



**HAL**  
open science

# Approche métabolomique pour l'étude du dépérissement de la lavande : application aux composés non-volatils

Bénédicte Héral

► **To cite this version:**

Bénédicte Héral. Approche métabolomique pour l'étude du dépérissement de la lavande : application aux composés non-volatils. Chimie organique. Université Côte d'Azur, 2021. Français. NNT : 2021COAZ4009 . tel-03267843

**HAL Id: tel-03267843**

**<https://theses.hal.science/tel-03267843v1>**

Submitted on 22 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



$$\rho \left( \frac{\partial v}{\partial t} + v \cdot \nabla v \right) = -\nabla p + \nabla \cdot T + f$$

$$e^{i\pi} + 1 = 0$$

# THÈSE DE DOCTORAT

Approche métabolomique pour l'étude  
du dépérissement de la lavande :  
application aux composés non-volatils

**Bénédicte Héral**

Institut de Chimie de Nice

**Présentée en vue de l'obtention  
du grade de docteur en Chimie  
de l'Université Côte d'Azur**

**Dirigée par :**

Dr. Thomas Michel

**Co-encadrée par :**

Pr. Xavier Fernandez

**Soutenue le :** 9 mars 2021

**Devant le jury, composé de :**

Anne-Emmanuelle Hay-de Bettignies,  
Maître de conférences, Université de Lyon 1

Alain Muselli, Professeur,  
Université de Corse

Serge Antonczak, Professeur,  
Université Côte d'Azur

Marie-Virginie Salvia, Maître de conférences,  
Université de Perpignan

Thomas Michel, Maître de conférences HDR,  
Université Côte d'Azur

Xavier Fernandez, Professeur,  
Université Côte d'Azur



CHANEL



RÉGION  
SUD



PROVENCE  
ALPES  
CÔTE D'AZUR



Thèse présentée pour l'obtention **du grade de docteur en Chimie**

**Université Côte d'Azur**

École Doctorale de Sciences Fondamentales et Appliquées

# Approche métabolomique pour l'étude du dépérissement de la lavande : application aux composés non-volatils

Présentée par **Bénédicte HÉRAL**

le 9 mars 2021

Sous la direction du Dr. Thomas MICHEL et l'encadrement du Pr. Xavier  
FERNANDEZ

Devant le jury composé de :

Dr. Anne-Emmanuelle Hay-de Bettignies	Université de Lyon 1	Rapporteur
Pr. Alain Muselli	Université de Corse	Rapporteur
Pr. Serge Antonczak	Université Côte d'Azur	Examineur
Dr. Marie-Virginie Salvia	Université de Perpignan	Examineur
Dr. Thomas Michel	Université Côte d'Azur	Directeur de thèse
Pr. Xavier Fernandez	Université Côte d'Azur	Co-encadrant de thèse



## Approche métabolomique dans le dépérissement de la lavande (CONVs)

La lavande et le lavandin sont des plantes à parfums très appréciées des industries cosmétiques et de la parfumerie, et sont un des symboles de la Provence. Seulement, les lavanderaies françaises subissent un fort déclin depuis une quinzaine d'années lié à des épisodes de sécheresse et à une maladie communément appelée « le dépérissement de la lavande ». Cette maladie est provoquée par une bactérie sans paroi le phytoplasme du Stolbur (*Candidatus Phytoplasma solani*) et véhiculée par un insecte la cicadelle (*Hyalestes obsoletus*). Cette thèse se propose d'étudier les interactions tripartites entre l'insecte, la plante et la bactérie à travers les modifications induites sur les composés organiques non volatiles (CONVs).

Pour répondre à cette problématique, une approche métabolomique non ciblée a été développée et réalisée sur quatre variétés de lavande (7713, maillette) et de lavandin (abrial, grosso). Pour ce faire, une campagne d'échantillonnage a été menée en champs et a conduit au prélèvement de 480 échantillons. Des triplicats d'extraction ont ensuite été réalisés par extraction solide-liquide assistée par ultra-sons avec un mélange eau-éthanol (50:50, v/v). Les analyses ont été conduites sur une UPLC-HRMS (XevoG2 QTOF, Waters). Le traitement des données a été effectué sur la plateforme W4M-Galaxy pour l'extraction des données spectrales, et sur MetaboAnalyst pour les corrections et les tests statistiques (ACP et PLS-DA).

Les résultats obtenus ont mis en évidence des différences significatives entre des variétés sensibles et tolérantes au phytoplasme du Stolbur pour chaque organe étudié (parties chlorophylliennes / inflorescences). Des signatures chimiques distinctes en fonction de l'état symptomatique des plants (symptomatique/asymptomatique) ont également pu être révélées pour plusieurs variétés et parties de la plante. En fonction des organes, des composés discriminants du caractère symptomatique communs à plusieurs variétés ont été détectés. Plusieurs dérivées de chlorophylles et des caroténoïdes ont été trouvés comme probables marqueurs de l'état sanitaires des plantes.

**Mots-clés :** *Lavandula*, CONVs, *Hyalestes obsoletus*, LC-MS, Phytoplasme, métabolomique

---

## Metabolomic approach for yellow disease (nVOCs)

Fine lavender and lavandin are historic and emblematic crops of Provence (South-East France). They are cultivated for their aromatic and perfume properties. They represented an important economic value in France. Nevertheless, over the last decades, lavender's field decline because of the lack of rain fall and a disease name "yellow disease". This infection lead to phloem-limited bacterial pathogen (*Candidatus Phytoplasma solani*) and transmitted by a planthopper (*Hyalestes obsoletus*). The purpose of this PhD is to study induced chemical defences (i.e. non-volatile organic compounds) in order to better understand interaction between insect, plant and bacteria.

To answer this issue, an untargeted metabolomic approach was developed. Four varieties chosen for their sensibility or tolerance against phytoplasma, were used: two lavender (7713, maillette) and two lavandin (abrial, grosso), and 480 samples were collected. Compound extraction was performed by a solid-liquid extraction assisted by ultrasounds with an ethanol-water mixture (50:50, v/v). Afterwards, 1429 extracts were analysed by an UPLC-HRMS (XevoG2 QTOF, Waters). Spectral data was first handle using W4M-Galaxy to obtain a peak table fitting by statistical tools. Next, MetaboAnalyst was conducted to correct data and performed multivariate analyses. PCA was used to visualize trends and outliers. PLS-DA was applied to highlight biological differences between subgroups, to discover the most relevant factors and detected biomarkers. The most discriminant compounds specific to sensitivity/tolerant species as well as to plant symptomatic status (asymptomatic/symptomatic) were annotated. Chlorophyll derivatives, under-expressed in symptomatic plants, are probably health plant status biomarkers.

**Keywords :** *Lavandula* spp., nVOCs, *Hyalestes obsoletus*, LC-MS, Phytoplasma, metabolomics



*À ma marraine,*



# Remerciements

Je tiens à remercier, en tout premier lieu, les membres du jury et particulièrement les rapporteurs :

- Mme Anne-Emmanuelle Hay-de Bettignies maître de conférences à l'université Lyon 1,
- Mr Alain Muselli, professeur à l'université Pasquale Paoli de Corse

Je vous remercie sincèrement d'avoir accepté d'évaluer ce manuscrit, ainsi que pour vos retours précieux et avisés sur ce travail. J'exprime ma gratitude à Mme Marie-Virginie Salvia maître de conférences à l'université de Perpignan et au professeur Serge Antonczak de l'université Côte d'Azur pour avoir accepté de faire partie de ce jury.

Je tiens également à remercier les directeurs successifs de l'Institut de Chimie de Nice : Mme Elisabet Dunach et Mr. Uwe Meierhenrich pour m'avoir accueillie et permis de réaliser ma thèse au sein du laboratoire.

Mes remerciements vont ensuite à mes encadrants, en particulier à mon directeur Thomas Michel, pour sa disponibilité et sa réactivité. Il a su donner de sa personne dans cette thèse, en sauvant mes échantillons au mépris de ses genoux ; et en étant le directeur de thèse le plus rapide de France avec sa belle performance en 2 min et 39,49 s pour courir acheter un rouleau d'aluminium. Plus sérieusement, je tiens à le remercier pour sa compréhension face à ma « petite » santé. Je tiens également à remercier Xavier Fernandez pour m'avoir accueillie au sein de son équipe, pour ses conseils avisés et sa gentillesse tout au long de cette thèse.

Je souhaite, ensuite, exprimer ma reconnaissance aux différents financeurs qui ont permis que cette thèse voie le jour : la région Provence-Alpes-Côte d'Azur, le fonds de dotation de Sauvegarde du Patrimoine Lavandes en Provence (SPLP) avec la participation financière mais aussi technique des sociétés Chanel et Bontoux. Je tiens à adresser mes remerciements les plus sincères à Éric Chaisse directeur du fond SPLP qui a eu l'amabilité de nous faire partager son expérience sur le dépérissement à phytoplasme du Stolbur. Merci pour son aide précieuse. Je tiens aussi à remercier le Centre de Créativité et Innovation en Sciences des Odorants pour sa participation aux financements des analyses UPLC-HRMS. A ce titre, je tiens à remercier Mr Sylvain Antoniotti pour son aide financière, sa gentillesse et son amabilité.

Je remercie chaleureusement tous les partenaires du projet EUCLIDE. Je me sens chanceuse d'avoir pu les rencontrer et travailler à leurs côtés. Je tiens à remercier particulièrement toutes les personnes qui ont participé aux campagnes d'échantillonnage Florence, Thomas C., Thomas M., Émilie et Élisabeth. Je tiens à remercier tout spécifiquement Mme Florence Nicolè pour sa grande gentillesse, pour ses conseils avisés notamment en statistique et toujours avec le sourire !

Je tiens également à exprimer ma gratitude envers Mme Anne-Violette Lavoie et toute son équipe qui m'ont accueillie à bras ouverts au sein de l'INRA de Sophia-Antipolis. Je ne pourrais oublier Nathalie et son éternelle bonne humeur, que je soupçonne d'être parfois venue travailler dans le laboratoire uniquement parce que ça « sentait bon la lavande ». Merci également aux garçons de l'atelier qui m'ont débloqué un nombre incalculable de fois les cellules de broyage. Un grand merci à Anne-Violette pour m'avoir permis de réaliser tous mes broyages et de conserver les échantillons à  $-80^{\circ}\text{C}$ , et je vous prie de m'excuser pour mes piètres qualités de jardinière qui n'ont pas réussi à garder en vie le ginkgo biloba que vous m'aviez si gentiment offert.

J'en profite pour présenter mes excuses aux nombreux pieds de lavande que j'ai fait souffrir durant cette thèse. Merci à ma collègue Émilie Stierlin pour avoir de nombreuses fois palliées mes lacunes en matière de jardinage. Merci à toi pour ton soutien et ta bonne humeur (et tes manalas ^^).

Je tiens à remercier chaleureusement François-Xavier pour son aide, sa rigueur, sa maintenance parfaite du Xevo G2 qTof, et pour avoir supporté mes nombreuses boutades à son égard. Je remercie

également les personnes de Botanicert (Francis, Julien, Laura, Loïc, Mélissa(s), Valentin) pour leur accueil. Je remercie Mr Lionel Massi pour son aide en spectrométrie de masse.

Je tiens à remercier chaleureusement Mr Jean-François Martin pour son aide précieuse sur la plateforme Galaxy-W4M. Sans vous je serais certainement encore bloquée sur ce fichu bug qui a failli avoir raison de mes nerfs. Je remercie également Jody pour m'avoir aidée à écrire le script sous Python.

Je remercie l'Université de m'avoir permis d'avoir cette expérience d'enseignement que j'ai trouvée plaisante et enrichissante. Je remercie tous les enseignants-chercheurs que j'ai pu croiser lors de ces TPs.

Je tiens à remercier tous ceux qui ont contribué à rendre cette thèse plus agréable. Merci à JJC et Loïc pour leurs blagues. Merci à Florence pour avoir été une diva parfaite. Merci aux garçons, Raph et Adrien pour ces fous rires collectifs. « Adriiiiiiiiiien! Cuillère ! ». Et non Raph... je ne rentre pas entière dans un carton, je suis juste petite, pas contorsionniste. Je te propose cependant de réaliser une immersion dans le sud-ouest de la France afin d'améliorer tes compétences picturales en matière de canards. Et bien sûr « Imothep ». Merci à mon amie Jeanne pour les « tours de lac », les pauses thé et nos innombrables conversations. Merci à Anne-Sophie pour sa bonne humeur et sa gentillesse sans faille, qui ont rendu toutes mes visites à Grasse très agréables, surtout ne change pas. Merci à Anne pour sa gentillesse. Merci à Papaye pour son humour et sa positive attitude. Merci aux deux Émilie pour leur gentillesse et leur bonne humeur. Merci à l'équipe MAPEC pour sa bonne humeur et les pauses déjeuners. Merci aux stagiaires qui ont su apporter de la joie et de la bonne humeur dans notre bureau (Coline, Elodie, Djamila, JC, João Mariam, Mathieu, Tara, ...). Je remercie particulièrement Tara pour ses corrections en anglais et pour la belle personne qu'elle est. Merci également à Benji Z. Caro, Conny, Hahn, Haruna, Jana, Manue, Mathilde, ... et toutes les personnes qui ont rendu cette expérience de thèse joyeuse !

Et, je ne peux terminer sans remercier mon compagnon qui a dû supporter un confinement avec une thésarde en fin de thèse ; Anaïs avec qui j'ai pu partager cette expérience qu'est la thèse ; ma petite Audrey, ma sœur de cœur, toujours présente dans les moments difficiles, et ma famille.

# Table des matières

<b>Table des figures</b> .....	<b>15</b>
<b>Liste des tableaux</b> .....	<b>21</b>
<b>Liste des abréviations</b> .....	<b>23</b>
<b>Glossaire</b> .....	<b>24</b>
<b>Introduction générale</b> .....	<b>25</b>
<b>Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse</b> .....	<b>28</b>
<b>I. Contexte de l'étude</b> .....	<b>29</b>
I.A. L'écologie chimique.....	29
I.B. La maladie : le dépérissement .....	30
I.B.1. Le dépérissement .....	30
I.B.2. Le phytoplasme du Stolbur.....	30
I.B.3. La maladie et son impact.....	32
I.B.4. Moyens de lutte .....	33
I.B.5. Critères de sensibilité et tolérances des variétés .....	34
I.C. L'insecte vecteur : la cicadelle .....	35
I.D. Le sujet d'étude : la lavande.....	38
I.D.1. Description botanique et aire de répartition géographique.....	38
I.D.2. Les enjeux économiques .....	41
<b>II. Les composés organiques non volatils (CONVs) dans le genre <i>Lavandula</i></b> .....	<b>42</b>
II.A. Terpénoïdes .....	42
II.B. Flavonoïdes .....	45
II.B.1. Flavones.....	45
II.B.2. Anthocyanidines.....	46
II.B.3. Autres flavonoïdes.....	47
II.C. Acides phénoliques.....	47
II.C.1. Dérivés d'acide hydroxybenzoïques .....	47
II.C.2. Dérivés d'acide hydroxycinnamiques .....	47
II.D. Coumarines .....	49
II.E. Dérivés des benzofuranes.....	49
II.F. Autres composés .....	50
<b>III. La métabolomique</b> .....	<b>51</b>
III.A. Définitions, concept et limite de la métabolomique .....	51

III.A.1. Généralités et définitions .....	51
III.A.2. Approche métabolomique .....	52
III.A.3. Les enjeux et limites de la métabolomique .....	52
a. Facteur biologique et confondant .....	52
b. Limite technique .....	53
III.B. La méthodologie.....	54
III.B.1. Workflow /Design de l'étude.....	54
III.B.2. Echantillonnage .....	55
III.B.3. Préparation des échantillons .....	56
a. Généralités et bonnes pratiques de préparation des échantillons en métabolomique ....	56
b. Cas de l'extraction des CO <sub>n</sub> Vs dans le genre <i>Lavandula</i> .....	56
III.B.4. Analyse des échantillons .....	57
a. Généralités .....	57
b. Bonnes pratiques d'analyse en métabolomique .....	58
c. Cas de la spectrométrie de masse couplée à la chromatographie en phase liquide (LC-MS) .....	59
i. Système chromatographique.....	59
ii. Spectromètre de masse .....	60
III.B.5. Traitement des données.....	63
a. Le prétraitement des données.....	63
i. Généralités.....	63
ii. Cas du prétraitement des données générées par LC-MS .....	63
b. Analyses statistiques .....	68
i. Généralités.....	68
ii. Les analyses multivariées descriptives (non supervisées).....	68
iii. Les analyses multivariées discriminantes (supervisées).....	69
c. Annotation, identifications chimiques des métabolites et réseaux moléculaires .....	70
<b>IV. Objectifs de la thèse .....</b>	<b>74</b>
<b>Chapitre 2 : Conception et optimisation de l'étude métabolomique .....</b>	<b>76</b>
<b>I. Echantillonnage .....</b>	<b>77</b>
I.A. Design de l'échantillonnage .....	77
I.B. Campagne d'échantillonnage .....	79
<b>II. Détermination du solvant d'extraction .....</b>	<b>81</b>
<b>III. Préparation d'échantillon.....</b>	<b>87</b>
III.A. Séchage .....	87
III.A.1. Mise en place .....	87

III.A.2. Résultats.....	89
III.A.3. Conclusion.....	93
III.B. Broyage .....	93
III.C. Mise en place du protocole d'extraction .....	95
III.C.1. Adaptation du protocole : suppression de l'étape d'évaporation et remplacement par une étape de centrifugation .....	96
III.C.2. Suppression filtration et ajout deuxième centrifugation et précipitation à froid .....	98
<b>IV. Développement analytique .....</b>	<b>98</b>
IV.A. Développement préliminaire par HPLC-UV .....	99
IV.A.1. Optimisation conditions analytiques par HPLC pour les échantillons de feuilles .....	99
IV.A.2. Optimisation des conditions analytiques HPLC pour les échantillons d'inflorescence .....	103
IV.B. Evaluation de la variabilité intra plant par HPLC-UV sur des extraits hydro-éthanoliques .....	104
IV.C. Optimisation UPLC-HRMS .....	107
IV.C.1. Optimisation conditions chromatographiques .....	107
a. Modification colonne.....	107
b. Paramètres chromatographiques .....	108
c. Vérification de la séparation des profils en fonction de la variété, de l'état sanitaire du plant et de l'organe .....	109
IV.C.2. Optimisation conditions d'ionisation .....	111
IV.C.3. Séquence analytique .....	111
IV.D. Détermination standard interne .....	112
IV.E. Contrôle de la dérive analytique .....	113
IV.E.1. Test de robustesse du système analytique.....	113
IV.E.2. Contrôle de la dérive analytique au cours des séquences .....	115
a. Suivi stabilité du signal.....	115
b. Dérive temporelle et massique.....	116
<b>V. Design expérimental .....</b>	<b>117</b>
<b>Chapitre 3 : Traitement des données .....</b>	<b>119</b>
<b>I. Extraction des données spectrales .....</b>	<b>120</b>
I.A. Choix du logiciel de retraitement .....	120
I.B. Stratégie adoptée pour la mise en place du <i>workflow</i> .....	120
I.C. Description et mise en place du <i>workflow</i> de prétraitement des données sous Galaxy.....	121
I.C.1. Préparation et entrée des données et de la matrice de design dans le <i>workflow</i> .....	121
I.C.2. Lecture des données (module MSnbase readMSData).....	122
I.C.3. Détection des ions (module findChromPeak ou xcmsSet) .....	122
I.C.4. Groupement des fichiers (module Merger) .....	123

I.C.5. Alignement des ions (module group) .....	124
I.C.6. Corrections des temps de rétention (module adjustRtime) .....	125
I.C.7. Attribution des aires aux pics manquants (module fillPeaks) .....	127
I.C.8. Corrections de la dérive analytique (modules determine batch correction et batch correction) .....	127
I.D. <i>Workflow</i> et paramétrage finals .....	130
<b>II. Corrections et préparation aux analyses statistiques.....</b>	<b>132</b>
II.A. Préparation, téléchargement, vérification et filtration des données .....	132
II.B. Normalisation.....	133
II.B.1. Normalisation par rapport aux échantillons (« Sample Normalization »).....	133
a. Type de normalisation sur les échantillons proposé.....	133
b. Comparaison avec « batch correction » proposé par Galaxy .....	136
II.B.2. Normalisation par rapport aux ions.....	139
a. Réduction des données (« Data scaling ») .....	139
b. Transformation des données (« Data transformation ») .....	141
II.C. Paramétrage final .....	143
<b>III. Nettoyage des données.....</b>	<b>143</b>
III.A. Nettoyage sur les ions .....	143
III.A.1. Suppressions des données redondantes : isotopes, fragments et adduits.....	143
III.A.2. Suppression des contaminants.....	145
a. Extraction à blanc .....	146
b. Suppression des contaminants des matrices de données .....	146
III.A.3. Suppression des variables non significatives .....	148
III.A.4. Impact du nettoyage sur les ions.....	151
III.B. Nettoyage sur les échantillons .....	153
III.B.1. Suppression des données aberrantes .....	153
a. Procédure générale .....	153
b. Analyses supprimées .....	154
<b>IV. Analyses statistiques et obtention des composés discriminants .....</b>	<b>159</b>
IV.A. Statistiques multivariées.....	160
IV.A.1. Analyses descriptives ou exploratoires : ACP .....	160
IV.A.2. Analyses discriminantes : PLS-DA.....	160
IV.B. Biomarqueurs.....	161
IV.B.1. Obtention des biomarqueurs .....	161
IV.B.2. Détermination du caractère discriminant de chaque marqueur .....	162
IV.B.3. Priorisation des biomarqueurs .....	162

a. Détermination des composés cibles .....	163
b. Composés cibles par catégorie .....	163
IV.C. Identification : confrontation aux banques de données .....	164
<b>Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions</b>	<b>166</b>
<b>I. Etude de la tolérance des variétés au dépérissement .....</b>	<b>167</b>
I.A. Résultats des analyses statistiques.....	167
I.A.1. La lavande .....	167
I.A.2. Le lavandin .....	169
I.B. Composés discriminants cibles.....	171
<b>II. Etude de l'influence de l'état sanitaire des plants : caractère symptomatique ou asymptomatique .....</b>	<b>175</b>
II.A. Résultats des analyses statistiques.....	175
II.A.1. Lavande .....	175
a. Variété 7713 .....	175
b. Variété maillette.....	178
II.A.2. Lavandin .....	180
a. Variété abrial .....	180
b. Variété grosso.....	182
II.B. Composés discriminants cibles.....	184
II.B.1. Etude sur les parties chlorophylliennes (Statut F) .....	184
a. Composés cibles en fonction de l'état symptomatique des plants pour les extraits des parties chlorophylliennes .....	184
b. Proposition de structure.....	186
II.B.2. Etude sur les inflorescences (StatutInflo) .....	192
a. Composés cibles en fonction de l'état symptomatique des plants pour les extraits d'inflorescences.....	192
b. Proposition de structure.....	193
II.B.3. Comparaison des deux listes de biomarqueurs cibles en fonction de l'état symptomatique des plants .....	199
<b>III. Discussions et perspectives .....</b>	<b>199</b>
<b>Conclusion générale .....</b>	<b>203</b>
<b>Valorisation des travaux de recherches .....</b>	<b>207</b>
<b>Annexes .....</b>	<b>208</b>
<b>Références bibliographiques .....</b>	<b>274</b>





## Table des figures

Figure 1 : Interactions entre les organismes et composés associés .....	29
Figure 2 : Photographie obtenue par microscopie en transmission d'une coupe de la veine centrale d'une feuille d' <i>Arabidopsis thaliana</i> contenant des phytoplasmes (modifiée d'après (Pagliari et al. 2016)) .....	31
Figure 3 : Evolution de la charge en phytoplasme sur des lavandes (LA) ou des lavandins(LIN) positifs au Stolbur au cours d'une année (D : Décembre 2010, M : Mai , J : Juin, J : Juillet, S : Septembre, O : Octobre, N : Novembre 2011), avec N le nombre d'échantillon. (d'après (Sémétey et al. 2018)) .....	32
Figure 4 : Illustration des 4 stades de dépérissement des plants de lavande et de lavandin (Chatelard et al. 2017) .....	33
Figure 5 : A gauche, essais d'enherbement inter-rang en plein champ. A droite, production de plants sains en serre anti-insecte (photographies CRIEPPAM).....	34
Figure 6 : Bilan des larves capturées (Ho) lors de l'arrachage du bloc n°3 planté lors des essais d'attractivité en 2011 à la station expérimentale du CRIEPPAM (données internes du CRIEPPAM) ....	35
Figure 7 : Photomicrographie de <i>H. obsoletus</i> (Kunz et al. 2010) .....	36
Figure 8 : Schéma du cycle de développement de la cicadelle sur la lavande et le lavandin .....	36
Figure 9 : Schéma du processus de transmission du phytoplasme par <i>H. obsoletus</i> sur la lavande .....	37
Figure 10 : Classification du genre <i>Lavandula</i> .....	38
Figure 11 : Aire de répartition des sections du genre <i>Lavandula</i> (d'après (Guitton 2010)).....	41
Figure 12 : Exemples de plusieurs structures de triterpènes et de dérivés présents dans le genre <i>Lavandula</i> .....	44
Figure 13 : Structure des principales flavones reportées dans le genre <i>Lavandula</i> .....	46
Figure 14 : Exemple d'anthocyanes et anthocyanidines présents dans le genre <i>Lavandula</i> .....	46
Figure 15 : Exemple de structures d'acide hydroxybenzoïque et dérivés reportés dans le genre <i>Lavandula</i> .....	47
Figure 16 : Structures de plusieurs acides hydroxycinnamiques et des dérivés associés reportés dans le genre <i>Lavandula</i> .....	48
Figure 17 : Structure de plusieurs dérivés des coumarines reportés dans le genre <i>Lavandula</i> .....	49
Figure 18 : Structures des différents dérivés de benzofuranes identifiés dans le genre <i>Lavandula</i> .....	50
Figure 19 : Structures d'un composé azoté et d'un gallotanin reportés dans le genre <i>Lavandula</i> .....	50
Figure 20: Schéma généraliste représentant les 4 principales techniques dites « -omics » et leurs cibles d'étude associées .....	51
Figure 21: Histogramme des concentrations moyennes et écart-types du métabolite A des groupes témoins et affectés (d'après (Courant and Giraudeau 2018)) .....	53
Figure 22: Schéma des principales étapes d'une étude métabolomique .....	54
Figure 23 : Schématisation d'un couplage LC-MS (inspiré de (Banerjee and Mazumdar 2012; Bécard 2012)).....	60
Figure 24 : Schéma d'une chambre d'ionisation et du principe d'ionisation pour une source de type ESI (inspiré de (Gaskell 1997; Banerjee and Mazumdar 2012)) .....	61
Figure 25 : Schématisation de la stratégie d'échantillonnage réalisée durant cette étude .....	78
Figure 26 : Localisation des sites d'échantillonnage de lavande et de lavandin .....	79
Figure 27 : Variation de la température (°C) en orange, et du taux d'hygrométrie (%) en bleu, durant les heures d'échantillonnages du 4 et 5 juillet 2017 relevés à la station de Carpentras (Vaucluse, France) situé à 25 km du lieu de prélèvement (Barbery 2017).....	79
Figure 28 : <b>A</b> échantillonnage en champ, <b>B</b> trempe des prélèvements dans l'azote liquide, <b>C</b> et <b>D</b> transport des échantillons dans de la carboglace et boîtes en graphite. ....	81

