019).

A. Étude des populations : outil pour la compréhension de l'adaptation des plantes aux métaux lourds

L'évolution de la tolérance aux métaux dépend de l'exposition d'une espèce végétale à la teneur en métaux des sols. Il est limité à certains ordres, familles et genres non liés. Néanmoins, il existe une grande similitude dans l'implication d'un ou deux gènes par métal contrôlant la tolérance spécifique au métal et dans ses mécanismes physiologiques. Ils sont tous basés sur un déplacement des performances sous des niveaux élevés de métaux et sur la compartimentation des métaux dans les cellules et tissus physiologiques moins actifs. Malgré le potentiel évolutif des tolérances aux métaux, il existe des sols extrêmement enrichis en métaux sans aucune croissance végétale (Ernst, 2006).

1. Adaptation des plantes et réponse au stress métallique

Les zones contaminées par les métaux lourds existent généralement depuis l'ère industrielle (origines plus récentes) et occupent des zones bien définies. Ces habitats représentent un modèle de référence pour l'étude des processus évolutifs rapides adaptés à des facteurs écologiques très limités (Antonovics et al., 1971). Souvent, des espèces végétales non résistantes aux métaux ont colonisé ces milieux, acquérant une résistance aux métaux lourds. En raison des fortes pressions de sélection de ces environnements, les individus issus de populations nouvellement adaptées peuvent subir des changements rapides, en particulier dans les systèmes de reproduction ou les traits d'histoire de vie tels que la pérennité, le mode de dispersion ou la croissance. La divergence génétique due à des mécanismes de dérive ou d'isolement reproductif peut conduire à la spéciation (Ashraf et al., 1987).

Certaines plantes sont capables de se développer uniquement sur des sols à fortes concentrations en métaux lourds et sont appelées des Metallophytes (Antonovics et al., 1971). (Lambinon and Auquier, 1964) les ont réparties en deux groupes : (1) les metallophytes absolues, endémiques des sols métallifères (*Viola calaminaria*, *Noccaea calaminare*, *Minuartia verna ssp. Hercynica*) et (2) les pseudo-metallophytes, que l'on rencontre à la fois sur des sites métallifères (population métallicole « PM ») et sur des sites non pollués (population non métallicole « PNM »). Ainsi, les populations se développant sur les sites pollués sont qualifiées de populations métallicoles et celles se développant sur les sites non contaminés de populations non-métallicoles.

Les études de populations passent d'abord par une compréhension au niveau phénotypique puis au niveau génétique de la distribution des polymorphismes pour les caractères de tolérance et d'hyperaccumulation. L'étude de variation entre populations issues d'habitats différents pour ces caractères revêt un intérêt fondamental et permet de rechercher des corrélations entre le caractère d'accumulation ou de tolérance et les caractéristiques et les propriétés liées à l'habitat.

L'étude génétique nous permet de faire l'évaluation du statut génétique de la population à de nombreux objectifs. D'abord, elle aide à (i) concevoir la translocation des plantes pour les espèces en voie de disparition (Van Rossum et al., 2022) ; ou (ii) pour comprendre la composition génétique qui fait que la plante résistante tolère le stress par rapport à la plante non tolérante. Par exemple, *Agrostis tenuis* non tolérante a développé des individus avec des tolérances au cuivre proches de celles des populations totalement tolérantes lorsqu'elles sont soumises à un stress métallique dans un sol riche en cuivre (Walley et al., 1974).

Dans le domaine d'étude des populations, différents marqueurs moléculaires ont été utilisés. Par exemple, chez *Thlaspi caerulescens*, les séquences polymorphes amplifiées clivées (CAPS) et les microsatellites aux loci des gènes potentiellement impliqués dans les réponses d'hyperaccumulation et de tolérance aux métaux lourds et les indels dans les motifs non répétés ont été étudiés (Besnard et al., 2009). Des motifs AFLP ont été utilisés chez *Armeria maritima* s.l pour l'étude de la population (Abratowska et al., 2012). Les données RAPD ont été utilisées pour comparer les populations indigènes (sud de l'Europe) et exotiques (îles britanniques) de *Hirschfeldia incana* (Lee et al., 2004). Chacun de ses marqueurs a un intérêt spécifique.

Dans les études sur les plantes adaptées aux métaux lourds, les plantes sur des sols contaminés sont considérées comme la population descendante et celles sur le sol adjacent non contaminé sont considérées comme la population ancestrale ou source (Cappa and Pilon-Smits, 2014). Étant donné qu'il est possible, voire probable, que la population source ait également subi des changements, la différence entre les deux populations est plus précisément appelée divergence. La période de divergence se situe entre 20 et 1000 ans. Ce faible taux de divergence pourraient être dus au fait que la sélection pour la tolérance aux métaux est forte sur les mines et que la plupart des évolutions se produisent au cours des premières générations (Bone and Farres, 2001).

2. Variabilité de la tolérance et l'accumulation des éléments-traces métalliques (ETMs)

La tolérance aux métaux lourds a été décrite pour la première fois par Pratt en 1934, qui a observé ce phénomène chez *Silvestre Melandrium* (*Silene vulgaris*) (Pauwels et al., 2006). (Prat, 1934) a comparé la survie d'individus issus de graines provenant de déblais d'une mine de cuivre, à d'autres individus issus de graines provenant de sols non pollués. Sur des sols à forte concentration en cuivre, les individus provenant de sites normaux meurent rapidement tandis que les autres survivent. Par la suite, le même phénomène a été observé chez *Agrostis tenuis* pour la tolérance au Pb (BRADSHAW, 1952) et chez *Mimulus guttatus* (Macnair, 1983) pour la tolérance au cuivre. Ces travaux ont démontré que la tolérance aux métaux lourds est un phénomène spécifique des métaux présents dans le sol d'origine des populations et déterminé génétiquement.

Par la suite, de nombreux articles ont été publiés sur *Arabidopsis halleri*, car elle est métallophte tolérante au zinc (Willems et al., 2007). Ces études de population utilisent une population tolérante au stress métallique et une autre homologue qui ne tolère pas à la fois la croissance dans un environnement stressant pour examiner son effet sur l'expression génétique et la constitution génique de la plante afin de déterminer les gènes appliqués à la tolérance au métal.

Pour *Armeria maritima*, un ensemble de traits morphologiques (longueur des bractées involucrees externes), ou du moins sa fréquence au sein d'une population donnée, pourrait être utilisé pour différencier les plantes des sites M et NM mais l'utilisation de marqueurs moléculaires apporte un éclairage nouveau sur la taxonomie des populations métallicoles (Abratowska et al., 2012).

Une différenciation génétique liée aux concentrations de métaux lourds dans le sol a été mise en évidence au niveau de certains loci candidats de *Thlaspi caerulescens*, en particulier pour les gènes codant pour les transporteurs de métaux (Besnard et al., 2009). Pour *T. caerulescens*, il y avait une forte différenciation de la structure génétique entre les différents sites d'échantillonnage, en particulier dans les sites métallifères que dans les sites non métallifères. Cela suggère que la sélection naturelle limite le flux de gènes entre les sites métallifères et non métallifères à ces loci (Besnard et al., 2009). Mais un constat contraire chez les populations d'*A. maritima*, les populations tolérantes aux métaux ont évolué indépendamment des populations non tolérantes, et que le modèle de flux de gènes entre les populations est cohérent avec un

modèle d'isolement par la distance, avec un flux de gènes considérable entre les populations voisines. Même pour les populations géographiquement proches de NM et M, l'isolement par la distance plutôt que par le type de substrat est d'une importance cruciale pour façonner la structure génétique de la population au sein d'*A. maritima* (Abratowska et al., 2012).

À proximité des populations naturelles d'*Agrotis tenuis*, une fonderie de Zn a été mise en place, conduisant à une contamination du sol. Quelques années plus tard, les populations de la région ont développé des capacités de tolérance de zinc identique à celle des populations métallicoles, issues des sites beaucoup plus âgés, situés en zone post-minier (Al-Hiyaly et al., 1988). La même tendance a été observée chez *Noccaea caerulescens*, une espèce hyperaccumulatrice des métaux (Cd, Pb et Zn), où la tolérance est également un caractère constitutif (Meerts and Van Isacker, 1997) mais la sélection naturelle l'a amplifié chez les populations M. Il semble que la tolérance aux métaux est un caractère qui peut évoluer rapidement. Des études ont montré que la tolérance aux métaux pourrait avoir évolué indépendamment des populations géographiquement éloignées (Mengoni et al., 2001; Pauwels et al., 2005).

B. Etude transcriptomique

La technique de séquençage, arrivée dès les années 70, a donné accès à la composition en nucléotide des molécules d'ADN (Sanger and Coulson, 1975). Cette avancée technologique a eu un impact considérable sur l'acquisition de nouvelles connaissances en biologie moléculaire, en évolution, génomique environnementale, dans le domaine médical et bien d'autres, ainsi que dans le développement d'outils statistiques et informatiques adaptés à ce type de données. Depuis, les technologies de séquençage ont évolué, en particulier avec l'arrivée de la seconde génération de séquenceurs (NGS), donnant un accès massif et à moindre coût aux séquences génomiques. Ces technologies ont également permis le renouvellement des études transcriptomiques (Dong and Chen, 2013).

La transcriptomique consiste en l'étude de l'ensemble des ARN (ou transcrits), souvent plus particulièrement les ARN messagers (ARNm), qui sont utilisés comme intermédiaires pour la production de protéines. Le transcriptome étudié peut être celui d'un type cellulaire particulier ou d'un tissu spécifique. La transcriptomique constitue aujourd'hui un domaine de recherche à part entière (Lowe et al., 2017).

Le séquençage à haut débit de l'ARN (appelé RNA-seq) est actuellement la technologie la plus employée pour identifier et quantifier, à large échelle, les transcrits extraits d'un ou plusieurs individus, tissus ou types cellulaires, dans des conditions physiologiques données. Avec la production massive de données RNA-seq, des méthodes et outils spécifiques permettant l'analyse de ce type de données ont été et sont encore développés.

1. Technologies pour l'analyse à large échelle du transcriptome

Différentes technologies permettent l'étude à large échelle des transcriptomes. On peut séparer les puces à ADN, basées sur l'hybridation des séquences pour quantifier l'expression des transcrits, et les technologies de séquençage, en particulier le RNA-seq qui permettent à la fois l'accès à la séquence des transcrits, et, selon la profondeur de séquençage et leur niveau d'expression, un accès plus ou moins précis à la quantification de ceux-ci. Il existe également des technologies (dites "gène à gène") permettant d'étudier et d'analyser les transcrits et/ou les gènes au cas par cas. C'est le cas des RT-PCR et des RT-PCR quantitatives, qui sont toujours utilisées, en particulier pour valider spécifiquement certains résultats obtenus après analyse des données dites "haut débit" (Lowe et al., 2017).

1.1. Les puces à ADN

La puce à ADN est une technologie développée au cours des années 90 et qui est principalement utilisée afin de quantifier l'expression des gènes (ou transcrits). Il s'agit d'une petite surface (puce) sur laquelle sont fixées plusieurs milliers de molécules d'ADN (appelées sondes) dont la séquence en acide nucléique, ainsi que la position sur la puce, sont connues. Comme pour la grande majorité des technologies permettant l'étude du transcriptome, elle nécessite au préalable de rétro-transcrire les molécules d'ARN en ADN dit complémentaire (ADNc). Elle permet, via l'hybridation des sondes fixées sur la puce avec les brins d'ADNc présents dans l'échantillon étudié, de mesurer la concentration relative d'une séquence nucléotidique dans cet échantillon. Celle-ci est mesurée par la fluorescence émise par les brins d'ADNc, marqués avant hybridation. L'analyse des intensités mesurées permet ensuite d'identifier et de quantifier les transcrits présents dans l'échantillon et généralement de comparer plusieurs échantillons. La principale limite des puces à ADN pour l'étude des transcriptomes vient du fait que cette technologie ne donne pas accès à la séquence des gènes ou des transcrits et qu'elle nécessite des connaissances a priori sur les gènes (ou transcrits) à étudier. Elle est donc peu adaptée pour travailler sur des espèces non modèles (Larsson et al., 2006; Watson et al., 1998).

1.2. Le RNA-seq

Le RNA-seq est une approche relativement récente utilisant la seconde génération de séquenceur, appelés NGS (Next Generation Sequencing). Elle est actuellement la méthode la plus utilisée pour les analyses de transcriptome à large échelle et permet d'identifier et de quantifier les transcrits. Elle ne nécessite pas de connaître à priori les séquences des gènes ou des transcrits et peut donc être utilisée dans des études portant sur des espèces non modèles, c'est à dire dont le génome de référence n'est pas disponible. Il existe plusieurs technologies regroupées sous le terme "NGS". On peut citer les technologies Illumina, 454, Ion Torrent ou encore SOLiD. Celles-ci sont dites à "haut-débit", permettant de produire plus de séquences et à un prix plus bas que le séquençage Sanger, mais les séquences produites sont également plus courtes (Figure 5) (Marguerat and Bähler, 2010; Wang et al., 2009).

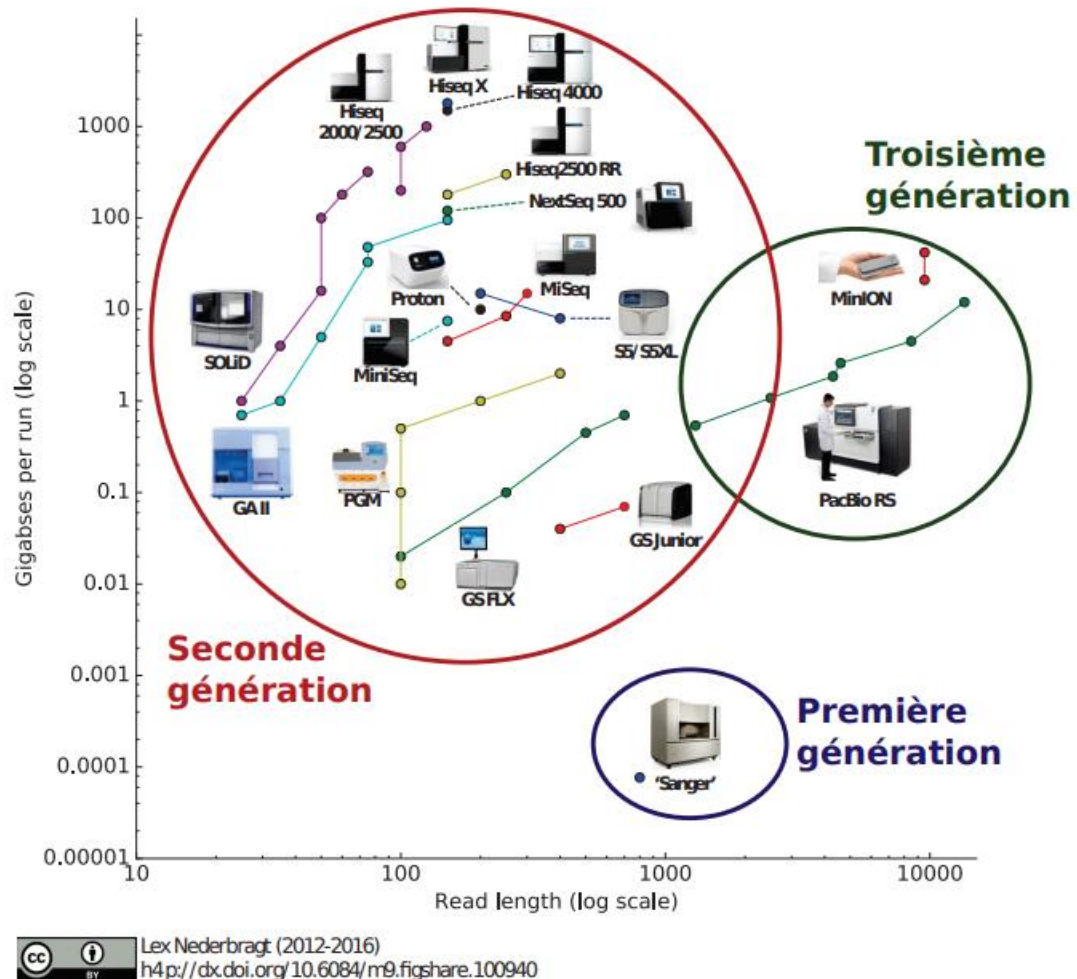


Figure 5 : Les différentes technologies de séquençage. Graphique représentant le nombre de lectures (reads) obtenues par run en fonction de la longueur des lectures pour les différentes technologies de séquençage de première, deuxième et troisième génération (Lex Nederbragt , 2012-2016)

Selon la machine utilisée, un run Illumina permet par exemple de produire plusieurs dizaines de millions de lectures (voire un peu plus d'un milliard) d'une centaine de nucléotides en moyenne. Les NGS possèdent plusieurs étapes communes, notamment au niveau de la préparation des librairies d'ADN qui seront séquencées. En effet, même dans le cadre d'études transcriptomiques, c'est de l'ADN qui est lu par les machines. Il est donc nécessaire de rétro-transcrire l'ARN en ADNc au préalable. La préparation des librairies et la suite du séquençage sont alors identiques (Figure 3), que ce soit pour des données génomiques (DNA-seq) ou transcriptomiques (RNA-seq). L'ADN à séquencer est ensuite fragmenté et il y a généralement sélection des fragments selon leur taille. Des adaptateurs sont ensuite ajoutés aux extrémités de chaque fragment. Ces adaptateurs, spécifiques à chaque technologie, permettent l'amplification et la fixation des fragments à séquencer. Dans le cas où plusieurs échantillons différents sont

séquencés, les adaptateurs peuvent également permettre l'identification de chaque échantillon (code barre). L'étape du séquençage est propre à chaque technologie. La plus répandue actuellement est la technologie Illumina (Bentley et al., 2008). Celle-ci permet d'obtenir des lectures assez courtes, aujourd'hui en moyenne de 100 à 150 pb et pouvant atteindre 300 pb. Le séquençage peut concerner une extrémité (single end) ou les deux extrémités (paired-end) des fragments. Il est également possible de préparer des bibliothèques dites brin spécifiques dans lesquelles l'information du brin d'origine des transcrits est conservée (Figure 6)(Wang et al., 2009).

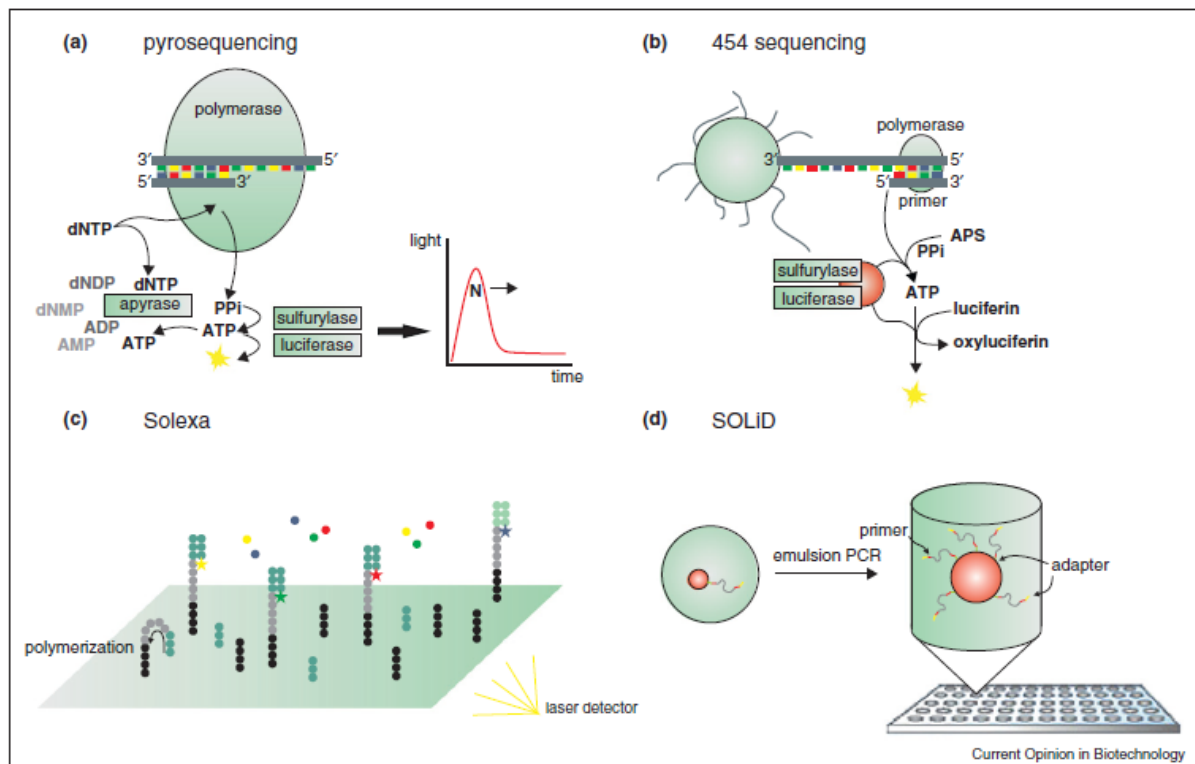


Figure 6 : Principes de base des techniques NGS. (a) pyroséquençage : l'incorporation d'un nouveau nucléotide génère une lumière détectable. (b) Séquençage 454 : l'incorporation de nucléotides est associée à la libération de pyrophosphate résultant en un signal lumineux. (c) Solexa : les fragments d'ADN forment des ponts double brin et après l'ajout des terminateurs marqués, le cycle de séquençage commence. (d) SOLiD : si les adaptateurs sont liés, une PCR en émulsion est effectuée pour générer ce que l'on appelle des clones de billes (Mutz et al., 2013).

Le RNA-seq permet, contrairement à la puce à ADN, d'avoir accès à la séquence, au nucléotide près, des transcrits de l'échantillon, et ne nécessite pas de connaissance a priori de ces séquences. Elle permet potentiellement, selon la profondeur de séquençage, d'avoir accès à l'ensemble des transcrits exprimés. La quantification relative des transcrits obtenue à partir de

données RNA-seq est également plus précise qu'avec la puce à ADN (Figure 7S) (Watson et al., 1998) .

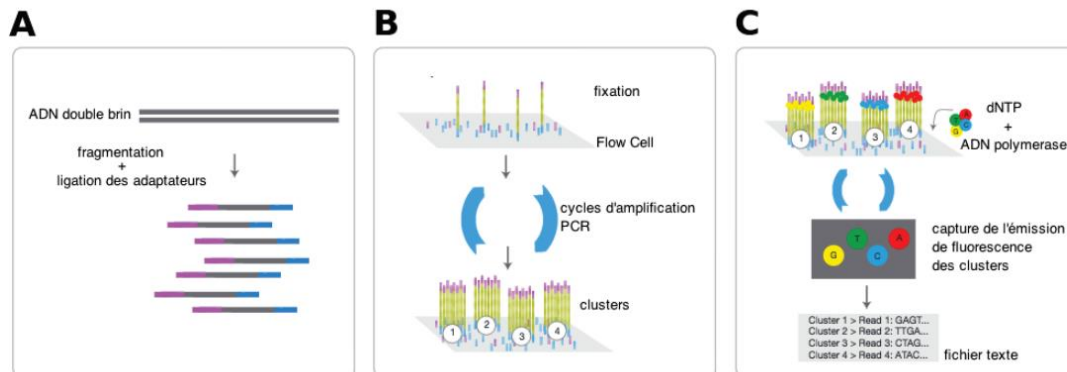


Figure 7 : Séquençage illumina. A) L'ADN à séquencer est fragmenté et des adaptateurs sont liées aux extrémités de chaque fragment obtenu. B) Les fragments à séquencer vont être amplifiés par PCR, après avoir été fixés sur la “flow cell” (lame de verre). On obtient des clusters correspondants chacun à un fragment initial et contenant environ 1000 copies de celui-ci. C) Dans chaque cluster, une base est incorporée dans chaque fragment, celle-ci émet un signal fluorescent qui est enregistré. Ces étapes sont répétées jusqu'à séquençage du fragment de taille souhaitée (Illumina 2016).