## Table des figures

Figure 29 : Chromatogrammes HPLC-UV d'extraits de tiges chlorophylliennes et feuilles de lavandin correspondant aux sept solvants testés. Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; volume d'injection de 10 µL ; λ = 280 nm ; T = 25°C. ....	83
Figure 30 : Chromatogrammes HPLC-UV (haut) et DEDL (bas) des extraits eau/éthanol (1/1) des différents organes (tiges chlorophylliennes et feuilles, des parties boisées et des racines) de lavandin grosso. Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C. Paramètres DEDL : pression du gaz de nébulisation (azote) : 3,5 bars ; température du tube d'évaporation 40°C et gain 4. ....	85
Figure 31 : Chromatogrammes HPLC-UV (haut) et DEDL (bas) des extraits acétate d'éthyle des différents organes (tiges chlorophylliennes et feuilles, des parties boisées et des racines) de lavandin grosso. Colonne Luna C18 (Annexe 4) ; gradient lipophile 2 (Annexe 5) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C. Paramètres DEDL : pression du gaz de nébulisation (azote) : 3,5 bars ; température du tube d'évaporation 40°C et gain 4. ....	86
Figure 32 : Schématisation du protocole des tests préliminaires pour déterminer le prétraitement des échantillons .....	89
Figure 33 : Chromatogrammes HPLC-UV des extraits d'eau/éthanol (1/1) de feuilles de lavandin grosso obtenus pour les six conditions testées ; colonne Kinetex PFP (Annexe 4) ; méthode Polaire 2 constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C.....	91
Figure 34 : Graphique de l'analyse en composante principale (ACP) des six conditions de prétraitement testées représentée avec les trois premières composantes (72,98%) .....	92
Figure 35 : Broyeur fabriqué par le CRIEPPAM dans son intégralité à gauche, et à droite la tête de broyage .....	94
Figure 36 : A gauche échantillons obtenus avec un broyeur de type couteau ; au centre échantillons broyés avec le broyeur du CRIEPPAM, à droite échantillon de feuille broyé avec un broyeur à billes. ....	94
Figure 37 : Broyeur à boules MM 301 de chez Retsch en fonctionnement .....	95
Figure 38 : Agrandissements des chromatogrammes HPLC-UV (haut) et DEDL (bas) des extraits eau/éthanol (1/1) des feuilles de lavandin grosso pour différents ratios plante sèche / solvant (20, 30, 40 et 50 mg/mL). Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5), constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C. ....	97
Figure 39 : Chromatogrammes HPLC-UV des extraits eau/éthanol (1/1) des feuilles de lavandin grosso obtenus avec le protocole d'extraction initiale (1) et sa version simplifiée (2). Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5), constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C. ....	98
Figure 40 : Chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso réalisés avec les cinq colonnes testées ; gradients généraliste (Luna C18, Synergi C18 et Synergi C12) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure, et Polaire 1 (Kinetex C18 et Kinetex PFP) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5); débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C. ....	100
Figure 41 : Comparaison des colonnes de type C18 – agrandissement des chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso. Colonne Luna C18, Kinetex C18 et Synergie C18 ; gradients Généraliste (Luna C18, Synergie C18) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure, et Polaire 1 (Kinetex C18) constitué d'eau	

## Table des figures

et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C.....	100
Figure 42 : Comparaison des colonnes ayant des phases stationnaires de nature chimique différentes - agrandissement des chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso. Colonne Synergi C18, Synergi C12 et Kinetex PFP (Annexe 4) ; gradients Généraliste (pour Synergi C18 et Synergi C12) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure, et Polaire 1 (pour Kinetex PFP) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C.	101
Figure 43 : Comparaison des gradients pour la séparation des extraits hydroalcoolique - chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso. Colonne Kinetex PFP (Annexe 4) ; gradients Polaire 1 et Polaire 2 constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C.	102
Figure 44 : Chromatogrammes HPLC-UV des extraits d'eau/éthanol (1/1) de feuilles et de fleurs de lavandin grosso asymptotique ; gradient d'élution Polaire 2 constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; colonne Kinetex PFP (Annexe 4) ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C.....	103
Figure 45 : Chromatogrammes HPLC-UV des extraits d'eau/éthanol (1/1) de fleurs de lavandin grosso asymptotique obtenus pour les trois gradient d'élution Polaire 1, Inflo 1 et 2 ; constitués d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; colonne Kinetex PFP; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C.....	104
Figure 46 : ACP des extraits eau/éthanol (1/1) de feuilles de lavandin grosso, résultats obtenus avec Xlstat à partir des résultats HPLC-UV à 280 nm.....	105
Figure 47 : ACP des extraits eau/éthanol (1/1) d'inflorescence de lavandin grosso, résultats obtenus avec Xlstat à partir des résultats HPLC-UV à 280 nm.....	106
Figure 48 : ACP des extraits eau/éthanol (1/1) d'inflorescence de lavandin grosso de la séquence n°1, résultats obtenus avec Xlstat à partir des résultats HPLC-UV à 280 nm.....	107
Figure 49 : Chromatogrammes UPLC-HRMS d'extraits d'eau/éthanol (1/1) de feuilles de lavande 7713 asymptotique sur les colonnes Kinetex PFP F5 et C18 HSS T3 (Annexe 4) ; gradient d'élution eau et acétonitrile à 0,1% d'FA Annexe 5; débit de 0,35 ml/min ; injection de 1 µL ; T = 40°C. ESI en mode positif.....	108
Figure 50 : Chromatogrammes UPLC-HRMS (BPI = Base Peak Ion) d'extraits d'eau/éthanol (1/1) de feuilles des quatre variétés étudiées (7713, maillette, abrial et grosso) ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; débit de 0,4 ml/min ; colonne C18 HSS T3 (Annexe 4) ; injection de 1 µL ; T = 40°C. ESI en mode positif.....	110
Figure 51 : Agrandissement des chromatogrammes UPLC-HRMS (BPI) d'extraits d'eau/éthanol (1/1) de feuilles d'abrial asymptotique et symptotique, et de feuilles et inflorescences de grosso asymptotique ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; débit de 0,4 ml/min ; colonne C18 HSS T3 (Annexe 4) ; injection de 1 µL ; T = 40°C. ESI en mode positif.....	110
Figure 52 : Chromatogrammes UPLC-HRMS (BPI) d'extraits eau/éthanol (1/1) de feuilles de lavandin grosso asymptotique en mode positif et négatif ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; colonne C18 HSS T3 (Annexe 4) ; débit de 0,4 ml/min ; injection de 1 µL ; T = 40°C. ....	111
Figure 53 : Schématisation d'une séquence d'analyse type.....	112
Figure 54 : Chromatogrammes UPLC-HRMS (BPI = Base Peak Ion) de la caféine à 10 et 50 ppm et d'un extrait eau/éthanol (1/1) de feuilles de grosso asymptotique contenant de la caféine à 5ppm ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; débit de 0,4 ml/min ; colonne C18 HSS T3 (Annexe 4) ; injection de 1 µL ; T = 40°C. ESI en mode positif. ....	113
Figure 55 : Suivi de l'aire du standard interne (caféine) en fonction du nombre d'injection.....	114

## Table des figures

Figure 56 : Suivi de l'aire du standard interne dans les pools au cours de chaque séquence d'analyse .....	115
Figure 57 : Modélisation suivant un polynôme de degrés 2 de l'évolution de l'aire du standard interne suivant l'ordre d'injection pour l'ensemble des séquences d'analyse .....	116
Figure 58 : Agrandissement des chromatogrammes UPLC-HRMS de la première injection et de la dernière injection du pool du lot B17 séquence 1 (à gauche) n'ayant aucune dérive temporelle et de la séquence 2 (à droite) dont les anomalies ont été entourées ou repérées par des flèches ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; colonne C18 HSS T3 (Annexe 4) ; débit de 0,4 ml/min ; injection de 1 µL ; T = 40°C. ESI en mode positif.....	117
Figure 59 : Schématisation du design expérimental utilisé pour l'étude métabolomique non ciblée	117
Figure 60 : Exemple de matrice de design réalisée sous excel .....	122
Figure 61 : Illustration de l'influence du paramètre « mzdifff » sur l'intégration des données.....	123
Figure 62 : Schématisation du fonctionnement de l'alignement des ions (figure adaptée d'après (Martin 2017)).....	124
Figure 63 : Exemple de deux fenêtres de $m/z$ consécutives pour la sortie graphique de l'outil « group » réalisé sur le jeu de données des feuilles de lavande (A17). .....	125
Figure 64 : Chromatogrammes obtenus après la correction des temps de rétention avec la méthode peak groups sur un jeu de données test contenant uniquement des pools des différentes séquences (T01).....	126
Figure 65 : Déviation des temps de rétention en fonction du temps pour le jeu de données des feuilles de lavande (lot A17).....	127
Figure 66 : Sortie graphique de l'outil « determine batch correction » correspondant au standard interne pour un jeu de données réduit contenant quatre séquences d'analyse d'échantillon de feuilles de lavande et de lavandin. ....	128
Figure 67 : Aire du standard interne en fonction de l'ordre d'injection avant et après correction, sur le jeu de données des feuilles de lavande (lot A17) .....	129
Figure 68 : A gauche les données avant rectification, à droite les résultats de la correction de la décroissance du signal intra et inter séquence pour le jeu de données des feuilles de lavande (lot A17) par l'outil « batch correction » de Galaxy .....	130
Figure 69 : Schématisation du <i>workflow</i> du prétraitement des données réalisé sous Galaxy-W4M..	130
Figure 70 : Agencement de la matrice de données adaptée à MetaboAnalyst .....	133
Figure 71 : Graphiques des échantillons obtenus après l'étape de normalisation sous MetaboAnalyst représentant à gauche le jeu de données avant, et à droite après normalisation. En bas, la modélisation de la répartition générale des échantillons en fonction de leurs aires ; et en haut, la distribution individuelle des plusieurs échantillons (matrice feuille-lavande et normalisation avec le standard interne) .....	134
Figure 72 : En haut, les graphiques représentant la distribution générale des données ( <i>features</i> ) ; au milieu, les graphiques représentant la distribution générale des échantillons ; et en bas, les graphiques des individus des ACP obtenus pour chacune des trois méthodes testées (normalisation par les pools, par le standard interne ou normalisation quantile) de MetaboAnalyst (matrice des variétés des feuilles de lavande, autres paramètres : normalisation automatique et transformation logarithmique).....	135
Figure 73 : Graphiques des échantillons, à gauche après une normalisation sur les pools sous MetaboAnalyst et à droite après l'application de l'outil « Batch correction » de Galaxy pour le jeu de données feuille-lavande (A17). Dans les deux cas, une transformation logarithmique et une normalisation automatique ont été appliquées.....	137
Figure 74 : Les analyses statistiques (ACP et PLS-DA) obtenues après correction de la dérive analytique sous MetaboAnalyst à gauche et sous Galaxy à droite. ....	138

## Table des figures

Figure 75 : Graphiques des échantillons des ACP pour les différentes méthodes de réduction des données de MetaboAnalyst sans transformation (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les pools). Le pourcentage entre parenthèse correspond à la variance totale exprimée par les deux premiers axes.....	140
Figure 76 : Graphiques des échantillons des ACP avec et sans transformation logarithmique avec centrage des données de MetaboAnalyst sans transformation (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les pools). Le pourcentage entre parenthèse correspond à la variance totale exprimée par les deux premiers axes.....	140
Figure 77 : Graphique représentant la distribution générale des données (features) obtenue pour chacune des méthodes de réduction des données de MetaboAnalyst (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les pools et transformation logarithmique). .....	141
Figure 78 : Tests des différentes transformations disponibles sous MetaboAnalyst (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les <i>pools</i> et <i>l'auto-scaling</i> ). .....	142
Figure 79 : Matrice de données prête pour la suppression des isotopes .....	145
Figure 80 : Chromatogrammes UPLC-HRMS de l'extraction à blanc, des solvant d'élution et d'extraction ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; colonne C18 HSS T3 (Annexe 4) ; débit de 0,4 ml/min ; injection de 1 µL ; T = 40°C. ESI en mode positif .....	146
Figure 81 : Illustration de la banque de contaminant élaborée.....	147
Figure 82 : Exemple d'annotation réalisée par Galaxy en utilisant la banque de contaminant (matrice inflorescence-lavande) .....	148
Figure 83 : Logigramme du processus utilisé pour la suppression des ions non significatifs.....	150
Figure 84 : Graphique des p-value obtenus pour chaque ion à l'issue du t-test pour la matrice des extraits d'inflorescence de lavandin et deux exemples de boîtes à moustaches pour un ion significatif et pour un ion non significatif. ....	151
Figure 85 : Exemple de l'influence de la suppression des ions lors du nettoyage sur la matrice des extraits de feuille de lavande. En haut, les graphiques des échantillons ( <i>score plot</i> ) ; en bas, les graphiques des ions ( <i>loading plot</i> ) ; à gauche, avant nettoyage ; et à droite, après suppression des ions.....	152
Figure 86 : ACP des matrices de feuilles (à gauche) et d'inflorescences (à droite) de lavande en fonction des variétés avant suppression des échantillons aberrants .....	154
Figure 87 : ACP de la matrice d'inflorescence de lavande en fonction des variétés (à gauche) et de la matrice d'inflorescence de maillette en fonction de l'état symptomatique du plant (à droite) avant suppression des échantillons aberrants .....	155
Figure 88 : ACP de la matrice d'inflorescence de lavande en fonction des variétés avant (à gauche) et après (à droite) suppression des échantillons aberrants.....	156
Figure 89 : ACP de la matrice d'inflorescence de lavandin en fonction des variétés (à gauche) et de la matrice d'inflorescence de grosso en fonction de l'état symptomatique du plant (à droite) avant suppression des échantillons aberrants .....	156
Figure 90 : ACP de la matrice de feuille de lavandin en fonction des variétés (à gauche) et de la matrice de feuille de grosso en fonction de l'état symptomatique du plant (à droite) avant suppression des échantillons aberrants .....	157
Figure 91 : ACP de la matrice de feuille (à gauche) et d'inflorescence d'abrial (à droite) en fonction de l'état symptomatique du plant avant suppression des échantillons aberrants .....	158
Figure 92 : Illustration du sous- et sur-apprentissage d'un modèle .....	160
Figure 93 : Deux exemples d'histogrammes obtenus à l'issue de tests de permutation effectués sous Metaboanalyst. A gauche, un test valide ; et à droite, un test ayant un sur-apprentissage du modèle. .....	161

## Table des figures

Figure 94 : Exemple d'un graphique contenant les ions les plus discriminants (en ordonnées) en fonction de leur VIP score (en abscisse) (réalisé avec Metaboanalyst) .....	162
Figure 95 : Boîtes à moustaches des composés discriminants cibles en fonction du caractère sensible (en rouge) ou tolérant (en vert) des variétés .....	172
Figure 96 : Comparaison entre le spectre MSe expérimental obtenu pour l'ion M361T283, en haut, et le spectre MS/MS de l'acide rosmarinique (Matsuda), en bas. ....	173
Figure 97 : Mise en évidence des analyses correspondant aux échantillons lyophilisés pendant le cycle n°5, sur les ACP de la matrice de feuille (à gauche) et d'inflorescence (à droite) de lavande 7713 en fonction de l'état symptomatique du plant après suppression des autres échantillons aberrants ....	177
Figure 98 : Boîtes à moustaches des composés discriminants cibles en fonction de l'état sanitaire (symptomatique en rouge et asymptotique en vert) des plants pour les parties chlorophylliennes .....	186
Figure 99 : Chromatogramme d'un extrait des organes chlorophylliens de lavandin annoté avec les composés cibles discriminants de l'état symptomatique des plants numérotés en fonction de leur pertinence ( <i>i.e.</i> en fonction de leur note puis de leur classement) .....	187
Figure 100 : Spectres UV nettoyés et MS de l'ion M593T819 dont la proposition de structure est celle du phéophorbide A .....	188
Figure 101: Structure des trois dérivés de chlorophylle potentiellement identifiés dans les extraits de feuilles.....	188
Figure 102 : Spectre MS/MS de la scopolétine obtenu en ESI en mode positive (Zeng et al. 2015) ...	189
Figure 103 : Boîtes à moustaches des composés discriminants cibles en fonction de l'état sanitaire des plants (symptomatique en rouge et asymptotique en vert) pour les inflorescences de lavande et de lavandin.....	193
Figure 104 : Chromatogramme des inflorescences de lavandin annoté avec les composés cibles discriminants de l'état symptomatique des plants. (A) O-coumaric acid; (B) Methyl caffeate; (C) Luteolin di-glucuronide; (D) Ferulic acid; (E) Caffeic acid; (F) Coumarin; (G) Herniarin .....	194
Figure 105 : Proposition de fragmentation de la lutéoline 7-O-glucuronide. ....	195
Figure 106 : Agrandissement du spectre de masse du pic correspondant à l'ion cible M262T571 ....	195
Figure 107 : A gauche, agrandissement du spectre MS annoté du pic présent à 8,52 min correspondant à l'étude des ions discriminants M169T511 et M109T511 ; et à droite, agrandissement du spectre MSe du même composé. ....	196
Figure 108 : De gauche à droite, photographie du bain à ultra-sons lors de la réalisation de deux jeux d'extraction en simultané, photographie d'une plaque d'extraction vu de dessous à droite, pool des extraits de feuille de lavande, et pool des extraits d'inflorescences de lavande. ....	248
Figure 109 : A gauche, <i>boxplot</i> de deux ions non significatifs ; à droite, <i>boxplot</i> de deux ions significatifs .....	252
Figure 110 : Tableur type préparé pour réaliser un test de normalité sous R .....	253

## Liste des tableaux

Tableau 1 : Système de notation visuel des symptômes dus au dépérissement pour la lavande et le lavandin (d'après (Stierlin 2020)) .....	33
Tableau 2 : Taxonomie du genre <i>Lavandula</i> d'après (Upson and Andrews 2004) .....	39
Tableau 3 : Procédés d'extraction et méthodes analytique employés dans le genre <i>Lavandula</i> .....	57
Tableau 4 : Les principales étapes de prétraitement pour une étude métabolomique en LC-MS.....	66
Tableau 5 : Plusieurs exemples de logiciels de traitement des données LC-MS en métabolomique (d'après (Spicer et al. 2017)) .....	67
Tableau 6 : Bases de données adaptées aux données MS et aux produits naturels (d'après (Wolfender et al. 2015)) .....	71
Tableau 7 : Relation entre les solvants envisagée et leur index de polarité (P') selon l'échelle de Snyder (Snyder 1974; Markom et al. 2007) .....	82
Tableau 8 : Rendements d'extraction obtenus en fonction des organes de la plante et du solvant d'extraction .....	84
Tableau 9 : Suivi de la perte massique pour les conditions contenant une étape de séchage .....	90
Tableau 10 : Moyenne et écart-type du nombre de pics et des aires des répliqués des six conditions de prétraitement testées .....	92
Tableau 11 : Durées et fréquences de broyage employées pour l'étude métabolomique selon le type d'échantillon .....	95
Tableau 12 : Capacité totale et nombre de pics pour les différentes colonnes testées pour la séparation des extraits polaires .....	102
Tableau 13 : Spécificités de la colonne ACQUITY UPLC HSS T3 .....	108
Tableau 14 : Echantillons utilisés pour la mise en place des conditions analytiques en UPLC-HRMS.	109
Tableau 15 : Le classement des séquences en fonction de la différence maximale d'aire de la caféine dans les pools sur l'intégralité de chaque séquence .....	115
Tableau 16 : Correspondances entre les lots d'analyse et d'extraction et les caractéristiques biologiques (organes de la plante et espèces).....	121
Tableau 17 : Paramètres finaux obtenus pour le traitement spectral des données sous Galaxy-W4M .....	130
Tableau 18 : Correction et paramétrages appliqués pour effectuer la comparaison entre les deux plateformes sur la correction de la dérive analytique .....	136
Tableau 19 : Méthodes de réduction des données proposées par MetaboAnalyst .....	139
Tableau 20 : Paramètres finaux obtenus pour la correction et la préparation des données aux analyses statistiques sous MetaboAnalyst .....	143
Tableau 21 : Paramétrage de l'outil « CAMERA.annotate » .....	144
Tableau 22 : Paramétrage de l'outil d'annotation banque de données interne de Galaxy .....	148
Tableau 23 : Dénomination des douze matrices conservées et leurs caractéristiques .....	149
Tableau 24 : Nombre d'ions avant et après le nettoyage des données .....	153
Tableau 25 : Récapitulatif des échantillons aberrants supprimés par matrice sur l'ensemble des jeux de données.....	159
Tableau 26 : Nombre d'ions contenus dans chaque matrice après nettoyage .....	162
Tableau 27 : Catégories d'étude et matrices associées .....	163
Tableau 28 : Nombre de composés discriminants en fonction de leur nombre d'apparition pour chaque catégorie.....	164
Tableau 29 : Analyses statistiques (ACP et PLS-DA) obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite de lavande. ....	168

## Liste des tableaux

Tableau 30 : Analyses statistiques (ACP et PLS-DA) obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite de lavandin .....	170
Tableau 31 : Composés discriminants en fonction du caractère sensible ou tolérant des variétés communs aux jeux de données correspondant aux extraits de feuille.....	171
Tableau 32 : Hypothèses d'identification pour les composés cibles discriminants du caractère sensible ou tolérant des plants contenus dans les extraits des parties chlorophylliennes .....	174
Tableau 33 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavande 7713 .....	176
Tableau 34 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavande maillette.....	179
Tableau 35 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavandin abrial .....	181
Tableau 36 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavandin grosso .....	183
Tableau 37 : Composés discriminants en fonction de l'état symptomatique des plants commun aux jeux de données correspondant aux extraits de feuilles .....	185
Tableau 38 : Hypothèses d'identification pour les composés cibles contenus dans les extraits des parties chlorophylliennes .....	190
Tableau 39 : Composés discriminants en fonction de l'état symptomatique des plants commun aux jeux de données correspondant aux extraits d'inflorescence .....	192
Tableau 40 : Hypothèses d'identification pour les composés cibles contenus dans les extraits d'inflorescences .....	197
Tableau 41 : Correspondance entre les extraits (solvant d'extraction) et les solvants pour solubiliser les extraits avant analyse. ....	241



# Liste des abréviations

- AcN : Acétonitrile
- ACP : Analyse en Composante Principale (*Principal Component Analysis = PCA*)
- DEDL : Détecteur Évaporatif à Diffusion de la Lumière (*Evaporative Light Scattering Detectors = ELSD*)
- ESI : Électronébulisation (*ElectroSpray Ionisation*)
- FA : Acide formique
- GC : Chromatographique en phase gazeuse (*Gaz Chromatography*)
- HPLC : Chromatographie en phase liquide à pression élevée (*High Performance Liquid Chromatography ou High Pressure Liquid Chromatography*)
- iPrOH : Isopropanol
- IR : Infra-rouge
- LC : Chromatographique en phase liquide (*Liquid Chromatography*)
- MAE : Extraction assistée par micro-ondes (*Microwave Assisted Extraction*)
- MS : Spectrométrie de Masse (*Mass Spectrometry*)
- PLE : Extraction par fluide pressurisée (*Pressurized Liquid Extraction*)
- PLSDA : Analyse discriminante utilisant la méthode de régression partielle des moindres carrés (*Partial Least Squares-Discriminant Analysis*)
- RMN : Résonance magnétique nucléaire
- SFE : Extraction par fluide supercritique (*Supercritical Fluid Extraction*)
- THF : Tétrahydrofurane
- UAE : Extraction assistée par ultra-sons (*Ultrasonic Assisted Extraction*)
- UPLC : Chromatographique en phase liquide à pression très élevée de chez Waters (*Ultra Performance Liquid Chromatography*)
- UV : Ultra-violet
- VIP : Importance de la variable sur la projection (*Variable Importance in Projection*)

# Glossaire

**Allomone** : composé allélochimique qui induit une réponse bénéfique uniquement pour l'organisme émetteur.

**Composé allélochimique** : substance produite par un organisme qui intervient dans les interactions interspécifiques. Elle induit chez l'espèce réceptrice une réponse comportementale ou physiologique.