Les données RNAseq permettent l'analyse des transcriptomes, que ce soit au niveau de leur séquence ou de leur abondance. Ces données nous permettent d'avoir accès à l'identification et la quantification les gènes exprimés dans l'échantillon étudié (Wang et al., 2009). Dans la plupart des cas, l'épissage alternatif produit plusieurs types de transcrits matures par gène. Il est possible d'approfondir l'analyse à l'échelle des transcrits, et d'identifier les variantes d'épissages alternatifs appartenant au même gène. Dans la pratique, l'association des différents variantes d'épissages alternatifs pour reconstruire l'ensemble des transcrits présents dans l'échantillon reste un problème complexe et ce même avec l'utilisation d'un génome de référence (Marguerat and Bähler, 2010). Le RNA-seq permet également d'avoir accès aux autres variations nucléotidiques (par exemple les SNP). Ces variations peuvent avoir lieu au sein des génomes ou bien pendant/après la transcription (dans ce cas on parle de RNA editing). Pour l'analyse de données en RNAseq, on peut globalement séparer les méthodes existantes en deux catégories : celles basées sur l'alignement des lectures sur un génome (ou un transcriptome) de référence, et celles basées sur l'assemblage de novo des lectures (Wang et al., 2009).

2. Méthodes d'assemblage (sans génome de référence)

Lorsque aucun génome ou transcriptome de référence n'est disponible (c'est le cas de notre étude avec la plante *d'Hirschfeldia incana*), il est possible d'assembler les lectures pour reconstruire les transcrits présents dans l'échantillon séquencé (Paszkiewicz and Studholme, 2010). On parle d'assemblage de novo. Ce type de méthode permet notamment l'analyse de données RNA-seq dans le cadre d'espèce dites "non-modèles" pour lesquelles il n'y a pas de génome de référence (Paszkiewicz and Studholme, 2010). L'utilisation de méthodes d'assemblage est également pertinente chez les espèces modèles lorsque l'on souhaite identifier de nouveaux gènes, de nouveaux variants d'épissages ou dans certains cas particuliers, lorsque le génome de référence est trop différent de celui étudié. L'assemblage consiste en l'utilisation des chevauchements entre les lectures afin de reconstituer les séquences. Il existe différentes méthodes d'assemblage et celles-ci sont généralement basées sur deux types de graphes : les graphes d'overlap et les graphes de de Bruijn (Robertson et al., 2010). Les différentes étapes d'assemblage sont décrites dans la figure 8, ci-dessous.

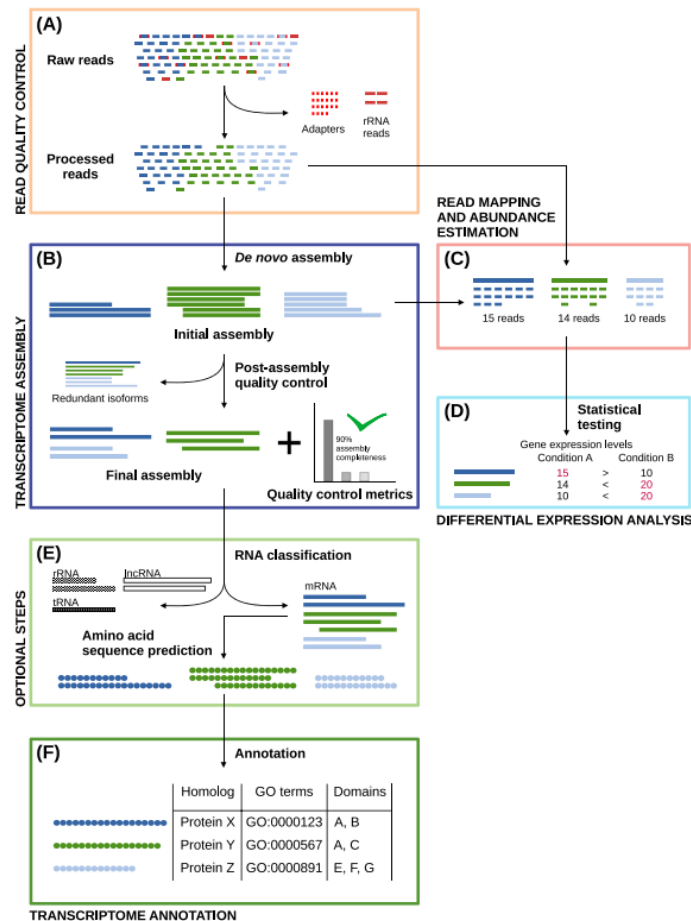


Figure 8 : Etapes de Novo assembly jusqu'à l'annotation des transcrits. (A) Contrôle de la qualité des lectures brutes en filtrant les lectures erronées. (B) Assemblage de séquences, y compris le regroupement en groupes d'isoformes et la suppression des séquences redondantes

(les isoformes sont des variants de transcription résultant d'un épissage alternatif). (C) Cartographier les lectures brutes sur les séquences assemblées pour le contrôle de la qualité de l'assemblage ou pour l'analyse de l'expression différentielle. (D) Application de tests statistiques pour l'identification des changements dans les niveaux d'expression. (E) Classification des séquences par espèces d'ARN et traduction en séquences protéiques avant annotation. (F) Annotation des séquences sur la base de la similarité des séquences, identification des caractéristiques des séquences (telles que les domaines fonctionnels) et annotation des termes d'ontologie génétique (Raghavan et al., 2022).

3. Quelques études transcriptomiques en réponse aux métaux lourds

L'étude de la transcriptomique est une étude comparative entre deux profils transcriptomiques dans deux conditions différentes. (Weber et al., 2006), ont été cultivées les plantes d'*Arabidopsis halleri* (plante tolérante) sous stress par Cd^{2+} et ont utilisées *Arabidopsis thaliana* comme plante sensible. Après le traitement, ils ont identifié trois catégories de gènes à l'aide de cette approche comparative en réponse au stress par le Cd: (1) les gènes en réponses communes, ce qui pourrait indiquer des changements stables et fonctionnellement pertinents conservés à travers les espèces végétales ; (2) des gènes dans les réponses spécifiques aux métallophytes ainsi que des transcrits régulés de manière différentielle entre les deux espèces, représentant des gènes candidats impliqués dans la tolérance au Cd^{2+} ; et (3) ceux spécifiquement sensibles au Cd^{2+} et donc indicatifs de mécanismes de toxicité ou potentiellement impliqués dans des cascades de signalisation (Weber et al., 2006). (Wang et al., 2020) ont trouvées plusieurs gènes impliqués de façon directe ou indirecte dans la réponse de *Fagopyrum tataricum* face au stress par le Pb. Plusieurs transcrits différentiellement exprimés ont été identifiés dans différentes parties de la plante. Ces transcrits impliqués dans différents processus biologiques et voies métaboliques sous le stress par le Pb. Les quatre voies les plus courantes chez *F. tataricum* répondant au stress au Pb étaient la « transduction du signal hormonal végétal », « l'interaction plante-pathogène », la « voie de signalisation MAPK - plante » et la « biosynthèse des phénylpropanoïdes » (Wang et al., 2020).

Plusieurs gènes impliqués dans la tolérance au Pb ont été trouvés chez *H. incana* tels que *ATM3* (ATP-binding cassette transporter of mitochondrial protein), *CNGC1* (cyclic nucleotide-gated channel), *GS2* (glutathione synthetase), *HMA4* (heavy metal ATPase), *MRP3* (protéine associée à la résistance multidrogue), *MT2a* (méthionine) et *PCSI* (phytochélatine synthase). Ces gènes étaient significativement plus exprimés ($\text{FC}=2$) sous stress par le Pb. Le gène *HiHMA4* a été noté et plus de huit fois l'expression de *HiMT2a* suggérant des rôles putatifs dans la tolérance au plomb et l'accumulation chez *H. incana* (Auguy et al., 2013)..

HMA4 fonctionne comme chargement du xylème métallique en bloquant la translocation conduisant à une accumulation du Pb dans les racines et à une réduction des pousses, ce qui est en concordance avec le rôle de *HMA4* dans l'efflux métallique cytosolique et comme c'est le cas chez d'autres espèces de la famille des Brassicacées (Auguy et al., 2013). Le gène *HiHMA4* est exprimé de la même manière chez les deux populations metallicoïles et non metallicoïles de *H. incana* ce qui fait que *HMA4* ne participe pas à la différenciation entre les deux populations en termes d'adaptation au Pb (Fahr, 2014).

Les gènes *MT2a* métallothionéines sont impliquées dans la réponse au stress oxydatif spécifiquement en tant que capteur de ROS dans le cytosol ainsi que dans la régulation de l'homéostasie des métaux, ces protéines ont une implication fonctionnelle dans la séquestration des métaux essentiels (Cu, Zn) et non essentiels (Cd, Pb, Cr) (Auguy et al., 2013). Le gène *HiGCN2* (General Control Non-repressible 2), appartient à la grande famille des ABC transporteurs et à la sous famille des *ABCF*. L'expression de ce gène conduit à une accumulation massive du Pb ce qui fait affecter directement le contenu en chlorophylle dans les plantes metallicoïles du *H. incana* (Fahr, 2014).

C. *Hirschfeldia incana* comme model d'étude de l'étude de la tolérance et l'accumulation du plomb

1. Description botanique de la plante « *Hirschfeldia incana* »

Hirschfeldia incana ou autrement nommé "Shortpod Mustard" ou "Mediterranean Mustard" est un *Spermatophyta* dans le sous-phylome des *Angiospermae* de la classe *Dicotyledonae* de la famille des *Brassicaceae* (CABI, 2022). C'est une moutarde herbacée, annuelle ou bisannuelle. Qui produit des tiges de 1 à 1,5 mètres de haut. Il produit des feuilles en rosette à la base de la plante et le long des parties inférieures des tiges. Les fleurs sont jaune pâle et regroupées jusqu'au bout des tiges, qui s'allongent beaucoup après la floraison. Les plantes produisent des gousses de fruits typiques ressemblant à de la moutarde (c'est-à-dire des siliques) qui mesurent environ 1 à 1,5 cm de long et sont appliquées sur les tiges. Les graines sont globuleuses à ovales-oblongues, d'environ 0,7-1,6 sur 0,7-0,9 mm. Les nombreux poils raides et blancs donnent aux plantes un aspect gris-vert (Agriculture, 2017).

2. Cycle de vie et habitat

Les graines présentent une dormance imposée par le tégument de maturation. A la suite d'une période de post-maturation qui dure environ un mois, les graines germent facilement en automne produisant une rosette de feuilles. A partir de mi-février, il y a apparition de la tige qui émerge du centre des feuilles. La floraison commence en septembre et se poursuit tant que l'humidité est suffisante (février). A l'arrivée de la saison sèche, la tige et les feuilles meurent alors que les racines persistent dans le sol en attendant l'arrivée de l'automne (Parsons et al., 2001).

3. *H. incana* : plante modèle

Hirschfeldia incana est une brassicaceae caractéristique de la région méditerranéenne. *H. incana* est connue pour sa capacité à se développer sur des sites fortement pollués en Pb (Del Río-Celestino et al., 2006; Gisbert et al., 2006; Siemens, 2011). *H. incana* présente également une capacité à accumuler le thallium dans ses fleurs et ses fruits (Madejón et al., 2007), et le zinc dans ses feuilles (Gisbert et al., 2006). En outre, *H. incana* est résistante au Cu et peut survivre dans des sols avec des concentrations extrêmement élevées en ce métal, allant de 5 000 à 16 800 mg / kg de cuivre extractible (Poschenrieder et al., 2001). Au Maroc, elle pousse naturellement sur des déblais miniers métallifères dans la région orientale avec des concentrations dans les sols varie entre 26 à 9479 mg.kg⁻¹. La concentration en Pb dans les

feuilles est en moyenne de 0,79 mg.g⁻¹ de MS, ce qui rend cette plante idéale pour les futurs projets de phytoremédiation à la pollution par le plomb (Auguy et al., 2013). En cultures hydroponiques, les plantes présentaient également une forte accumulation du Pb. La plus grande quantité de plomb accumulée dans les plantes reste au niveau des racines avec 23,7 mg/l'organe de Pb mais une importante translocation du Pb peut être observée des racines vers les feuilles et même vers les siliques avec 0,54 mg de Pb/l'organe (Auguy et al., 2013).

En outre, *H. incana* est proche génétiquement de *Brassica Nigra* et proche d'*Arabidopsis thaliana* (Auguy et al., 2013). Cette proximité génétique va permettre d'avoir accès aux outils mis en place pour cette plante modèle (génom séquencé et annoté...). Elle a une croissance rapide et un cycle court (2 mois, graine à graine). Elle peut également se cultiver aisément en chambre de culture (en culture in vitro ou en hydroponie). Elle est facile à manipuler (extractions des acides nucléiques).

Elle présente l'avantage également d'être une métallophyte facultative se développant à la fois sur les sites plombifères et sur les sites non pollués. Cette plante est donc un modèle idéal pour l'étude moléculaire et physiologique de la tolérance au Pb chez les plantes. Il est à noter qu'à nos jours, aucune plante-modèle pour l'étude de la tolérance et l'accumulation du Pb n'a été décrite.

H. incana est donc un modèle expérimental intéressant pour l'analyse transcriptomique afin d'identifier de nouveaux gènes impliqués dans les mécanismes moléculaires d'accumulation et de tolérance au Pb chez les plantes. L'étude de population peut nous donner une nouvelle vision de l'évolution de nouveaux gènes en comparant les populations non métallicoles et les populations métallicoles.

Matériels et méthode

I- Étude de la diversité intra et inter- populations d'*Hirschfeldia incana*

1. Matériel végétal

Les feuilles des individus correspondant à chaque population d'*Hirschfeldia incana* ont été récoltées sur deux sites :

- Site Métallifère de Oued el Heimer, aux abords d'une fonderie du plomb (coordonnées GPS : Alt: 959m, N 34° 24 88' W 001°54 03') pour la population métallicole.
- Site naturel non contaminé dans la région de Tetouan (35 ° 34'21 "N, 5 ° 21'17" W) pour la population non métallicole.

2. Analyse des produits de séquençage

Les séquences d'ADN correspondantes aux 4 gènes candidats (Tableau 2) ont été obtenues après séquençage de produits PCR selon la méthode de Sanger ou méthode des terminateurs de chaîne dont le principe est de rajouter, en plus des dNTPs, des ddNTPs fluorescents (avec une couleur différente par type de bases, soit 4 couleurs au total), qui arrêteront l'élongation de la copie du brin à séquencer, en obtenant alors des brins de différentes tailles.

Tableau 2 : Gène Candidats avec les amorces spécifiques d'*H.incana* correspondants

Séquence <i>H.incana</i>	Taille en pb	Amorce
Hi GCN2	466	F GCAGAAGGCCATAGAAGGAGCTGT
		R AGCTTCAACCAACTCTGCCTTCG
Hi ABC22	737	F GCTCCATGACATCGCCGAGGTT
		R TGGGTCTTCTCTGGTGGCAGTCTG
Hi CCH	545	F GCCGTGTCGGGGACTTTCCGA
		R TCAGCAGCCTCTGTTCGGAAGAAA
Hi MT2A	525	F TTGCTGTGGAGGAAACTGTGGCT
		R TCATCGATCACACACATGGCAGAG

Les produits de séquençage, une fois purifiés, ont été analysés par électrophorèse en gel capillaire à l'aide du séquenceur ABI PRISM® 3130XL Genetic Analyser (Applied Biosystems) à la plateforme IRD de Montpellier.

L'identification des bases s'est faite à partir des chromatogrammes bruts en utilisant le logiciel Geneious (Geneious Pro 4.5.5). L'ensemble des produits de PCR est séquencé dans les deux sens.

La qualité de chacune des séquences est vérifiée visuellement et les erreurs (pics anormaux sur le chromatogramme, différences entre le chromatogramme) sont corrigées en comparant les brins positif et négatif, permettant d'obtenir une séquence unique, de qualité, par individu.

3. Analyse bio-informatique

Les séquences sont débarrassées des parties vectorielles en utilisant le logiciel Vecscreen. (<http://www.ncbi.nlm.nih.gov/VecScreen/>).

Les séquences de base obtenues étaient organisées dans deux dossiers séparant les séquences des deux populations métallicoles (PM) et non métallicoles (PNM). Un contigs a été créé pour chaque deux séquences qui ont le même noms mais séquencer avec deux amorces (« forward », « revers ») via le logiciel ([DNA sequence alignment/DNA contig assembly software|Sequence assembly, analysis & mutation detection software \(dnabaser.com\)](#)), puis après obtention de tous les contigs, l'identification se fait avec l'outil blastn ([BLAST: Basic Local Alignment Search Tool \(nih.gov\)](#)) du Centre national d'information sur la biotechnologie (NCBI). Les séquences ont été classées en quatre gènes (« *ABCG22* », « *GCN2* », « *CCH* », « *Mt2a* »). Par la suite, un fichier Fasta a été créé pour chaque gène qui contient les séquences nucléotidiques des deux populations PM et PNM.

4. Test de la diversité génétique des populations d'*H. incana*

Pour le test de Hardy-Weinberg, le test F_s de F_u , le test D de Tajima, l'analyse de la variance moléculaire (AMOVA) et la différence par paires F_{ST} (index de fixation), le logiciel Arlequin ([Arlequin 3.5 - Downloads \(unibe.ch\)](#)) ont été utilisés pour notre analyse. Les séquences Fasta de chaque gène ont été converties en type de fichier ARP (ARlequin Project) pour que notre logiciel puisse les traiter en spécifiant dans le fichier les paramètres suivants : NbSamples = 2, DataType = DNA, GenotypicData = 1, LocusSeparator = NONE, GameticPhase = 1. Les paramètres de tous les tests ont été utilisés par défaut. Le type de test de Hardy-Weinberg a été modifié de « locus par locus » vers « haplotype entier ».

5. Les tests statistiques

L'analyse de variance multivariée (MANOVA) est réalisée avec le langage R en utilisant le paquet 'adeqenet' ([adeqenet on the web \(r-project.org\)](#)) pour l'Analyse Discriminante des

Composantes Principales (DAPC), le graphique de la composition génotypique et la détermination des SNPs critiques pour l'association au groupe.

L'arbre phylogénique est créé avec le langage R en utilisant les paquets : 'ape' ([ape – home page \(ird.fr\)](http://ape-home-page.ird.fr)) pour le calcul de la distance entre les séquences d'ADN utilisant le model K80. 'ggtree' ([Bioconductor - ggtree](http://Bioconductor-ggtree)) est utilisé pour la représentation de l'arbre phylogénique avec l'alignement des séquences. 'pegas' ([CRAN - Package pegas \(r-project.org\)](http://CRAN-Package-pegas-r-project.org)) est utilisé pour calculer la distance entre différents haplotypes et le paquets 'stats' ([CRAN - Package STAT \(r-project.org\)](http://CRAN-Package-STAT-r-project.org)) pour générer la carte du chaleur entre les haplotypes.

II- Étude transcriptomique

1. Matériel végétal

Les graines d'*Hirschfeldia incana* ont été collectées dans la zone minière du Pb d'Oued El Heimer, située au sud de la ville d'Oujda dans l'est du Maroc ($34^{\circ}26'88''\text{N}$ – $1^{\circ}54'03''\text{W}$) (Figure 9) (el Hasnaoui et al., 2020). Le sol de cette région est fortement contaminé par le Pb avec des concentrations allant jusqu'à 18,626 mg/kg DW (Abdelaziz et al., 2010; Auguy et al., 2013; el Hasnaoui et al., 2020).



Figure 9 : Plant d'*Hirschfeldia incana* poussant sur les scories d'Oued EL Heimer

2. Culture des plantes

Après deux semaines de germination des graines dans des boîtes de pétri, les plantules ont été transférées dans un milieu hydroponique (phase d'acclimatation) contenant un milieu Broughton et Dilworth (BD), pendant 15 jours. Après l'acclimatation, les plantes d'*H. incana* ont été transférées dans du milieu BD (sans phosphate) et additionnées de 100 μM de $\text{Pb}(\text{NO}_3)_2$. Cette concentration a été utilisée sur la base des réponses physiologiques et moléculaires de *H. incana* sous stress par le Pb (Auguy et al., 2013; Mouna et al., 2015). Des plantes témoins ont été cultivées dans du milieu hydroponique contenant un milieu BD sans Pb. Les plantes (3 répétitions par traitement) ont été récoltées après quinze jours de traitement.

3. Préparation des cDNA et séquençage (RNAseq)

Par la suite l'ARNm a été extrait à partir des parties racinaires et aériennes. Les cDNA ont été construits et le séquençage (RNAseq) a été réalisé par la plateforme génomique centre INRAE de Toulouse Occitanie – France. Toute cette partie a été faite et réalisé par Said EL HASNAOUI dans le cadre de son projet de thèse. Les séquences (Raw reads) obtenues après le séquençage (Illumina HiSeq) font l'objet d'une série d'analyses bio-informatiques.

4. Assemblage a novo (*De Novo* assembly)

Après le séquençage, le contrôle qualité des reads a été effectué à l'aide d'outil FastQC v0.11.9. Les séquences ont été utilisées pour l'assemblage du transcriptome à l'aide de l'outil Trinity avec les paramètres par défaut : K-mers = 25 et la longueur minimale égal 200 pb. Afin de regrouper les contigs provenant du même gène ou de la même protéine, un assemblage secondaire a été réalisé à l'aide de l'outil CD-HIT-EST v4.6.4 (<http://weizhong-lab.ucsd.edu/cdhit-web-server>) avec un pourcentage d'identité de 95% (Fu et al. 2012). La qualité de l'assemblage a été évaluée à l'aide de l'outil BUSCO version 5.2.2 (The Benchmarking Universal Single-Copy Orthologs) (Waterhouse et al., 2018). Cet outil d'évaluation de la qualité fournit des quantifications à haute résolution pour les génomes, les ensembles de gènes et les transcriptomes et vérifie si chacun des groupes BUSCO est complet, dupliqué, fragmenté ou manquant dans l'assemblage du génome ou du transcriptome (Waterhouse et al., 2018). De plus, pour évaluer la cohérence de l'assemblage et le taux d'alignement, les séquences générées lors de séquençage d'ARNm (Raw reads) ont été réaligner avec les transcriptomes générés lors du deuxième assemblage à l'aide de l'outil d'alignement Bowtie2.

5. Annotation et profil fonctionnel

L'annotation des transcrits assemblés a été effectuée par BLASTX (E-Value $1e^{-10}$) sur un serveur local à l'aide de séquences protéiques *d'Arabidopsis thaliana* (http://ftp.ensemblgenomes.org/pub/plants/release-52/fasta/arabidopsis_thaliana/pep/Arabidopsis_thaliana.TAIR10.pep.all.fa.gz). Les résultats BLASTX obtenue sous forme d'un fichier texte qui représente chaque transcrit avec leur ID TAIR. Ces résultats ont été combinés avec le fichier d'annotations Gene Ontology *d'Arabidopsis thaliana* à l'aide du logiciel Excel 2021 pour récupérer les termes GO de chaque transcrit. Avec ces termes GO, tous les transcrits sont identifiés en trois catégories telles que le processus biologique (BP), la fonction moléculaire (MF) et le composant cellulaire (CC). Ensuite, l'outil ShinyGO v0.75 (Ge et al., 2020) et REVIGO (Supek et al., 2011) ont été utilisés pour classer fonctionnellement les termes GO et représenter graphiquement les fonctions de transcrit.

6. Estimation de l'abondance des transcrits et analyse de l'expression des gènes

Pour quantifier l'abondance des transcrits, le serveur Galaxy (<https://usegalaxy.org>) a été utilisé. Dans cette analyse, RNA-Seq by Expectation-Maximization (RSEM) (Haas et al., 2013) a été utilisé comme méthode d'estimation de l'abondance et bowtie2 (<https://usegalaxy.org>) a été choisi pour l'alignement des séquences brutes sur le transcriptome assemblé.

7. Analyse des gènes différentiellement exprimés (DEGs)

L'analyse de l'expression différentielle a été déduite sur la base des comptages normalisés avec le logiciel statistique R à l'aide du package DESeq2, disponible en ligne : <https://bioconductor.org/packages/release/bioc/html/DESeq2.html> et les résultats ont été extraits à l'aide de méthode restrictive "Bonferroni" avec une valeur de p-adj $<0,05$. Ensuite, le package VennDiagram a été utilisé pour représenter le nombre des gènes différentiellement exprimés partagés entre les deux conditions (Contrôle x Pb) dans les racines et les parties aériennes (Chen and Boutros, 2011). Un schéma qui résume l'approches utiliser pour cette analyse est dans la figure 10.

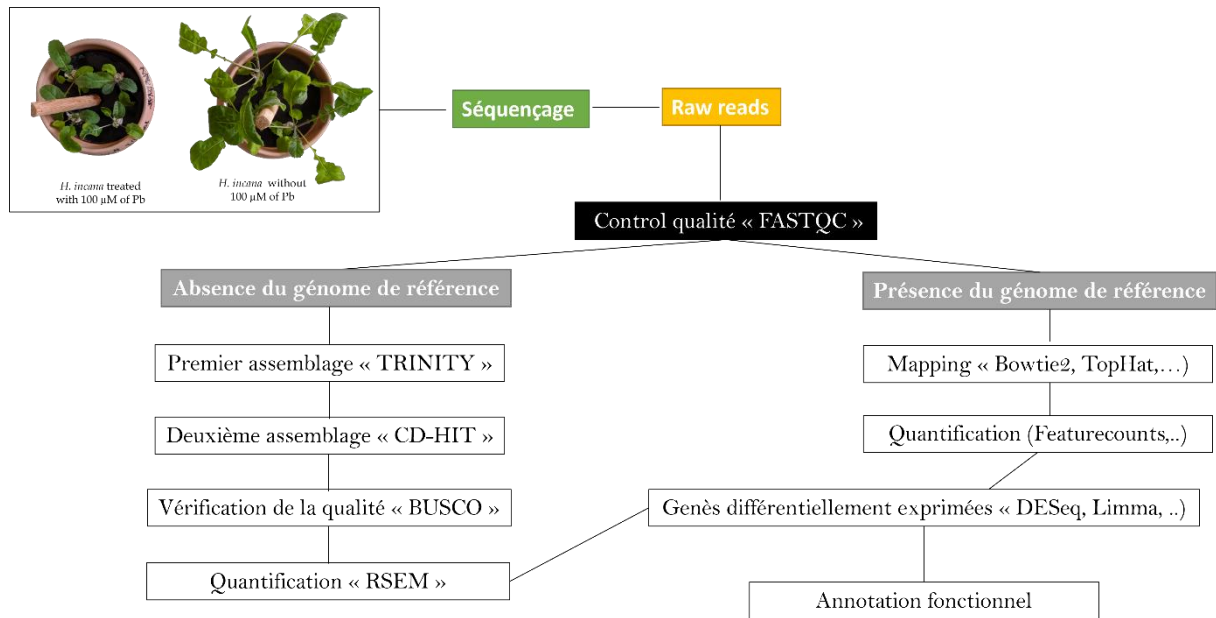


Figure 10 : Les différentes étapes d'analyses des données RNA-seq

Résultats et discussion

I- L'analyse de la diversité intra et inter population d'*Hischfeldia incana*

Les modèles de l'histoire évolutive sont enregistrés dans les séquences d'ADN, et l'application de la génétique des populations révèle de nombreux secrets sur le passé évolutif. À mesure que les méthodes de « génomique fonctionnelle » se développent pour identifier les rôles de nouveaux gènes, une attention inévitablement croissante est accordée à l'importance de la variation génétique dans les populations.

Les plantes (hyper)accumulatrices de métaux sont certainement un outil intéressant pour étudier le phénomène de sélection et d'adaptation aux conditions environnementales. *H. incana* est une pseudométallophyte qui se développe naturellement dans deux habitats contrastés en teneurs du Pb. Les populations PM et PNM sont bien adaptées à leur environnement et ont montré une capacité à accumuler des concentrations élevées du Pb en culture contrôlée (Fahr et al., 2015). Toutefois, il existe une variation substantielle naturelle entre les deux populations, conduisant à des différences d'accumulation et de tolérance du Pb (Fahr et al., 2015). Plusieurs gènes (*ABCG22*, *CCH*, *GCN2* et *Mt2a*), susceptibles d'être impliqués dans la tolérance et l'accumulation du Pb, ont été sélectionnés pour une analyse de la diversité intra et inter-population chez *H. incana*.

1. Choix des gènes candidats

Les gènes candidats étaient choisis en fonction de leur surexpression en présence du Pb suite aux résultats de l'hybridation hétérologue des transcrits d'*H. incana* et un microarray (Agilent...) (Auguy et al., 2016) et de leur expression spécifique chez *H. incana* et de la possibilité de dessiner des amorces spécifiques pour l'étude de la diversité génétique des deux populations (Tableau 3).

Tableau 3 : Liste des gènes candidats

AGI	FC	localisation	Description TAIR
AT5G09930	4,69/ 2,95	PA + R	GCN2 (ATP binding cassette F2)
At3g09390	4,58	PA	MT2A (METALLOTHIONEIN 2A)
At5g06530	2,65	R	ABCG22 (ABC transporter family protein G 22)
At4g08570	13,92	R	CCH (heavy-metal-associated domain-containing protein / copper chaperone)

Avec : PA: Partie aérienne ; R: Racine

Ces gènes ont des fonctions importantes dans la réponse au stress métallique. Les transporteurs de cassettes de liaison à l'ATP (*ABCG22*) catalysent le transport d'une variété de substrats à travers les membranes biologiques, y compris l'acide abscissique qui joue un rôle vital en réponse aux stress abiotiques tels que la sécheresse, la salinité, la lumière élevée, la carence en nutriments et les métaux lourds (Hu et al., 2020). Lors d'une évolution expérimentale de *Noccaea caerulea* dans des conditions polluées par le zinc, *ABCG22* est l'un des gènes ayant évolué à partir du variant non métallicole (PROST, 2018).

La *CCH* a été la première métallochaperone décrite chez les plantes. Il contient les résidus typiques de lysine et le motif de liaison au cuivre (Andrés-Colás et al., 2006) qui fait des chaperons de cuivre une classe omniprésente de protéines qui jouent un rôle important à la fois dans l'apport de cuivre et dans la protection cellulaire contre l'exposition au cuivre dans des conditions métaboliques normales (Harrison et al., 1999). Pour *Arabidopsis thaliana*, *CCH* était l'un des gènes qui sont régulés (surexprimé) en réponse à 1 mM de Cd (Semane et al., 2010).

Le gène *GCN2*, General Control Non-repressible 2, appartient à la grande famille des ABC transporteurs et à la sous famille des ABCF. Dans une étude chez *Arabidopsis thaliana* sous stress de Cd (Plaza and Bovet, 2008), le gène *AtGCN5* a montré une expression préférentielle chez les feuilles que les racines. Et grâce à son homologie avec le gène *GCN2* on peut suggérer que sa localisation est dans le chloroplaste.

H. incana montre une surexpression du gène *Mt2a* dans les feuilles en réponse au Pb (Auguy et al., 2016). *Mt2a* est connu pour être localisé dans le cytosol et pourrait chélater les métaux lourds mais il protège également l'apex de la racine en réponse au Cu chez *A. thaliana* (Guo et al., 2003). *Mt2a* n'est pas directement impliqué dans la tolérance aux métaux lourds chez *A. thaliana* et sa surexpression n'est pas corrélée à une amélioration de la tolérance aux métaux lourds, ce qui implique que dans le cas de *MT2a*, un transport intercellulaire de facteur(s) transcriptionnel(s) pourrait éventuellement être impliqué dans la livraison de métaux lourds (Mineta et al., 2000).

2. Équilibre de Hardy-Weinberg

Pour étudier l'organisation de la variation génétique en génotypes intra-population d'*H. incana*, le test de l'équilibre de Hardy-Weinberg a été réalisé. Le modèle de Hardy-Weinberg est basé sur ces hypothèses (Hartl et al., 1997):

- L'organisme est diploïde.
- La reproduction est sexuée.
- Les générations ne se chevauchent pas.
- Le gène considéré à deux allèles.
- Les fréquences alléliques sont identiques chez les mâles et les femelles.
- L'accouplement est aléatoire.
- La taille de la population est très grande (en théorie, infinie).
- La migration est négligeable.
- La mutation peut être ignorée.
- La sélection naturelle n'affecte pas les allèles considérés.

L'équilibre de Hardy-Weinberg n'est atteint que lorsque ces hypothèses sont vérifiées. Les individus pour chaque gène avec le même allèle en était regroupé en haplotype pour réaliser le test de Hardy-Weinberg et les tests de neutralité (annexe 1).

En raison du faible nombre d'individus, les haplotypes en commun ont été utilisés car ils peuvent donner plus d'informations et un aperçu sur l'action de la recombinaison, de la sélection, des changements dans la taille de la population et de la généalogie sous-jacente ainsi que l'histoire démographique (Schraiber and Akey, 2015; Song et al., 2017).

Les résultats montrent que toutes les valeurs de p sont inférieures à 0.05 (tableau 4). L'hypothèse nulle est donc rejetée. Donc l'équilibre de Hardy-Weinberg n'est pas atteint. Ce qui montre qu'il y a un effet de mutation ou de sélection naturelle ou une combinaison de ces forces évolutives exercées sur les 4 gènes des deux populations métallicoles et non métallicoles.

Tableau 4 : Les valeurs P d'équilibre de Hardy Weinberg.

Gène	<i>ABCG22</i>		<i>CCH</i>		<i>GCN2</i>		<i>Mt2a</i>	
Population	PM	PNM	PM	PNM	PM	PNM	PM	PNM
Valeur de	0.00300	0.00000	0.00000	0.00000	0.00000	0.00000	0.00005	0.00066
P	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	0.00102	0.00000	0.00000	0.00000	0.00000	0.00000	0.00005	0.00050

3. Tests de neutralité

Les tests de neutralité sont utilisés pour tester l'équilibre intra- population et pour vérifier s'il y a une sélection sur un allèle parmi un grand nombre d'allèles au sein de la population. Le test de Tajima est l'un des tests de neutralité qui est généralement le plus puissant pour vérifier les hypothèses alternatives de balayage sélectif, de goulots d'étranglement de la population ou de subdivision de la population (Hartl et al., 1997).

Les valeurs de D peut varier en trois résultats (Dogan and Dogan, 2016) :

$D < 0$: Allèles rares présents à basses fréquences / Balayage sélectif récent, expansion de la population après un goulot d'étranglement récent, liaison à un gène balayé,

$D = 0$: Variation observée similaire à la variation attendue / Population évoluant selon l'équilibre mutation-dérive. Aucune preuve de sélection,

$D > 0$: Présence d'allèles multiples, certains à basse fréquence, d'autres à haute fréquence / Sélection équilibrante, contraction brutale de la population,

Chez les deux populations, les valeurs D et Fs n'étaient pas significatifs pour les quatre gènes étudiés. Les valeurs varient entre négative et positive et sont ainsi considérées comme zéro. Le D de Tajima indique que la variation observée est similaire à la variation attendue et la population évoluant selon l'équilibre mutation-dérive (tableau 5). Ce qui implique qu'il n'y a aucune preuve de sélection. Le nombre d'allèles perdus par dérive génétique est égal au nombre d'allèles nouveaux produits par mutation. Nos résultats sont similaires à ceux de *Noccaea* car il n'y avait aucune preuve de sélection sous stress de nickel (Maestri et al., 2013).

Le test de F_s du F_u suggère d'estimer la probabilité d'observer un échantillon aléatoire avec un nombre d'allèles égale ou inférieur à la valeur observée compte tenu du niveau de diversité observé et de l'hypothèse que tous les allèles sont sélectivement neutres (Holsinger, 2012).

Les valeurs F_s du F_u pour les 4 gènes étudiés varient de -1,82 à 11,29 (Tableau 5). Une valeur négative de F_s est la preuve d'un nombre excessif d'allèles, ce qui signifie qu'il y a une expansion démographique récente ou une dérive génétique (hitchhiking selection). Ceci sous-entend que lorsqu'un allèle change de fréquence non pas parce qu'il est lui-même sous sélection naturelle, mais parce qu'il est proche d'un autre gène qui subit un balayage sélectif et qui se trouve sur la même chaîne d'ADN. Une valeur positive de F_s est la preuve d'un déficit d'allèles, comme on pourrait s'y attendre d'un récent goulot d'étranglement de la population qui est une épisode de réduction sévère de la population, suivi d'une nouvelle expansion démographique (Holsinger, 2012). Dans notre cas, la p-value de F_s n'est pas significative (supérieur à 0.02). Ce que nous fait réfuter les deux hypothèses. Ainsi, les deux populations ne sont pas affectées par le flux génétique ou une diminution démographique.

Il y'a donc une variation significative entre les différents haplotypes dans les deux populations des quatre gènes étudiés et différentes histoires démographiques parmi les régions analysées ce qui fait qu'ils évoluent de manière neutre.

Tableau 5 : Les résultats de test de D de Tajima et F_s de F_u .

Gène	Population	Nombre d'individu	Nombre de sites avec substitution	Nombre d'allèles	D	P de valeur D	F_s	P de valeur F_s
<i>ABCG22</i>	PM	26	35	16	0.19241	0.64100	1.39177	0.75800
	PNM	17	26	10	0.71350	0.80900	3.56097	0.91800
<i>CCH</i>	PM	24	16	10	-0.68272	0.28900	-0.64432	0.41400
	PNM	28	16	12	-0.65294	0.28300	-1.82763	0.23600
<i>GCN2</i>	PM	19	34	13	1.04419	0.86800	2.51924	0.84600
	PNM	24	26	10	1.14537	0.91700	6.82634	0.96300
<i>Mt2a</i>	PM	31	32	15	1.43037	0.95100	11.29175	0.99200
	PNM	23	38	18	0.59848	0.78800	9.08938	0.98600

4. Analyse de la structure génétique

L'analyse de la variance moléculaire (AMOVA) est utilisée pour calculer le niveau de différenciation génétique entre différentes populations. Elle estime les indices de structure génétique en utilisant des informations sur le contenu allélique des haplotypes, ainsi que leurs fréquences (Excoffier and Lischer, 2010). Des similitudes génétiques élevées ont été trouvées/signalées entre les deux populations avec un taux de variation inférieur à 10 % (tableau 6). A l'inverse, la variation génétique intra-population est supérieure à 90 % (tableau 6). Cela implique qu'il n'y a pas de différenciation significative inter-population. Ainsi, la PM et la PNM ont probablement évolué à partir de différents ancêtres locaux.

L'indice de fixation (F_{ST}) est une mesure de la différenciation des populations due à la structure génétique. Il est utile pour examiner le niveau global de divergence génétique entre les sous-populations. F_{ST} a un minimum théorique de 0 (indiquant l'absence de divergence génétique) et un maximum théorique de 1 (indiquant la fixation d'allèles alternatifs dans différentes sous-populations) (Hartl et al., 1997) et par conséquent les deux populations examinées sont éloignées génétiquement (Frankham et al., 2002). Nos résultats montrent que pour les 4 gènes, les valeurs de F_{ST} sont comprises entre 0,05 et 0,15 (tableau 6). Ce qui signifie qu'il s'agit d'une différenciation génétique modérée indiquant un flux de gènes restreint. Des résultats similaires ont été trouvés chez *Thlaspi caerulescens* avec une valeur F_{ST} comprise entre 0,05 et 0,15 mais dans leurs cas la différence n'est pas distincte pour les allozymes qui sont supposés représenter des allèles neutres (Dubois et al., 2003). Cela peut indiquer que la différenciation éco-typique est relativement récente et/ou que les flux de gènes sont toujours efficaces et que l'isolement reproductif entre les populations PM et PNM n'est pas atteint.

L'indice de tri généalogique (gsi) a été utilisé pour quantifier le degré de regroupement des taxons partageant un attribut commun qui est le type de la population étant donné une phylogénie relative à ces taxons. Les valeurs trouvées ne se rapprochent pas ce qui fait que les taxons associés à un des deux populations ne forment pas un groupe monophylétique. Alors plus le degré de dispersion des taxons associés à l'attribut est élevé dans l'arbre (compte tenu du nombre de taxons et de la taille de l'arbre), plus le gsi sera petit pour cet attribut (Cummings et al., 2008). Donc pour le gène *ABCG22*, les deux populations métallicoles et non métallicoles ont montré un taux de divergence généalogique qui est modéré et tous les haplotypes d'une population de gène *ABCG22* n'ont pas un ancêtre exclusif. Le même cas a été observé pour la population non métallicoles pour les gènes *GCN2* et *Mt2a* (Tableau 6). Pour le gène *GCN2*

(population métallicole) et *Mt2a* (population non métallicoles), les valeurs *gsi* sont proches de 0 ce qui implique que la relation phylogénique entre les haplotypes est très divergente et le degré de coalescence est faible (Zimmers et al., 2017).

Tableau 6 : La structure génétique des populations et l'indice de fixation et l'indice de tri généalogique.

Gènes	Entre les populations	Au sein d'une population	FST	P de FST	Valeur <i>gsi</i> des PM	Valeur <i>gsi</i> des PNM
<i>ABCG22</i>	9.72%	90.28%	0.09717	0	0.3496	0.4267
<i>CCH</i>	6.73%	93.27%	0.06732	0.00098+-0.00098		
<i>GCN2</i>	10.08%	89.92%	0.10084	0	0.034	0.4545
<i>Mt2a</i>	6.06%	93.94%	0.06058	0	0.4375	0.0594

5. Analyse multivariée de variance des 4 gènes

L'analyse de variance multivariée (MANOVA) peut avoir une ou plusieurs variables de regroupement par rapport à l'ANOVA. L'intérêt d'une MANOVA avec une variable de regroupement est la comparaison de groupe. Les groupes sont comparés par rapport aux moyennes sur un ou plusieurs composites linéaires des variables de résultat (Keselman et al., 1998). Nous avons utilisé l'analyse discriminante des composants principaux (DAPC) pour maximiser la séparation entre les groupes tout en minimisant la variation au sein du groupe (Jombart et al., 2010). Les quatre gènes ont montré une nette différenciation entre les deux populations (figure 11). Le diagramme de composition génotypique utilisé pour représenter la probabilité d'affectation de groupe d'individus à plusieurs groupes. Donc ce test représente la probabilité de prendre n'importe quelle séquence et de représenter le bon groupe auquel elle appartient. Les gènes *ABCG22* et *GCN2* ont montré une réelle distinction de la majorité des séquences nucléotidiques (haplotype) entre les deux populations (figure 12). Cependant, quelques haplotypes sont partagés entre les deux populations. Pour les gènes *CCH* et *Mt2a*, il y avait un grand échec d'attribuer correctement les haplotypes à leurs populations d'origine (figure 12).

Les SNP du gène *ABCG22* en position 214 (entre T/ C) et 417 (entre A/G) reflètent le plus la distinction entre les deux populations (figure 13). Pour le gène *CCH*, les SNP en position 377 (entre T/C) et 508 (entre C/G) reflète le plus la distinction entre les deux populations (figure 13). Pour le *GCN2*, trois positions de SNP étaient les principaux distinctifs du type de

population à la position 163 (entre T/C), la position 250 (entre T/C) et la position 411 (entre A/G) (figure 13). Pour le gène *Mt2a*, les SNP à 244 (entre A/G) et 475 (entre A/G) sont les points réfléchissants du type de population (figure 13).