**Interaction abiotique** : interaction entre un organisme et son environnement (non vivant).

**Interaction biotique** : interaction entre les êtres vivants.

**Kairomone** : composé allélochimique qui induit une réponse bénéfique uniquement pour l'organisme récepteur.

**Métabolite** : molécule de faible poids moléculaire, c'est-à-dire compris entre 50 et 1500 Da, contenue dans un organisme vivant.

**Métabolome** : ensemble des métabolites.

**Métabolomique** : ensemble des démarches analytiques permettant d'analyser le métabolome d'un système vivant (tissus, cellules, organes, ...).

**Phéromone** : substance produite par un organisme qui intervient dans les interactions intraspécifiques, c'est-à-dire entre sujet d'une même espèce.

**Puissance statistique** : est la probabilité que l'hypothèse nulle soit rejetée alors qu'elle est vraie. En d'autres termes, elle correspond à la probabilité d'obtenir un résultat statistiquement significatif. Elle dépend de plusieurs critères tels que la taille de l'échantillonnage ou l'intensité du phénomène à observer. (Cohen 1992)

**Synomone** : composé allélochimique qui induit une réponse bénéfique pour les deux organismes (émetteur et récepteur).

# Introduction générale

La lavande (*Lavandula angustifolia* Miller) et le lavandin (*Lavandula x intermedia*) sont des plantes emblématiques et historiques de la Provence et de la Haute Provence. Elles y sont cultivées principalement pour la production d'huile essentielle qui intéresse aussi bien les secteurs de la parfumerie que de la cosmétique. La France est le deuxième producteur mondial d'huile essentielle de lavande, et le premier pour le lavandin avec environ 1400 tonnes produites en 2016 (Giray 2018).

Néanmoins, depuis une vingtaine d'années, les cultures de lavande française sont affectées par une maladie communément dénommée « dépérissement ». Elle a des conséquences économiques importantes sur la filière lavandicole car elle engendre une diminution des rendements des plants et une réduction de la durée d'exploitation des champs. Ce déclin est lié aux effets de la maladie et des événements de sécheresse climatique répétés et intenses. Le principal agent responsable de ce dépérissement a pu être identifié comme le phytoplasme du Stolbur (*Candidatus Phytoplasma solani*) (Sémétey et al. 2018) une bactérie sans paroi se développant dans les tubes criblés de la plante. Ce pathogène est transmis par un insecte vecteur : la cicadelle (*Hyalesthes obsoletus*). Cet insecte piqueur-suceur se nourrit de la sève de la lavande transmettant la bactérie de plant infecté en plant sain (Yvin 2010). Le phytoplasme vient ensuite progressivement obstruer les vaisseaux conducteurs de sève élaborée (Pagliari et al. 2016) induisant un assèchement et un jaunissement des pieds, menant à la mort prématurée des plantes (Sémétey et al. 2018).

A l'heure actuelle, les moyens de lutte contre cette maladie sont inexistantes ou restreints. En effet, les antibiotiques autorisés en France ne sont pas efficaces contre le phytoplasme. D'autre part, l'utilisation d'insecticides contre la cicadelle n'est pas concevable du fait de la présence d'abeilles. De plus, des traces d'insecticides pourraient être retrouvées dans les huiles essentielles ou le miel de lavande, ce qui est inenvisageable.

Par ailleurs, beaucoup de zones d'ombre perdurent autour de cette problématique et notamment sur les interactions entre la cicadelle, le phytoplasme et la lavande. Ces inconnues ont conduit les acteurs du secteur à lancer un vaste projet de recherche intitulé EUCLIDE. Cette étude s'inscrit dans ce projet, comprenant plusieurs partenaires et faisant intervenir de nombreuses disciplines (agronomie, écologie, biologie, biostatistique, chimie, ...). Les différents acteurs du projet sont : le Centre Régionalisé Interprofessionnel d'Expérimentation en Plantes à Parfum, Aromatiques et Médicinales (CRIEPPAM) implanté à Manosque, le Laboratoire de Biotechnologies Végétales appliquées aux Plantes Aromatiques et Médicinales (LBVpam) de l'université Jean-Monnet de St-Étienne, le Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) basé à Montpellier, l'INRA d'Avignon, ainsi que l'équipe Métabolome et Valorisation de la Biodiversité Végétale (MVBV) de l'Institut de Chimie de Nice (ICN) au sein de laquelle a été réalisée cette thèse et celle de Mlle Émilie Stierlin. Cette thèse a pu voir le jour grâce aux financements de la région Provence-Alpes-Côte d'Azur et du fonds de dotation « Sauvegarde du Patrimoine Lavandes en Provence » (Fonds SPLP) soutenu par les sociétés Chanel et Bontoux.

Les actions du projet se déclinent en cinq points majeurs répartis entre les différents partenaires :

- Action 1 : étude des composés organiques volatils (COVs) émis et stockés par des plantes saines/infectées par le phytoplasme du Stolbur (thèse d'Émilie Stierlin)
- Action 2 : étude métabolomique des Composés Organiques non Volatils (CONVs), c'est-à-dire la caractérisation des molécules impliquées dans les phénomènes de défenses induites à la suite de l'infection du phytoplasme
- Action 3 : étude de la perception olfactive des composés volatiles par *H. obsoletus* (LBVpam, CEFE et CRIEPPAM)

## Introduction générale

- Action 4 : étude de la perception des composés volatiles et non-volatiles par *H. obsoletus* via des tests comportementaux (CRIEPPAM et LBVpam)
- Action 5 : étude de l'expression des gènes en fonction du statut des plantes saines/infectées (LBVpam)

Cette thèse rentre dans le cadre de l'action numéro 2. Le principal objectif a été de mettre en place une étude métabolomique permettant d'étudier les composés non volatils du métabolome de la lavande et le lavandin. Pour ce faire, il a été nécessaire :

- De mettre au point une méthode de préparation des échantillons, d'extraction et d'analyse des composés non volatils présents dans la lavande et le lavandin.
- De réaliser une étude métabolique, permettant de mettre en avant les différences de composition chimique qualitative et quantitative entre les plants symptomatiques et asymptomatiques des variétés étudiées (maillette et 77-13 pour la lavande, et abrial et grosso, pour le lavandin).
- De mettre en avant des différences de composition chimique entre les variétés dites « sensible » (77-13 et abrial) et « tolérante » (maillette et grosso)

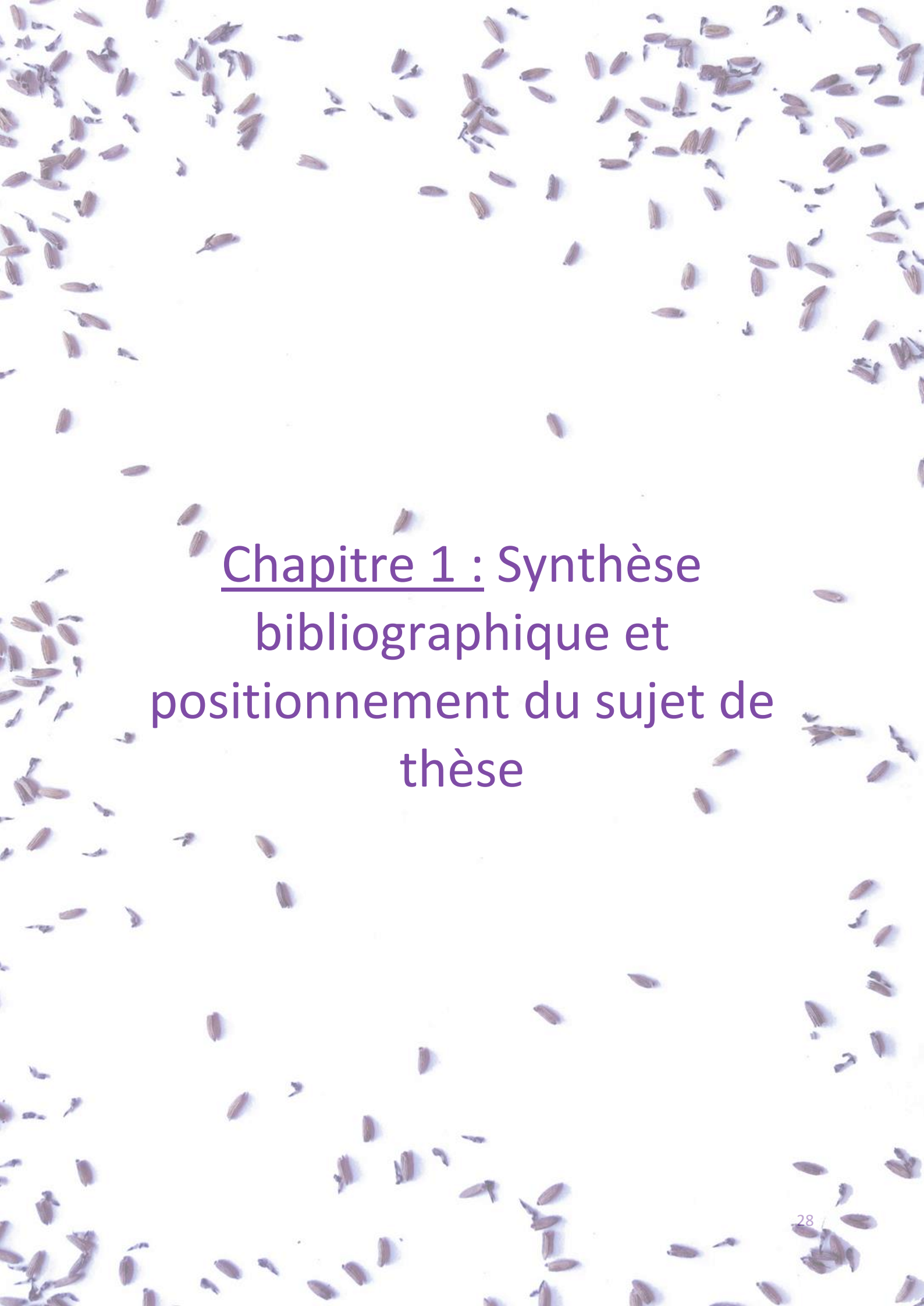
Dans un premier chapitre, seront présentés le contexte de l'étude avec une présentation des trois intervenants dans la problématique du dépérissement de la lavande à savoir le pathogène (le phytoplasme du Stolbur), l'insecte vecteur (la cicadelle) et enfin la plante étudiée (la lavande et le lavandin). Dans une seconde partie, nous nous intéresserons à la richesse et la diversité des composés contenus dans le genre *Lavandula* qui a donné lieu à une banque de données et à la publication d'une revue sur le sujet. Dans un troisième temps, nous nous attarderons sur les fondamentaux de la métabolomique. Nous verrons ce qu'est une étude métabolomique, quelles sont les différentes étapes qui la composent, les biais qui peuvent être introduits à travers chaque étape et les bonnes pratiques à mettre en place pour les éviter. Nous détaillerons également les principes de fonctionnement des instruments analytiques employés durant cette étude. Enfin, dans une dernière partie, nous exposerons les différents objectifs de cette thèse.

Dans un second chapitre, nous verrons à la fois la conception et la réalisation de la partie expérimentale de l'étude jusqu'à l'étape d'analyse des échantillons. Dans une première partie seront présentées le design et le protocole d'échantillonnage. Dans un second temps, nous verrons la détermination du protocole de préparation des échantillons (procédé, solvant, conservation et broyage des échantillons). Par la suite, nous nous attarderons sur le développement analytique avec des tests préliminaires par HPLC-UV puis une optimisation en UPLC-HRMS. Nous aborderons également dans cette partie le contrôle de la dérive analytique face au nombre important d'échantillons analysés (1429). Dans une dernière partie sera présentée le design expérimental final de l'étude.

Le troisième chapitre portera sur le développement méthodologique du traitement des données. La première partie sera consacrée à l'extraction des données spectrales réalisée sur la plateforme en ligne W4M-Galaxy. Par la suite, nous verrons la correction et la préparation des différents jeux de données aux analyses statistiques. Cette partie a été effectuée sur une deuxième plateforme dénommée MetaboAnalyst. Les données ont également été nettoyées avec, entre autres, la suppression des contaminants ou des données telles que les isotopes ou les adduits. Enfin, dans une dernière partie seront abordées les analyses statistiques multivariées employées.

Dans un quatrième chapitre, nous présenterons et discuterons des résultats obtenus à l'issue de cette étude métabolomique non ciblée, ainsi que des perspectives du projet.





Chapitre 1 : Synthèse  
bibliographique et  
positionnement du sujet de  
thèse

## I. Contexte de l'étude

### I.A. L'écologie chimique

Les organismes vivants, quels qu'ils soient, communiquent entre eux. Un des moyens de communiquer est la médiation chimique. Ce mode de communication a longtemps été sous-évalué et une science s'applique désormais à l'étudier. Cette science transdisciplinaire est l'écologie chimique. Elle se situe à la frontière des deux disciplines qu'elle regroupe et fait intervenir à la fois des biologistes, des écologues, des chimistes, des bio-informaticiens, ... Cette science récente s'est développée avec les progrès de la chimie analytique et a mené à une nouvelle vision plus complexe et encore méconnue des réseaux de communication, des interactions et des symbioses pouvant intervenir dans un écosystème (Bagnères and Hossaert-Mckey 2016).

De nombreuses interactions peuvent intervenir (voir Figure 1), elles sont classifiées selon les types suivants (Francis 2003):

- Les interactions biotiques. Elles correspondent aux interactions entre être vivants et sont regroupées en deux sous-catégories :
  - Intraspécifiques : entre sujets d'une même espèce. Les molécules associées à ce type d'interactions sont nommées phéromones.
  - Interspécifiques : entre des sujets d'espèces différentes. Dans ce cas, on parle de composés allélochimiques. Il en existe trois types définis selon leur impact sur l'émetteur et le récepteur du signal chimique (allomone, kairomone et synomone respectivement ayant un effet positif pour l'émetteur et négatif pour l'émetteur, un effet négatif pour l'émetteur et un effet positif pour l'émetteur et le récepteur), et enfin un effet positif pour les deux intervenants.
- Les interactions abiotiques, sont celles intervenant entre l'organisme et son environnement (non vivant). Un exemple de stress abiotique est le stress hydrique subi par une plante (Beck et al. 2018).

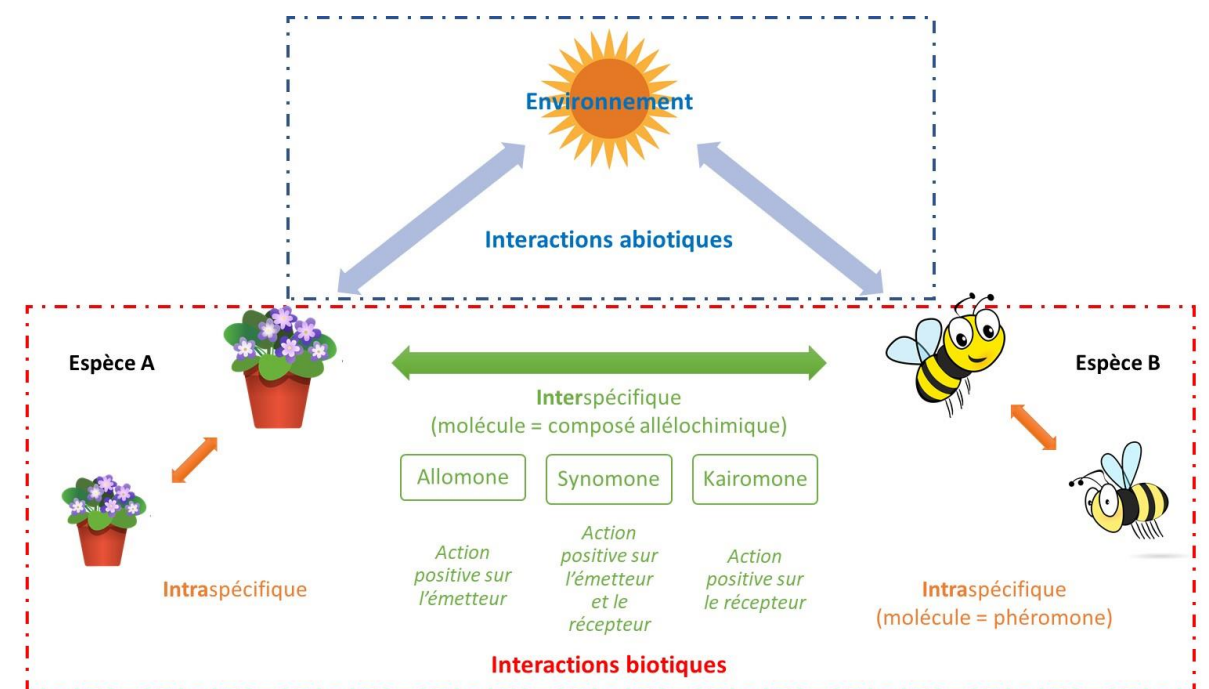


Figure 1 : Interactions entre les organismes et composés associés

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

Le signal chimique impliqué dans ces interactions peut être quantitatif ou qualitatif (activation ou inhibition d'une voie biochimique) et peut dépendre d'une seule molécule ou d'un mélange plus complexe contenant plusieurs composés (avec ou sans seuil de concentration limite). Les sémiochimiques sont généralement actifs à de très faible dose. Néanmoins, le signal doit être suffisamment élevé pour être détecté par le récepteur malgré « le bruit de fond » émis par l'environnement. (Bagnères and Hossaert-McKey 2016)

L'intégralité des rôles des médiateurs chimiques est certainement très vaste et méconnue à ce jour. Cependant, nous savons d'ores et déjà qu'ils jouent un rôle important dans la pollinisation des plantes, dans la prédation (émission de signaux d'alarme pour les congénères, émission de toxines pour se défendre face à l'attaque d'un ravageur). Le message transmis peut être aussi bien à caractère attractifs ou répulsifs. Ainsi certaines variétés d'arum sont bien connues pour émettre une odeur de putréfaction qui attire les mouches à l'intérieur des fleurs et permettent leur pollinisation. Dans un registre défensif, l'*Epipactis helleborine* lorsqu'elle est attaquée par un herbivore (puceron) émet un bouquet de composés volatils (octanal, acétate d'hexyle, hexèn-1-ol) qui attire les guêpes, prédateur naturel du puceron. (Nishida 2014; Bagnères and Hossaert-McKey 2016)

L'écologie chimique pourrait modifier nos pratiques culturales en offrant de nouveaux moyens de lutte grâce à une meilleure compréhension des interactions entre les plantes, les insectes et les microbes (Beck et al. 2018). Dans cette thèse nous nous intéresserons à l'étude des signaux chimiques potentiellement impliqués dans une maladie bactérienne : le dépérissement de la lavande.

### I.B. La maladie : le dépérissement

#### I.B.1. Le dépérissement

Dès les années 1920, des premiers signalements mentionnent un dépérissement des lavandes, mais c'est dans les années 1980 que le phénomène devient préoccupant en France. Dans les premiers temps, cette maladie est désignée sous le nom de « dépérissement ». Ce terme est utilisé de façon large, abusive et désigne « un état des plants conduisant à une perte de leur longévité ». En réalité, il dissimule plusieurs causes d'affaiblissements des plants telles que le gel ou l'impact de la cécidomyie (*Resseliella lavandulae* qui est un insecte ravageur). (Yvin 2010)

En 1995, *Candidatus Phytoplasma solani* appelé également phytoplasme du Stolbur est détecté pour la première fois dans un plant symptomatique et dans une cicadelle (Yvin 2010). Il sera par la suite formellement identifié comme étant le facteur responsable du dépérissement de la lavande et du lavandin en France (Sémétey et al. 2018). Les spécialistes parlent désormais de « dépérissement à phytoplasme » pour les différencier des autres causes de mortalité (Yvin 2010).

Cependant, le nombre de plants dépérissant négatifs au phytoplasme du Stolbur a conduit les chercheurs à suspecter l'existence d'autres agents pathogènes. Des études complémentaires ont été menées dans ce sens afin de rechercher la présence d'autres phytoplasmes et bactéries, sans succès. La présence d'autres phytoplasmes a été testée sur 71 échantillons correspondants à des plants symptomatiques mais négatifs au Stolbur. Seulement, deux d'entre eux ce sont avérés positifs. Le phytoplasme en question a été identifié comme étant le « *Candidatus Phytoplasma trifolii* ». La flore bactérienne de la lavande s'est avérée riche mais aucune différence significative entre les plants symptomatiques et asymptomatiques n'a pu être mise en avant. (Sémétey et al. 2018)

#### I.B.2. Le phytoplasme du Stolbur

Les phytoplasmes sont des bactéries sans paroi parasitant de nombreux végétaux (Figure 2) et se développant uniquement dans le phloème (les tissus conducteurs de sève élaborée) (Gasparich 2010; Yvin 2010). Ils sont responsables de nombreuses maladies telles que la « maladie du Bois Noir » chez la vigne (Yvin et al. 2018).



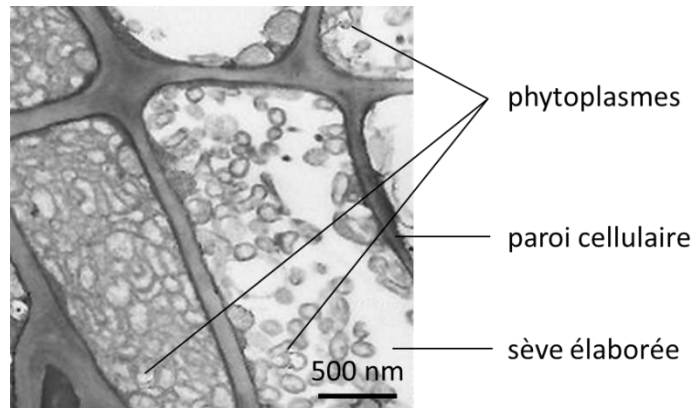


Figure 2 : Photographie obtenue par microscopie en transmission d'une coupe de la veine centrale d'une feuille d'*Arabidopsis thaliana* contenant des phytoplasmes (modifiée d'après (Pagliari et al. 2016))

La détection des phytoplasmes se fait par PCR (*Polymerase Chain Reaction*) (Gasparich 2010; Yvin 2010). Ce test permet d'amplifier un génome cible en multipliant l'ADN et ainsi déceler les fragments caractéristiques de la bactérie (Croville 2017). Plusieurs souches de phytoplasme du Stolbur ont été détectées sur les lavandes françaises. Précisément, dix-sept génotypes ont été mis en évidence dans des plants infectés dont trois sont majoritaires (S14, S16 et S17). Cette grande diversité de génotypes laisse penser que le pathogène est présent depuis longtemps dans les lavanderaies françaises (Yvin 2010). En particulier, S14 et S16 qui appartiendraient à d'anciennes épidémies. La proportion de ces deux souches décline au profit de S17 qui est actuellement la souche majoritaire dans les plants et dans la cicadelle.

La prédominance des génotypes S14, S16 et S17 s'accroît au fil des années alors qu'ils ne sont pas ou peu présents dans les autres cultures en Europe. Ils n'ont pas été non plus détectés dans le liseron présent à proximité des champs alors qu'il constitue un réservoir naturel de cicadelle. Le génotype S14 proviendrait d'Europe de l'ouest et serait transmis par bouturage. A l'inverse, les génotypes S17 et S16 auraient évolué localement et seraient des souches spécifiquement françaises. (Sémétey et al. 2018)

#### **Charge en phytoplasme :**

Des études ont montré que la charge en phytoplasme dans les cultures de lavande varie au cours de l'année et atteint un maximum durant l'automne (Figure 3). Les raisons de ces fluctuations sont encore inconnues. (Sémétey et al. 2018)

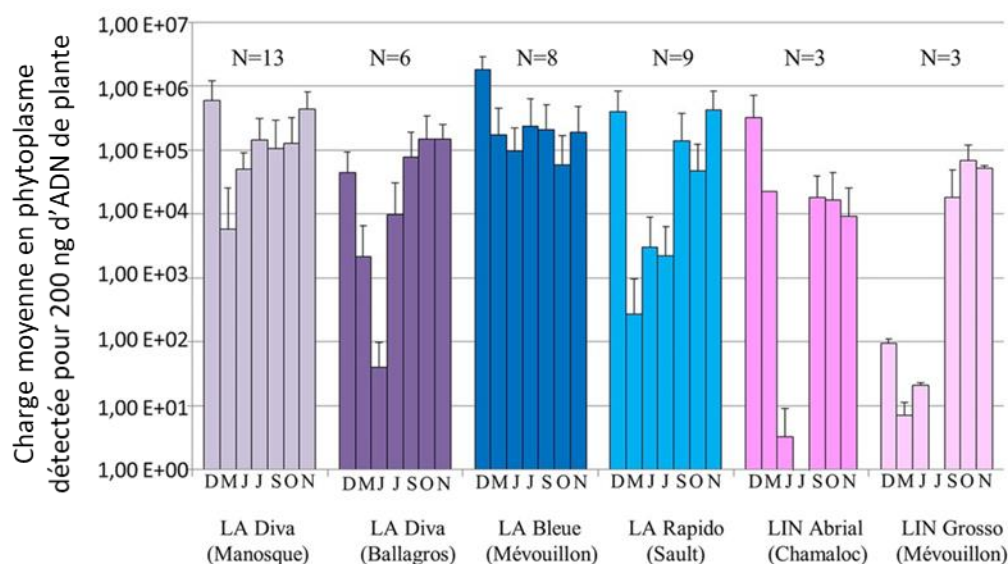


Figure 3 : Evolution de la charge en phytoplasme sur des lavandes (LA) ou des lavandins (LIN) positifs au Stolbur au cours d'une année (D : Décembre 2010, M : Mai, J : Juin, J : Juillet, S : Septembre, O : Octobre, N : Novembre 2011), avec N le nombre d'échantillon. (d'après (Sémétey et al. 2018))

De plus, l'aggravation de la sévérité des symptômes de dépérissement affectant un pied de lavande est corrélée avec l'augmentation de la charge en phytoplasme porté par le plant (CRIEPPAM 2013; Sémétey et al. 2018).