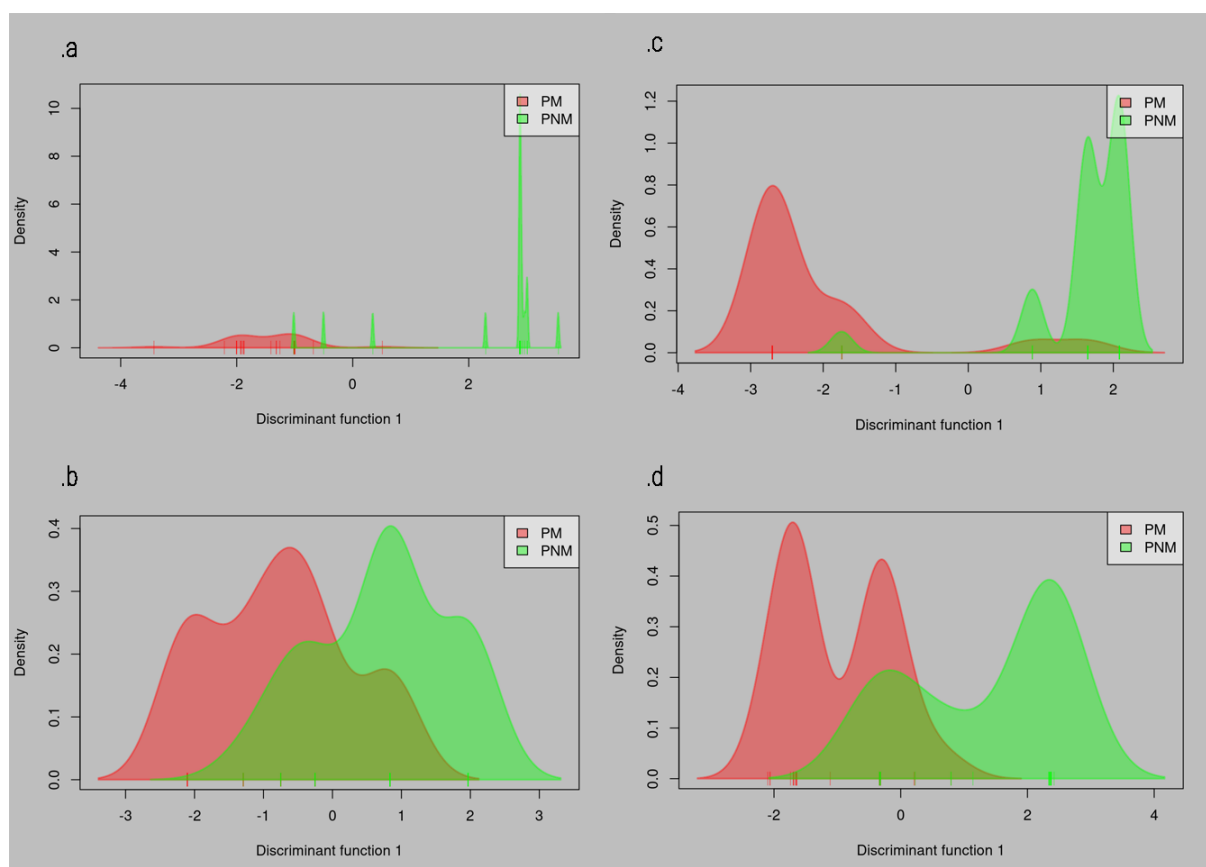


Figure 11 : Analyse discriminante des composantes principales des 4 gènes (a : gène *ABCG22*, .b : gène *CCH*, .c : gène *GCN2*, .d : gène *Mt2a*)

En rouge : la population metalicoles ;

En vert : la population non metalicoles

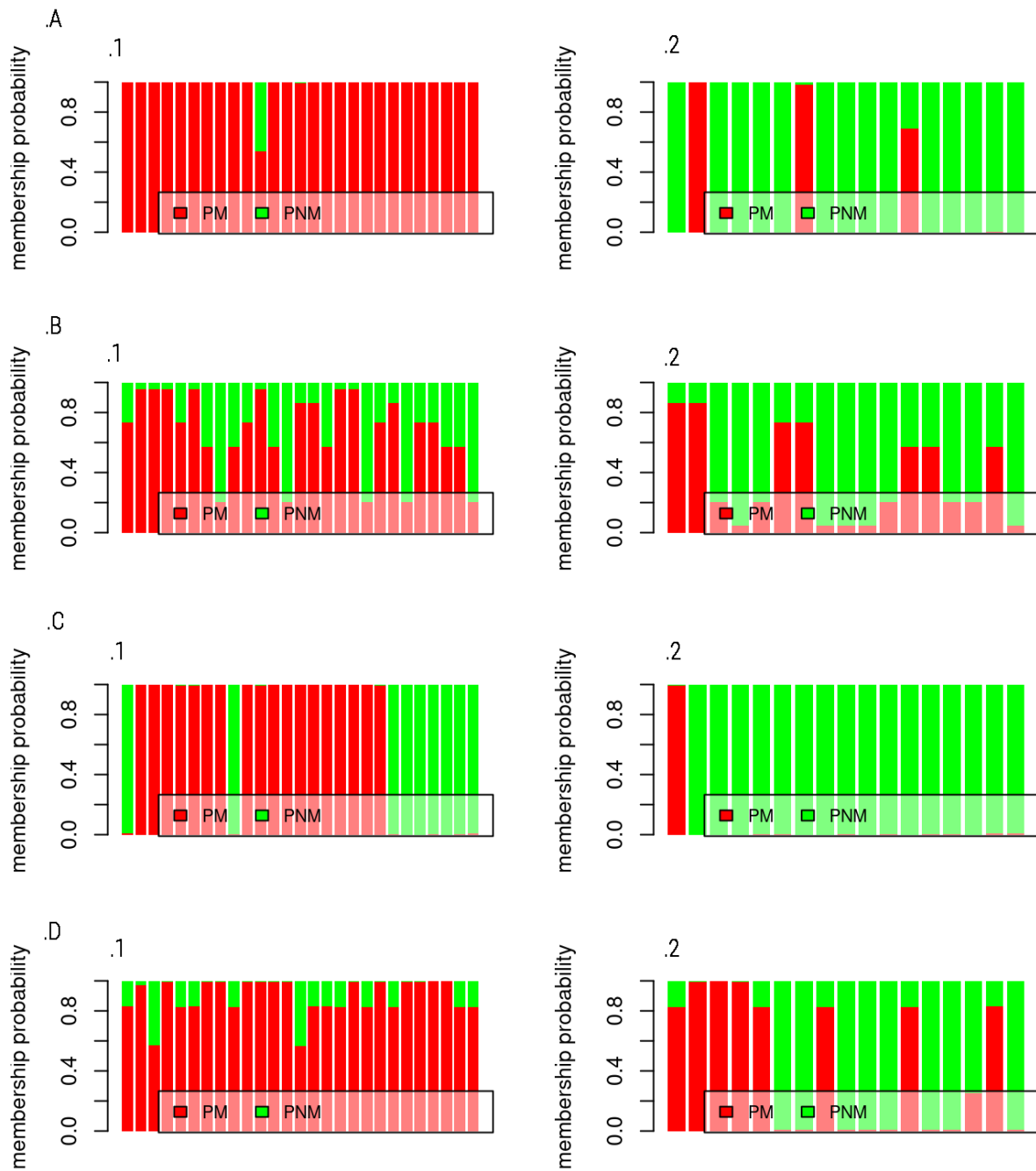


Figure 12 : Diagramme de la composition génotypique de chaque population pour les quatre gènes (A : gène *ACG22*, B : gène *CCH*, .C : gène *GCN2*, D : gène *Mt2a*) (1. : Population métallicoles, .2 : population non métallicoles).

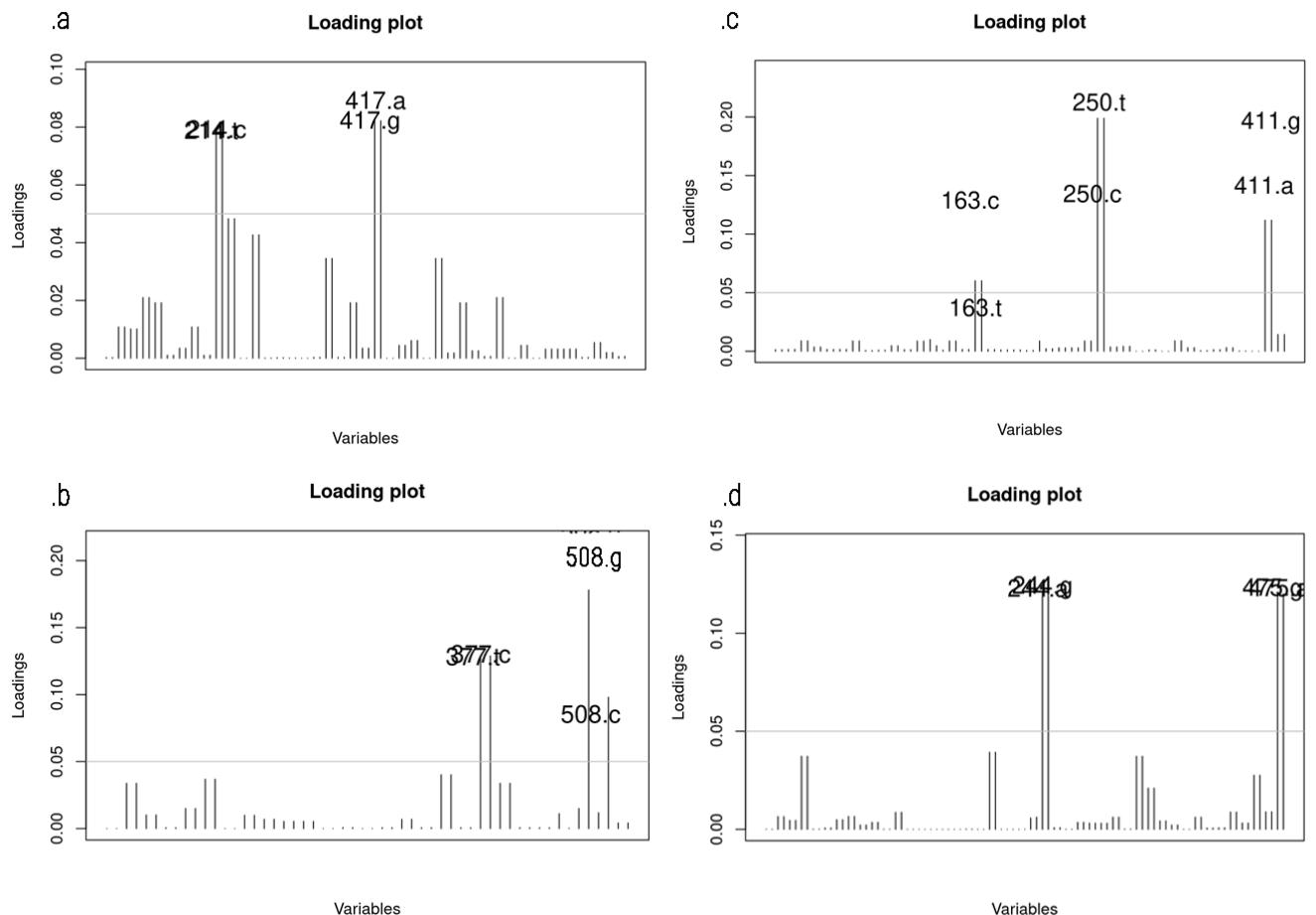


Figure 13 : la position des SNP indicative de la différenciation entre les deux populations (a : gène *ABCG22*, b : gène *CCH*, c : gène *GCN2*, d : gène *Mt2a*)

6. Analyse phylogénétique et généalogique des gènes

Pour l'analyse phylogénétique, le neighbor joining nous donne une idée sur la relation phylogénétique entre les séquences des deux populations. Pour le gène *ABCG22*, il y'a formation des groupes distincts comportant un individu des deux populations (figure 14). Ce qui est renforcé par la Heat map du gène *ABCG22* qui nous montre qu'il y'a une distinction entre quatre groupes des haplotypes (annexe 2).

Le graphique de généalogie nous donne une idée sur la relation ancestrale entre les haplotypes. Les tailles des nœuds sont dessinées proportionnellement aux fréquences des haplotypes et la distance entre les nœuds adjacents n'est pas représentée par le nombre minimum de mutations mais la distance de Hamming est calculée comme le nombre de différences de base entre deux séquences qui peuvent être exprimées en nombre ou en proportion (Salzburger et al., 2011). Cette méthode utilisée pour la construction d'arbre phylogénétique en utilisant le Fitch algorithme a été la première méthode de parcimonie pour déduire les états ancestraux qui nécessite le moins

de changements évolutifs (Yang et al., 2011). Pour le gène *ABCG22*, dans le graphique de généalogie des haplotypes, les nœud les plus internes vides sont entourés par les haplotypes qui contiennent les individus de la population métallicoles tandis que ceux dans l'extrémité sont les haplotypes qui contiennent les individus de la population non métallicoles (Figure 15 et annexe 6).

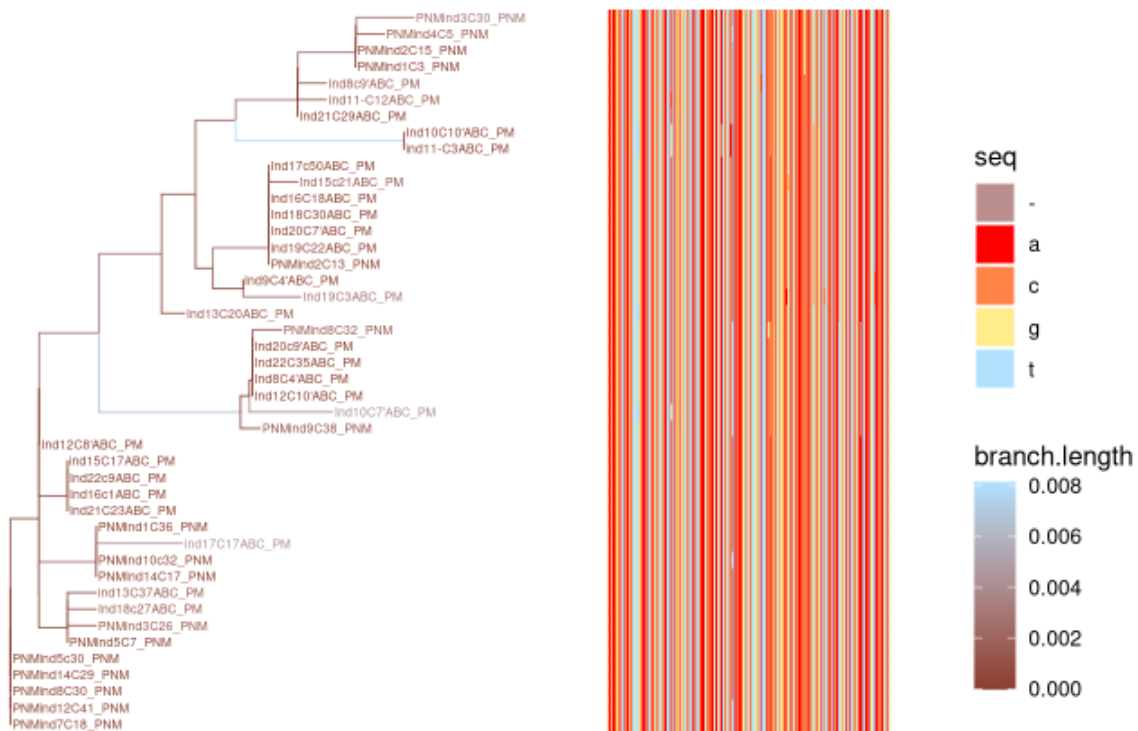


Figure 14 : Arbre phylogénétique d'alignement de séquences multiples du gène *ABCG22* avec Neighbor Joining algorithm.

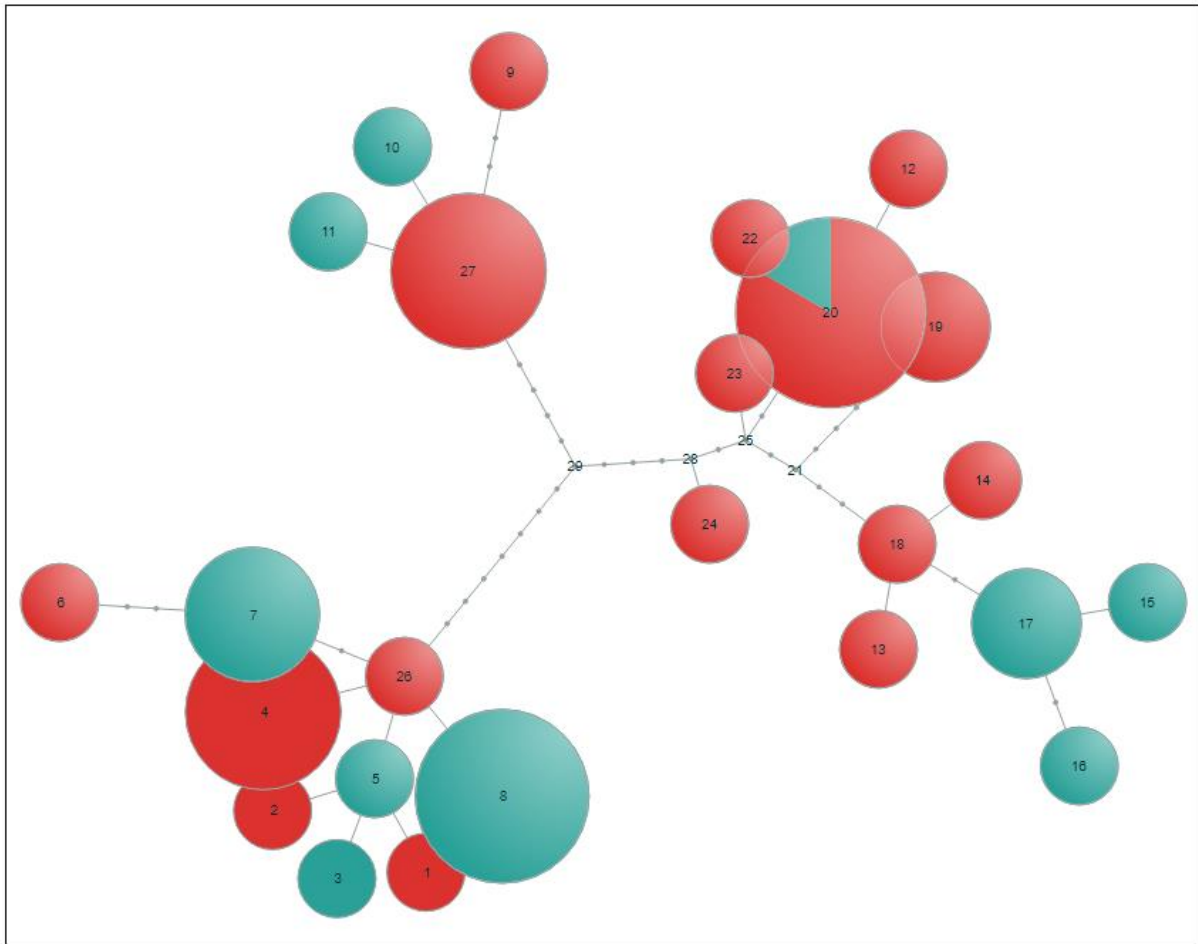


Figure 15 : Graphique de généalogie des haplotypes du gène *ABCG22*.

En rouge : les individus de la population metallicoles (PM) ; En bleu : la population non metallicoles (PNM).

Pour la gène *CCH*, tous les individus sont plutôt similaires vu la relation phylogénétique avec une distinction d'un sous-groupe (figure 16) qui est renforcé par la carte de fréquentation (Heat map) (annexe 3).

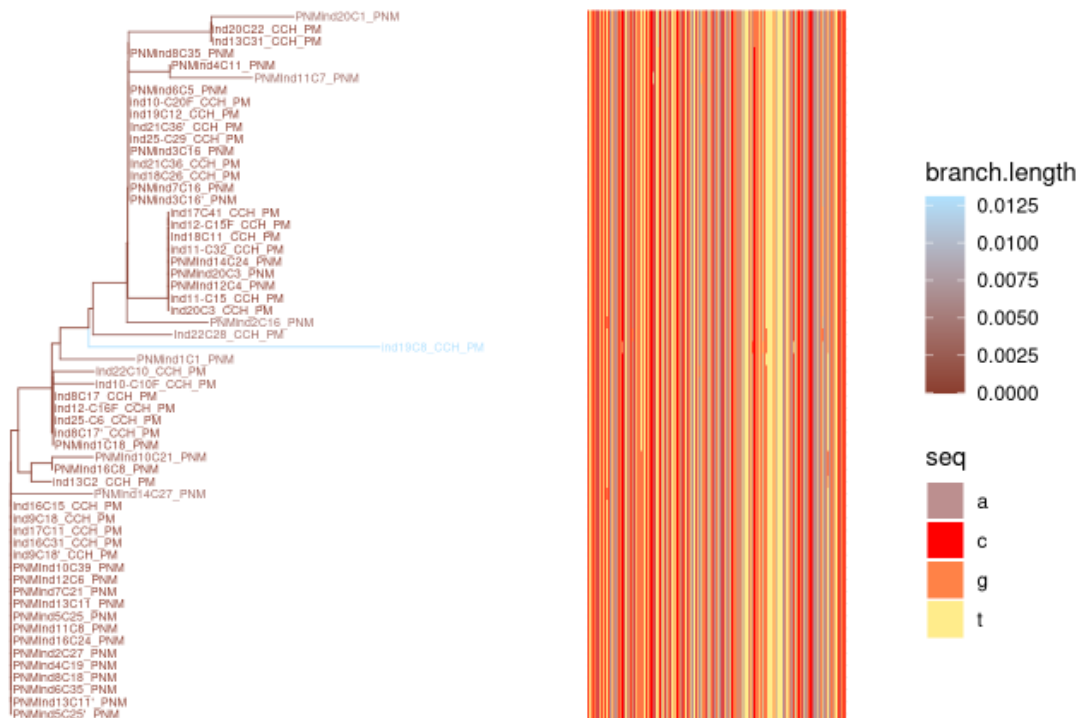


Figure 16 : Arbre phylogénétique d'alignement de séquences multiples du gène *CCH* avec Neighbor Joining algorithm.

Pour le gène *GCN2*, l'arbre phylogénétique et la carte de fréquentation (heat map) nous montrent qu'il y a formation de trois groupes distincts (Figure 17 et annexe 4). En ce qui concerne la relation ancestrale, le gène *GCN2* montre un résultat pareil au gène *ABCG22*. Les nœuds central sont de la population métallicoles et la population non métallicoles est divergé depuis les individus de la PM mais encore les nœuds les plus extrême appartient aux individus de la population PM (Figure 18 et annexe 6).

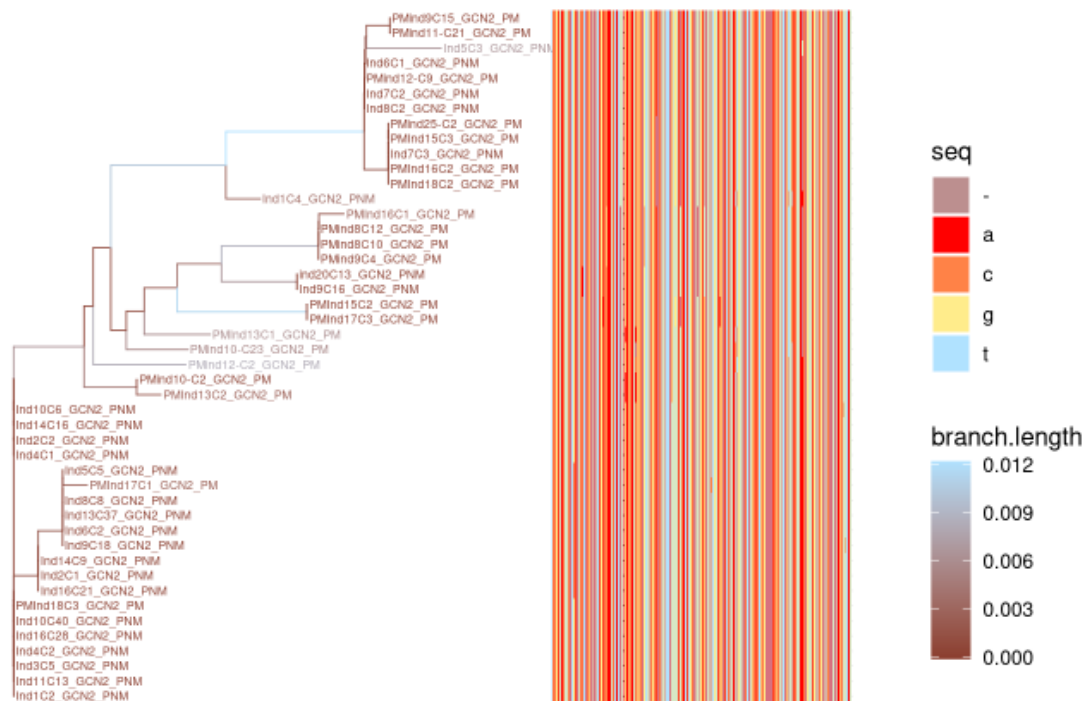


Figure 17 : Arbre phylogénétique d'alignement de séquences multiples du gène *GCN2* avec Neighbor Joining algorithme.