#### Transmission :

Les phytoplasmes sont incapables de survivre sans un hôte (Yvin 2010). Leur transmission se fait donc soit directement par le matériel végétal (par exemple par un bouturage), soit par un vecteur. Ces transmetteurs sont des insectes piqueurs-suceurs appartenant à la famille des Cicadellidae (les cicadelles) ou plus rarement à celle des Fulgoridae (Gasparich 2010).

Dans le cas du phytoplasme du Stolbur, la cicadelle (*cixiide Hyalesthes obsoletus*) (voir description partie I.C ci-dessous) est suspectée d'être le vecteur principal dès 1996. Cette hypothèse sera confirmée par la suite (Yvin 2010). Les scientifiques suspectent l'existence d'autres insectes vecteurs car la période de contamination des plantes est supérieure à la période de vol de *H. obsoletus*. Un autre insecte *C. quadrinotatus* a effectivement été identifié comme vecteur du Stolbur mais de façon très minoritaire (Yvin 2010).

#### I.B.3. La maladie et son impact

Le dépérissement à phytoplasme constitue aujourd'hui le principal problème sanitaire pour la lavande et le lavandin (Yvin et al. 2018). Cette maladie a de graves répercussions économiques sur la filière lavandicole française du fait d'une diminution importante du rendement. Cette perte est imputable à l'affaiblissement des plants, à la mortalité accrue (Sémétey et al. 2018) et aux arrachages précoces de parcelles, engendrant pernicieusement le recul des surfaces cultivées (Yvin et al. 2018). D'autres études portant sur le vin (Sémétey et al. 2018) ou d'autres huiles essentielles (Marcone et al. 2016), ont prouvé que le phytoplasme modifie la qualité des produits obtenus à partir de plante touchée. Il est donc probable que le phytoplasme du Stolbur impacte la qualité des huiles essentielles de lavande, cependant aucune étude n'a été menée sur le sujet (Sémétey et al. 2018).

La pathologie du phytoplasme du Stolbur se traduit par un jaunissement et un étiolement progressif des plants conduisant à leur mort prématurée. De plus, les lavandes infectées sont généralement chétives et comportent pas ou peu d'inflorescence. Ces symptômes résulteraient de l'obstruction des

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

tubes criblés alimentant les lavandes en sève (Chatelard et al. 2017) qui découlerait de la prolifération de *Ca. Phytoplasma solani* dans le phloème et de l'épaississement de la paroi des tissus conducteurs de sève en réaction à l'infection (Pagliari et al. 2016).

Afin d'évaluer la sévérité de ces symptômes, le Centre Régionalisé Interprofessionnel d'Expérimentation en Plantes à Parfum Aromatiques et Médicinales (CRIEPPAM) a mis en place un système de notation reposant sur quatre stades présentés ci-dessous (Tableau 1 et Figure 4).

Tableau 1 : Système de notation visual des symptômes dus au dépérissement pour la lavande et le lavandin (d'après (Stierlin 2020))

Note	Etat sanitaire du plant	Critères visuels
1	Plant asymptotique	Plant vigoureux, verdoyant, absence de symptôme
2	Plant symptomatique	Jaunissement partiel Plant chétif Flétrissement (des feuilles ou des tiges) partiel Présence de bois mort < 50%
3	Plant symptomatique	Jaunissement sur l'ensemble du plant Plant très chétif Flétrissement (des feuilles ou des tiges) sur l'ensemble du plant Présence de bois mort > 50%
4	Plant mort	100% de bois sec sur la plante



Figure 4 : Illustration des 4 stades de dépérissement des plants de lavande et de lavandin (Chatelard et al. 2017)

L'intensité des symptômes des plants est étroitement liée aux conditions climatiques (sécheresse et canicule). Trois possibilités peuvent expliquer ce phénomène : les plants sont soumis à un stress supplémentaire (climatique), la hausse des températures provoque l'augmentation des populations de *H. obsoletus* et la multiplication du phytoplasme. (Yvin 2010; Sémétey et al. 2018)

### I.B.4. Moyens de lutte

Les stratégies de lutte peuvent être directes, c'est-à-dire contre le pathogène, ou indirectes. Actuellement, il n'existe aucun moyen de lutte directe en France. En effet, aucun antibiotique autorisé dans l'Union Européenne ne s'est avéré efficace contre les phytoplasmes (ANSES 2012). D'autres pistes ont été envisagées mais se sont avérées infructueuses comme la thérapie (destruction de la

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

bactérie par la chaleur) utilisée pour lutter contre la maladie du bois noir (vigne), car la lavande est trop sensible pour résister aux fortes chaleurs imposées par ce procédé (Yvin 2010).

Quant à la lutte indirecte visant à employer des insecticides contre *H. obsoletus*, ce n'est pas une option envisageable. Du fait de la coïncidence du temps de vol de la cicadelle avec la période de floraison des lavanderaies, l'utilisation d'insecticide aurait un impact dévastateur sur les abeilles et l'économie qui en découle.

En revanche, certains procédés indirects alternatifs ont montré leur effet pour temporiser la propagation de la maladie. Le moyen de lutte privilégié aujourd'hui reste la sélection variétale, mais aucun cultivar de lavande n'est résistant au phytoplasme du Stolbur. Afin de réduire l'épidémie, la méthode la plus efficace actuellement est la production de plant certifié sain et l'arrachage systématique des plants infectés. Les autres méthodes envisagées sont des modifications ou adaptations des pratiques culturales (voir Figure 5). L'élimination des plantes réservoirs tels que le liseron a très bien fonctionné pour la maladie du bois noir mais n'a montré qu'un impact très limité sur la lavande. Une autre méthode est l'emploi de barrière physique pour limiter la colonisation par *H. obsoletus* des nouveaux champs mais la construction de « cage » protectrice est coûteuse voire non viable économiquement pour les agriculteurs. Les méthodes alternatives actuellement en test sont la plantation des inter-rangs, la pulvérisation de kaolinite et la couverture du sol (les femelles *H. obsoletus* préfèrent des sols chauds et nus pour pondre). (Yvin et al. 2018; Sémétey et al. 2018)



Figure 5 : A gauche, essais d'enherbement inter-rang en plein champ. A droite, production de plants sains en serre anti-insecte (photographies CRIEPPAM)

L'influence d'un enherbement inter-rang (orge) a montré une diminution par deux du nombre d'insectes collectés. Plusieurs hypothèses peuvent expliquer cette diminution d'insectes. Des tests visant à déterminer la hauteur de vol maximal de la cicadelle, ont montré que 80 % des individus ont été capturés entre 0 et 40 cm du sol. Il est donc possible que l'orge (hauteur 80 cm) ait joué un rôle de barrière physique en limitant le déplacement des insectes. Néanmoins, le couvert végétal inter-rang a provoqué une concurrence avec le lavandin qui a eu un développement moindre (Rapport interne du CRIEPPAM 2012). Les essais sont toujours poursuivis par le CRIEPPAM et montrent des résultats prometteurs pour limiter la propagation. Cependant, cette technique rencontre des réticences au niveau des agriculteurs (espacement des rangs de lavande, introduction d'un deuxième type de culture, modification visuelle des champs, ...).

### I.B.5. Critères de sensibilité et tolérances des variétés

Actuellement aucune variété de lavande ou de lavandin n'est résistante au phytoplasme (Sémétey et al. 2018). Cependant, une étude menée par le CRIEPPAM et basée sur les observations faites par les producteurs et les techniciens, a mis en avant des différences de sensibilités face au dépérissement en fonction des variétés. L'évaluation du caractère sensible ou tolérant des variétés a été fait à partir de deux critères : le nombre d'individus de *H. obsoletus* et les symptômes de dépérissement observés.

Les résultats ont montré un lien entre le nombre d'individus (larves ou adultes) présents et la tolérance de la variété au dépérissement. Plus précisément, ce sont les variétés sensibles, c'est-à-dire présentant les plus forts symptômes de dépérissement, qui dénombrent le plus grand nombre d'individus (voir Figure 6). A noter que toutes les variétés sont porteuses de *H. obsoletus*. En revanche, aujourd'hui rien n'explique cette différence de population. (Rapport interne du CRIEPPAM 2012) De plus, les derniers résultats confirment une plus grande sensibilité de la lavande face au dépérissement par rapport au lavandin (Yvin et al. 2018).

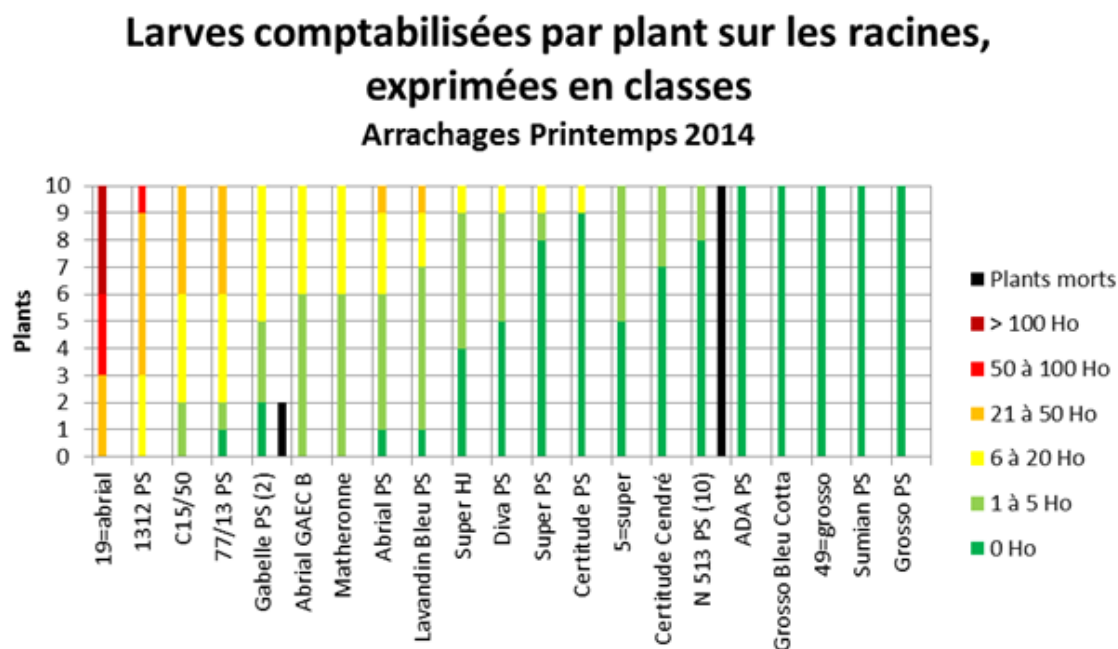


Figure 6 : Bilan des larves capturées (Ho) lors de l'arrachage du bloc n°3 planté lors des essais d'attractivité en 2011 à la station expérimentale du CRIEPPAM (données internes du CRIEPPAM)

Dans le cadre de cette thèse, deux variétés dites « sensibles » (7713 et abrial) et deux variétés dites « tolérantes » (maillette et grosso) ont été choisies dans le but de mettre en avant des différences de composition chimique en fonction du caractère de sensibilité des variétés. Les variétés 7713 et maillette sont des lavandes, alors qu'abrial et grosso sont des variétés de lavandin (voir partie I.D.1 Description botanique et aire de répartition géographique ci-dessous).

### I.C. L'insecte vecteur : la cicadelle

#### Généralités : morphologie, répartition et plantes hôtes

Le principal vecteur en Europe de l'ouest du phytoplasme est *H. obsoletus* (Figure 7) communément désigné sous le nom de cicadelle (Sémétey et al. 2018). *H. obsoletus* appartient à la famille des Cixiidae et au genre *Hyalesthes* Signoret. Ce genre contient 34 espèces dont l'aire de répartition s'étend du pourtour Méditerranéen jusqu'à l'Allemagne et la Russie. Ce sont des insectes piqueurs-suceurs se nourrissant de la sève contenue dans le phloème des végétaux. Vingt-cinq variétés de plantes ont été identifiées comme étant des plantes hôtes sur lesquels *H. obsoletus* réalise son cycle de vie complet. En France, les plantes hôtes les plus importantes sont le liseron, la passerage drave, la lavande et le lavandin. (Sforza et al. 1999; Kosovac et al. 2018)



Figure 7 : Photomicrographie de *H. obsoletus* (Kunz et al. 2010)

### Cycle de développement :

Habituellement, chez les Cixiidae les femelles pondent leurs œufs dans la terre au pied de leur plante hôte. Lorsqu'elles se développent, les larves se nourrissent des feuilles situées à la base de la plante ou de ses racines. Ensuite, les adultes émergent du sol et volent sur la partie supérieure des végétaux. Des observations effectuées en champs ont permis de confirmer et de détailler ce cycle chez la cicadelle (Figure 8). *H. obsoletus* est une espèce dont les larves passent l'hiver enterrées le long des racines axiales très profondément entre novembre et mars. L'émergence des adultes se fait dans le sol à quelques centimètres de la surface (1-5cm). Leur durée de vol est d'au moins 10 semaines et leur pic de population se situe entre fin juin et août. Le cycle complet de la cicadelle s'effectue sur une année, et une seule génération voit le jour durant cette période. Sur la lavande, la période au stade œuf est d'environ 7 semaines et la durée totale pour arriver au stade adulte est approximativement de 27 semaines. (Sforza et al. 1999)

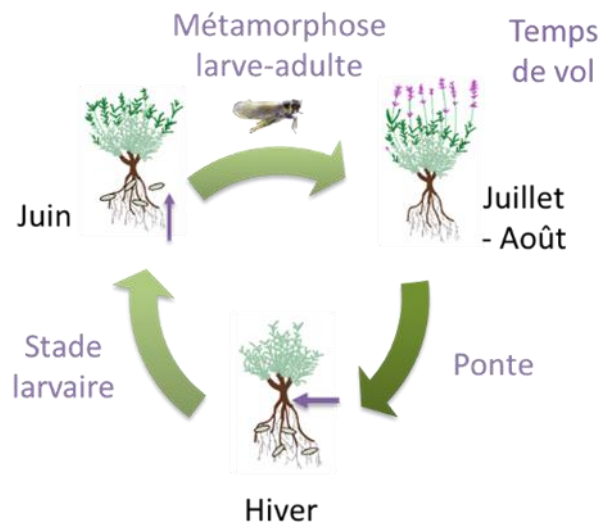


Figure 8 : Schéma du cycle de développement de la cicadelle sur la lavande et le lavandin

### Transmission du phytoplasme :

La cicadelle adulte se nourrit de la sève en piquant avec son stylet les plantes afin d'atteindre le phloème. Lors d'une ponction alimentaire dans un végétal infecté, celle-ci va également ingérer et se charger en phytoplasme (Figure 9). Elle va devenir porteuse mais deviendra vectrice seulement lorsque la charge bactérienne sera suffisante. Le processus qui intervient est alors le suivant : une fois dans le tube digestif, le parasite va migrer vers les glandes salivaires où il va se multiplier. Il passe ensuite dans la salive, prêt à être injecté dans le phloème d'une nouvelle plante à la prochaine ponction alimentaire. (Gasparich 2010; Yvin 2010) Il est à noter que la phase de maturation peut durer un mois. Entre ces deux phases le test PCR ne peut faire la différence entre les deux états de l'insecte. (Yvin 2010)

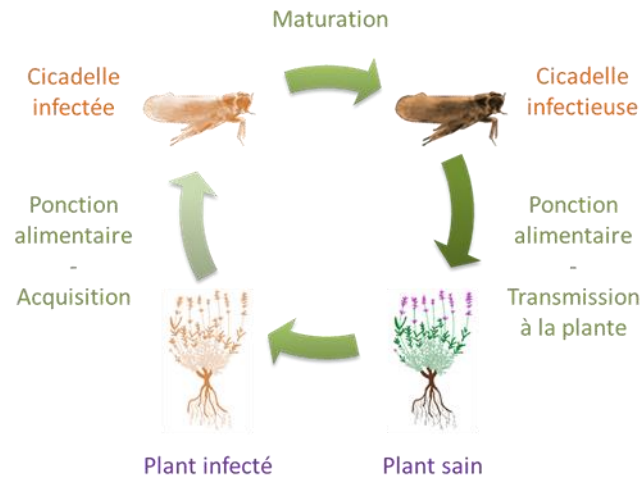


Figure 9 : Schéma du processus de transmission du phytoplasme par *H. obsoletus* sur la lavande

A noter que les larves peuvent également acquérir le phytoplasme, car elles s'alimentent sur les racines des plants. En revanche, l'infection se fait exclusivement par l'alimentation car il n'y a pas de transmission entre les mères et les larves. (Gonella et al. 2011)

#### Élevage :

Les connaissances sur le comportement de la cicadelle et de ses interactions avec la lavande et le phytoplasme du Stolbur sont à l'heure actuelle encore lacunaires. Afin de palier à ce manque, plusieurs essais d'élevage de *H. obsoletus* ont été conduits sur la lavande, le liseron et l'ortie. L'élevage sur ortie est fonctionnel mais reste instable et sans explications. Malgré de nombreux essais, aucun élevage sur lavande n'est actuellement opérationnel. (Yvin et al. 2018) La majorité des tests a donc été réalisée sur des insectes capturés en milieu naturel. Nous ignorons l'influence de la nature du plant natif de l'insecte sur son comportement.

Néanmoins les derniers essais réalisés par le CRIEPPAM sont prometteurs (CRIEPPAM 2019). Les captures en champs faites en 2018 ont permis l'émergence en 2019 d'une nouvelle génération indiquant que les adultes ont pu se reproduire ou au moins pondre. Cet élevage issu des insectes « sauvages » a été réalisé sur trois types de plantes hôtes : la lavande gabelle, le lavandin abrial et la sauge sclarée. Un autre type d'élevage dit « sain » car issu uniquement des œufs a également été mis en place. Malheureusement, aucun adulte n'a été obtenu sur la lavande gabelle et le lavandin abrial. En revanche, des émergences ont eu lieu sur la sauge sclarée. Des nouvelles captures ont été effectuées durant l'été 2019 afin d'alimenter ces différents élevages et les pérenniser.

#### Perception des composés organiques volatiles (COVs) :

Des études préliminaires ont été menées sur les composés organiques volatils (COVs) de lavande et de lavandin. Bien que des différences significatives n'aient pu être obtenues, les tests ont mis en évidence une sur-expression du  $\beta$ -caryophyllène et du  $\beta$ -farnésène chez les plants symptomatiques. De plus, une étude (Riolo et al. 2012) a montré que le  $\beta$ -caryophyllène et le  $\beta$ -farnésène induisaient une réponse antennaire chez *H. obsoletus* dans le cas de l'ortie et du gattilier (comme plante hôte) et quel que soit le sexe de l'insecte. Des études comportementales sur la cicadelle ont été effectuées par le CRIEPPAM à l'aide de tests de choix en cage. Les deux composés relevés comme étant prometteurs ont été testés : le  $\beta$ -caryophyllène a montré une tendance répulsive alors que le  $\beta$ -farnésène a été plutôt attractif pour *H. obsoletus*.

Pour compléter ces premiers résultats prometteurs, une étude métabolomique sur les COVs (émis et stockés) a été menée (Stierlin 2020). La thèse a mis en avant des différences significatives entre les

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

plants symptomatiques et asymptomatiques pour les COVs émis et stockés et a mis en évidence plusieurs marqueurs communs. Le  $\beta$ -farnésène a effectivement été trouvé comme biomarqueur de l'état symptomatique ainsi que quatre autres composés (l'acétate de linalyle, le bornéol, l' $\alpha$ -bergamotène et le germacrène D). Quant au  $\beta$ -caryophyllène, il a été identifié comme discriminant uniquement dans les extraits d'inflorescences symptomatiques de lavandin abrial.

Parallèlement à cette étude, la perception des molécules par la cicadelle a été approfondie en étudiant la réponse antennaire de *H. obsoletus* aux COVs de lavandin grosso (Chatelard et al. 2017). Ces analyses ont montré que la cicadelle était capable de détecter plusieurs COVs émis par le lavandin, parmi lesquels les biomarqueurs communs trouvés dans l'étude métabolomique.

### I.D. Le sujet d'étude : la lavande

#### I.D.1. Description botanique et aire de répartition géographique

##### Taxonomie :

Les différentes espèces de lavande appartiennent au genre *Lavandula*. Ce genre fait partie de la famille des *Lamiaceae* qui est divisée en 7 sous-familles, dont celle des *Nepetoideae*, elle-même répartie en 3 tribus. Le genre *Lavandula* appartient à la tribu des *Ocimeae*.



Figure 10 : Classification du genre *Lavandula*

Le genre *Lavandula* contient 39 espèces, ainsi que des hybrides et de nombreuses sous-espèces et variétés. Il est subdivisé en 3 sous-genres *Fabricia*, *Lavandula* et *Sabaudia* et 8 sections (voir tableau ci-dessous). (Guitton 2010)



Tableau 2 : Taxonomie du genre *Lavandula* d'après (Upson and Andrews 2004)

Genre	Sous-genre	Section	Espèce	Sous-espèce ou variété <sup>(b)</sup>
<b>Lavandula</b>	<b>Lavandula</b>	<b>Lavandula</b>	<i>L. angustifolia</i> Mill.	subsp. <i>angustifolia</i> (synonyme <i>L. spica</i> ) subsp. <i>pyrenaica</i> (synonyme <i>L. vera</i> )
			<i>L. x intermedia</i> Emeric ex Loiseleur <sup>(b)</sup> (synonyme lavandin), (hybride <i>L. angustifolia</i> x <i>L. latifolia</i> )	var. abrial var. grosso var. sumian var. super
			<i>L. lanata</i> Medik.	
			<i>L. latifolia</i> Boiss	
			<i>L. x heterophylla</i> <sup>(b)</sup> (synonyme <i>L. x allardii</i> ), (hybride <i>L. dentata</i> x <i>L. latifolia</i> )	
			<i>L. dentata</i> L.	var. <i>candicans</i> var. <i>dentata</i>
		<b>Stoechas</b>	<i>L. pedunculata</i> (Mill) Cav.	
			<i>L. stoechas</i> L.	subsp. <i>atlantica</i> subsp. <i>cariensis</i> subsp. <i>luiseiri</i> subsp. <i>lusitanica</i> subsp. <i>pedunculata</i> subsp. <i>sampaiana</i> subsp. <i>stoechas</i>
			<i>L. viridis</i> L'Hér.	
			<i>L. multifida</i> L.	
			<i>L. canariensis</i> Mill.	
			<i>L. x christiana</i> Gattef. & Maire <sup>(b)</sup> (hybride <i>L. canariensis</i> x <i>L. pinnata</i> )	
			<i>L. minutolii</i> Bolle.	var. <i>minutolii</i> var. <i>tenuipinna</i>
	<b>Fabricia</b>	<b>Pterostoechas</b>	<i>L. bramwellii</i> Upson & S.Andrews <sup>(a)</sup>	
			<i>L. pinnata</i> L.f.	
			<i>L. buchii</i> Webb & Berthel.	var. <i>buchii</i> var. <i>gracile</i>
			<i>L. rotundifolia</i> Benth.	
			<i>L. maroccana</i> Murb.	
			<i>L. tenuisecta</i> Coss. Ex Ball	
			<i>L. rejadalii</i> Upson & Jury (synonyme <i>L. brevidens</i> )	
			<i>L. mairei</i> Humbert	var. <i>antiatlantica</i> var. <i>maieri</i>
			<i>L. coronopifolia</i> Poir.	
			<i>L. saharica</i> Upson & Jury <sup>(a)</sup>	
<i>L. antinea</i> Maire				

		<i>L. pubescens</i> Decne.
		<i>L. citriodora</i> A.G. Mill. <sup>(a)</sup>
	<b>Subnudae</b>	<i>L. subnuda</i> Benth.
		<i>L. macra</i> Baker <sup>(a)</sup>
		<i>L. dhofarensis</i> A.G. Mill.
		<i>L. samhanensis</i> Upson & S.Andrews <sup>(a)</sup>
		<i>L. setifera</i> T. Anderson <sup>(a)</sup>
		<i>L. qishnensis</i> Upson & S.Andrews <sup>(a)</sup>
		<i>L. nimmoi</i> Benth.
		<i>L. galgalloensis</i> A.G. Mill. <sup>(a)</sup>
		<i>L. aristibracteata</i> A.G. Mill.
		<i>L. somaliensis</i> Chaytor <sup>(a)</sup>
	<b>Chaetostachys</b>	<i>L. bipinnata</i> (Roth) Kuntz
		<i>L. gibsonii</i> J. Graham <sup>(a)</sup>
	<b>Hasikenses</b>	<i>L. haskinensis</i> A.G. Mill. <sup>(a)</sup>
		<i>L. sublepidota</i> Rech.f. <sup>(a)</sup>
	<b>Sabaudia</b>	<i>L. atriplicifolia</i> Benth.
		<i>L. erythraeae</i> (Chiov.) Cufod
<p>(a) Espèces n'ayant pas fait l'objet d'une étude phytochimique à ce jour (avril 2020)</p> <p>(b) Seuls les sous-espèces, variétés et hybrides mentionnés dans ce manuscrit sont reportés dans ce tableau</p>		

Dans ce projet, quatre variétés de la section *Lavandula* (7713, abrial, grosso et maillette) ont été étudiées. Elles appartiennent à l'espèce *L. angustifolia* Mill. (7713 et maillette) communément appelée lavande fine, ou à l'espèce *L. x intermedia* Emeric ex Loiseleur (abrial et grosso) couramment dénommée lavandin.