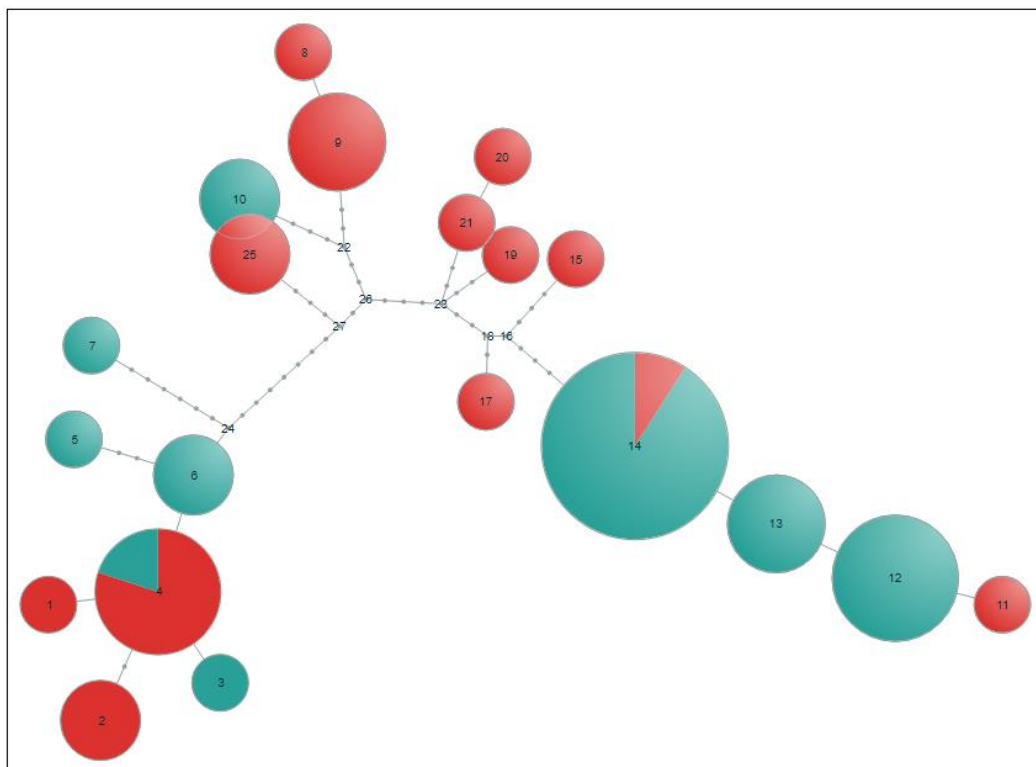


Figure 18 : Graphique de généalogie des haplotypes du gène *GCN2*

En rouge : les individus de la population metallicoïles (PM) ; En bleu : la population non metallicoïles (PNM).

Pour la gène *Mt2a*, l'arbre phylogénétique nous montre qu'il y a 2 groupes distincts (figure 19). Par contre la carte de frequentation montre que est divisé en trois sous-groupes (annexe 5). Cependant, la relation ancestral englobant les deux populations du gène *Mt2a* (Figure 20 et annexe 6) montre que le nœud centrale appartient à l'individus de la population non métallicoles et ce nœuds se dérive vers les individus métallicoles. La PNM est presque monophylogénique et la PM montre une grande diversité sans avoir une origine en commun.

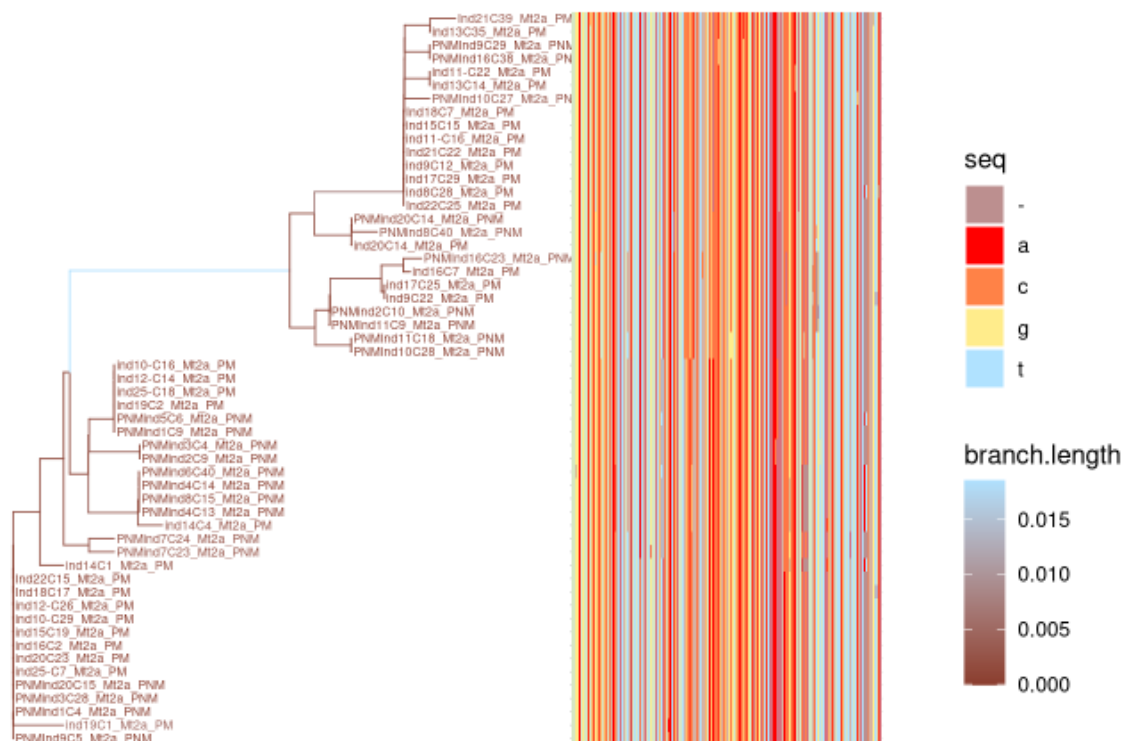


Figure 19 : Arbre phylogénétique d'alignement de séquences multiples de gène *Mt2a* avec Neighbor Joining algorithm.

II- Étude transcriptomique d'*Hirschfeldia incana*

La recherche sur *H. incana* est limitée et les mécanismes moléculaires derrière sa tolérance et son accumulation de Pb ne sont pas clairs en raison du manque de connaissances sur le génome de cette plante. L'assemblage de novo du transcriptome rend l'étude des ARN possible, même en l'absence d'un génome de référence, en fournissant un grand nombre de transcriptomes de recherche, caractérisant et identifiant les gènes qui sont exprimés dans les racines et les parties aériennes. Dans cette étude, nous avons réalisé pour la première fois, l'assemblage de novo du transcriptome RNA-seq et l'annotation des gènes du transcriptome d'*Hirschfeldia incana* et caractériser des changements transcriptionnels qui se produisent dans cette espèce en réponse au Pb.

1. Contrôle qualité des reads

Les 24 fichiers obtenus après le séquençage par illumina HiSeq était de 60.4 Gb de taille, il contenait des séquences de longueur de 150 pb. Chacun des fichiers obtenus comprenait en moyen 35 millions séquences avec un contenu de GC moyen de 45% (l'annexe 7).

L'analyse de la qualité des reads a été réalisée par l'outil FastQC qui donne un graphique pour chacun des reads (figure 21). Le résultat était similaire à tous les reads avec un score de qualité de chaque base de nos séquences supérieures à 28 (Phred-score > 28). Ceci signifie que l'ensemble des fichiers obtenus sont de bonne qualité (il n'est y a pas de contamination par les adaptateurs utilisés lors du séquençage par illumina Hiseq).

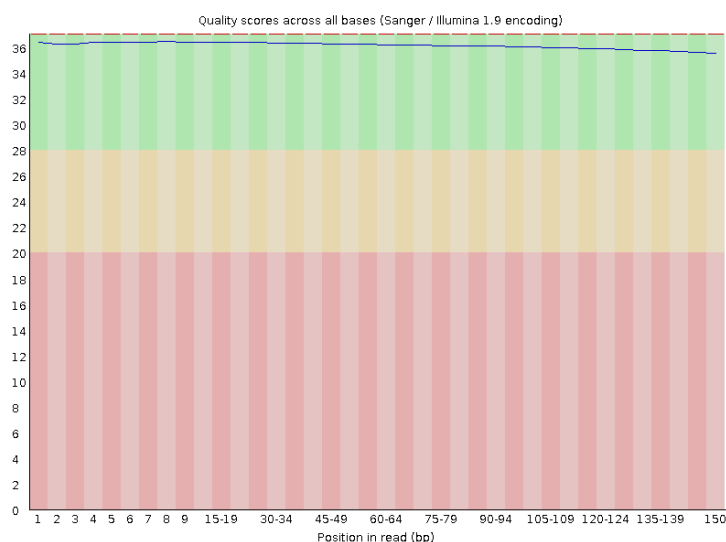


Figure 21 : Exemple d'un résultat de contrôle de la qualité de nos reads.

2. Contrôle de la qualité du transcriptome d'*H.incana*

Pour voir si la reconstruction du transcriptome obtenu est correcte (*de Novo* assembly), l'exactitude de l'assemblage a été évalué en utilisant BUSCO tool (Benchmarking Universal Single-Copy Orthologs). Le résultat de cette analyse montre un score BUSCO élevé (Figure 22) d'environ 94,2 % étaient complets, 2,8 % étaient fragmentés et 3,0 % étaient manquants (Figure 22). Le deuxième critère, pour vérifier la qualité de notre transcriptome assembly, est de réaligner les Raw reads avec le *novo* assembly en utilisant Bowtie2 comme tool d'alignement. Le résultat montre un pourcentage d'alignement global supérieur à 93 % (Tableaux 7). Les résultats obtenus par BUSCO et par l'alignement, signifient que l'assemblage (*De novo* assembly) a été très bien fait, et donc le transcriptome généré est de très bonne qualité.

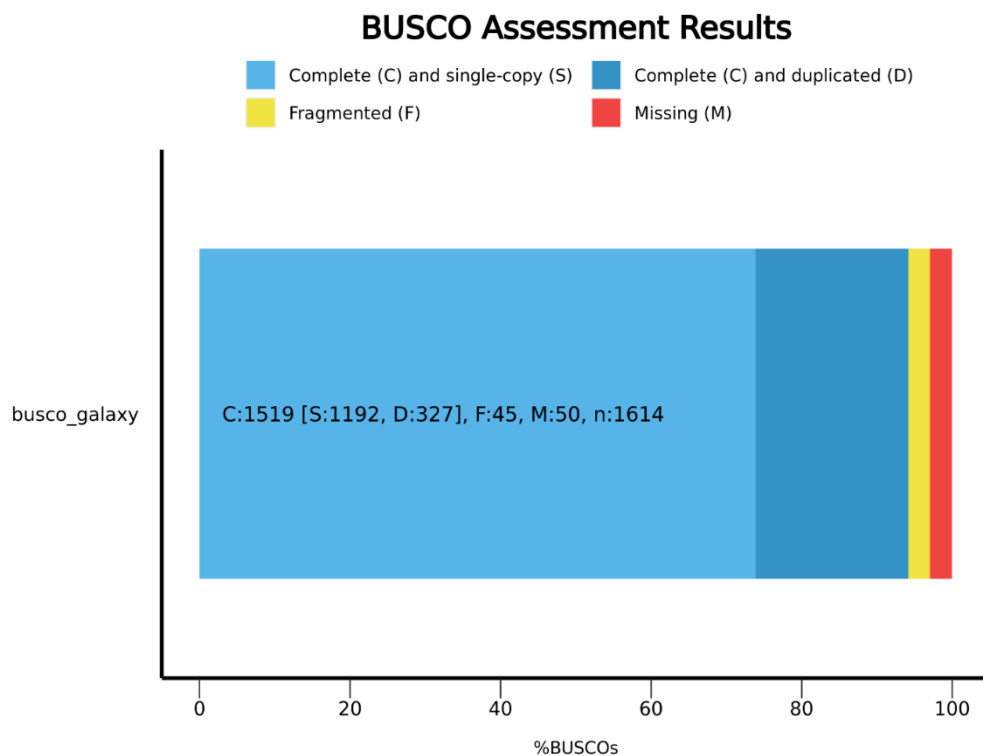


Figure 22 : Résultat de l'évaluation de la qualité de transcriptome par BUSCO tool.

Tableau 7 : Résultats d'alignement avec Bowtie2 tool

Transcrits	Taux d'alignement global	Aligné en concordance 0 fois	Aligné concordant exactement 1 fois	Aligné de façon concordante > 1 fois
HI-A-P-rep1	96.57%	17.15%	28.20%	54.65%
HI-A-P-rep2	96.01%	20.37%	27.01%	52.62%
HI-A-P-rep3	97.02%	14.98%	28.96%	56.06%
HI-A-T-rep1	97.02%	19.65%	28.17%	52.18%
HI-A-T-rep2	97.12%	15.99%	28.76%	55.25%
HI-A-T-rep3	96.89%	18.55%	28.08%	53.37%
HI-R-P-rep1	94.56%	23.21%	26.73%	50.06%
HI-R-P-rep2	93.53%	27.56%	24.08%	48.36%
HI-R-P-rep3	94.01%	26.62%	25.76%	47.62%
HI-R-T-rep1	95.38%	21.92%	27.28%	50.81%
HI-R-T-rep2	94.80%	20.13%	27.82%	52.04%
HI-R-T-rep3	93.71%	29.00%	24.35%	46.65%

Avec : HI : *H. incana*, R : racine, A : aérienne, P : non traité, T : traité, rep : répétition.

3. Annotation du transcriptome d'*H. incana*

La haute qualité de l'assemblage du transcriptome d'*H. incana* obtenu fournit des données de référence utiles qui peuvent également être utilisées pour clarifier les mécanismes de réponse au stress par le Pb et mener des recherches ultérieures en génomique fonctionnelle chez *Hirschfeldia incana*. L'annotation du 77 491 transcrits a été effectuée à l'aide de BLASTX sur la base de données de protéines d'*Arabidopsis thaliana*. Cela a généré environ 72 306 transcrits annotés, avec 11 862 transcrits en double qui ont été supprimés et 60 444 transcrits uniques. Pour identifier les GO des transcrits d'*H. incana*, le fichier d'annotation d'Ontologie génétique d'*Arabidopsis thaliana* a été utilisé. Par la suite, les annotations GO ont été classées en trois domaines essentiels importants : BP (biological process), CC (cellular component) et MF (molecular function). Sur 60 444 transcrits annotés, 60 395 transcrits étaient avec GO. Environ 13 046 étaient avec des termes GO uniques dont 7048 GO dans le domaine BP, 4535 GO dans le MF et 1463 GO dans le domaine CC.

Les GO les plus représentés en fonction du nombre des gènes dans les trois domaines sont présentés dans la figure 23. Dans le domaine BP, « signal transduction », « small molecule metabolic process » sont les plus représentés avec un nombre de gènes autour de 1600 gènes par processus. D'autres sub-domaines sont liés avec le stress métallique comme par exemple « Response to metal ion » (GO:0010038) qui contenait 392 gènes et « Metal ion homeostasis » (GO:0055065) avec 144 gènes. « Non-membrane-bounded organelle » et « mitochondrie » sont les lieux où il y a plus de gènes parmi les GOs trouvés dans le domaine CC. Pour le domaine MF, plus de 1600 et 1400 gènes sont impliqués dans « RNA binding » et « kinase activity », respectivement, tandis que le sous domaine « Metal cluster binding » (GO :0051540) contient uniquement 158 gènes.

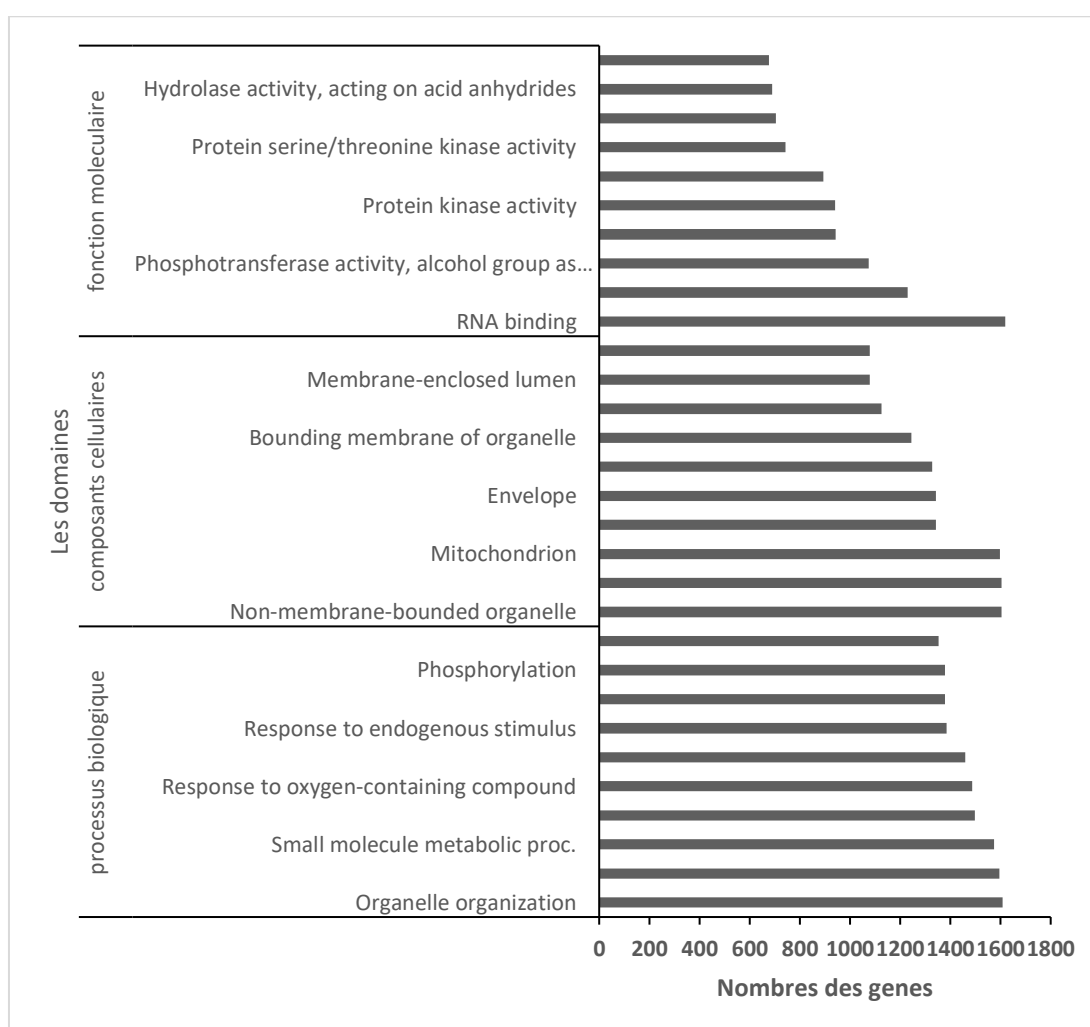


Figure 23 : Les dix GO les plus dominants en fonction du nombre de gènes selon les trois domaines fonctionnels : Processus Biologique (BP), Composants cellulaires (CC) et Fonction moléculaire (MF).

4. Analyse de l'expression des gènes en réponse au plomb

Dans cette étude, les résultats du diagramme de Venn indiquent que la plupart des transcrits étaient partagés entre les conditions traitées (Pb) et témoins (T). Environ 38 721 et 41 772 transcrits ont été partagés entre les deux conditions dans les parties aériennes et les racines, respectivement (Figure 24 a et b). Les gènes différentiellement exprimés sont caractérisés en utilisant le critère supérieur ou égal à 2 fold-change (FC) pour les gènes surexprimés et inférieur ou égal -2 FC pour les gènes sous-exprimés avec la méthode restrictive de "Bonferroni" avec une p valeur < 0,05. Les gènes ont été exprimés de manière différentielle (DE) dans les racines et les parties aériennes, respectivement (Figure 24 c) après 15 jours de traitement avec 100 μ M du Pb. Pour les racines, parmi 577 DEGs, 508 transcrits étaient surexprimés, tandis que 69 étaient sous exprimés sous le stress par le Pb (figure 24 c). Sans utiliser la méthode restrictive, plusieurs études ont détecté un nombre élevé de gènes en réponse au stress par les métaux lourds. Par exemple, un total de 4614 DEG ont été détectés dans les racines de *Raphanus sativus* L, avec 2154 gènes surexprimés et 2460 gènes sous exprimés sous stress par le Pb (Wang et al., 2013). Dans notre étude, parmi les 270 gènes, 244 étaient surexprimés, tandis que 26 gènes étaient inhibés dans les parties aériennes d'*H. incana* en réponse au Pb (Figure 24 c). D'autres auteurs ont rapporté que sous stress par le Pb, environ 1641 gènes surexprimés et 2884 gènes inhibés dans les parties aériennes de l'espèce hyper-tolérante au Pb *Fagopyrum tataricum* (Wang et al., 2020).

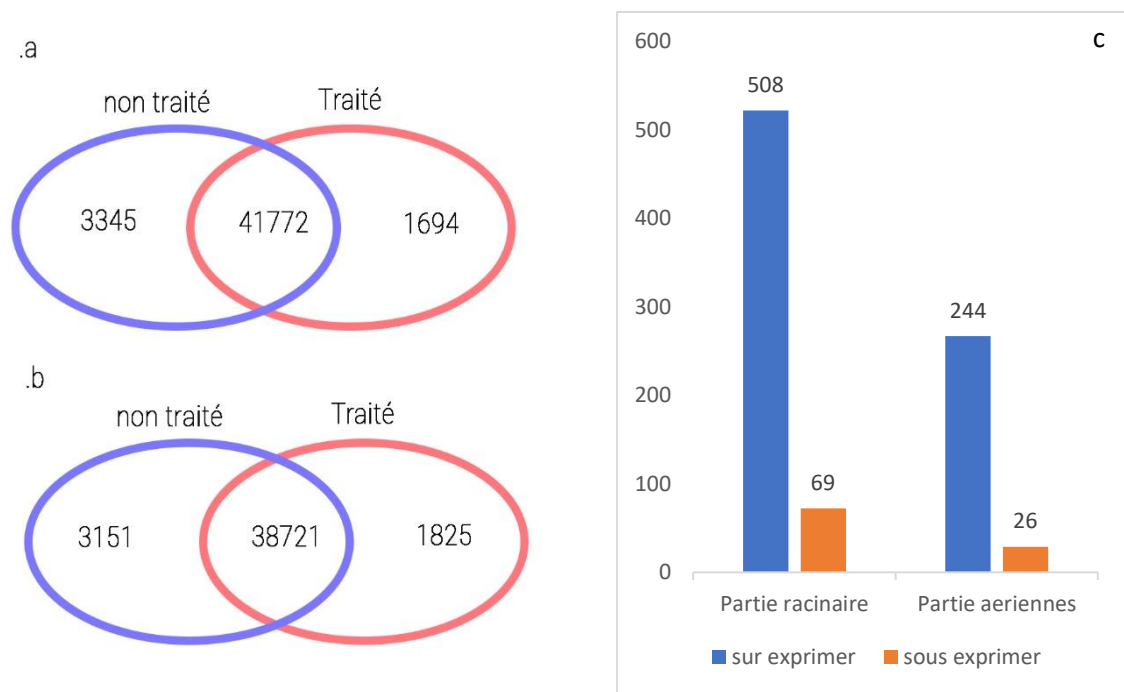


Figure 24 : Diagramme de venn des gènes communs entre les plantes traitées au plomb et non traitées (a. transcrit de la partie racinaire, b. transcrit de la partie aérienne) et c : l’histogramme de nombre des gènes différentiellement exprimés

5. Exemple de quelques gènes différentiellement exprimés en réponse au Pb

Quelques gènes régulés positivement dans la partie racinaire chez *H. incana* étaient trouvées surexprimés dans d'autres études sur des plantes soumises à un stress métallique. Le gène *PR-3-like* « pathogenèse liée à la protéine 3 », était surexprimé en réponse à un stress par d'aluminium (Al) (Arasimowicz-Jelonek et al., 2014). Le gène *HA9* « H(+) -ATPase » était surexprimée dans la partie racinaire d'*H. incana* en réponse à un stress par le Pb avec un logFold change égal à 10.29 (Tableau 8). Janicka-Russak et al., (2008) ont indiqué que le même gène a été surexprimé en réponse au Cd, Cu et Ni. Les mêmes auteurs indiquent que le gène *HA9* joue un rôle central dans la régulation de l'homéostasie ionique et la régulation de la croissance (Janicka-Russak et al., 2008).

Le gène *LRR-RLK* « sérine/thréonine protéine kinase de type récepteur LRR » était sur-exprimé dans la partie racinaire chez *H. incana* en réponse au stress par le Pb (Tableau 8). D'autres études indiquent que ce gène a un rôle important dans la tolérance aux métaux lourds. Par exemple, Ye et al., (2017) mentionnent que l'inhibition de l'expression de ce gène conduit à la diminution de la sensibilité aux métaux chez *Arabidopsis* (Ye et al., 2017).

Tableau 8 : Les 5 premier gènes (surexprimés et sous exprimés) en réponse au Pb dans la partie racinaire chez *H. incana*.

	transcript id	LogFCan ge	pval ue	padj	TAIR id	geneNa me	GeneFullName
Sur exprim é	TRINITY_DN6451_c 0_g2	11.68794 975	2.6E- 11	7.91E- 07	AT2G43 590	<i>PR-3- like genes</i>	pathogenesis related 3 proteine (chitinase)
	TRINITY_DN64488_ c0_g1	10.29081 573	1.53 E-07	0.0046 43	AT1G80 660	<i>HA9</i>	H(+)-ATPase 9
	TRINITY_DN5848_c 0_g1	9.896278 631	5.01 E-07	0.0152 48	AT2G34 690	<i>ACD11</i>	ACCELERATED CELL DEATH 11
	TRINITY_DN27683_ c0_g1	9.740509 936	2.39 E-13	7.26E- 09	AT3G62 210	<i>EDA32</i>	embryo sac development arrest 32
	TRINITY_DN6940_c 0_g1	9.471312 198	2.56 E-10	7.79E- 06	AT3G55 510	<i>RBL</i>	REBELOTE
Sous exprim é	TRINITY_DN36451_ c0_g1	- 21.69725 281	2.88 E-08	0.0008 76	AT2G28 960	<i>LRR- RLK</i>	LRR receptor-like serine/threoninepr otein kinase
	TRINITY_DN836_c1 _g1	- 11.89973 497	1.38 E-22	4.21E- 18	AT1G07 360	<i>MAC5A</i>	MOS4-associated complex subunit 5A
	TRINITY_DN24010_ c0_g1	- 9.258992 965	5.78 E-12	1.76E- 07	AT4G31 470	<i>CAP</i>	CAP (Cysteine- rich secretory proteins, Antigen 5, and Pathogenesis- related 1 protein) superfamily protein
	TRINITY_DN27612_ c0_g1	- 9.163154 928	3.85 E-08	0.0011 72	AT1G34 047	<i>DEFL</i>	Defensin-like (DEFL) family protein
	TRINITY_DN55982_ c0_g1	- 9.075984 985	2.11 E-11	6.41E- 07	AT4G05 320	<i>UBQ10</i>	polyubiquitin 10

Un gène qui n'est pas cité dans le tableau « monogalactosyldiacylglycérol (*MGDG*) » était également surexprimé dans la partie racinaire d'*H. incana* en réponse au stress par le Pb. Liu et

al., (2020) indiquent que l'absence de *MGDG* chez le mutant *Arabidopsis* a montré une inhibition de la croissance racinaire et la perte de l'intégrité membranaire sous stress par Al (Liu et al., 2020). La partie aérienne d'*H. incana* exprimer des gènes de façon différentiel sous le stress métallique du Pb et on a cité les 5 premier gènes les plus réguler par la plante dans le tableau 9 et tableau 10.

Tableau 9 : Les 5 premier gènes sur-exprimés en réponse au stress par le Pb dans la partie aérienne d'*H. incana*.

	transcript id	logFC	pvalue	padj	TAIR id	geneName	GeneFullName
sur expri mé	TRINITY_DN694 0_c0_g1	10.5673 4533	5.43 E-08	0.001 4	AT3G5 5510	<i>RBL</i>	REBELOTE
	TRINITY_DN646 66_c0_g1	8.55570 1231	6.97 E-09	0.000 18	AT1G4 3730	<i>RNA- directed DNA polymera se (reverse transcrip tase)</i>	RNA-directed DNA polymerase (reverse transcriptase)
	TRINITY_DN546 30_c0_g1	7.31561 7434	2.02 E-07	0.005 21	AT4G3 0690	<i>Translati on initiation factor 3 protein</i>	Translation initiation factor 3 protein
	TRINITY_DN387 97_c0_g1	6.07674 4269	1.57 E-11	4.05 E-07	AT5G3 9580	<i>PRX62</i>	PRX62
	TRINITY_DN823 3_c0_g1	5.65972 4087	6.74 E-11	1.74 E-06	AT1G1 7710	<i>PEPCI</i>	phosphoethanolamine/phosphocholine phosphatase 1

Tableau 10 : Les 5 premier gènes sous-exprimés en réponse au stress par le Pb dans la partie aérienne d'*H. incana*.

	transcript id	logFC	pvalue	padj	TAIR id	geneName	GeneFullName
sous exprimé	TRINITY_DN836_c1_g1	##### ##	5.63 E-21	1.45E- 16	AT1G073 60	MAC5A	MOS4-associated complex subunit 5A
	TRINITY_DN10297_c0_g1	##### ##	9.05E -14	2.33E- 09	AT1G100 00	Ribonuclease H-like superfamily protein	Ribonuclease H-like superfamily protein
	TRINITY_DN621_c1_g2	##### ##	1.63E -09	4.21E- 05	AT3G196 40	MGT4	magnesium transporter 4
	TRINITY_DN47752_c0_g2	##### ##	4.34E -07	0.0111 82	AT2G072 00	Cysteine proteinases superfamily protein	Cysteine proteinases superfamily protein
	TRINITY_DN8435_c0_g1	##### ##	1.62E -12	4.19E- 08	AT1G745 90	GSTU10	glutathione S-transferase TAU 10

Conclusion générale

Conclusion générale

Le plomb (Pb) est parmi les métaux lourds les plus toxiques. Il a une origine naturelle (éruption volcaniques, roches mères...) mais principalement anthropique tels que les éruptions volcaniques. En tant qu'élément non essentiel, le Pb n'a aucune fonction biologique pour les plantes, les animaux et les micro-organismes même à de faibles concentrations. Cependant, il existe des plantes tolérantes et capables de se développer et se reproduire sur des concentrations élevées du Pb dans le sol. Ces plantes présentent un potentiel pour leur utilisation dans des procédés de réhabilitation (phytoremédiation) des sols pollués par le Pb.

Hirshfeldia incana est une plante tolérante et hyperaccumulatrice du Pb, capable de concentrer des concentrations dépassant 1000ppm du Pb dans ses parties aériennes. *H. incana* est une pseudométallophyte, ce qui signifie qu'il existe une population homologue se développant sur des sols non contaminés par le Pb.

Cette étude a pour objectif d'apporter les premiers éléments de compréhension du fonctionnement moléculaire et génétique et du processus d'adaptation de cette espèce aux fortes concentrations en Pb.

Suite à l'approche génétique des populations, quatre séquences introniques des gènes candidats (ABCG22, CCH, GCN2, Mt2a) ont été examinées. Les résultats montrent qu'il n'y a pas un équilibre de Hardy-Weinberg et il existe au moins une des forces qui affecte la structure de nos populations pour les 4 gènes étudiés. Le résultat non significatif du Tajima D et la valeur F_s de Fu a informé de l'évolution neutre des deux populations. En outre, une grande différenciation entre les populations avec une valeur F_{ST} comprise entre 0,05 et 0,15 a été signalé. Ceci a été également confirmé par l'analyse DAPC. Des mutations dans des points clé ont été identifiées différenciant ainsi entre les deux populations. Cela incite à étendre la recherche et à explorer la raison de la différenciation entre les populations, qui pourrait être due à la grande distance géographique entre les deux populations ou liée à la présence ou absence du Pb dans le sol.

Le séquençage du transcriptome d'*H. incana* par RNAseq (Illumina) dans deux conditions (traitement au Pb et témoin) a été réalisé. Dans un premier temps, le novo assembly et l'annotation fonctionnelle des gènes identifiés ont été menés. Le transcriptome assemblé contient 77,491 transcrits avec plus de 60 444 transcrits annotés chez *H. incana*. La majorité des gènes appartenait au domaine de BP (biological process) dont essentiellement les processus de « signal transduction » et « small molecule metabolic process ». Pour le domaine de MF (molecular function), la majorité des gènes sont impliqués dans « RNA binding » et « kinase

activité ». Par la suite, les gènes différentiellement exprimés ont été identifiés au niveau de la partie aérienne et racinaire. Parmi les gènes surexprimés chez *H. incana*, les gènes *LRR-RLK* et *MGDG* pourraient être impliqués en réponse au Pb.

Les deux approches suivies dans cette étude ont ouvert la voie à d'autres questions et perspectives. Concernant l'étude de population, il serait judicieux de (i) réaliser une analyse inter et intra population des séquences des gènes régulateurs de nos 4 gènes étudiés ; (ii) refaire la même étude intra et interpopulation en utilisant les séquences codant de notre 4 gènes candidats ; (iii) cibler d'autres gènes régulés par le Pb à partir de nos résultats d'étude transcriptomique ; (iv) faire une analyse combinée des résultats génétiques et physiologiques en réponse au Pb. Par rapport aux résultats transcriptomiques, il serait intéressant de (i) de valider l'expression des gènes différentiellement exprimés par Q-PCR et de (ii) réaliser une validation fonctionnelle par génétique inverse pour certains gènes candidats.

Références bibliographiques

Références bibliographiques

- Abdelaziz, S., Ater, M., Auguy, F., Laplaze, L., El Mzibri, M., Berhada, F., Abdelkarim, F.-M., and Patrick, D. (2010). Evaluation de la contamination par les éléments-traces métalliques dans une zone minière du Maroc oriental = Assessment of contamination by metallic trace elements in a mining area of eastern Morocco.
- Abratowska, A., Wąsowicz, P., Bednarek, P., Telka, J., and Wierzbicka, M. (2012). Morphological and genetic distinctiveness of metallicolous and non-metallicolous populations of *Armeria maritima* s.l. (Plumbaginaceae) in Poland. *Plant Biology* **14**, 586-595.
- Adriano, D. C. (2001). "Trace elements in terrestrial environments: biogeochemistry, bioavailability, and risks of metals," Springer.
- Agriculture, U. S. D. o. (2017). Weed Risk Assessment for *Hirschfeldia incana* (L.) Lagr.-Foss (Brassicaceae) – Shortpod mustard. Vol. 2022.
- Al-Hiyaly, S., McNeilly, T., and Bradshaw, A. (1988). The effects of zinc contamination from electricity pylons—evolution in a replicated situation. *New phytologist* **110**, 571-580.
- Alford, É. R., Pilon-Smits, E. A., and Paschke, M. W. (2010). Metallophytes—a view from the rhizosphere. *Plant and Soil* **337**, 33-50.
- Andra, S. S., Datta, R., Sarkar, D., Makris, K. C., Mullens, C. P., Sahi, S. V., and Bach, S. B. (2010). Synthesis of phytochelatins in vetiver grass upon lead exposure in the presence of phosphorus. *Plant and soil* **326**, 171-185.
- Andrés-Colás, N., Sancenón, V., Rodríguez-Navarro, S., Mayo, S., Thiele, D. J., Ecker, J. R., Puig, S., and Peñarrubia, L. (2006). The Arabidopsis heavy metal P-type ATPase HMA5 interacts with metallochaperones and functions in copper detoxification of roots. *The Plant Journal* **45**, 225-236.
- Antonovics, J., Bradshaw, A. D., and Turner, R. (1971). Heavy metal tolerance in plants. In "Advances in ecological research", Vol. 7, pp. 1-85. Elsevier.
- Arasimowicz-Jelonek, M., Floryszak-Wieczorek, J., Drzewiecka, K., Chmielowska-Bąk, J., Abramowski, D., and Izbiańska, K. (2014). Aluminum induces cross-resistance of potato to *Phytophthora infestans*. *Planta* **239**, 679-694.
- Arshad, M. (2009). Phytoextraction du plomb par les *Pélargoniums* odorants: Interactions sol-plante et mise en place d'outils pour en comprendre l'hyperaccumulation, Toulouse, INPT.
- Ashraf, M., McNeilly, T., and Bradshaw, A. (1987). Selection and Heritability of Tolerance to Sodium Chloride in Four Forage Species 1. *Crop Science* **27**, 232-234.
- Auguy, F., Fahr, M., Moulin, P., Brugel, A., Laplaze, L., Mzibri, M. E., Filali-Maltouf, A., Dumas, P., and Smouni, A. (2013). Lead tolerance and accumulation in *Hirschfeldia incana*, a Mediterranean Brassicaceae from metalliferous mine spoils. *PLoS One* **8**, e61932.
- Auguy, F., Fahr, M., Moulin, P., El Mzibri, M., Smouni, A., Filali-Maltouf, A., Béna, G., and Dumas, P. (2016). Transcriptome changes in *Hirschfeldia incana* in response to lead exposure. *Frontiers in plant science* **6**, 1231.
- Baker, A. J., and Brooks, R. (1989). Terrestrial higher plants which hyperaccumulate metallic elements. A review of their distribution, ecology and phytochemistry. *Biorecovery*. **1**, 81-126.
- Balafrej, H., Bogusz, D., Triqui, Z.-E. A., Guedira, A., Bendaou, N., Smouni, A., and Fahr, M. (2020). Zinc hyperaccumulation in plants: A review. *Plants* **9**, 562.
- Becerra-Castro, C., Monterroso, C., Prieto-Fernández, A., Rodríguez-Lamas, L., Loureiro-Viñas, M., Acea, M. J., and Kidd, P. S. (2012). Pseudometallophytes colonising Pb/Zn mine tailings: A description of the plant–microorganism–rhizosphere soil system and isolation of metal-tolerant bacteria. *Journal of Hazardous Materials* **217-218**, 350-359.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., and Bignell, H. R. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *nature* **456**, 53-59.

- Besnard, G., Basic, N., Christin, P. A., Savova-Bianchi, D., and Galland, N. (2009). *Thlaspi caerulescens* (Brassicaceae) population genetics in western Switzerland: is the genetic structure affected by natural variation of soil heavy metal concentrations? *New Phytologist* **181**, 974-984.
- Bhargava, A., Carmona, F. F., Bhargava, M., and Srivastava, S. (2012). Approaches for enhanced phytoextraction of heavy metals. *Journal of environmental management* **105**, 103-120.
- Boldyrev, M. (2018). Lead: Properties, history, and applications. *WikiJournal of Science* **1**, 1-23.
- Bone, E., and Farres, A. (2001). Trends and rates of microevolution in plants. *Genetica* **112**, 165-182.
- BRADSHAW, A. D. (1952). Populations of *Agrostis tenuis* resistant to lead and zinc poisoning. *Nature* **169**, 1098-1098.
- Brännvall, M.-L., Bindler, R., Renberg, I., Emteryd, O., Bartnicki, J., and Billström, K. (1999). The Medieval metal industry was the cradle of modern large-scale atmospheric lead pollution in northern Europe. *Environmental Science & Technology* **33**, 4391-4395.
- Burton, K., Morgan, E., and Roig, A. (1984). The influence of heavy metals upon the growth of sitka-spruce in South Wales forests. *Plant and soil* **78**, 271-282.
- CABI (2022). *Hirschfeldia incana* (shortpod mustard). (I. S. Compendium, ed.), Wallingford, UK: CAB International.
- Cappa, J. J., and Pilon-Smits, E. A. (2014). Evolutionary aspects of elemental hyperaccumulation. *Planta* **239**, 267-275.
- Carpenter, K. E. (1924). A study of the fauna of rivers polluted by lead mining in the Aberystwyth district of Cardiganshire. *Annals of Applied Biology* **11**, 1-23.
- Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics* **12**, 1-7.
- Cheng, H., and Hu, Y. (2010). Lead (Pb) isotopic fingerprinting and its applications in lead pollution studies in China: a review. *Environmental pollution* **158**, 1134-1146.
- Çimrin, K. M., Turan, M., and Kapur, B. (2007). Effect of elemental sulphur on heavy metals solubility and remediation by plants in calcareous soils. *Fresenius Environmental Bulletin* **16**, 1113-1120.
- Cummings, M. P., Neel, M. C., and Shaw, K. L. (2008). A genealogical approach to quantifying lineage divergence. *Evolution: International Journal of Organic Evolution* **62**, 2411-2422.
- Cunha Neto, A. R. d., Ambrósio, A. d. S., Wolowski, M., Westin, T. B., Govêa, K. P., Carvalho, M., and Barbosa, S. (2020). Negative effects on photosynthesis and chloroplast pigments exposed to lead and aluminum: a meta-analysis. *Cerne* **26**, 232-237.
- De Silva, S., Bernett, C., Meaklim, J., Abeywardane, E., and Reichman, S. M. (2019). Probing the effects of different lead compounds on the bioavailability of lead to plants. *Chemosphere* **230**, 24-28.
- Del Río-Celestino, M., Font, R., Moreno-Rojas, R., and De Haro-Bailón, A. (2006). Uptake of lead and zinc by wild plants growing on contaminated soils. *Industrial Crops and Products* **24**, 230-237.
- Dogan, I., and Dogan, N. (2016). Statistical Tests for Neutrality: A Review. *Turkiye Klinikleri Journal of Biostatistics* **9**.
- Dong, Z., and Chen, Y. (2013). Transcriptomics: advances and approaches. *Science China Life Sciences* **56**, 960-967.
- Dubey, S., Shri, M., Misra, P., Lakhwani, D., Bag, S. K., Asif, M. H., Trivedi, P. K., Tripathi, R. D., and Chakrabarty, D. (2014). Heavy metals induce oxidative stress and genome-wide modulation in transcriptome of rice root. *Functional & integrative genomics* **14**, 401-417.
- Dubois, S., Cheptou, P. O., Petit, C., Meerts, P., Poncelet, M., Vekemans, X., Lefebvre, C., and Escarré, J. (2003). Genetic structure and mating systems of metallicolous and nonmetallicolous populations of *Thlaspi caerulescens*. *New Phytologist* **157**, 633-641.
- el Hasnaoui, S., Mouna, F., Keller, C., Levard, C., Angeletti, B., Chaurand, P., Triqui, Z., Abdelkarim, G., Rhazi, L., colin, f., and Abdelaziz, S. (2020). Screening of Native Plants Growing on a Pb/Zn Mining Area in Eastern Morocco: Perspectives for Phytoremediation. *Plants* **9**.

- Epelde, L., Hernández-Allica, J., Becerril, J. M., Blanco, F., and Garbisu, C. (2008). Effects of chelates on plants and soil microbial community: comparison of EDTA and EDDS for lead phytoextraction. *Science of the total environment* **401**, 21-28.
- Ernst, W. H. (2006). Evolution of metal tolerance in higher plants. *For Snow Landsc Res* **80**, 251-274.
- Excoffier, L., and Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564-7.
- Fahr, M. (2014). Étude physiologique et moléculaire de la tolérance et l'accumulation du Pb chez une population métallicole et une population non métallicole d'*Hirschfeldia incana* (Brassicacée récoltée au Maroc): Perspectives pour la phytoremédiation.
- Farber, D. A. (1992). Politics and procedure in environmental law. *JL Econ. & Org.* **8**, 59.
- Frankham, R., Ballou, S. E. J. D., Briscoe, D. A., and Ballou, J. D. (2002). "Introduction to conservation genetics," Cambridge university press.
- Ge, S. X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628-2629.
- Ghnaya, T., Zaier, H., Baioui, R., Sghaier, S., Lucchini, G., Sacchi, G. A., Lutts, S., and Abdelly, C. (2013). Implication of organic acids in the long-distance transport and the accumulation of lead in *Sesuvium portulacastrum* and *Brassica juncea*. *Chemosphere* **90**, 1449-1454.
- Gisbert, C., Clemente, R., Navarro-Avinó, J., Baixauli, C., Ginér, A., Serrano, R., Walker, D. J., and Bernal, M. P. (2006). Tolerance and accumulation of heavy metals by Brassicaceae species grown in contaminated soils from Mediterranean regions of Spain. *Environmental and Experimental Botany* **56**, 19-27.
- Guo, W.-J., Bundithya, W., and Goldsbrough, P. B. (2003). Characterization of the Arabidopsis metallothionein gene family: tissue-specific expression and induction during senescence and in response to copper. *New Phytologist*, 369-381.
- Gupta, D., Huang, H., and Corpas, F. (2013). Lead tolerance in plants: strategies for phytoremediation. *Environmental Science and Pollution Research* **20**, 2150-2161.
- Gupta, D., Nicoloso, F., Schetinger, M., Rossato, L., Pereira, L., Castro, G., Srivastava, S., and Tripathi, R. (2009). Antioxidant defense mechanism in hydroponically grown Zea mays seedlings under moderate lead stress. *Journal of Hazardous Materials* **172**, 479-484.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., and Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494-1512.
- Harrison, M. D., Jones, C. E., and Dameron, C. T. (1999). Copper chaperones: function, structure and copper-binding properties. *JBIC Journal of Biological Inorganic Chemistry* **4**, 145-153.
- Hartl, D. L., Clark, A. G., and Clark, A. G. (1997). "Principles of population genetics," Sinauer associates Sunderland.
- Hettiarachchi, G. M., and Pierzynski, G. M. (2004). Soil lead bioavailability and in situ remediation of lead-contaminated soils: A review. *Environmental progress* **23**, 78-93.
- Holsinger, K. (2012). Tajima's D, Fu's FS, Fay and Wu's H, and Zeng et al.'s E. *Lecture notes in population genetics*, 239-244.
- Hu, B., Deng, F., Chen, G., Chen, X., Gao, W., Long, L., Xia, J., and Chen, Z.-H. (2020). Evolution of abscisic acid signaling for stress responses to toxic metals and metalloids. *Frontiers in Plant Science* **11**, 909.
- Imadi, S. R., Kazi, A. G., Ahanger, M. A., Gucel, S., and Ahmad, P. (2015). Plant transcriptomics and responses to environmental stress: an overview. *J Genet* **94**, 525-37.
- Islam, E., Yang, X., Li, T., Liu, D., Jin, X., and Meng, F. (2007). Effect of Pb toxicity on root morphology, physiology and ultrastructure in the two ecotypes of *Elsholtzia argyi*. *Journal of Hazardous Materials* **147**, 806-816.
- Jamla, M., Khare, T., Joshi, S., Patil, S., Penna, S., and Kumar, V. (2021). Omics approaches for understanding heavy metal responses and tolerance in plants. *Current Plant Biology* **27**, 100213.