### Morphologie :

Les caractéristiques morphologiques du genre *Lavandula* sont mixtes et divergentes. La morphologie générale varie de l'arbuste ligneux à l'herbacé non pérenne. Les inflorescences sont regroupées dans un épi terminal compact porté sur un long pédoncule (tige de la fleur), les étamines sont déclinantes (*i.e.* courbées vers le bas). (Lis-Balchin 2002)

La section *Lavandula* se caractérise par des plantes ligneuses ayant des feuilles linéaires à lancéolées (Despinasse 2015). La lavande fine est un arbuste au feuillage persistant pouvant atteindre 40-80 cm de hauteur. La tige est de forme quadrangulaire et accueille des feuilles de forme oblongue, simple, généralement linéaire et opposées. Elle possède des inflorescences portant plusieurs fleurs généralement de couleur violette (Grand 2009; Despinasse 2015).

Le lavandin est, quant à lui, plus massif que la lavande fine et peut atteindre jusqu'à 1 m à 1 m 50 de haut. Ces inflorescences comprennent trois fourches, une tige centrale portant l'inflorescence principale et deux fourches secondaires. (Despinasse 2015)

### Aire de répartition :

Les lavandes sont cultivées ou employées comme fleurs d'ornement dans le monde entier. L'aire de répartition naturelle du genre est vaste et s'étend de l'Europe, en passant par le nord de l'Afrique jusqu'à l'Inde (voir Figure 11). De nombreuses espèces sont endémiques, c'est par exemple le cas pour *L. buchii* et *L. minutolii* endémiques des îles Canaries. (Moja et al. 2016) Le bassin principal du genre *Lavandula* est celui de la Méditerranée. Les sections *Lavandula* et *Stoechas* sont prédominantes sur les rives nord du bassin méditerranéen. Les autres sections sont réparties en Afrique, au Proche-Orient et en Asie (voir Figure 11 ci-dessous). (Guitton 2010)



Figure 11 : Aire de répartition des sections du genre *Lavandula* (d'après (Guitton 2010))

Plus spécifiquement, l'aire de distribution de la lavande fine et du lavandin est la partie nord du bassin méditerranéen, en particulier en France, en Italie et en Espagne. Ils se développent généralement sur des sols calcaires et arides à des altitudes comprises entre 500 et 1800 m pour la lavande fine et à des altitudes inférieures pour le lavandin. (Despinasse 2015)

#### I.D.2. Les enjeux économiques

L'économie mondiale de la lavande et du lavandin repose principalement autour du marché des huiles essentielles. Les principaux consommateurs de ces huiles sont l'industrie de la parfumerie, de la cosmétique et la pharmaceutique. Près de 30 types d'huiles essentielles de « lavande » pures ou sous forme de mélange sont présentes sur le marché. Ce chiffre important est imputable au grand nombre d'espèces appartenant au genre *Lavandula*. Cependant, la majorité des espèces ont des productions anecdotiques et les plus commercialisées sont la lavande fine (*L. angustifolia*), le lavandin (*L. x intermedia*) et la lavande aspic (*L. latifolia*). (Giray 2018)

Une étude sur le marché des huiles essentielles du genre *Lavandula* a été réalisée par Giray en 2018. La majorité de la production mondiale est imputée au lavandin dont le marché représente un volume cinq fois supérieur à celui de la lavande. Les trois pays producteurs les plus importants sont la Bulgarie en premier avec 56 % de la production mondiale (4500 ha et 200 tonnes d'huile essentielle), suivie de la France avec 26 % puis la Chine avec 12 % (chiffres 2017). Les autres pays ne représentent que 10 % de la production mondiale, parmi lesquels le Royaume-Uni, l'Espagne, la Russie, l'Italie, le Maroc ou encore les Etats-Unis ou l'Afrique du Sud. (Giray 2018)

La France est traditionnellement le plus gros producteur dans le monde. Cependant depuis les années 2000, les producteurs français sont confrontés à une chute de la production d'environ 300 t (imputable au dépérissement et accentuée par des épisodes de sécheresse), soit environ 23 % de perte, en 5 ans. (Conord et al. 2012; Chaillan 2013) La France est passée au deuxième rang mondial mais reste le

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

premier producteur de lavandin avec environ 1400 tonnes d'huile essentielle produite en 2016 (Giray 2018). En 2018, la production de lavande et lavandin confondus, représentait 25 620 ha de surfaces cultivées en France, avec un total de 1593 producteurs. La culture se concentre dans le Sud-Est de la France et en particulier sur trois départements que sont les Alpes-de-Haute-Provence avec 11 456 ha, la Drôme avec 6840 ha et le Vaucluse avec 5575 ha. Plus précisément, le plateau d'Albion (Vaucluse) représente le principal bassin de la culture de la lavande. Quant au lavandin, il se concentre sur deux secteurs : le plateau de Valensole (Alpes-de-Haute-Provence) et le Tricastin (Drôme et Vaucluse). Les surfaces cultivées sont à la hausse ces quatre dernières années (29 %) et la proportion de surface labellisée bio augmente très fortement même si sa représentativité reste faible avec 2081 ha en 2018. La production d'huile essentielle en 2018, de lavande est de 138 t et du lavandin est de 1470 t. (Délégation nationale de FranceAgriMer (Volx) 2020) Le cultivar dominant en France est le grosso avec 89 % de la production (1322 t). Les autres variétés exploitées sont sumian (4 %), super (4 %) et abrial (3 %) (Giray 2018; Délégation nationale de FranceAgriMer (Volx) 2020).

La lavande et le lavandin sont aussi écoulés sur des marchés de niche ou locaux toujours sous forme d'huile essentielle pour des utilisations en aromathérapie par exemple. Ils sont également vendus sous forme de plantes d'ornement, de bouquets, de produits séchés et ont une valeur touristique forte. (Giray 2018) En Provence, les champs de lavande attirent les touristes, et les retombées économiques liées à cet agro-tourisme sont évaluées à 1,7 milliard d'euros par an (chiffre 2004). (Chaillan 2013)

## II. Les composés organiques non volatils (CONVs) dans le genre *Lavandula*

Une veille bibliographique a été effectuée tout au long de cette thèse, dans le but de constituer une base de données des composés organiques non volatils (CONVs) présents dans le genre *Lavandula* et quel que soit l'organe de la plante. De nombreuses études existent sur la composition des huiles essentielles, qui est extrêmement bien connue. En revanche, la connaissance sur les CONVs est bien plus ténue, néanmoins des études récentes comme Tang *et al.* (2017), Zhou *et al.* (2017), Yadikar *et al.* (2017 et 2018), Contreras *et al.* (2018), ou encore Ablikim *et al.* (2019), sont venues enrichir ces connaissances révélant la complexité et la diversité chimique contenues dans le genre *Lavandula*.

Cette synthèse bibliographique, complétée par la composition en COVs (composés organiques volatils), a fait l'objet d'une publication « *Phytochemicals from the genus Lavandula: a review* » dans *Phytochemistry reviews* (Héral *et al.* 2020). Dans cette partie, nous présenterons exclusivement la diversité chimique et les grandes familles des CONVs. La banque de données correspondante est présentée dans l'Annexe 1. Dans les métabolites spécialisés non volatils, les familles chimiques les plus représentées pour le genre *Lavandula* sont les composés phénoliques avec notamment les acides hydroxycinnamiques, suivi des triterpènes.

### II.A. Terpénoïdes

Les terpénoïdes sont des terpènes fonctionnalisés avec, par exemple, des groupes tels que des alcools ou des cétones. Les terpènes sont des hydrocarbures résultant de la condensation de deux précurseurs l'isopentényl-pyrophosphate (IPP) et son isomère le diméthylallyl-pyrophosphate (DMAPP) sous l'action d'une enzyme la prényltransférase. La répétition de condensation du motif isoprénique permet la formation des différents terpènes, à savoir : les monoterpènes (C<sub>10</sub>), les sesquiterpènes (C<sub>15</sub>), les diterpènes (C<sub>20</sub>), les triterpènes (C<sub>30</sub>) ou encore les tetraterpènes (C<sub>40</sub>) tels que les caroténoïdes. (Crozier *et al.* 2006; Woronuk *et al.* 2011)

La lavande et le lavandin contiennent de très nombreux mono- et sesquiterpènes mais nous n'évoquerons que quelques cas particuliers du fait de la volatilité de ces composés. Notamment, nous pouvons aborder le cas d'un monoterpène glycosylé, le nepetansoïde B, qui a été retrouvé dans un extrait hydroéthanolique de *L. dentata*. Dans la même étude, un sesquiterpène le (1R, 6R, 7R, 8S, 11R)-

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

1,6-dihydroxy-4,11-diméthyl-germacran- 4(5),10(14)-dien-8,12-olide a également été identifié (Contreras et al. 2018).

Actuellement, un seul diterpène qui est l'abiétatriène a été identifié dans le genre *Lavandula* (Stierlin et al. 2019).

En revanche, un total de 37 triterpènes a été reporté comprenant des composés tétra-, pentacyclique et des dérivés de stéroïdes. Nous pouvons citer par exemple les  $\alpha$ - et  $\beta$ -amyrines, les acides oléanolique et ursolique ou encore l'uvaol. Les principales structures triterpéniques identifiées dans le genre *Lavandula* sont présentées sur la Figure 12 ci-dessous. Dans les études ayant identifiées des triterpènes, celle menée par Topçu et ses collaborateurs (Topçu et al. 2001), se démarque par la prospection des composés racinaires de *L. stoechas*. Cette étude a d'ailleurs permis d'identifier 13 molécules parmi lesquelles deux nouveaux composés (l'acide 18-hydroxy-27-noroléan-12,14-dièn-30-al-28-oïque et l'acide 3 $\beta$ -hydroxy-1-oxo-oléan-12-ène-30-al-28-oïque) et cinq molécules qui n'avaient encore jamais été identifiées dans le genre *Lavandula* (16 $\beta$ -hydroxylupéol-3-*O*-myristate, 16 $\beta$ -hydroxylupéol-3-*O*-palmitate, 11-oxo- $\beta$ -amyrine, monogynol A *cis*-coumaryl ester, et monogynol A *trans*-coumaryl ester).

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

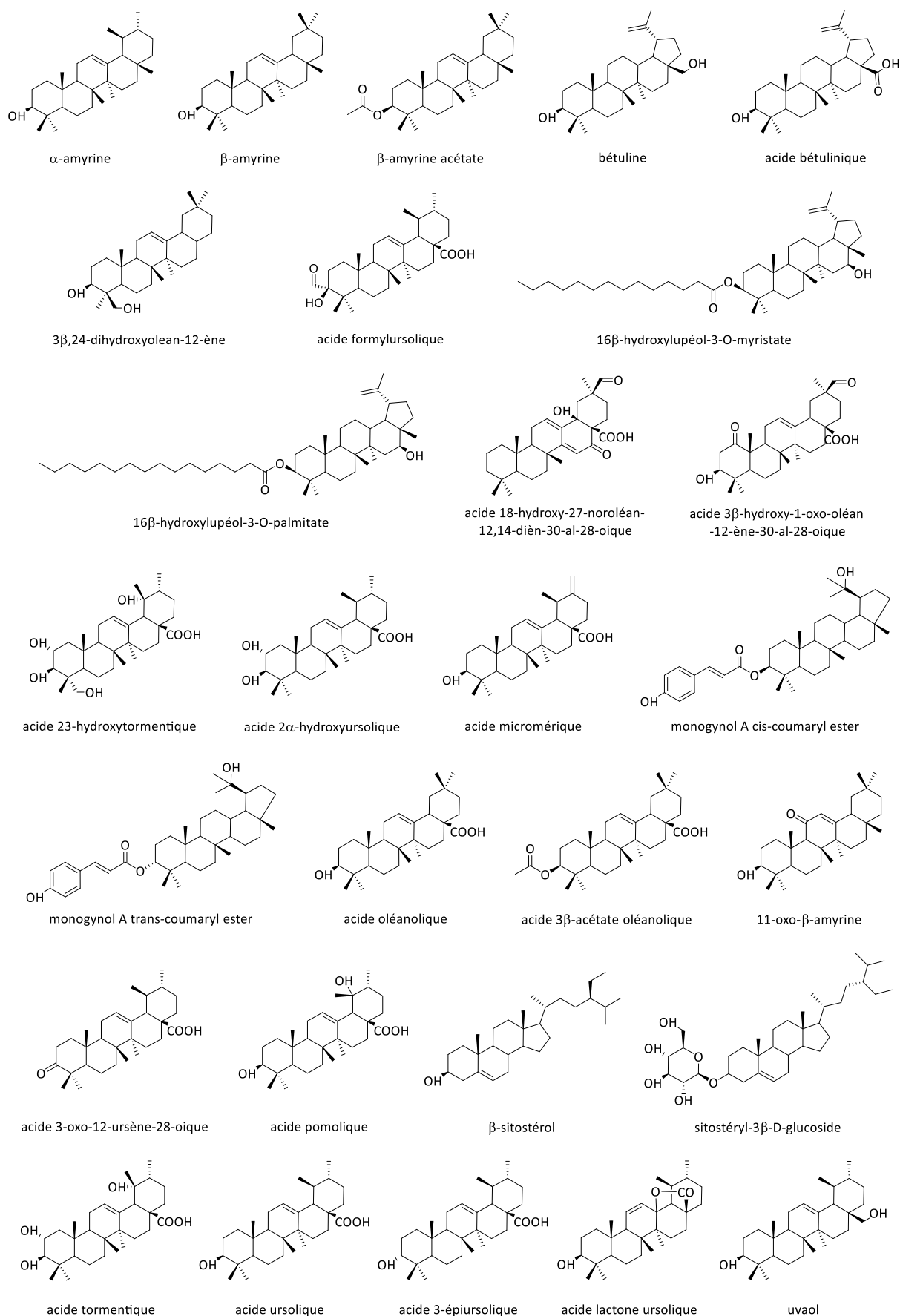


Figure 12 : Exemples de plusieurs structures de triterpènes et de dérivés présents dans le genre *Lavandula*

### II.B. Flavonoïdes

Les flavonoïdes sont des composés phénoliques très répandus dans le règne végétal. Ils se caractérisent par une structure de base en C<sub>6</sub>-C<sub>3</sub>-C<sub>6</sub>, c'est à dire deux cycles aromatiques reliés par un pont contenant trois atomes de carbone (Crozier et al. 2006). Les différentes classes de flavonoïdes retrouvées dans le genre *Lavandula* sont les flavones, les flavonols, les flavanones et les anthocyanidines. Ils peuvent être présents sous forme aglycone ou sous forme O- ou C-glycosylée. Les sucres sont généralement des glucosides, glucuronides ou rutinosides. Plusieurs acylations avec des acides coumarique ou malonique ont également été reportées notamment dans la catégorie des anthocyanidines.

#### II.B.1. Flavones

Les flavones sont la classe la plus abondante des flavonoïdes avec plus de 50 % du nombre de ces composés (Contreras et al. 2018). Les flavones reportées dans le genre *Lavandula* sont dérivées essentiellement de l'apigénine, la lutéoline, l'hypolaétine, la scutellaréine et l'isoscutellaréine (El-Garf et al. 1999; Upson et al. 2000; Lis-Balchin 2002; Yadikar et al. 2017). Des exemples de structures sont présentés ci-dessous (Figure 13).

La plupart des études sont conduites sur les parties aériennes de la plante de manière indistincte (El-Garf et al. 1999; Torras-Claveria et al. 2007; Algieri et al. 2016; Yadikar et al. 2017; Contreras et al. 2018). Cependant, certaines recherches ont été conduites uniquement sur des extraits de feuilles (Upson et al. 2000). Upson *et al.* ont mené une étude intéressante sur le profil phytochimique de plusieurs variétés dans le but d'identifier des composés spécifiques à chaque section du genre *Lavandula*. Cette étude a permis d'identifier 21 flavones et de classer en trois catégories les sections du genre en fonction de leur chimie. Le premier groupe comprend les sections *Lavandula*, *Dentata* et *Stoechas* et se caractérise par une accumulation de flavones 7-glycosylés. La deuxième catégorie comporte les sections *Pterostoechas*, *Subnuda* et *Chaetostachys*, qui se distingue par la présence de flavones 7 et 8-glycosylés ainsi que 8-hydroxylés. Enfin la section *Sabaudia* contient de façon prépondérante des flavones 8-hydroxylés 7-glycosylés, et 7-glycosylés.

Quant à la composition des inflorescences, elle a fait l'objet de deux études (Banthorpe et al. 1985; Areias et al. 2000) qui ont mis en évidence la présence du chrysoériol, de l'apigénine, l'apigénine 7-O-glucoside et de la lutéoline, également présents dans les extraits de feuilles.

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

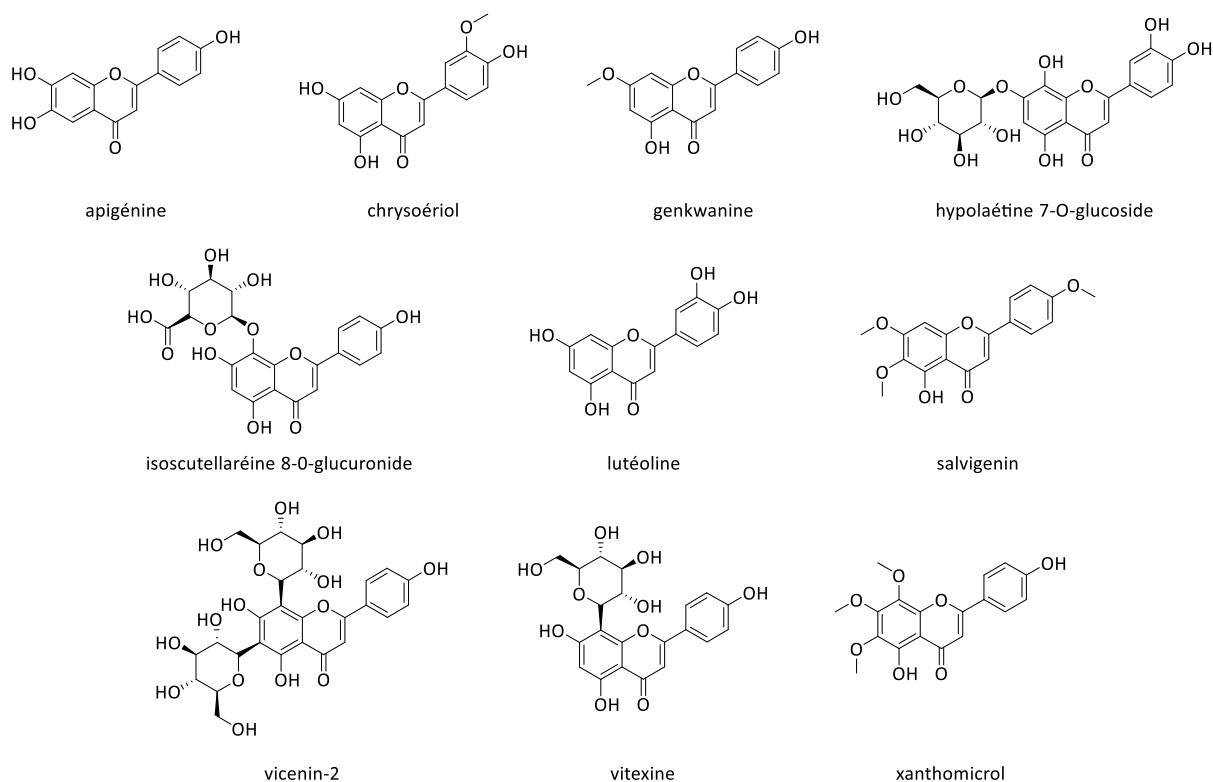


Figure 13 : Structure des principales flavones reportées dans le genre *Lavandula*

### II.B.2. Anthocyanidines

Les anthocyanidines sont des composés hydroxylés basés sur la structure du 2-phénylbenzopyriliium. Elles sont généralement présentées dans les plantes sous forme d'anthocyanes (forme glycosylée). Ce sont des pigments responsables de la couleur des fleurs et des fruits. Ils peuvent prendre des teintes pouvant aller du rouge au noir en passant par le violet. (Yoshida et al. 2009)

Très peu d'études ont été conduites sur les anthocyanidines dans le genre *Lavandula* (Banthorpe et al. 1985; Lis-Balchin 2002). Uniquement huit composés sont reportés dont la delphinidine et la malvidine, ainsi que des anthocyanes dérivés de ces deux molécules et de la cyanidine (voir Figure 14 ci-dessous).

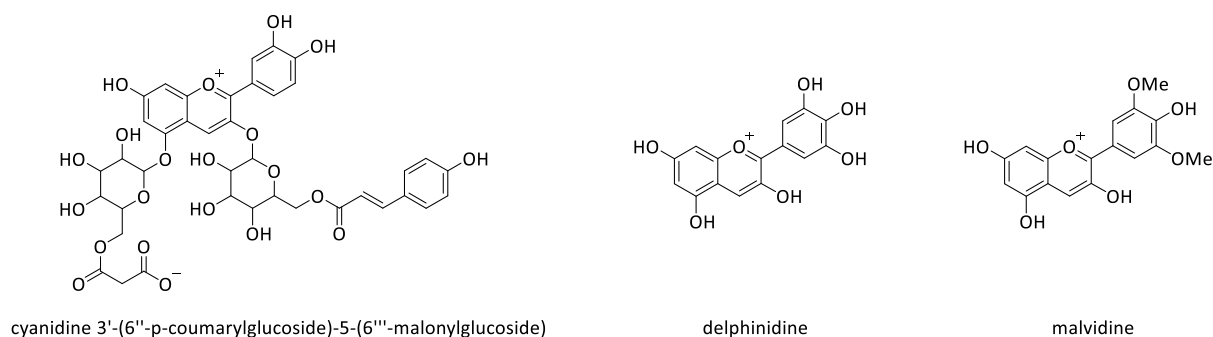


Figure 14 : Exemple d'anthocyanes et anthocyanidines présents dans le genre *Lavandula*



### II.B.3. Autres flavonoïdes

Les autres flavonoïdes reportés dans le genre *Lavandula* sont trois flavonols (kaempferol malonyl glucoside, kaempferol-3-coumaryl glucoside et la quercétine-3-O-glucoside) et une flavanone partiellement identifiée comme étant l'ériodictyol-O-hexoside. (Torras-Claveria et al. 2007; Spiridon et al. 2011)

### II.C. Acides phénoliques

Les acides phénoliques constituent la deuxième grande catégorie avec les flavonoïdes des composés phénoliques présents dans la lavande et le lavandin. Leur structure se caractérise par la présence d'au moins un cycle phénol et d'une fonction acide carboxylique (Goleniowski et al. 2013). Les deux sous-catégories les plus représentées dans le genre *Lavandula* sont les acides hydroxybenzoïques et les acides hydroxycinnamiques.

#### II.C.1. Dérivés d'acide hydroxybenzoïques

Seulement une dizaine d'acides hydroxybenzoïques ont été recensés dans le genre *Lavandula* (Torras-Claveria et al. 2007; Yadikar et al. 2017, 2018; Contreras et al. 2018). Parmi eux, nous pouvons citer l'acide protocatéchuique ou l'acide vanilique. De façon plus surprenante des dérivés soufrés à savoir l'acide 4-hydroxybenzoïque 4-(6-O-sulfo)glucoside et l'acide vanilique sulfoquinovoside ont été potentiellement identifiés sur la base de leur spectre de masse (Contreras et al. 2018). Des exemples d'acide hydroxybenzoïques sont présentés sur la Figure 15.

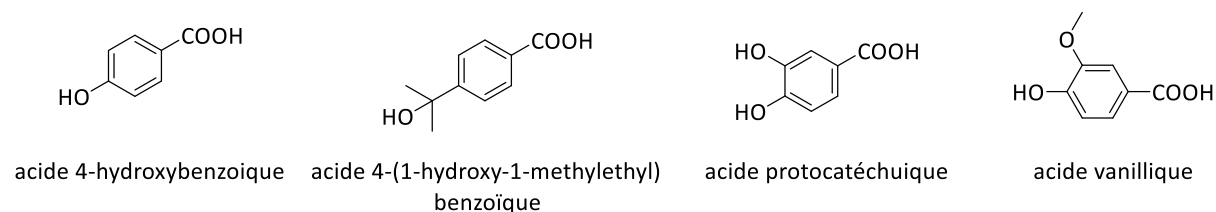


Figure 15 : Exemple de structures d'acide hydroxybenzoïque et dérivés reportés dans le genre *Lavandula*

#### II.C.2. Dérivés d'acide hydroxycinnamiques

Les acides hydroxycinnamiques possèdent une structure en C<sub>6</sub>-C<sub>3</sub> (Crozier et al. 2006). Ils sont la classe de composés organiques non volatils la plus représentée dans le genre *Lavandula* avec un total de 78 molécules identifiées. De nombreux exemples sont présentés sur la Figure 16. Les structures les plus courantes sont celles qui dérivent des acides coumarique, férulique et caféique sous forme aglycone ou glycosylée. Il est intéressant de noter que plusieurs molécules correspondent à des dimères (les acides rosmarinique et chlorogénique), des trimères (des isomères des acides salvianolique, isosalvianolique et yunnanéique) et même des tétramères (l'acide sagérinique) de l'acide caféique ou de sa forme hydroxylée (l'acide 3,4-dihydroxyphenyllactique) (Lis-Balchin 2002; Spiridon et al. 2011; Contreras et al. 2018).

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

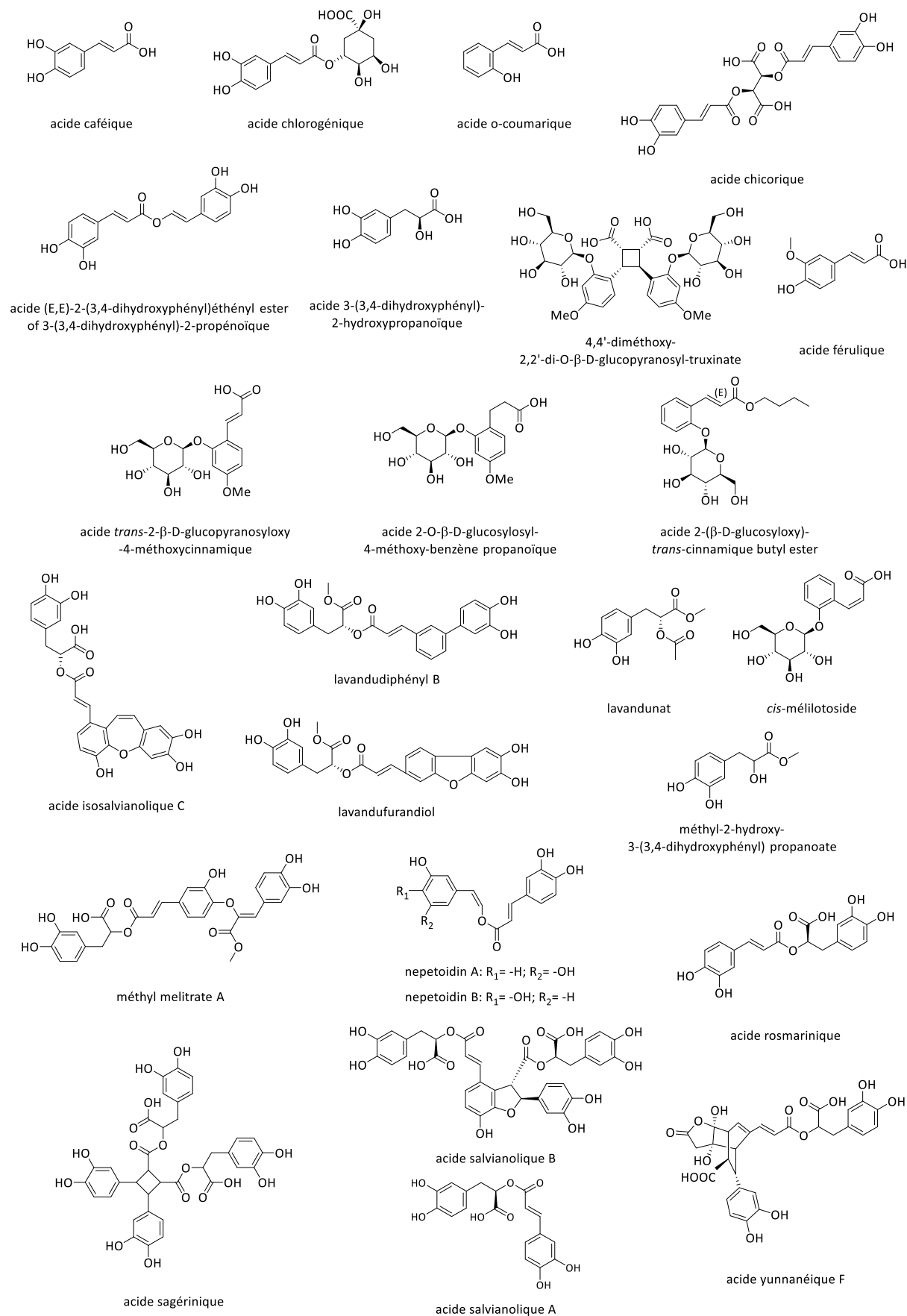


Figure 16 : Structures de plusieurs acides hydroxycinnamiques et des dérivés associés reportés dans le genre *Lavandula*

### II.D. Coumarines

La structure de base des coumarines correspond au 2*H*-1-benzopyran-2-one (Buckingham and Munasinghe 2015). Les coumarines majoritairement retrouvées dans le genre *Lavandula* sont la coumarine et son dérivé 7-méthoxylé : l'herniarine (Banthorpe et al. 1985; Areias et al. 2000; Lis-Balchin 2002; Yadikar et al. 2017). Les autres coumarines identifiées sont l'umbelliféone (7-hydroxycoumarine), l'esculétine (6,7-dihydroxycoumarine) ainsi que plusieurs formes glycosylées (Lis-Balchin 2002; Contreras et al. 2018). Dans une étude récente, deux nouvelles coumarines, appelées lavandupyronone A et B, ont été isolées d'un extrait hydroéthanolique de *L. angustifolia* (Yadikar et al. 2018).

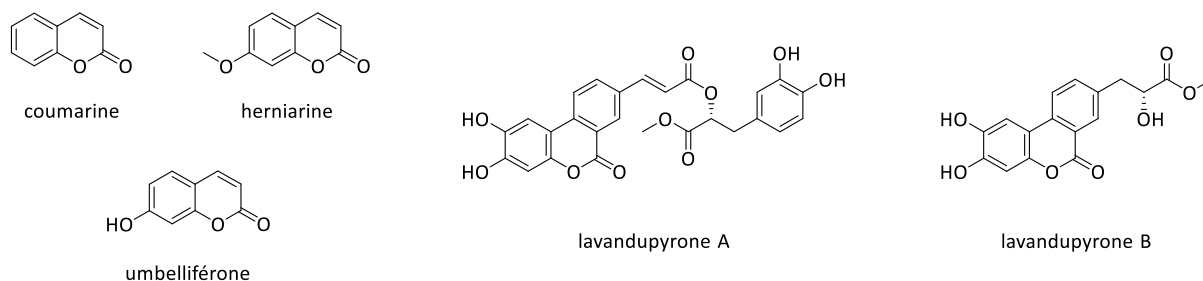


Figure 17 : Structure de plusieurs dérivés des coumarines reportés dans le genre *Lavandula*

### II.E. Dérivés des benzofuranes

Les benzofuranes sont des composés hétérocycliques comprenant un cycle benzène et un cycle furane (Khanam and Shamsuzzaman 2015). Deux études ont mis en évidence la présence de 12 dérivés benzofuranes (Tang et al. 2017; Zhou et al. 2017) dans les parties aériennes de la lavande fine. Les structures de ces composés sont décrites dans le Figure 18 ci-dessous. Parmi lesquels six nouveaux composés le 4-(5-(2-hydroxyethyl)-7-methoxy-3-methylbenzofuran-2-yl) phénol, le 2-(6-methoxy-2-(4-methoxyphenyl)-3-methylbenzofuran-5-yl) éthanol, le 2-(7-methoxy-2-(4-methoxyphenyl)-3-methylbenzofuran-5-yl) éthanol (Tang et al. 2017), et trois autres dénommés par les auteurs lavandulactone A-C (Zhou et al. 2017).

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

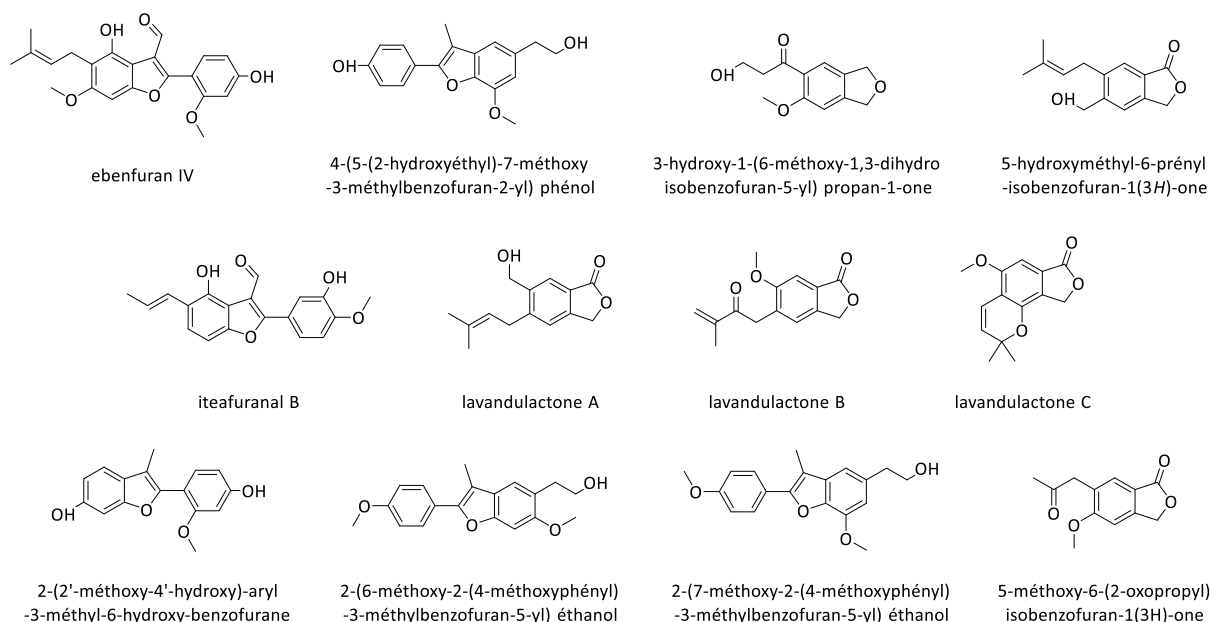


Figure 18 : Structures des différents dérivés de benzofuranes identifiés dans le genre *Lavandula*

### II.F. Autres composés

Dans les autres catégories de composés, un lignane (isolariciresinol hexoside), un disaccharide, trois acides organiques (les acides citrique, isocitrique, gluconique ou galactonique), trois dérivés de l'acide jasmonique et l'acide trihydroxy-octadécadiénoïque ont été détectés dans des extraits hydrométhanolique de *L. dentata* et *L. stoechas* (Contreras et al. 2018). Un gallotanin, le galloyl-HHDP-glucose a potentiellement été identifié dans un extrait de *L. angustifolia* (Spiridon et al. 2011). Enfin, un composé azoté le 2-N-phénylaminonaphthalène a été caractérisé dans l'espèce *L. vera* (Papanov et al. 1985).

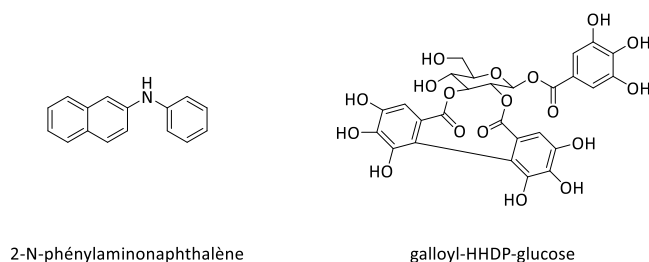


Figure 19 : Structures d'un composé azoté et d'un gallotanin reportés dans le genre *Lavandula*

Cette synthèse bibliographique a permis d'évaluer la grande diversité chimique des CONVs répertoriés dans le genre *Lavandula*. Dans le contexte qui nous préoccupe d'écologie chimique et face à cette grande variété chimique, nous avons fait le choix de recourir à la métabolomique. L'utilisation de la métabolomique se développe pour répondre aux questions posées en écologie et est parfois désignée sous le terme « Eco-métabolomique » (Peters et al. 2018).

### III. La métabolomique

Les plantes produisent un grand nombre de molécules appelées métabolites. Il en existe deux types : les métabolites primaires entrant dans les processus de photosynthèse, de croissance, et de reproduction de la plante (Courant et al. 2014); et les métabolites dits secondaires ou spécialisés. Les métabolites primaires sont des composés du type protéine, lipide et chlorophylle. Les métabolites secondaires qui interviennent dans des rôles de défense, de communication, ... chez les végétaux, peuvent être d'une grande diversité chimique (comme celle du genre *Lavandula* présentée dans la partie II ci-dessus). Ces molécules ont depuis très longtemps suscité l'intérêt pour leurs propriétés pharmaceutiques (Wolfender et al. 2013).

Les études phytochimiques conduites sur les métabolites secondaires sont généralement ciblées ou cherchent à l'être ; afin, par exemple, de découvrir un composé responsable d'une activité biologique donnée. La métabolomique, quant à elle, est généralement sur une échelle plus large. Elle s'efforce d'offrir une vision globale du métabolome, et recherche les variations chimiques induites par les facteurs d'étude.

#### III.A. Définitions, concept et limite de la métabolomique

##### III.A.1. Généralités et définitions

Depuis plusieurs décennies, nous voyons se développer les techniques dites « -omics ». Les deux premières à être apparues sont la génomique et la transcriptomique dans les années 80. La génomique correspond à l'étude des gènes. Elle est suivie dans le processus cellulaire par la transcriptomique qui s'intéresse à l'expression des gènes, c'est-à-dire à l'ARN. Dans les années 90 est survenue la protéomique qui étudie les protéines et les peptides. La dernière de ces techniques est apparue au début des années 2000, c'est la métabolomique qui s'intéresse aux métabolites (Figure 20). Elle arrive en bout de chaîne des autres techniques et du processus cellulaire, ce qui en fait l'observatrice de l'expression finale des changements qui ont pu s'opérer dans l'organisme. (Courant et al. 2014; Favre 2017)

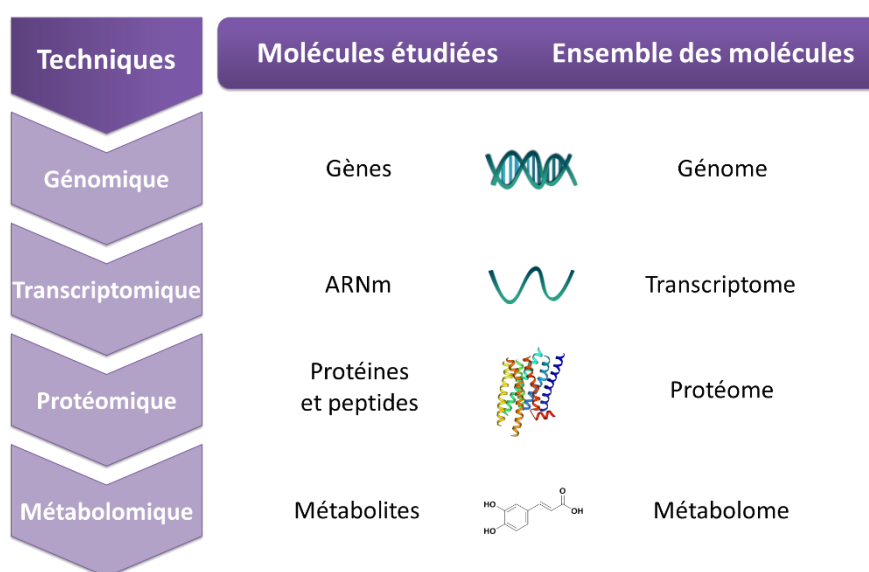


Figure 20: Schéma généraliste représentant les 4 principales techniques dites « -omics » et leurs cibles d'étude associées

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

Plus précisément, la métabolomique se définit comme l'intégralité des démarches analytiques permettant d'analyser le métabolome d'un système vivant (tissus, cellules, organes, ...). Les métabolites sont des molécules de faible poids moléculaire, c'est-à-dire compris entre 50 et 1500 Da, contenues dans un organisme vivant, et l'ensemble des métabolites est appelé métabolome. (Courant et al. 2014; Wolfender et al. 2015; Favre 2017)

Le but de la métabolomique n'est pas d'identifier les molécules une par une mais de comparer des empreintes chimiques pour identifier les changements opérés dans un système biologique donné. Les composés les plus affectés par les changements du système sont appelés des biomarqueurs. En pratique, il est impossible d'obtenir actuellement une image exhaustive du métabolome, que ce soit d'un point de vue quantitatif ou qualitatif.

Une étude métabolomique peut avoir trois finalités différentes. Elle peut-être à visée descriptive (et souvent comparative), dans ce cas ce sont les spécificités entre les différents échantillons qui sont recherchées. Une étude prédictive aura pour objectif la découverte d'un marqueur caractéristique d'une situation biologique (ex : pathologie) pouvant être transposé à un panel plus important d'individus (ex : réaliser le dépistage d'une maladie à partir de ce marqueur). Enfin, elle peut avoir un but cognitif visant à une meilleure compréhension du système d'étude tel que la réactivité d'un système biologique donné à une modification de son environnement (ex : activation d'une voie biochimique spécifique). (Bonfille 2017)

Le champ d'application de la métabolomique est vaste. Le principal domaine d'application est le domaine du médical que ce soit pour la découverte de biomarqueurs précoces, pour la compréhension des modifications du métabolisme engendré par une maladie, la découverte de nouveaux médicaments ou la compréhension des mécanismes d'action. Elle est également employée dans l'agronomie, en nutrition, en criminalistique (Liu and Locasale 2017) ou pour lutter contre le dopage (Raro et al. 2015) ou l'adultération de produits naturels (Saint-Lary 2015). Dans notre cas, nous nous intéresserons à la métabolomique appliquée aux plantes et plus particulièrement à l'écologie chimique.

### III.A.2. Approche métabolomique

Il est d'usage de distinguer deux types d'approche en métabolomique et de parler d'approche ciblée ou non ciblée.

L'approche globale ou non ciblée a pour but d'être exhaustive et sans a priori. Aucune hypothèse sur la nature des molécules n'est générée en amont de l'étude (Royer et al. 2017). L'expérimentateur a pour but de comparer deux ensembles d'échantillon ou de générer des hypothèses.

Dans l'approche ciblée, l'étude se concentre sur une famille de molécules déterminée au préalable. Elle permet de tester une hypothèse de départ. Cette approche permet de faciliter l'élucidation structurale. Un autre avantage est la possibilité de réaliser une étude quantitative. Sa réussite dépend des connaissances préalables et des hypothèses générées sur les composés à cibler. (Liu and Locasale 2017)

Dans certains cas, on peut parler d'approche semi-ciblée lorsqu'une liste de métabolites est ciblée spécifiquement et quantifiée (comme pour une approche ciblée) mais qu'aucune hypothèse n'a été émise au préalable. Cette approche est intéressante lorsque l'on souhaite suivre l'influence d'une perturbation sur l'organisme étudié pour une voie métabolomique connue donnée. (Liu and Locasale 2017)

### III.A.3. Les enjeux et limites de la métabolomique

#### a. Facteur biologique et confondant

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

Lors d'une étude métabolomique, le but est de mettre en évidence l'influence d'un facteur biologique sur l'organisme d'étude. Ceci va se traduire par observer des différences de concentrations entre différents groupes d'individus, avec au minimum un groupe témoin et un groupe affecté comme représenté dans la Figure 21.

Lorsque l'on souhaite mener une étude métabolomique, il faut être conscient que chaque étape de l'étude peut induire une variabilité, on parle de facteur confondant. Si ces facteurs confondants ne sont pas connus et maîtrisés, ils peuvent conduire à masquer la variabilité due au facteur biologique de l'étude. (Courant et al. 2014) Il existe deux grands types de source de variabilité :

- La variabilité technique
- La variabilité biologique

La variabilité technique peut-être liée à l'opérateur ou aux instruments de mesure. A chaque étape de l'étude, il est important d'essayer de minimiser la dérive analytique ou à défaut de la connaître, et de veiller à la robustesse du protocole et des gestes expérimentaux. La multiplicité des étapes augmentera le nombre de sources de variabilité technique, un protocole simple est à privilégier.

La variabilité biologique peut-être inter ou intra-individus (Courant et al. 2014). Prenons le cas d'un plant de tomate, la concentration en métabolites du fruit sera différente entre un plant exposé au soleil et un planté à l'ombre, même si tous les deux appartiennent au groupe témoin. De même, pour un plant donné la concentration en métabolites varie au cours de la journée. Pour contrer la variabilité inter-individuelle, les prélèvements devront être effectués à la même heure, et pour palier à la variabilité intra-individuelle, il sera conseillé de prendre uniquement des plants ayant eu le même ensoleillement.

Il est important de prêter attention à la conception de la campagne d'échantillonnage en amont de l'étude. En n'oubliant pas de choisir un nombre suffisant d'individus afin de satisfaire la puissance statistique. (Courant et al. 2014)

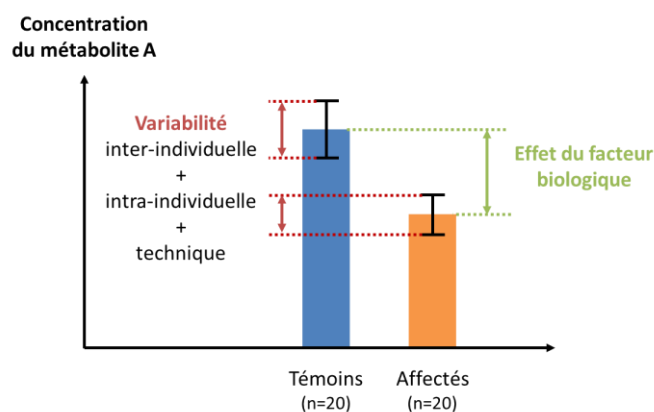


Figure 21: Histogramme des concentrations moyennes et écart-types du métabolite A des groupes témoins et affectés (d'après (Courant and Giraudeau 2018))

### b. Limite technique

Aujourd'hui, aucune technique analytique ne permet de couvrir l'intégralité de la diversité chimique du métabolome. Pour remédier à ce problème des études combinent plusieurs techniques d'analyse, telles que la LC-MS et la GC-MS (Heuberger et al. 2014), ou la MS et la RMN (Chen et al. 2014), afin d'obtenir une image la plus complète du métabolome. Néanmoins, cette approche n'est réalisable que sur des jeux de données restreints et devient difficile à mettre en place sur des échantillonnages de grande envergure (Courant et al. 2014).

### III.B. La méthodologie

La métabolomique repose sur une démarche opérationnelle dont nous allons décrire les différentes étapes dans cette partie. Avant de commencer une étude métabolomique, il est important d'avoir spécifié la question biologique et les différentes hypothèses relatives au problème d'étude, puis pensé et défini cette méthodologie qui contient de nombreuses étapes. Cette multiplicité impose une rigueur dans la mise en place et la réalisation de chaque étape. Des bonnes pratiques en métabolomique tendent à émerger dans une volonté d'uniformisation des usages, plusieurs seront présentées dans cette partie.

#### III.B.1. Workflow /Design de l'étude

La métabolomique suit une méthodologie définie qui consiste à une succession d'étapes dénommée par l'anglicisme « workflow » (voir Figure 22). Même si la communauté scientifique de la métabolomique tente depuis plusieurs années de codifier celle-ci, le protocole d'une nouvelle étude dépendra surtout de la nature de la matrice étudiée.

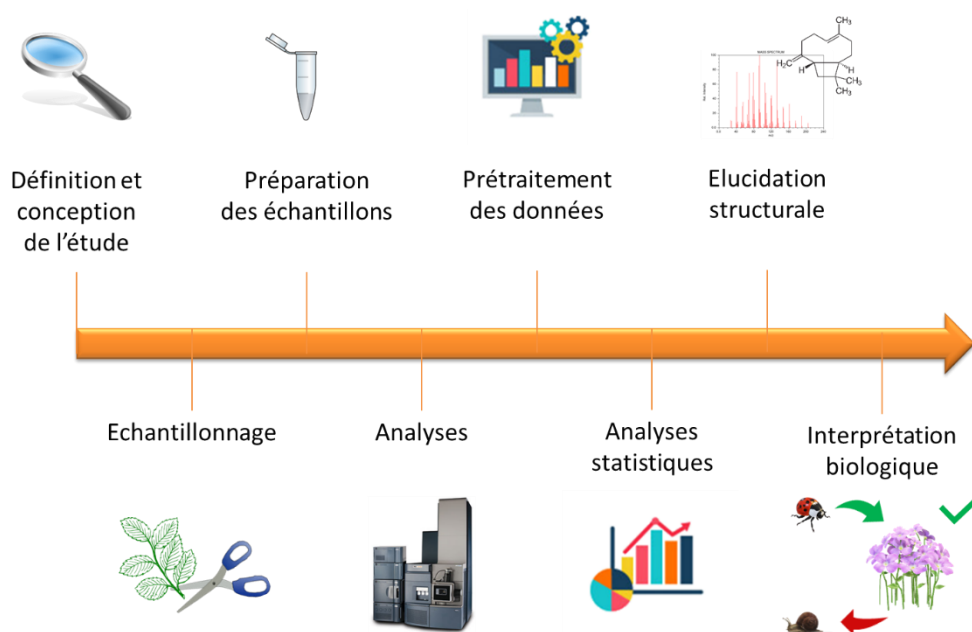


Figure 22: Schéma des principales étapes d'une étude métabolomique

En amont de la réalisation d'une étude métabolomique, il est important de définir et conceptualiser celle-ci. Le premier choix à réaliser est celui de la matrice en fonction de la question biologique. Ce choix conditionnera l'intégralité des étapes suivantes. De manière générale, chaque choix dans une étape impactera les suivantes. Le design de l'étude n'est donc pas anodin.

Après cette étape de réflexion et de conception, vient la mise en œuvre de l'expérimentation avec en premier lieu la collecte des échantillons. Plusieurs paramètres sont à prendre en compte ici tels que le nombre d'échantillons à prélever, l'âge des individus, les conditions de récolte (moment, durée, composition des flacons de collectes, ...) (Royer et al. 2017). La deuxième grande étape est celle de la préparation des échantillons. Elle peut être très simple avec seulement une étape de dilution ou comprendre de nombreuses phases (précipitation, extraction, ...) (Royer et al. 2017). Dans tous les cas, le protocole va fortement dépendre de la matrice et de l'approche (ciblée ou globale) choisies. La troisième étape est l'analyse des échantillons. La méthode analytique doit être robuste pour soutenir de longue séquence analytique car la métabolomique est un outil à haut débit, c'est-à-dire générant un nombre de données important. Le choix de la technique analytique dépendra, entre autres, de la



## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

nature de l'échantillon telle que sa capacité à se volatiliser ; et du choix de l'approche. Par exemple, dans le cas d'une étude ciblée effectuée par spectrométrie de masse, le type de source d'ionisation retenu dépendra des propriétés de la famille de molécules cibles. Après l'expérimentation s'ensuit une étape de prétraitement des données. Les données acquises sont « brutes » et nécessitent plusieurs transformations pour réaliser les analyses statistiques. Cette étape est réalisable sur divers supports informatiques (gratuit ou payant, en ligne ou sur machine). Elle n'est pas à négliger car elle peut introduire des biais voire rendre les données acquises inexploitable. (Royer et al. 2017) Sa complexité diffère selon la technique analytique employée. A l'issue de ces étapes, les données sont exploitables et des analyses statistiques peuvent être générées. Les tests appliqués dépendront de la nature des données et de la problématique initiale. Selon les résultats générés par les statistiques et la finalité de l'étude, un ou des composés doivent faire l'objet d'une élucidation structurale. Cette étape est indispensable pour transformer les données en résultats biologiques concrets. La complexité de cette étape dépendra fortement des bases de données à disposition et des connaissances scientifiques sur la chimie de l'objet d'étude. (Wolfender et al. 2013) A l'issue de ce processus, les acteurs pourront interpréter les résultats obtenus et répondre à la question biologique de départ.