- Janicka-Russak, M., Kabała, K., Burzyński, M., and Kłobus, G. (2008). Response of plasma membrane H⁺-ATPase to heavy metal stress in *Cucumis sativus* roots. *Journal of Experimental Botany* **59**, 3721-3728.
- Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics* **11**, 1-15.
- Joshi, R., Pareek, A., and Singla-Pareek, S. L. (2016). Plant metallothioneins: classification, distribution, function, and regulation. In "Plant Metal Interaction", pp. 239-261. Elsevier.
- Jusselme, M. D., Poly, F., Miambi, E., Mora, P., Blouin, M., Pando, A., and Rouland-Lefèvre, C. (2012). Effect of earthworms on plant *Lantana camara* Pb-uptake and on bacterial communities in root-adhering soil. *Science of the Total Environment* **416**, 200-207.
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., and Keselman, J. C. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of educational research* **68**, 350-386.
- Kevresan, S., Petrovic, N., Popovic, M., and Kandrac, J. (2001). Nitrogen and protein metabolism in young pea plants as affected by different concentrations of nickel, cadmium, lead, and molybdenum. *Journal of Plant Nutrition* **24**, 1633-1644.
- Krämer, U. (2010). Metal hyperaccumulation in plants. *Annual review of plant biology* **61**, 517-534.
- Krämer, U., Talke, I. N., and Hanikenne, M. (2007). Transition metal transport. *FEBS letters* **581**, 2263-2272.
- Kumar, A., Kumar, A., MMS, C.-P., Chaturvedi, A. K., Shabnam, A. A., Subrahmanyam, G., Mondal, R., Gupta, D. K., Malyan, S. K., and Kumar, S. S. (2020). Lead toxicity: health hazards, influence on food chain, and sustainable remediation approaches. *International journal of environmental research and public health* **17**, 2179.
- Lambinon, J., and Auquier, P. (1964). La flore et la végétation des terrains calaminaires de la Wallonie septentrionale et de la Rhénanie aixoise. *Types chorologiques et groupes écologiques. Natura Mosana* **16**, 13-131.
- Larsson, O., Wennmalm, K., and Sandberg, R. (2006). Comparative microarray analysis. *Omics: a journal of integrative biology* **10**, 381-397.
- Leach, D. L., Bradley, D. C., Huston, D., Pisarevsky, S. A., Taylor, R. D., and Gardoll, S. J. (2010). Sediment-hosted lead-zinc deposits in Earth history. *Economic Geology* **105**, 593-625.
- Lee, P. L., Patel, R., Conlan, R., Wainwright, S., and Hipkin, C. (2004). Comparison of genetic diversities in native and alien populations of hoary mustard (*Hirschfeldia incana* [L.] Lagreze-Fossat). *International Journal of Plant Sciences* **165**, 833-843.
- Liang, J., and Mao, J. (2015). Source analysis of global anthropogenic lead emissions: their quantities and species. *Environmental Science and Pollution Research* **22**, 7129-7138.
- Likens, G. E. (2010). "Biogeochemistry of inland waters," Academic press.
- Liu, C., Liu, Y., Wang, S., Ke, Q., Yin, L., Deng, X., and Feng, B. (2020). Arabidopsis mgd mutants with reduced monogalactosyldiacylglycerol contents are hypersensitive to aluminium stress. *Ecotoxicology and Environmental Safety* **203**, 110999.
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. *PLoS computational biology* **13**, e1005457.
- Lynch, J. M., Brimecombe, M. J., and De Leij, F. A. (2001). Rhizosphere. *e LS*.
- Macnair, M. (1983). The genetic control of copper tolerance in the yellow monkey flower, *Mimulus guttatus*. *Heredity* **50**, 283-293.
- Madejón, P., Murillo, J., Marañón, T., and Lepp, N. (2007). Factors affecting accumulation of thallium and other trace elements in two wild Brassicaceae spontaneously growing on soils contaminated by tailings dam waste. *Chemosphere* **67**, 20-28.
- Maestri, E., Pirondini, A., Visioli, G., and Marmioli, N. (2013). Trade-off between genetic variation and ecological adaptation of metallicolous and non-metallicolous *Noccaea* and *Thlaspi* species. *Environmental and Experimental Botany* **96**, 1-10.

- Malecka, A., Jarmuszkiewicz, W. a., and Tomaszewska, B. (2001). Antioxidative defense to lead stress in subcellular compartments of pea root cells. *Acta Biochimica Polonica* **48**, 687-698.
- Małkowski, E., Kita, A., Galas, W., Karcz, W., and Kuperberg, J. M. (2002). Lead distribution in corn seedlings (*Zea mays* L.) and its effect on growth and the concentrations of potassium and calcium. *Plant Growth Regulation* **37**, 69-76.
- Marguerat, S., and Bähler, J. (2010). RNA-seq: from technology to biology. *Cellular and molecular life sciences* **67**, 569-579.
- Meerts, P., and Van Isacker, N. (1997). Heavy metal tolerance and accumulation in metallicolous and non-metallicolous populations of *Thlaspi caerulescens* from continental Europe. *Plant ecology* **133**, 221-231.
- Mengoni, A., Barabesi, C., Gonnelli, C., Galardi, F., Gabbrielli, R., and Bazzicalupo, M. (2001). Genetic diversity of heavy metal-tolerant populations in *Silene paradoxa* L.(Caryophyllaceae): a chloroplast microsatellite analysis. *Molecular ecology* **10**, 1909-1916.
- Mills, R. F., Krijger, G. C., Baccarini, P. J., Hall, J., and Williams, L. E. (2003). Functional expression of AtHMA4, a P1B-type ATPase of the Zn/Co/Cd/Pb subclass. *The Plant Journal* **35**, 164-176.
- Mineta, K., Akahoshi, E., Takahashi, K., Horiguchi, T., Hayashi, H., Ohyama, T., Ikarashi, T., Fujiwara, T., and Chino, M. (2000). Localization of MT2a gene promoter expression is different from the site (s) of copper accumulation in the roots of transgenic *Arabidopsis thaliana*. *Soil science and plant nutrition* **46**, 189-197.
- Mishra, A., and Choudhuri, M. (1998). Amelioration of lead and mercury effects on germination and rice seedling growth by antioxidants. *Biologia Plantarum* **41**, 469-473.
- Mouna, F., Laplaze, L., El Mzibri, M., Patrick, D., Bendaou, N., Hoher, V., Bogusz, D., and Abdelaziz, S. (2015). Assessment of lead tolerance and accumulation in metallicolous and non-metallicolous populations of *Hirschfeldia incana*. *Environmental and Experimental Botany* **109**, 186-192.
- Mudd, G. M., Jowitt, S. M., and Werner, T. T. (2017). The world's lead-zinc mineral resources: scarcity, data, issues and opportunities. *Ore Geology Reviews* **80**, 1160-1190.
- Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G., and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Current opinion in biotechnology* **24**, 22-30.
- Papoyan, A., and Kochian, L. V. (2004). Identification of *Thlaspi caerulescens* genes that may be involved in heavy metal hyperaccumulation and tolerance. Characterization of a novel heavy metal transporting ATPase. *Plant Physiology* **136**, 3814-3823.
- Parsons, W. T., Parsons, W. T., and Cuthbertson, E. (2001). "Noxious weeds of Australia," CSIRO publishing.
- Paszkiwicz, K., and Studholme, D. J. (2010). De novo assembly of short sequence reads. *Briefings in bioinformatics* **11**, 457-472.
- Patterson, C. C., and Settle, D. M. (1987). Magnitude of lead flux to the atmosphere from volcanoes. *Geochimica et Cosmochimica Acta* **51**, 675-681.
- Pauwels, M., Frerot, H., Bonnini, I., and SAUMITOU-LAPRADE, P. (2006). A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). *Journal of evolutionary biology* **19**, 1838-1850.
- Pauwels, M., Saumitou-Laprade, P., Holl, A. C., Petit, D., and Bonnini, I. (2005). Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. *Mol Ecol* **14**, 4403-14.
- Peng, J.-S., Guan, Y.-H., Lin, X.-J., Xu, X.-J., Xiao, L., Wang, H.-H., and Meng, S. (2021). Comparative understanding of metal hyperaccumulation in plants: a mini-review. *Environmental Geochemistry and Health* **43**, 1599-1607.
- Plaza, S., and Bovet, L. (2008). 18 Cadmium Detoxification in Plants: Involvement of ABC Transporters. *Trace Elements as Contaminants and Nutrients: Consequences in Ecosystems and Human Health*, 449.

- Poschenrieder, C., Bech, J., Llugany, M., Pace, A., Fenés, E., and Barceló, J. (2001). Copper in plant species in a copper gradient in Catalonia (North East Spain) and their potential for phytoremediation. *Plant and soil* **230**, 247-256.
- Pourrut, B., Shahid, M., Douay, F., Dumat, C., and Pinelli, E. (2013). Molecular mechanisms involved in lead uptake, toxicity and detoxification in higher plants. *Heavy metal stress in plants*, 121-147.
- Prat, S. (1934). Die Erbllichkeit der Resistenz gegen kupfer. *Berichte der Deutschen Botanischen Gesellschaft* **52**, 65-67.
- Prendergast, W. D. (1910). The classification of the symptoms of lead poisoning. *British Medical Journal* **1**, 1164.
- PROST, L. (2018). Genetic responses during an experimental evolution of *Noccaea caerulea* in zinc-polluted conditions.
- Raghavan, V., Kraft, L., Mesny, F., and Rigerte, L. (2022). A simple guide to de novo transcriptome assembly and annotation. *Brief Bioinform* **23**.
- Rauser, W. E. (1999). Structure and function of metal chelators produced by plants. *Cell biochemistry and biophysics* **31**, 19-48.
- Reddy, C., and Patrick Jr, W. (1977). "Effect of redox potential and pH on the uptake of cadmium and lead by rice plants," Rep. No. 0047-2425. Wiley Online Library.
- Reeves, R., and Baker, A. (2000). Metal-accumulating plants. *Phytoremediation of Toxic Metals: Using Plants to Clean Up the Environment*. John Wiley **6**.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., Mungall, K., Lee, S., Okada, H. M., and Qian, J. Q. (2010). De novo assembly and analysis of RNA-seq data. *Nature methods* **7**, 909-912.
- Ruley, A. T., Sharma, N. C., Sahi, S. V., Singh, S. R., and Sajwan, K. S. (2006). Effects of lead and chelators on growth, photosynthetic activity and Pb uptake in *Sesbania drummondii* grown in soil. *Environmental pollution* **144**, 11-18.
- Salzburger, W., Ewing, G. B., and Von Haeseler, A. (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular ecology* **20**, 1952-1963.
- Sanger, F., and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* **94**, 441-448.
- Sauvé, S., McBride, M., and Hendershot, W. (1998). Lead phosphate solubility in water and soil suspensions. *Environmental Science & Technology* **32**, 388-393.
- Schraiber, J. G., and Akey, J. M. (2015). Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics* **16**, 727-740.
- Semane, B., Dupae, J., Cuypers, A., Noben, J.-P., Tuomainen, M., Tervahauta, A., Kärenlampi, S., Van Belleghem, F., Smeets, K., and Vangronsveld, J. (2010). Leaf proteome responses of *Arabidopsis thaliana* exposed to mild cadmium stress. *Journal of plant physiology* **167**, 247-254.
- Sengar, R. S., Gautam, M., Sengar, R. S., Garg, S. K., Sengar, K., and Chaudhary, R. (2008). Lead stress effects on physiobiochemical activities of higher plants. *Rev Environ Contam Toxicol* **196**, 73-93.
- Seregin, I., and Ivanov, V. (2001). Physiological aspects of cadmium and lead toxic effects on higher plants. *Russian journal of plant physiology* **48**, 523-544.
- Seregin, I., Shpigun, L., and Ivanov, V. (2004). Distribution and toxic effects of cadmium and lead on maize roots. *Russian Journal of Plant Physiology* **51**, 525-533.
- Shahid, M., Pinelli, E., and Dumat, C. (2012). Review of Pb availability and toxicity to plants in relation with metal speciation; role of synthetic and natural organic ligands. *Journal of hazardous materials* **219**, 1-12.
- Sharma, P., and Dubey, R. S. (2005). Lead toxicity in plants. *Brazilian journal of plant physiology* **17**, 35-52.

- Siemens, J. (2011). *Hirschfeldia*. In "Wild Crop Relatives: Genomic and Breeding Resources", pp. 171-176. Springer.
- Song, S., Sliwerska, E., Emery, S., and Kidd, J. M. (2017). Modeling human population separation history using physically phased genomes. *Genetics* **205**, 385-395.
- Sunkar, R., Kaplan, B., Bouche, N., Arazi, T., Dolev, D., Talke, I. N., Maathuis, F. J., Sanders, D., Bouchez, D., and Fromm, H. (2000). Expression of a truncated tobacco NtCBP4 channel in transgenic plants and disruption of the homologous Arabidopsis CNGC1 gene confer Pb²⁺ tolerance. *The Plant Journal* **24**, 533-542.
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* **6**, e21800.
- Thomas, N. E., Feng, W., and Henzler-Wildman, K. A. (2021). A solid-supported membrane electrophysiology assay for efficient characterization of ion-coupled transport. *Journal of Biological Chemistry* **297**.
- Thomine, S., Lelièvre, F., Debarbieux, E., Schroeder, J. I., and Barbier-Brygoo, H. (2003). AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *The Plant Journal* **34**, 685-695.
- Usman, K., Abu-Dieyeh, M. H., Zouari, N., and Al-Ghouti, M. A. (2020). Lead (Pb) bioaccumulation and antioxidative responses in *Tetraena qataranse*. *Scientific Reports* **10**, 17070.
- Van Rossum, F., Le Pajolec, S., Raspé, O., and Godé, C. (2022). Assessing population genetic status for designing plant translocations.
- Verbruggen, N., Hermans, C., and Schat, H. (2009). Molecular mechanisms of metal hyperaccumulation in plants. *New phytologist* **181**, 759-776.
- Verma, S., and Dubey, R. (2003). Lead toxicity induces lipid peroxidation and alters the activities of antioxidant enzymes in growing rice plants. *Plant science* **164**, 645-655.
- Verret, F., Gravot, A., Auroy, P., Leonhardt, N., David, P., Nussaume, L., Vavasseur, A., and Richaud, P. (2004). Overexpression of AtHMA4 enhances root-to-shoot translocation of zinc and cadmium and plant metal tolerance. *FEBS letters* **576**, 306-312.
- Walley, K., Khan, M., and Bradshaw, A. (1974). The potential for evolution of heavy metal tolerance in plants. *Heredity* **32**, 309-319.
- Wang, L., Zheng, B., Yuan, Y., Xu, Q., and Chen, P. (2020). Transcriptome profiling of *Fagopyrum tataricum* leaves in response to lead stress. *BMC plant biology* **20**, 1-14.
- Wang, Y., Xu, L., Chen, Y., Shen, H., Gong, Y., Limera, C., and Liu, L. (2013). Transcriptome profiling of radish (*Raphanus sativus* L.) root and identification of genes involved in response to lead (Pb) stress with next generation sequencing. *PLoS One* **8**, e66539.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* **10**, 57-63.
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., and Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution* **35**, 543-548.
- Watson, A., Mazumder, A., Stewart, M., and Balasubramanian, S. (1998). Technology for microarray analysis of gene expression. *Current opinion in biotechnology* **9**, 609-614.
- Weber, M., Trampczynska, A., and Clemens, S. (2006). Comparative transcriptome analysis of toxic metal responses in *Arabidopsis thaliana* and the Cd²⁺-hypertolerant facultative metallophyte *Arabidopsis halleri*. *Plant, Cell & Environment* **29**, 950-963.
- Willems, G., Dräger, D. r. B., Courbot, M., Godé, C., Verbruggen, N., and Saumitou-Laprade, P. (2007). The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. *Genetics* **176**, 659-674.
- Williams, L. E., and Mills, R. F. (2005). P1B-ATPases—an ancient family of transition metal pumps with diverse functions in plants. *Trends in plant science* **10**, 491-502.
- Wong, C. K. E., and Cobbett, C. S. (2009). HMA P-type ATPases are the major mechanism for root-to-shoot Cd translocation in *Arabidopsis thaliana*. *New phytologist* **181**, 71-78.

- Wood, E., and Nimmo, K. (1994). In search of new lead-free electronic solders. *Journal of Electronic Materials* **23**, 709-713.
- Yang, J., Li, J., Dong, L., and Grünewald, S. (2011). Analysis on the reconstruction accuracy of the Fitch method for inferring ancestral states. *BMC bioinformatics* **12**, 1-9.
- Ye, Y., Ding, Y., Jiang, Q., Wang, F., Sun, J., and Zhu, C. (2017). The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant cell reports* **36**, 235-242.
- Zhang, X., Yang, L., Li, Y., Li, H., Wang, W., and Ye, B. (2012). Impacts of lead/zinc mining and smelting on the environment and human health in China. *Environmental monitoring and assessment* **184**, 2261-2273.
- Zhou, J., Zhang, Z., Zhang, Y., Wei, Y., and Jiang, Z. (2018). Effects of lead stress on the growth, physiology, and cellular structure of privet seedlings. *PloS one* **13**, e0191139.
- Zhu, J., Fang, X. Z., Dai, Y. J., Zhu, Y. X., Chen, H. S., Lin, X. Y., and Jin, C. W. (2019). Nitrate transporter 1.1 alleviates lead toxicity in Arabidopsis by preventing rhizosphere acidification. *Journal of experimental botany* **70**, 6363-6374.
- Zientara, K., Wawrzyńska, A., Łukomska, J., López-Moya, J. R., Liszewska, F., Assunção, A. G., Aarts, M. G., and Sirko, A. (2009). Activity of the AtMRP3 promoter in transgenic Arabidopsis thaliana and Nicotiana tabacum plants is increased by cadmium, nickel, arsenic, cobalt and lead but not by zinc and iron. *Journal of Biotechnology* **139**, 258-263.
- Zimmers, J., Thomas, M., Yang, L., Bombarely, A., Mancuso, M., Wojciechowski, M., and Smith, J. (2017). Species boundaries in the Astragalus cusickii complex delimited using molecular phylogenetic techniques. *Molecular Phylogenetics and Evolution* **114**, 93-110.

Annexes

Annexe 1

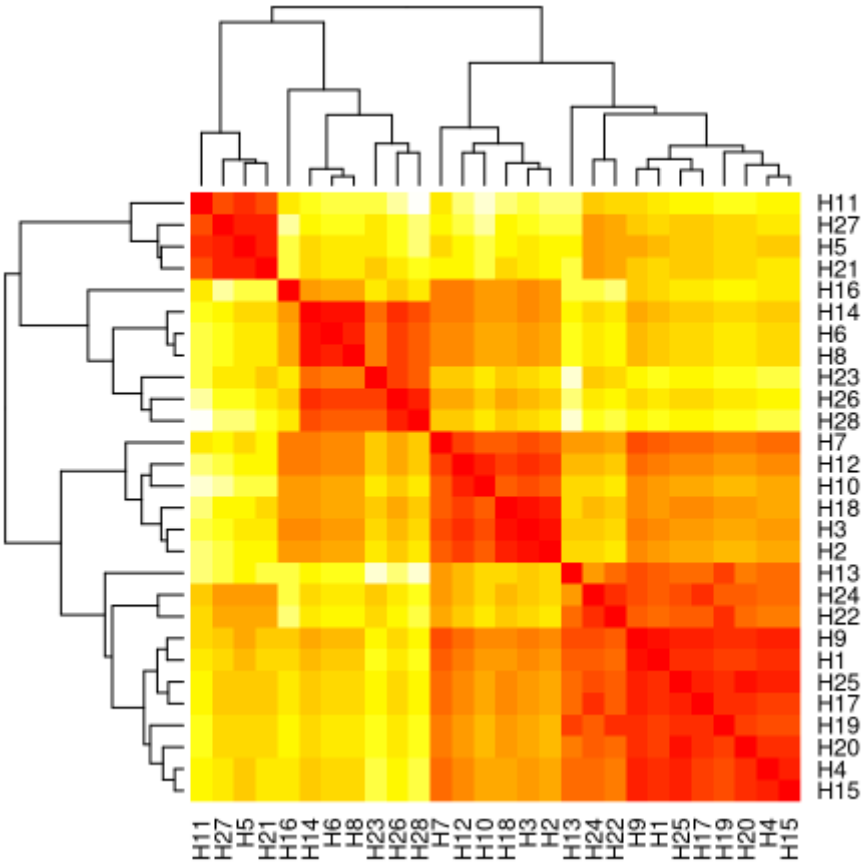
Liste des séquences nucléotidique similaire regrouper dans un haplotypes des quatre gènes.