Quels que soient les choix réalisés à chaque étape, la mise en place d'un protocole en métabolomique doit soigneusement prendre en compte et minimiser au maximum toutes les sources de variabilités présentes à chaque étape du processus sous peine de rendre les données inexploitable.

### III.B.2. Echantillonnage

Le premier choix à faire est celui de la matrice. Le choix de la matrice dépend fortement de la question biologique à laquelle l'étude souhaite répondre, et de la facilité à se procurer tel ou tel type d'échantillon. Il est plus aisé de prélever une feuille sur un arbre qu'une de ses racines. Mais est-il certain que le ou les marqueurs à mettre en évidence soient présents dans tous les organes ? Le choix de la matrice est d'autant moins anodin qu'il conditionnera partiellement ou intégralement le choix de la préparation des échantillons.

La deuxième question à se poser est celle de la taille de l'échantillonnage. Sa dimension doit être suffisante pour satisfaire la puissance statistique de l'étude, qui dépend aussi de l'amplitude de l'effet induit par le facteur biologique et le niveau seuil de significativité statistique souhaité. Il est donc difficile d'établir la taille de l'échantillonnage sans étude préalable permettant d'estimer la variation du facteur d'étude car celle-ci est inconnue (Royer et al. 2017).

#### **Protocole et bonnes pratiques d'échantillonnage en métabolomique :**

Les protocoles de prélèvement et de conservation des échantillons dépendent de la nature de la matrice biologique. Ces deux étapes conditionnent la qualité et la fiabilité de l'étude. Quels que soient les protocoles adoptés, ils devront être appliqués rigoureusement à chaque échantillon afin de ne pas introduire de variabilité.

Il est recommandé lors de l'échantillonnage pour une étude métabolomique de plonger immédiatement les échantillons dans l'azote liquide (« quenching ») afin de figer le métabolisme. (Royer et al. 2017) Puis de maintenir les échantillons à très basse température, c'est-à-dire à  $-80^{\circ}\text{C}$ . Cette trempe permet l'arrêt de toutes les réactions chimiques et enzymatiques (Salem et al. 2020).

Il est également préconisé d'éviter l'utilisation de certains contenants notamment les plastiques qui peuvent relarguer des contaminants tels que les phtalates ou les plastifiants (Royer et al. 2017). Les conditions de stockage sont à contrôler. Elles ne doivent pas induire de réaction chimique ou enzymatique pouvant entraîner une dénaturation de la matrice d'origine (température) et ne doivent pas différer d'un échantillon à l'autre (exemple : durée de stockage). (Royer et al. 2017; Salem et al. 2020)

### III.B.3. Préparation des échantillons

#### a. Généralités et bonnes pratiques de préparation des échantillons en métabolomique

La préparation des échantillons est l'ensemble des étapes après l'échantillonnage, permettant de transformer les échantillons en une solution d'analyte directement injectable pour l'analyse.

La préparation des échantillons peut contenir une ou plusieurs étapes. Elle dépendra de la nature de la matrice, c'est-à-dire de l'échantillon biologique. Il n'y a pas de méthode universelle permettant l'extraction de l'ensemble du métabolome, ni de consensus de la communauté scientifique sur un protocole d'extraction. Cette étape est d'autant plus critique que le métabolome a une grande diversité chimique. Les molécules à extraire ont des propriétés physico-chimiques très variées. Elles peuvent avoir des solubilités très différentes selon les solvants, des polarités différentes, des stabilités différentes, ... ce qui rend l'obtention d'un protocole complexe. Il est également important que le protocole soit répétable afin de ne pas biaiser les résultats. Généralement, l'échantillon biologique est filtré et/ou extrait. (Courant et al. 2014; Royer et al. 2017)

Dans le cas d'une étude métabolomique, lorsque la matrice d'étude est un végétal, une première étape de séchage est parfois employée. Généralement, les échantillons sont broyés pour faciliter l'extraction des analytes et homogénéiser les échantillons. Des extractions solides-liquides sont réalisées. Le choix de la technique et du solvant d'extraction est important et va conditionner les familles de molécules extraites. Des techniques « rapides » telles que l'extraction assistée par ultra-sons (UAE : *Ultrasonic Assisted Extraction*), l'extraction assistée par micro-ondes (MAE : *Microwave Assisted Extraction*), ou l'extraction par liquide pressurisée (PLE : *pressurised liquid extraction*), sont généralement favorisées pour réaliser un très grand nombre d'échantillons (Wolfender et al. 2013) (voir Chapitre 2 section III.C Mise en place du protocole d'extraction). Quant au solvant d'extraction ce sont souvent des mélanges qui sont plébiscités afin d'extraire une gamme de polarité la plus étendue possible (voir Chapitre 2 section II Détermination du solvant d'extraction) (Mushtaq et al. 2014).

Dans le cas d'approche métabolomique ciblée, la SPE (« solid-phase extraction ») peut être utilisée afin de réaliser une adsorption sélective des composés (Wolfender et al. 2013; Mushtaq et al. 2014; Royer et al. 2017).

Comme explicité précédemment, les contenants employés lors de l'extraction peuvent relarguer des polluants. Par conséquent, une bonne pratique à adopter est la réalisation d'extraction à « blanc » qui consiste à réaliser toutes les étapes d'extraction sans la matrice solide puis d'analyser « l'extrait » obtenu. Cette étape permet d'identifier des potentiels contaminants afin de les supprimer ou, à minima, les annoter lors du retraitement des données. (Broadhurst et al. 2018)

Lorsqu'une ou plusieurs étapes telles qu'une extraction sont nécessaires, il est bienvenu de randomiser l'ordre de préparation des échantillons afin de ne pas induire un biais supplémentaire.

#### b. Cas de l'extraction des COnVs dans le genre *Lavandula*

Pour les métabolites non volatils contenus dans les espèces du genre *Lavandula*, les procédés employés sont des extractions solide-liquide classiques, à savoir la macération, l'extraction à l'aide d'un montage de Soxhlet et par liquide pressurisé. La méthode la plus usitée est celle de macération (voir Tableau 3). La matrice est généralement séchée et broyée au préalable ou l'extraction est directement effectuée sur le matériel végétal frais.

Afin d'extraire les composés phénoliques, les solvants les plus employés sont des mélanges aqueux de méthanol et d'éthanol. D'après Areias *et al.* (Areias et al. 2000), l'éthanol est dans ce cas le solvant le plus approprié. Des solvants de polarité plus faible (chloroforme), voire même apolaire (hexane) ont été utilisés pour extraire les triterpènes de la lavande.

Tableau 3 : Procédés d'extraction et méthodes analytique employés dans le genre *Lavandula*

Procédé d'extraction	Solvant	Familles de composés extraits	Références
Solide-liquide extraction (Macération)	Méthanol-Eau (30 %, 80 %, 100 %)	Coumarines, flavonoïdes, acides hydroxycinnamiques et dérivés, composés azotés, acides phénolique	(Papanov et al. 1985; El-Garf et al. 1999; Upson et al. 2000; Torras-Claveria et al. 2007)
	Ethanol-Eau (70 %, 95 %, 100 %)	Coumarines, flavones, acides phénolique, isobenzofuranes	(Areias et al. 2000; Zhou et al. 2017; Yadikar et al. 2017, 2018; Ablikim et al. 2019)
	Acétone-Eau 70 %	Dérivés de stilbènes (arylbenzofuranes)	(Tang et al. 2017)
	Acétate d'éthyle	Flavonoïdes (flavone, flavonol), dérivés de galloylglucose, acides hydroxycinnamiques et dérivés, triterpénoïdes	(Banthorpe et al. 1985; Papanov et al. 1992; Spiridon et al. 2011)
	Ether diéthylique	Anthocyanidins, coumarines, acides hydroxycinnamique et dérivés	(Banthorpe et al. 1985; Grayer et al. 2003)
	Chloroforme	Terpénoïdes (diterpénoïdes et triterpénoïdes)	(Ulubelen et al. 1988; Topçu et al. 2001)
	Hexane	Terpénoïdes	(Breton Funes and Jaudenes Ruiz de Aauri 1986)
Solide-liquide extraction (Soxhlet)	Ether diéthylique	Acides hydroxycinnamique et dérivés, acides phénoliques	(Banthorpe et al. 1985)
Extraction par liquide pressurisé	Méthanol-eau 50 % (v/v)	Acides hydroxycinnamiques, acides phénoliques, coumarines, flavonoïdes, terpénoïdes	(Algieri et al. 2016; Contreras et al. 2018)

### III.B.4. Analyse des échantillons

#### a. Généralités

Comme vu précédemment aucune technique analytique ne permet de couvrir l'intégralité du métabolome. Les deux techniques majoritairement employées pour réaliser du profilage en métabolomique sont la résonance magnétique nucléaire (RMN) et la spectrométrie de masse (MS) (Alonso et al. 2015). Elles peuvent être couplées ou non avec une technique chromatographique. D'autres méthodes comme l'infra-rouge (IR) ou les ultra-violets (UV), sont parfois employées (Engel et al. 2013). Le choix de la technique d'analyse dépendra des caractéristiques physico-chimiques de la matrice d'étude (volatilité, polarité, ...).

Un des avantages de la RMN en métabolomique est sa répétabilité. Elle présente également l'avantage d'être simple, quantitative et à haut débit. De plus, la RMN est un outil souvent utilisée pour obtenir l'élucidation structurale des biomarqueurs. Ses limites les plus considérables sont son manque de sensibilité et de résolution. Généralement, la RMN du proton est préférée en métabolomique, cependant il existe des études conduites sur le carbone 13 ou en 2D. (Wolfender et al. 2013)

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

La spectrométrie de masse, quant à elle, offre une très grande sensibilité et permet la détection de composés minoritaires même dans un mélange complexe. Avec la MS l'élucidation des biomarqueurs est souvent incomplète, seules des hypothèses d'identification peuvent être générées. L'élucidation structurale requiert d'avoir une base de données interne ou de faire appel à d'autres techniques telles que la RMN. La MS détecte une large gamme de composés, néanmoins elle n'est pas universelle car les molécules doivent être ionisables pour être analysées. La répétabilité de la technique d'analyse est un critère important en métabolomique. La spectrométrie de masse possède une bonne répétabilité sur de courtes séquences d'analyse. Il faut cependant être vigilant sur la perte de répétabilité du système avec l'allongement des séquences, ce qui est fréquent en métabolomique. La MS est généralement couplée à un système chromatographique, soit en phase gazeuse (GC) soit en phase liquide (LC) permettant la séparation des analytes. (Wolfender et al. 2013)

Il est à noter qu'une nouvelle tendance se développe depuis quelques années, celle de l'approche « multi-instrumental ». En effet, l'idéal pour un métabolicien est d'obtenir l'ensemble du métabolome dans sa large diversité, d'un point de vue technique cela nécessite de mettre en œuvre plusieurs instruments analytiques. Cependant cette approche est confrontée à des difficultés notamment celle du traitement des données. Pour y palier, des analyses dites multiblocs sont développées (Smilde et al. 2005; Martin et al. 2015; Riekeberg and Powers 2017). Un bel exemple d'étude multi-instrumentale est une étude métabolomique portant sur l'huile essentielle de citron (Mehl et al. 2014) qui a combiné trois plateformes analytiques et quatre types d'analyse : GC couplée à un détecteur à ionisation de flamme (GC-FID), UHPLC-MS (en mode positif et négatif) et de la RMN du proton ( $^1\text{H}$ ). Cette étude a développé des tests statistiques sur l'ensemble des données et a ainsi différencié les 64 échantillons étudiés en fonction des cinq protocoles d'extraction mis en œuvre. De plus, cette approche « multi-instrumentale » a permis d'obtenir une vision complète de la composition de l'huile essentielle et faciliter l'identification des composés.

### b. Bonnes pratiques d'analyse en métabolomique

#### **Randomisation des échantillons :**

Une bonne pratique à adopter en métabolomique est la randomisation des échantillons. Cette randomisation est réalisable à différentes étapes (extraction, analyse ...) mais est plus critique au moment de l'analyse des échantillons. Le principe de la randomisation consiste à analyser les échantillons dans un ordre aléatoire. Cela permet d'éviter ici qu'une variabilité se confonde avec la dérive analytique. (Gorrochategui et al. 2016; Royer et al. 2017; Reisdorph et al. 2019)

#### **Création des échantillons QC ou pool :**

Il est de plus en plus courant dans les études métabolomiques de grande envergure, c'est-à-dire nécessitant plusieurs centaines d'analyses, d'utiliser des échantillons de contrôle qualité (QC) (Broadhurst et al. 2018). La finalité de ces échantillons est de contrôler la dérive analytique au sein de chaque séquence. Ces QC doivent être constitués de façon à être représentatifs des échantillons de la séquence analytique, c'est-à-dire qu'ils doivent contenir la variabilité biologique totale afin d'obtenir une image de la variabilité du système analytique. C'est pour cette raison que les QC sont souvent un mélange de l'ensemble des échantillons de la séquence analytique. Ce mélange est habituellement désigné sous le terme de « pool ». L'échantillon est ensuite injecté régulièrement au cours de la séquence pour effectuer le suivi et la correction. La fréquence d'injection est un compromis à faire entre la qualité de la correction et la durée de la séquence analytique (augmentation du coût, de la dérive et perte de temps). Liu et al. recommandent une fréquence d'injection tous les 10-15 échantillons (Liu et al. 2019). En post-expérimental, la réalisation de tests descriptifs tels que l'ACP va permettre d'évaluer la qualité d'une séquence. En effet, plus les injections du QC apparaîtront resserrées sur le « score plot », plus l'analyse est reproductible et fiable. (Wolfender et al. 2013; Royer et al. 2017; Liu et al. 2019)

### c. Cas de la spectrométrie de masse couplée à la chromatographie en phase liquide (LC-MS)

Les approches non ciblées sont les plus fréquentes dans les études en métabolomique. Elles utilisent généralement des technologies de séparation chromatographique très efficaces couplées avec des spectromètres de masse hautes résolutions. Seul le couplage LC-MS sera détaillé mais des études ont été menées par injection directe DIMS (Direct Infusion Mass Spectrometry) (Wolfender et al. 2013, 2015). Par exemple, l'étude de Gougeon *et al.* portant sur la qualité du vin vieilli en fût de chêne a été menée par injection directe. Elle a permis de mettre en évidence un lien entre les propriétés organoleptiques du vin et la provenance géographique des chênes ayant servi à la fabrication des fûts (Gougeon et al. 2009). Cette technique présente l'avantage d'être en haut débit cependant le couplage de la spectrométrie de masse avec des techniques séparatives augmente la sélectivité et facilite le retraitement des données. De plus, les avancés techniques ont permis d'améliorer les performances des systèmes de chromatographie en phase liquide notamment en augmentant les débits avec l'HPLC et plus récemment l'UHPLC (Ultra High Performance Liquid Chromatography), conduisant à réduire fortement les temps d'analyses. (Wolfender et al. 2013; Alonso et al. 2015; Liu et al. 2019)

### i. Système chromatographique

La chromatographie est une méthode physico-chimique permettant de séparer les différents constituants d'un mélange en mettant à profit leurs différences de répartition entre une phase stationnaire et une phase mobile. (Ettre 1993)

Les systèmes de type HPLC (High Performance ou High Pressure Liquid Chromatography) sont basés sur ce principe. Un système d'HPLC est composé de plusieurs parties (McMaster 2005) :

- une partie réservoir à solvant, comprenant généralement deux ou quatre voies, un filtre et un dégazeur
- un système de pompage assurant la circulation de la phase mobile dans l'ensemble de l'HPLC à pression et débit constant
- un injecteur qui assure l'injection de l'échantillon dans le système avec un volume précis et reproductible
- une colonne contenant la phase stationnaire et souvent précédée d'une colonne de garde ou précolonne.
- selon les modèles, un four pour chauffer ou thermostatier la colonne
- un ou des détecteurs

Dans cette thèse, le système chromatographique employé est une UPLC® (Ultra Performance Liquid Chromatography) qui est une marque déposée de chez Waters.

L'UHPLC est une version plus performante de l'HPLC, les volumes morts au sein de l'instrument ont été optimisés et un débit et une pression plus élevés sont accessibles. La possibilité de supporter des pressions élevées a permis l'emploi de colonnes avec des diamètres de particules inférieurs à 2 µm. Cette amélioration confère à l'appareil une meilleure séparation, la possibilité de réduire la longueur et la largeur des colonnes, qui vont concourir à augmenter la rapidité d'analyse ainsi que diminuer la quantité de solvant employé. (Guillarme 2014)

Le système de chromatographie peut-être couplé à un ou plusieurs détecteurs. Dans notre cas, l'UPLC® était couplée à un détecteur UV à barrettes de diodes (DAD = Diode Array Detector) et à un spectromètre de masse haute résolution de type QTOF.

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

### ii. Spectromètre de masse

Un spectromètre de masse se décompose en trois grandes parties montées en série dans l'ordre suivant (voir Figure 23) :

- La source, son but est d'ioniser les molécules de l'échantillon (sous forme positive ou négative et en phase gazeuse)
- L'analyseur dont l'objectif est de séparer les ions selon leur rapport  $m/z$
- Le détecteur, qui va amplifier et convertir le flux d'ions qui sort de l'analyseur

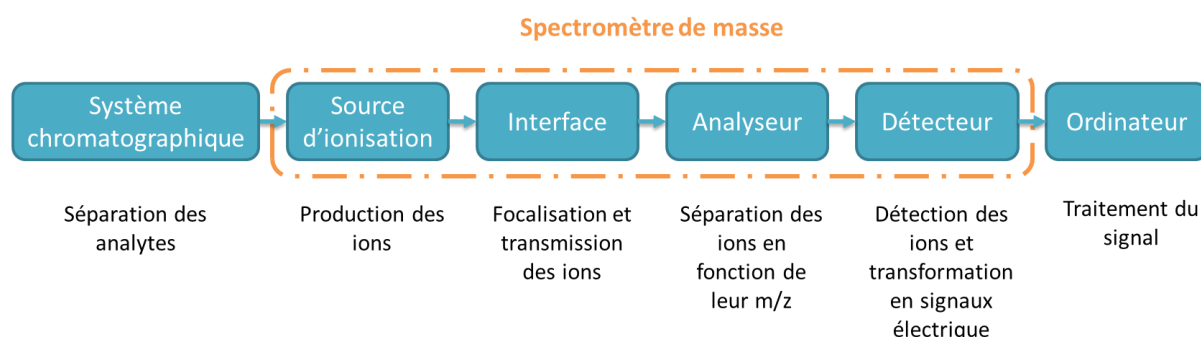


Figure 23 : Schématisation d'un couplage LC-MS (inspiré de (Banerjee and Mazumdar 2012; Bécard 2012))

Des lentilles électromagnétiques sont placées entre la source et l'analyseur, la première lentille focalise et extrait les ions de la source et la deuxième oriente le flux d'ion dans l'analyseur. A la suite du détecteur, un ordinateur permet de traiter le signal.

#### Les sources :

Les méthodes d'ionisation adaptées à la LC-MS sont l'APCI (*Atmospheric Pressure Chemical Ionization*), l'APPI (*Atmospheric Pressure Photoionization*), et l'ESI (*Electrospray Ionisation*). Elles sont toutes les trois employées en métabolomique mais la plus fréquente est l'ESI. La force de l'électronébullisateur est de couvrir une large gamme de métabolites à la fois à caractère hydrophile et hydrophobe. Il est à noter que des études commencent à apparaître avec des sources nano-ESI, qui est une technologie dérivée de la précédente et qui a l'avantage d'être plus sensible (Aszyk et al. 2018; Chetwynd and David 2018). L'ESI est adaptée à de nombreux composés mais à tendance plutôt hydrophile. Dans le cas des analytes à caractère lipophile, thermostables de poids moléculaire inférieur, les sources APCI et APPI seront à favoriser. A noter que les sources DESI (*Desorption Electrospray Ionisation*), FAB (*Fast Atom Bombardement*) et MALDI (*Matrix Assisted Laser Desorption-Ionisation*) peuvent être directement utilisées sur des matrices solides et sont aussi employées en métabolomique mais rarement appliquées aux plantes. Elles sont adaptées à des molécules de haut poids moléculaire. Dans le cas du MALDI, l'échantillon est cristallisé et ionisé par un laser. Cependant, cette technique souffre de son manque de reproductibilité. (Wolfender et al. 2013; Aszyk et al. 2018)

#### Principe de l'électronébullisation (ESI = *electrospray ionisation*)

L'ESI est une méthode d'ionisation dite indirecte, c'est-à-dire que l'ionisation se fait par désolvatation d'ions préexistants en solution. Une grande quantité d'énergie est fournie aux molécules dans un court laps de temps (afin de limiter la dégradation thermique), pour ioniser et vaporiser simultanément les analytes. Ce processus est adapté aux molécules plutôt polaires et permet de couvrir une large gamme de poids moléculaires.

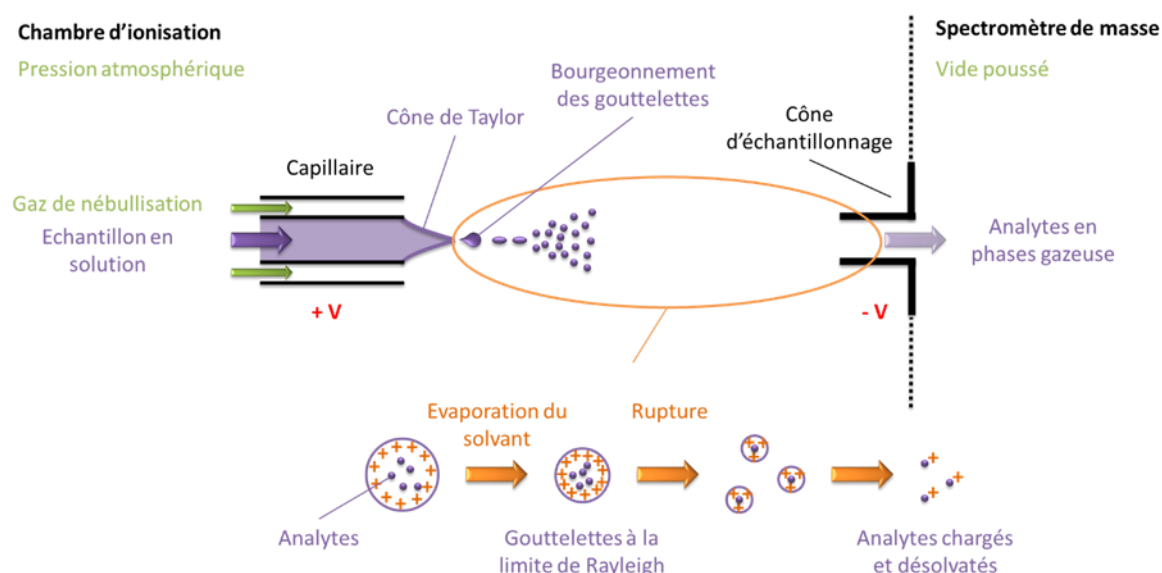


Figure 24 : Schéma d'une chambre d'ionisation et du principe d'ionisation pour une source de type ESI (inspiré de (Gaskell 1997; Banerjee and Mazumdar 2012))

L'ionisation dans un électronébuliseur se déroule à pression atmosphérique. L'échantillon en solution circule dans un tube capillaire chauffé auquel un champ électrique puissant est appliqué. Ce champ entraîne la migration des charges à la surface du liquide et l'expulsion du liquide hors du tube. Le liquide va alors former un cône dit de Taylor jusqu'à se transformer en spray. Ce phénomène est dû à l'accumulation des charges électriques à la surface du liquide occasionnant de fortes contraintes jusqu'à atteindre la limite de Rayleigh et conduire à la rupture du liquide en gouttelettes hautement chargées. La limite de Rayleigh est le point d'équilibre entre la cohésion du liquide (tension superficielle) et la répulsion des charges. Le processus est accéléré par un flux de gaz, généralement de l'azote, qui va entraîner l'évaporation du solvant provoquant un rétrécissement des gouttelettes et donc une concentration des charges jusqu'à aboutir à la désorption des ions en phase gazeuse. (Gaskell 1997; El-Aneed et al. 2009; Banerjee and Mazumdar 2012)

### Les analyseurs de masse :

Le principe de base de tous les analyseurs de masse consiste à séparer les ions selon leur rapport masse/nombre de charge ( $m/z$ ). Une fois les molécules ionisées au sein de la source, les ions sont déviés par l'application d'un champ électromagnétique. La déviation de chaque ion va dépendre de son rapport  $m/z$ . Les principaux types d'analyseurs de masse sont le quadripôle, le tube à temps de vol (TOF), la trappe ionique 2D ou 3D, l'orbitrap et le FT-ICR (*Fourier Transform Ion Cyclotron Resonance*). (El-Aneed et al. 2009)

Des études métabolomiques sont conduites sur des analyseurs haute et basse résolution. Néanmoins, les approches non ciblées favorisent des analyseurs haute résolution, ayant une rapidité d'acquisition des données et une très grande précision en masse (généralement  $< 5\text{ppm}$ ), tels que le LTQ-Orbitrap (*Linear Trap Quadrupole-Orbitrap*), le FT-ICR ou le Q-TOF (*Quadrupole-Time-Of-Flight*). Le triple quadripôle (QQQ) ou le triple quadripôle couplé avec trappe ionique (QqQ-LIT) sont quant à eux plutôt employés pour des approches ciblées. (Wolfender et al. 2013; Aszyk et al. 2018; Liu et al. 2019)

### Le QTOF :

Le QTOF est un spectromètre de masse hybride combinant un quadripôle (Q) et un tube de temps de vol (TOF).