Haplotypes	ABC	CCH	GCN2	Mt2a
1	Ind22c9_ABC_PM Ind16c1_ABC_PM Ind21C23_ABC_PM ind15C17_ABC_PM	ind12-C15F_CCH_PM ind17C41_CCH_PM ind11-C15_CCH_PM ind20C3_CCH_PM Ind18C11_CCH_PM ind11-C32_CCH_PM PNMInd14C24_CCH PNMInd20C3_CCH PNMInd12C4_CCH	PMInd12-C9_GC2N2_PM Ind6C1_GC2N2_PNM Ind7C2_GC2N2_PNM Ind8C2_GC2N2_PNM	ind12-C14_Mt2a_PM ind10-C16_Mt2a_PM ind25-C18_Mt2a_PM ind19C2_Mt2a_PM PNMInd5C6_Mt2a PNMInd1C9_Mt2a
2	Ind15c21_ABC_PM	ind10-C10F_CCH_PM	PMInd8C10_GC2N2_PM PMInd9C4_GC2N2_PM PMInd8C12_GC2N2_PM	ind14C4_Mt2a_PM
3	Ind17c50_ABC_PM ind16C18_ABC_PM Ind18C30_ABC_PM Ind20C7'_ABC_PM ind19C22_ABC_PM PNMInd2C13_ABC	ind13C2_CCH_PM	PMInd17C1_GC2N2_PM	ind16C7_Mt2a_PM
4	Ind18c27_ABC_PM	Ind22C10_CCH_PM	PMInd15C3_GC2N2_PM PMInd16C2_GC2N2_PM PMInd25-C2_GC2N2_PM PMInd18C2_GC2N2_PM Ind7C3_GC2N2_PNM	ind11-C16_Mt2a_PM ind15C15_Mt2a_PM Ind21C22_Mt2a_PM ind9C12_Mt2a_PM ind17C29_Mt2a_PM ind8C28_Mt2a_PM Ind18C7_Mt2a_PM Ind22C25_Mt2a_PM
5	Ind20c9'_ABC_PM Ind22C35_ABC_PM Ind8C4'_ABC_PM Ind12C10'_ABC_PM	ind20C22_CCH_PM ind13C31_CCH_PM	PMInd13C1_GC2N2_PM	Ind22C15_Mt2a_PM Ind18C17_Mt2a_PM ind12-C26_Mt2a_PM ind10-C29_Mt2a_PM ind15C19_Mt2a_PM ind16C2_Mt2a_PM ind20C23_Mt2a_PM ind25-C7_Mt2a_PM PNMInd1C4_Mt2a PNMInd20C15_Mt2a PNMInd3C28_Mt2a PNMInd9C5_Mt2a
6	Ind8c9'_ABC_PM	ind25-C29_CCH_PM Ind18C26_CCH_PM Ind21C36_CCH_PM ind10-C20F_CCH_PM ind19C12_CCH_PM Ind21C36'_CCH_PM PNMInd8C35_CCH PNMInd6C5_CCH PNMInd3C16_CCH PNMInd7C16_CCH PNMInd3C16'_CCH	PMInd10-C23_GC2N2_PM	ind13C14_Mt2a_PM ind11-C22_Mt2a_PM
7	Ind13C20_ABC_PM	ind16C15_CCH_PM ind9C18_CCH_PM	PMInd18C3_GC2N2_PM Ind10C40_GC2N2_PNM	ind20C14_Mt2a_PM PNMInd20C14_Mt2a

		ind17C11_CCH_PM ind16C31_CCH_PM ind9C18'_CCH_PM PNMInd10C39_CCH PNMInd12C6_CCH PNMind7C21_CCH PNMInd13C11_CCH PNMind5C25_CCH PNMInd11C8_CCH PNMInd16C24_CCH PNMind2C27_CCH PNMind4C19_CCH PNMind8C18_CCH PNMind6C35_CCH PNMInd13C11'_CCH PNMind5C25'_CCH	Ind16C28_GCN2_PNM Ind4C2_GCN2_PNM Ind3C5_GCN2_PNM Ind11C13_GCN2_PNM Ind10C6_GCN2_PNM Ind14C16_GCN2_PNM Ind2C2_GCN2_PNM Ind4C1_GCN2_PNM Ind1C2_GCN2_PNM	
8	Ind11-C12_ABC_PM	ind19C8_CCH_PM	PMind11-C21_GCN2_PM PMind9C15_GCN2_PM	ind13C35_Mt2a_PM
9	Ind12C8'_ABC_PM	ind8C17_CCH_PM ind12-C16F_CCH_PM ind25-C6_CCH_PM ind8C17'_CCH_PM PNMind1C18_CCH	PMind12-C2_GCN2_PM	Ind21C39_Mt2a_PM
10	Ind19C3_ABC_PM	Ind22C28_CCH_PM	PMInd17C3_GCN2_PM PMInd15C2_GCN2_PM	ind9C22_Mt2a_PM ind17C25_Mt2a_PM
11	Ind10C7'_ABC_PM	PNMInd11C7_CCH	PMInd16C1_GCN2_PM	ind19C1_Mt2a_PM
12	Ind9C4'_ABC_PM	PNMInd16C8_CCH	PMInd13C2_GCN2_PM	ind14C1_Mt2a_PM
13	ind17C17_ABC_PM	PNMind2C16_CCH	PMind10-C2_GCN2_PM	PNMInd10C28_Mt2a PNMInd11C18_Mt2a
14	Ind21C29_ABC_PM	PNMind4C11_CCH	Ind5C5_GCN2_PNM Ind8C8_GCN2_PNM Ind13C37_GCN2_PNM Ind6C2_GCN2_PNM Ind9C18_GCN2_PNM	PNMInd11C9_Mt2a PNMind2C10_Mt2a
15	ind13C37_ABC_PM	PNMind1C1_CCH	Ind14C9_GCN2_PNM Ind2C1_GCN2_PNM Ind16C21_GCN2_PNM	PNMInd16C38_Mt2a PNMInd9C29_Mt2a
16	ind11-C3_ABC_PM Ind10C10'_ABC_PM	PNMInd10C21_CCH	Ind9C16_GCN2_PNM ind20C13_GCN2_PNM	PNMind7C23_Mt2a
17	PNMind5c30_ABC PNMind14C29_ABC PNMind8C30_ABC PNMind12C41_ABC PNMind7C18_ABC	PNMInd14C27_CCH	Ind1C4_GCN2_PNM	PNMind2C9_Mt2a PNMind3C4_Mt2a
18	PNMind1C36_ABC PNMind10c32_ABC PNMInd14C17_ABC	PNMind20C1_CCH	Ind5C3_GCN2_PNM	PNMind8C40_Mt2a
19	PNMind3C26_ABC			PNMind4C14_Mt2a PNMind6C40_Mt2a PNMind8C15_Mt2a PNMind4C13_Mt2a
20	PNMind9C38_ABC			PNMInd10C27_Mt2a
21	PNMind4C5_ABC			PNMInd16C23_Mt2a

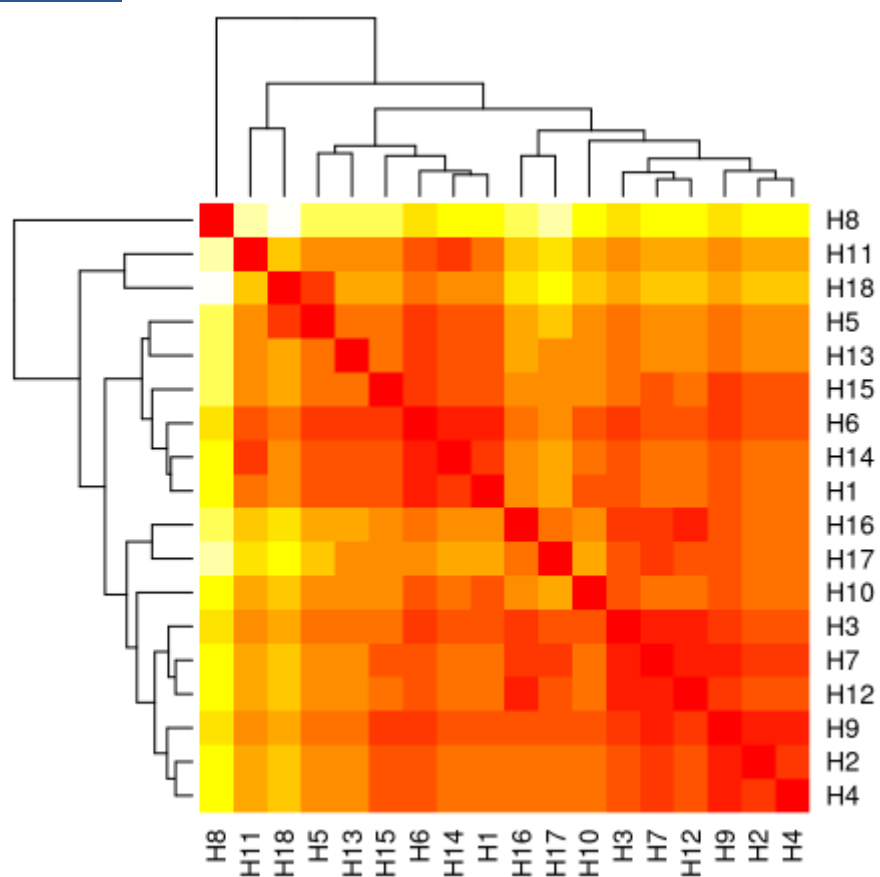
22	PNMind5C7_ABC		PNMind7C24_Mt2a
23	PNMind2C15_ABC PNMind1C3_ABC		
24	PNMind8C32_ABC		
25	PNMind3C30_ABC		

Annexe 2



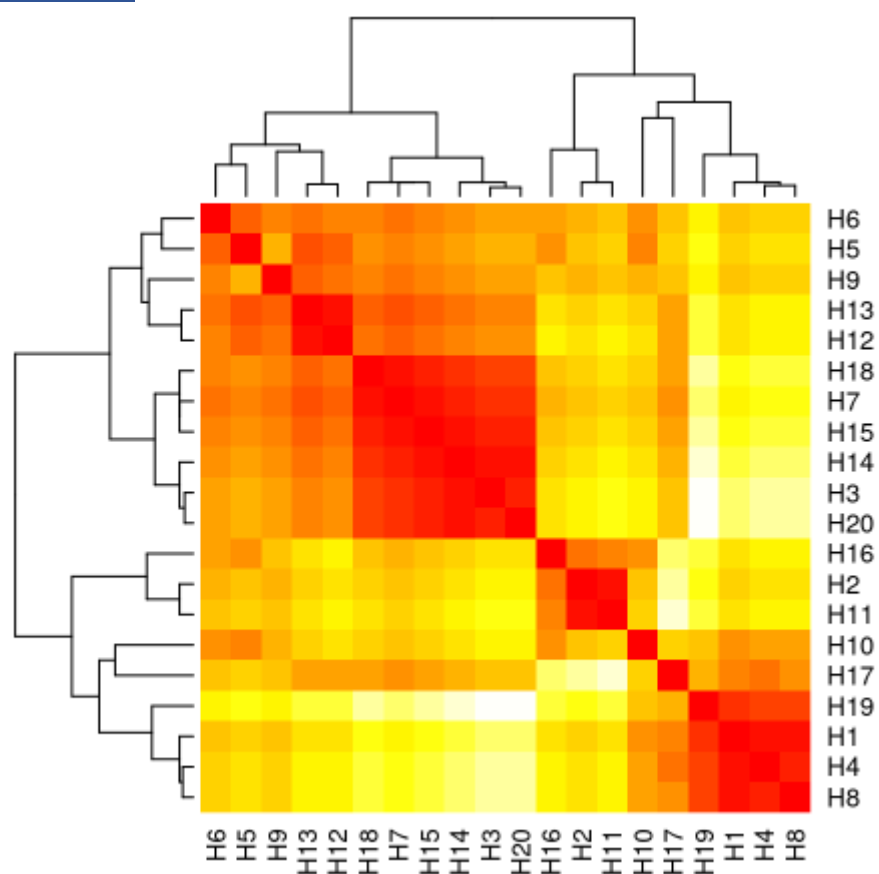
Carte de chaleur de la gène ABC.

Annexe 3



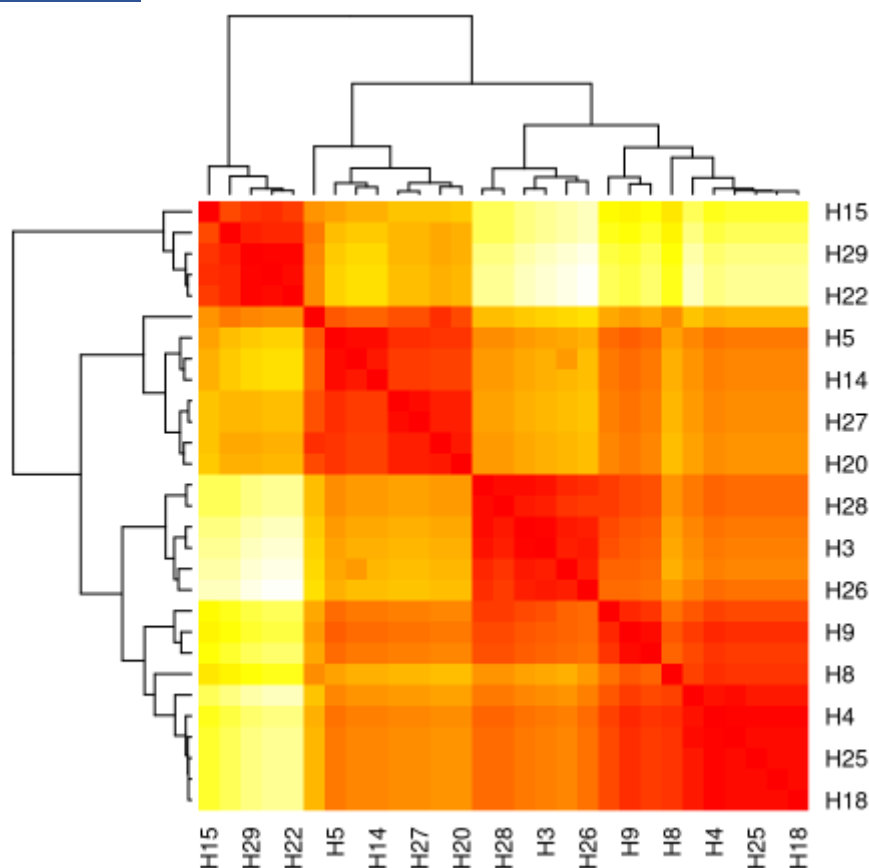
Carte de chaleur de la gène CCH.

Annexe 4



Carte de chaleur de la gène GCN2.

Annexe 5



Carte de chaleur de la gène Mt2a.

Annexe 6

Liste des individus dans chaque nœud de chaque gène.

Nœud	ABC	GCN2	Mt2a
1	Ind18c27ABC_PM	PMind12-C9_GC2N2_PM	PNMind7C24_Mt2a_PNM
2	ind13C37ABC_PM	PMind11-C21_GC2N2_PM PMind9C15_GC2N2_PM	PNMind7C23_Mt2a_PNM
3	PNMind3C26_PNM	Ind6C1_GC2N2_PNM	ind19C1_Mt2a_PM
4	Ind16c1ABC_PM Ind21C23ABC_PM Ind22c9ABC_PM ind15C17ABC_PM	Ind7C3_GC2N2_PNM PMind15C3_GC2N2_PM PMind16C2_GC2N2_PM PMind18C2_GC2N2_PM PMind25-C2_GC2N2_PM	ind14C1_Mt2a_PM
5	PNMind5C7_PNM	Ind5C3_GC2N2_PNM	Ind18C17_Mt2a_PM Ind22C15_Mt2a_PM PNMind9C5_Mt2a_PNM PNMind1C4_Mt2a_PNM PNMind20C15_Mt2a_PNM PNMind3C28_Mt2a_PNM ind10-C29_Mt2a_PM ind12-C26_Mt2a_PM ind15C19_Mt2a_PM

			ind16C2_Mt2a_PM ind20C23_Mt2a_PM ind25-C7_Mt2a_PM
6	ind17C17ABC_PM	Ind7C2_GCN2_PNM Ind8C2_GCN2_PNM	-
7	PNMInd14C17_PNM PNMind10c32_PNM PNMind1C36_PNM	Ind1C4_GCN2_PNM	-
8	PNMind12C41_PNM PNMind14C29_PNM PNMind5c30_PNM PNMind7C18_PNM PNMind8C30_PNM	PMInd16C1_GCN2_PM	ind14C4_Mt2a_PM
9	Ind10C7ABC_PM	PMind8C10_GCN2_PM PMind8C12_GCN2_PM PMind9C4_GCN2_PM	PNMind4C13_Mt2a_PNM PNMind4C14_Mt2a_PNM PNMind6C40_Mt2a_PNM PNMind8C15_Mt2a_PNM
10	PNMind9C38_PNM	Ind9C16_GCN2_PNM ind20C13_GCN2_PNM	PNMind1C9_Mt2a_PNM PNMind5C6_Mt2a_PNM ind10-C16_Mt2a_PM ind12-C14_Mt2a_PM ind19C2_Mt2a_PM ind25-C18_Mt2a_PM
11	PNMind8C32_PNM	PMInd17C1_GCN2_PM	PNMind2C9_Mt2a_PNM PNMind3C4_Mt2a_PNM
12	Ind15c21ABC_PM	Ind13C37_GCN2_PNM Ind5C5_GCN2_PNM Ind6C2_GCN2_PNM Ind8C8_GCN2_PNM Ind9C18_GCN2_PNM	-
13	Ind8c9ABC_PM	Ind14C9_GCN2_PNM Ind16C21_GCN2_PNM Ind2C1_GCN2_PNM	PNMind8C40_Mt2a_PNM
14	Ind11-C12ABC_PM	Ind10C40_GCN2_PNM Ind10C6_GCN2_PNM Ind11C13_GCN2_PNM Ind14C16_GCN2_PNM Ind16C28_GCN2_PNM Ind1C2_GCN2_PNM Ind2C2_GCN2_PNM Ind3C5_GCN2_PNM Ind4C1_GCN2_PNM Ind4C2_GCN2_PNM PMInd18C3_GCN2_PM	PNMInd10C27_Mt2a_PNM
15	PNMind4C5_PNM	PMInd13C1_GCN2_PM	PNMInd16C38_Mt2a_PNM PNMInd9C29_Mt2a_PNM
16	PNMind3C30_PNM	-	Ind21C39_Mt2a_PM
17	PNMind2C15_PNM	PMind10-C23_GCN2_PM	ind13C35_Mt2a_PM
18	Ind21C29ABC_PM	-	ind11-C22_Mt2a_PM ind13C14_Mt2a_PM
19	Ind10C10ABC_PM ind11-C3ABC_PM	PMind12-C2_GCN2_PM	PNMind20C14_Mt2a_PNM ind20C14_Mt2a_PM
20	Ind17c50ABC_PM Ind18C30ABC_PM Ind20C7ABC_PM PNMind2C13_PNM	PMInd13C2_GCN2_PM	Ind18C7_Mt2a_PM Ind21C22_Mt2a_PM Ind22C25_Mt2a_PM ind11-C16_Mt2a_PM ind15C15_Mt2a_PM ind17C29_Mt2a_PM ind8C28_Mt2a_PM ind9C12_Mt2a_PM

	ind16C18ABC_PM ind19C22ABC_PM		
21	-	PMind10- C2_GC22_PM	PNMInd16C23_Mt2a_PNM
22	Ind19C3ABC_PM	-	ind16C7_Mt2a_PM
23	Ind9C4ABC_PM	-	ind17C25_Mt2a_PM ind9C22_Mt2a_PM
24	Ind13C20ABC_PM	-	PNMInd10C28_Mt2a_PNM PNMInd11C18_Mt2a_PNM
25	-	PMInd15C2_GC22_PM PMInd17C3_GC22_PM	PNMInd11C9_Mt2a_PNM PNMInd2C10_Mt2a_PNM
26	Ind12C8ABC_PM	-	-
27	Ind12C10ABC_PM Ind20c9ABC_PM Ind22C35ABC_PM Ind8C4ABC_PM	-	-
28	-		-
29	-		-
30			-
31			-

Annexe 7

Résultats du contrôle de qualité.

ID	TRANSCRITS	NOMBRES DES SEQUENCES	TAILS DES SEQUENCES	GC (%)
1	HI-A-P- rep1_R1	23934480	150	46
2	HI-A-P- rep1_R2	23934480	150	46
3	HI-A-P- rep2_R1	35774380	150	46
4	HI-A-P- rep2_R2	35774380	150	46
5	HI-A-P- rep3_R1	31349162	150	46
6	HI-A-P- rep3_R2	31349162	150	47
7	HI-A-T- rep1_R1	54306920	150	47
8	HI-A-T- rep1_R2	54306920	150	46
9	HI-A-T- rep2_R1	47743835	150	47
10	HI-A-T- rep2_R2	47743835	150	46
11	HI-A-T- rep3_R1	39982852	150	47
12	HI-A-T- rep3_R2	39982852	150	46

13	HI-R-P- rep1_R1	24105543	150	45
14	HI-R-P- rep1_R2	24105543	150	45
15	HI-R-P- rep2_R1	24289410	150	46
16	HI-R-P- rep2_R2	24289410	150	44
17	HI-R-P- rep3_R1	26314171	150	46
18	HI-R-P- rep3_R2	26314171	150	45
19	HI-R-T- rep1_R1	41073505	150	46
20	HI-R-T- rep1_R2	41073505	150	46
21	HI-R-T- rep2_R1	37103633	150	46
22	HI-R-T- rep2_R2	37103633	150	45
23	HI-R-T- rep3_R1	36873721	150	46
24	HI-R-T- rep3_R2	36873721	150	44

Résumé :

La pollution par le plomb (Pb) est un problème environnemental très répandu en raison de sa toxicité pour les organismes vivants. *Hirschfeldia incana*, une espèce de la famille des *Brassicaceae*, communément trouvée dans les régions méditerranéennes, est caractérisée par sa capacité à tolérer et à accumuler de fortes concentrations du Pb. C'est une pseudométallophyte qui se développe aussi bien sur des sols pollués que sur des sites naturels non contaminés. Cette plante a été proposée comme un excellent modèle pour évaluer la réponse et l'adaptation des plantes au Pb.

L'objectif de cette étude est d'identifier et de comprendre les mécanismes moléculaires de la tolérance, l'adaptation et l'accumulation du Pb chez *H. incana*. Pour cela, une analyse génétique a été réalisée d'une part en vue d'apprécier l'effet du stress métallique sur la spéciation de cette espèce via une étude comparative des séquences introniques de 4 gènes candidats entre une population métallicole (PM) et une population non métallicole (PNM) d'*H. incana*. D'autre part, le profil transcriptomique de cette espèce a été également assemblé et annoté, et les gènes différentiellement exprimés en réponse au Pb ont été identifiés.

Les gènes *MT2a*, *CCH*, *ABCG22* et *GCN2* ont été choisis pour une étude génétique. Suivant le test AMOVA, l'analyse de séquences a montré une variation significative au niveau intrapopulation. Tandis que la variation interpopulation n'est pas importante pour ces gènes. Ces gènes ont montré une évolution suivant un modèle neutre avec une valeur D de Tajima et Fs de Fu non significative. L'analyse phylogénique a montré que pour les 4 gènes qu'il y'a une distinction des groupes avec des similarités génétiques entre les individus des deux populations PM et PNM. L'analyse généalogique a révélé que pour les gènes *ABCG22* et *GCN2*, la PNM est issue de la PM (origine ancestrale) alors que l'étude du gène *Mt2a* a montré l'inverse. A ce niveau d'analyse, nous n'avons pas détecté d'effet de sélection sur les gènes étudiés chez les deux populations. Et l'ambiguïté de la relation ancestrale entre les populations nous incite à utiliser d'autres marqueurs moléculaires pour étudier les origines ancestrales de la population métallicole. La différence en termes d'adaptation au Pb des deux populations manifesterait probablement niveau de la régulation des gènes étudiés.

Le séquençage du transcriptome (RNA-seq) d'*H. incana* dans deux conditions, contrôle (sans Pb) et traité avec 100 μM de $\text{Pb}(\text{NO}_3)_2$ en hydroponie pendant 15 jours a été réalisé. Un total de 842 millions de reads a été généré. 77491 séquences de transcrits ont été assemblées. Les analyses de similarité de séquence et l'annotation de ces transcrits ont été effectuées par rapport à la base de données des protéines d'*Arabidopsis thaliana* et de Gene Ontology (GO). En réponse au Pb, 577 et 270 gènes ont été exprimés de manière différentielle dans les racines et les parties aériennes, respectivement.

Mot clé : *Hirschfeldia incana*, plomb (Pb), tolérance, accumulation, phytoremediation, population, métallicole, RNAseq.