### Quadripôle :

Le quadripôle est composé de quatre barres rondes ou hyperboliques disposées symétriquement selon le même axe qui servent d'électrodes. Les paires d'électrodes sont soumises à un potentiel électrique de signe opposé de manière à créer un champ électrique capable de guider les ions à travers l'analyseur. Les ions adoptent une trajectoire oscillante. Les variations de tension permettent de moduler l'amplitude d'oscillation des ions. Lorsque celle-ci est trop grande, les ions viennent percuter les électrodes, se déchargent et sont perdus. De cette façon, il est possible de sélectionner les ions qui traversent l'analyseur en fonction de leur rapport  $m/z$  (El-Aneed et al. 2009). Un balayage de la gamme de voltage permet la transmission de l'ensemble des masses. Le quadripôle permet soit la sélection d'un ion spécifique soit la transmission de l'ensemble des masses qui nécessite un balayage de la gamme de voltage. (Chernushevich et al. 2001; Jabot 2017; Allen and McWhinney 2019)

### Le tube à temps de vol (TOF = *time-of-flight*) :

Actuellement, le TOF est l'analyseur le plus rapide et sa très grande précision ( $10^{-4}$  Da) en fait un analyseur de masse haute résolution ou HRMS (*High-Resolution Mass Spectrometer*).

Le principe de mesure du TOF repose sur le lien entre le temps nécessaire à un objet pour parcourir une distance donnée et la masse de cet objet. A l'entrée du tube, le faisceau d'ions est accéléré par l'application d'une différence de potentielle ( $V_0$ ) dans le but de conférer une énergie cinétique à chaque ion. Les ions sont alors projetés dans le tube qui est sous vide afin de permettre leur libre parcours. Ces ions finissent leur course sur le détecteur qui va transformer le courant ionique en courant électrique. Le rapport  $m/z$  est directement proportionnel au temps de vol ( $t$ ) de l'ion, selon la formule suivante :

$$\frac{m}{z} = 2V_0 \frac{t^2}{d^2}$$

$d$ , étant la distance parcourue par l'ion qui est caractéristique de l'instrument (dépend de la longueur du tube de vol).

Les analyseurs TOF modernes ont tous introduit un « reflectron » qui est un miroir électrostatique. Son but est de dévier la trajectoire du faisceau d'ions. Il limite ainsi la dispersion des ions d'une même population et contribue à améliorer la résolution de l'analyseur. La déviation induite ne touche que les ions, le reflectron a donc un deuxième avantage celui d'éliminer les neutres (El-Aneed et al. 2009). Cette éviction augmente également la résolution des TOF la portant désormais à 20 000 en moyenne.

Les analyseurs à temps de vol n'ont en théorie aucune limite pour la gamme de masse détectée. En pratique, la gamme de masse est comprise entre 20 et 100 000 Da (pour le XEVO G2 QTOF utilisé dans le cadre de ce projet). Son deuxième avantage est de posséder une très grande sensibilité car il détecte la totalité des ions produits (absence de balayage). (Waters Corporation 2010; Lin et al. 2015; Allen and McWhinney 2019)

### **Les modes d'acquisition des données :**

Le spectromètre de masse utilisé dans cette thèse (XEVO G2 QTOF) gère trois modes d'acquisition des données : la MS, la  $MS^F$  et la MS/MS.

Le mode balayage ou « full scan » (dénommé mode MS sur le XEVO G2 QTOF) correspond à l'acquisition des spectres en continu. Les ions sont simplement transmis au détecteur qui balaye une gamme de masse définie. Ce mode sans fragmentation volontaire vise à déterminer l'ion moléculaire.

La spectrométrie de masse en tandem couramment désignée sous l'acronyme MS/MS ou  $MS^2$ , consiste à sélectionner dans un analyseur un ion cible (appelé ion parent) puis à le fragmenter (ions fils) dans une cellule de collision. Le spectre obtenu est le spectre de fragmentation de l'ion choisi. Ce mode nécessite deux analyseurs en série. (El-Aneed et al. 2009)



## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

Les deux modes d'acquisition explicités précédemment sont fréquents et présents sur de nombreux spectromètres, le troisième mode d'acquisition appelé  $MS^E$  a été développé et est spécifique au constructeur Waters. Le mode d'acquisition  $MS^E$  permet d'obtenir à la fois les spectres de masse en full scan et une fragmentation « simultanément » pour l'ensemble des analytes (sans sélection), c'est pour cette raison qu'il est parfois appelé pseudo MS/MS. La  $MS^E$  est obtenue grâce à la permutation continue et très rapide (10 spectres à la seconde) du spectromètre entre deux fonctions. Dans le premier mode ( $MS1$ ), le faisceau d'ions en sortie de la source est simplement transmis à l'analyseur en étant soumis à une faible énergie (5eV). Ce mode correspond au mode « full scan » et va permettre de visualiser les ions précurseurs. Pour le deuxième mode ( $MS2$ ), le faisceau d'ions est soumis à une rampe linéaire d'énergie élevée allant de 15 à 35 eV, afin de fragmenter les ions. ((Plumb et al. 2006) et documentations constructeurs)

A noter qu'il existe d'autres modes d'acquisition comme la  $MS^n$  nécessitant des analyseurs tel que l'orbitrap ou le piège à ions (*Ion Trap*) (El-Aneed et al. 2009; Lin et al. 2015). La  $MS^n$  ou MS multiple est la succession d'étapes de fragmentation des ions produits, avec n correspondant au nombre de génération d'ions produits. (Louris et al. 1990)

### III.B.5. Traitement des données

Une étude métabolomique génère une quantité massive de données. Le traitement est quasi-impossible de façon manuelle et requiert l'utilisation d'algorithmes et de logiciels dédiés (Courant et al. 2014). Le traitement des données regroupe trois grandes étapes : le prétraitement, les analyses statistiques et l'élucidation structurale.

#### a. Le prétraitement des données

##### i. Généralités

Le prétraitement des données regroupe l'ensemble des étapes permettant de passer des données brutes en tableau de données à deux dimensions adaptées à la réalisation des tests statistiques uni- ou multivariés. C'est une étape importante et cruciale en métabolomique qui permet de conserver l'information appropriée et rendre les données exploitables pour la suite du processus. (Wolfender et al. 2013)

La méthodologie et les logiciels employés vont dépendre de la méthode analytique utilisée. L'utilisation de techniques chromatographiques, telles que la LC ou la GC, permet de séparer les molécules d'un mélange avant la détection. Cependant, elles complexifient l'étape de prétraitement car elles introduisent une troisième dimension (le temps de rétention). Le format du jeu de données devient multidimensionnel, or pour les analyses statistiques, il doit être sous forme d'une matrice à deux dimensions. Par exemple, en LC-MS chaque ion ou analyte est caractérisé par un rapport m/z et un temps de rétention. Chaque échantillon comporte plusieurs analytes. Le jeu de données globales sera donc une matrice en trois dimensions (échantillon, rapport m/z et temps de rétention) et il devra être réarrangé pour être adapté à la suite du protocole. Dans le cas des techniques non séparatives, telles que la RMN ou l'injection MS directe, les données sont déjà en deux dimensions. Le prétraitement en est facilité. En revanche, l'identification des biomarqueurs ne pourra se faire directement et nécessitera une déconvolution (*i.e.* la séparation des pics par un algorithme) des spectres au préalable. (Wolfender et al. 2013)

##### ii. Cas du prétraitement des données générées par LC-MS

Etant donné la spécificité du prétraitement en fonction de la technique analytique, nous allons prendre le cas de données générées par spectrométrie de masse couplée à la chromatographie en phase liquide. Il comprend plusieurs étapes incluant la conversion des données, la suppression du bruit de fond, la détection et l'intégration des pics, l'alignement des pics chromatographiques, et la normalisation. (Wolfender et al. 2013)

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

La terminologie sur les différentes étapes du prétraitement ne fait pas consensus, nous avons choisi de découper le prétraitement pour l'expliquer en trois catégories. L'extraction des données spectrales qui consiste à rendre les données obtenues à l'issue des analyses en tableau en deux dimensions. Ensuite, nous ferons la différence entre le nettoyage et la transformation des données, qui eux consistent à alléger le jeu de données afin de ne visualiser que l'information biologique pertinente, notamment en supprimant les données inutiles et la variabilité non souhaitée introduite lors des étapes précédentes.

### **Extraction des données spectrales :**

Dans le cas où le traitement des données ne s'effectue pas à l'aide d'un logiciel constructeur dépendant de l'instrument analytique, les données brutes doivent être converties pour être compatibles avec le logiciel de traitement. Les formats les plus couramment acceptés par les supports informatiques de retraitement sont mzML, mzXML et netCDF pour les données obtenues par spectrométrie de masse. La plupart des logiciels constructeurs proposent des modules de conversion des données, dans le cas contraire, il existe des logiciels en libre accès comme msconvert qui est issu de la protéomique (module de proteowizard). (Spicer et al. 2017)

La première étape du traitement spectrale en LC-MS est la filtration des données inutiles générées par la variabilité technique (ex : dérive de la ligne de base, bruit de fond, ...). Ces données peuvent alourdir le traitement voire nuire à celui-ci en masquant la variabilité du facteur d'étude (Courant et al. 2014). Cette étape va consister à la suppression du bruit de fond qui a pu être généré lors de l'expérimentation. La suppression du bruit de fond repose sur le lissage du signal. Les deux algorithmes les plus couramment employés sont Savitzky-Golay basé sur l'ajustement du signal à une fonction polynomiale sur des petites portions du signal, et wavelets qui utilise le même principe que la transformée de Fourier en supprimant les hautes fréquences du signal (Engel et al. 2013). Certains logiciels permettent également de restreindre la fenêtre de masse ou de corriger la ligne de base. Habituellement, la correction de la ligne de base se fait en soustrayant au spectre la valeur de la dérive (Engel et al. 2013).

La deuxième étape est celle de la détection des pics ou « peak picking » en anglais. L'algorithme va chercher tous les pics présents dans tous les échantillons du jeu de données et intégrer ses pics afin d'obtenir leur hauteur ou leur aire (Courant et al. 2014).

La troisième étape du prétraitement est l'alignement des pics à travers les échantillons. Une fois les pics détectés dans l'ensemble des échantillons, l'algorithme va chercher à faire correspondre les pics à travers tous les échantillons. Si cette étape est mal paramétrée, elle peut conduire l'algorithme à associer deux analytes différents à un même pic ou inversement à différencier en deux pics un même composé. Sans cette étape, la validité des données n'est pas assurée. L'alignement des variables à travers les échantillons va permettre de corriger l'écart de temps de rétention entre les échantillons. (Wolfender et al. 2013) En pratique, l'algorithme va dilater ou compresser les signaux afin de les aligner. Selon l'algorithme, l'alignement se fera de façon globale c'est-à-dire sur l'ensemble du spectre, soit de façon locale en découpant le signal en petits intervalles. Dans tous les cas, les temps de calcul sont importants et peuvent même devenir problématiques en métabolomique du fait des volumes de données importants. De plus, cette quantité élevée de données complexifie le paramétrage de cette étape. (Engel et al. 2013)

Si la dérive analytique est trop importante pour permettre l'alignement correct des variables, les logiciels de traitement de données en métabolomique permettent généralement de réaliser une étape spécifique de correction des temps de rétention. (Courant et al. 2014)

A l'issue de ces étapes, le jeu de données brutes a été transformé en une matrice 2D. En théorie, le prétraitement des données pourrait s'achever là. Cependant, plusieurs étapes sont encore nécessaires pour la bonne mise en place des tests statistiques.

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

A la suite de l'étape d'alignement des pics, une étape de remplissage des données manquantes, ou « gap filling » en anglais, est généralement effectuée. Cette étape consiste à attribuer une valeur aux pics qui ne sont pas détectés dans la totalité des échantillons. Elle permet d'éviter la présence de zéro dans la matrice qui aurait un effet néfaste sur les analyses statistiques (Wolfender et al. 2013). L'algorithme va dans un premier temps rechercher les pics manquants, puis calculer la masse et le temps que devrait avoir ces pics et lui attribuer une intensité. La valeur attribuée aux pics manquants dépend de la programmation de l'algorithme. Une des méthodes employées consistent à chercher la valeur de la ligne de base de part et d'autre du pic manquant et de lui attribuer cette valeur (ex : fillpeak de Galaxy). Une autre méthode consiste à attribuer aux pics manquants une valeur correspondant à la moitié ou au tiers de l'ion en question. (Courant et al. 2014)

### **Nettoyage des données :**

En ce qui concerne le nettoyage des données, on distingue deux types de nettoyages. Les nettoyages effectués sur les variables (dans notre cas, les ions) et les nettoyages effectués sur les échantillons.

Pour le nettoyage des variables, le principe est d'enlever du jeu de données les variables qui n'apportent aucune information utile vis-à-vis du problème biologique étudié et qui vont simplement produire « un bruit de fond » non informatif dans les analyses statistiques. En effet, les isotopes, les adduits, les dimères et les fragments des ions conduisent à la redondance d'informations car ils appartiennent au même métabolite. La présence de ces ions secondaires peut perturber les analyses statistiques et ralentir l'identification des marqueurs en augmentant le nombre de variables à traiter. Il existe des outils tels que CAMERA (Kuhl et al. 2012), PUTMEDID-LCMS (Brown et al. 2011), et mzMatch (Scheltema et al. 2011) qui permettent leur annotation automatique. Les ions sont uniquement annotés par ces logiciels mais pas supprimés. (Courant et al. 2014)

Pour le nettoyage des échantillons, le principe est de supprimer un échantillon qui aurait été identifié comme aberrant. Lors de la réalisation de l'étude, l'opérateur n'est pas à l'abri d'une erreur de manipulation. Les échantillons clairement identifiés comme erronés doivent être supprimés du jeu de données, s'ils ne peuvent pas être corrigés. La suppression d'échantillon doit toujours être justifiée pour ne pas biaiser les résultats, et elle doit se faire avec parcimonie car elle est dangereuse pour la puissance statistique de l'étude. Elle peut se faire à différent moment du protocole de traitement des données. En amont du prétraitement si l'erreur est déjà connue ou après la réalisation d'analyse permettant la visualisation du jeu de données tel que l'ACP. (Xia and Wishart 2011)

### **Transformations des données :**

Les transformations des données ont pour objectifs de faire ressortir la variance due au facteur biologique étudiée. Ces étapes ne sont pas nécessaires mais fortement recommandées. Il faut cependant prendre garde à ne pas obtenir l'effet inverse en surexprimant des données inutiles.

Au terme stricto sensu, la normalisation en statistique désigne la transformation d'une série de valeur afin qu'elle se conforme à une distribution normale. En métabolomique, la normalisation a pour but de réduire la variabilité non souhaitée ou des biais systématiques afin de permettre la comparaison entre les échantillons et ce quelle que soit l'intensité des métabolites. Dans les études métabolomiques non-ciblées en LC-MS, on distingue deux types de normalisation : la normalisation en fonction des échantillons et celle par rapport aux ions (Xia and Wishart 2011; Wolfender et al. 2013). De nombreuses normalisations existent (Li et al. 2016), pour la normalisation des échantillons, on peut citer celles *via* un standard interne, des analyses QC (cf. III.B.4.b Bonnes pratiques d'analyse en métabolomique) ou un paramètre spécifique à chaque échantillon tel que le poids ou le volume. (Xia and Wishart 2011)

La normalisation par rapport aux ions a pour but de réduire l'hétéroscédasticité des ions. Les données sont généralement centrées et réduites. Le centrage des données est une translation du barycentre de chaque composé à 0, c'est-à-dire on soustrait à chaque composé sa moyenne sur l'ensemble des

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

échantillons (Courant et al. 2014; Hervé et al. 2018). La réduction consiste à diviser la valeur de chaque composé par une valeur donnée sur l'ensemble des échantillons. Ainsi, la variance de tous les composés est homogène. Elle permet de donner le même poids à chaque composé et d'éviter que les composés majoritaires pèsent plus dans l'analyse statistique que les composés minoritaires (Hervé et al. 2018). Les réductions les plus souvent employées sont « l'autoscaling », dont le facteur de division est l'écart-type, et Pareto où le facteur est alors la racine carrée de l'écart-type. La méthode Pareto a été développée pour atténuer la trop grande importance conférée aux données inutiles, telles que du bruit de fond, par la méthode « d'autoscaling ». (Courant et al. 2014)

Des transformations comme la fonction log, ou sa variante clr (« centered log ratio ») ou racine sont également employées pour homogénéiser la variance des composés (Courant et al. 2014; Hervé et al. 2018).

Tableau 4 : Les principales étapes de prétraitement pour une étude métabolomique en LC-MS

Etapes prétraitement	Description	
<b>Extraction des données spectrales</b>	Conversion	Passer les fichiers au format constructeurs dans un format compatible avec le logiciel de prétraitement (ex : netCDF)
	Filtration	Suppression de données inutiles (bruit de fond) et corrections de dérive analytique (ligne de base)
	Détection et intégration	Détection des pics et intégration des aires
	Alignement	Alignement des pics d'un même métabolite à travers les échantillons
	(Correction tr)	Correction d'une dérive analytique des temps de rétention (si alignement insuffisant)
	Gestion des données manquantes	Attribution d'une valeur (faible) aux pics non détectés
<b>Nettoyage</b>	Sur les analytes	Suppression des variables redondantes (isotopes, adduits, dimères, fragments)
	Sur les échantillons	Suppression des échantillons aberrants
<b>Normalisation</b>	Sur les échantillons	Normalisation des échantillons via un standard interne, des analyses QC ou un paramètre spécifique (poids, volume)
	Sur les analytes	Réduction de l'hétéroscédasticité des ions par centrage et réduction (autoscaling, clr, log, Pareto, ...)

A l'issue du prétraitement de données générées par LC-MS, nous obtenons un tableau à deux dimensions contenant les intensités de chaque ion intégré pour chaque échantillon. Les ions sont définis par leur rapport de masse  $m/z$  ( $M_x$ ) et leur temps de rétention ( $T_y$ ) souvent présenté sous le code  $M_xT_y$  (ex : M205T783). (Courant et al. 2014)

### Les logiciels :

La LC-MS est la technique d'analyse la plus employée en métabolomique et pour laquelle le nombre de logiciels développés pour le traitement des données est le plus important. On trouve actuellement de nombreux logiciels (voir Tableau 5) en accès libre tels que : XCMS Online, MAVEN (Metabolomics Analysis and Visualization Engine), MZmine2, MetaboAnalyst, Galaxy workflow-GalaxyM, Workflow4metabolomics (W4M). Ils permettent de réaliser l'intégralité du processus de traitement des données. Nous pouvons également citer les logiciels OpenMS, MetAlign, MS-DIAL (Mass Spectrometry-Data Independent Analysis) qui eux sont dédiés au prétraitement. (Spicer et al. 2017)

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

Tableau 5 : Plusieurs exemples de logiciels de traitement des données LC-MS en métabolomique (d'après (Spicer et al. 2017))

Etapes protocole	Applications	Nature des données d'entrée	Site internet	Références
Prétraitement	XCMS	LC-MS, GC-MS	<a href="http://bioconductor.org/packages/release/bioc/html/xcms.html">http://bioconductor.org/packages/release/bioc/html/xcms.html</a>	(Smith et al. 2006)
	OpenMS-Feature Finder Metabo	LC-MS	<a href="http://ftp.mi.fu-berlin.de/pub/OpenMS/release-documentation/html/TOPP_FeatureFinderMetabo.html">http://ftp.mi.fu-berlin.de/pub/OpenMS/release-documentation/html/TOPP_FeatureFinderMetabo.html</a>	(Bertsch et al. 2011)
	MetAlign	LC-MS	<a href="http://www.wageningenur.nl/en/show/MetAlign-1.htm">http://www.wageningenur.nl/en/show/MetAlign-1.htm</a>	(Lommen and Kools 2012)
	MS-DIAL	LC-MS	<a href="http://prime.psc.riken.jp/Metabolomics_Software/MSDIAL/index.htm">http://prime.psc.riken.jp/Metabolomics_Software/MSDIAL/index.htm</a>	(Tsugawa et al. 2015)
	mzMatch	LC-MS	<a href="http://mzmatch.sourceforge.net/index.php">http://mzmatch.sourceforge.net/index.php</a>	(Scheltema et al. 2011)
	IDEOM	LC-MS	<a href="http://mzmatch.sourceforge.net/ideom.php">http://mzmatch.sourceforge.net/ideom.php</a>	(Creek et al. 2012)
	MeltDB	LC-MS, GC-MS	<a href="https://meltdb.cebitec.uni-bielefeld.de/cgi-bin/login.cgi">https://meltdb.cebitec.uni-bielefeld.de/cgi-bin/login.cgi</a>	(Kessler et al. 2013)
Analyses statistiques	Ionwinze	LC-MS	<a href="https://sourceforge.net/projects/ionwinze">https://sourceforge.net/projects/ionwinze</a>	(Kokubun and D'Costa 2013)
	MetabolAnalyze	MS, NMR	<a href="https://cran.r-project.org/web/packages/MetabolAnalyze/index.html">https://cran.r-project.org/web/packages/MetabolAnalyze/index.html</a>	(Nyamundanda et al. 2010)
	MetaboLyser	MS, NMR	<a href="https://sites.google.com/a/georgetown.edu/fornace-lab-informatics/home/metabolizer">https://sites.google.com/a/georgetown.edu/fornace-lab-informatics/home/metabolizer</a>	(Mak et al. 2014)
	Muma	MS, NMR	<a href="https://cran.r-project.org/web/packages/muma/">https://cran.r-project.org/web/packages/muma/</a>	(Gaude et al. 2013)
	SIMCA-P	MS, NMR	<i>Logiciel commercial (Umetrics)</i>	(Triba et al. 2015)
	ropls	MS, NMR	<a href="https://www.bioconductor.org/packages/release/bioc/html/ropls.html">https://www.bioconductor.org/packages/release/bioc/html/ropls.html</a>	(Thévenot et al. 2015)
Multi-étapes	Workflow4metabolomics	LC-MS, GC-MS	<a href="http://workflow4metabolomics.org">http://workflow4metabolomics.org</a>	(Giacomoni et al. 2015)
	Galaxy-M	LC-MS, DIMS	<a href="https://github.com/Viant-Metabolomics/Galaxy-M">https://github.com/Viant-Metabolomics/Galaxy-M</a>	(Davidson et al. 2016)
	XCMS Online	LC-MS, GC-MS	<a href="https://xcmsonline.scripps.edu/landing_page.php?pgcontent=mainPage">https://xcmsonline.scripps.edu/landing_page.php?pgcontent=mainPage</a>	(Tautenhahn et al. 2012)
	MetaboAnalyst 4.0	LC-MS	<a href="http://www.metaboanalyst.ca">http://www.metaboanalyst.ca</a>	(Chong et al. 2019)
	MAVEN	LC-MS	<a href="http://genomics-pubs.princeton.edu/mzroll/index.php">http://genomics-pubs.princeton.edu/mzroll/index.php</a>	(Melamud et al. 2010)
	MAIT	LC-MS	<a href="https://www.bioconductor.org/packages/release/bioc/html/MAIT.html">https://www.bioconductor.org/packages/release/bioc/html/MAIT.html</a>	(Fernández-Albert et al. 2014)
	MZmine 2	LC-MS	<a href="http://mzmine.github.io/">http://mzmine.github.io/</a>	(Pluskal et al. 2010)

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

Le tableau n'est pas exhaustif et ne comporte que des logiciels adaptés aux analyses LC-MS. Pour plus d'information <https://raspicer.github.io/MetabolomicsTools/>.

### b. Analyses statistiques

#### i. Généralités

Les tests statistiques peuvent être uni- ou multivariés. Les tests univariés tels que le test de *Student* sont moins usités en métabolomique. Nous aborderons dans cette partie uniquement les tests statistiques multivariés, c'est-à-dire les tests réalisés sur des données multidimensionnelles.

Le but des analyses statistiques en métabolomique est d'évaluer l'impact du facteur biologique étudié sur les valeurs des données métabolomiques. Comme pour les autres étapes, le choix des tests statistiques dépend de la nature des données et de ce que l'on souhaite mettre en évidence donc de la question biologique de départ. On distinguera deux grands types de tests :

- Les tests descriptifs qui permettent de visualiser le jeu de données qui rappelons-le est une matrice de grande dimension (ACP, heatmap)
- Les tests discriminants qui permettent de séparer au mieux le jeu de données en fonction des métadonnées de ce dernier (PLS-DA)

Les métadonnées sont les données relatives aux échantillons, cela peut-être la nature du prélèvement (sang, urine, ...), le sexe, l'âge de l'individu prélevé, l'organe de la plante... Le tableau regroupant les métadonnées est souvent appelé matrice de design (Y), en parallèle à la matrice de données (X).

#### ii. Les analyses multivariées descriptives (non supervisées)

Les analyses descriptives permettent de visualiser le jeu de données. Elles sont généralement utilisées en métabolomique comme première approche du jeu de données. Elles servent également à la correction en visualisant les éventuelles données aberrantes présentes. Elles ont aussi un rôle dans l'évaluation de la qualité des données obtenues avec la stratégie des QC. Enfin, elles permettent de faire des corrélations et de se donner un premier aperçu sur les liens potentiels entre les variables.

Les analyses descriptives les plus connues en métabolomique sont certainement l'analyse en composante principale (ou ACP), et la classification hiérarchique ascendante permettant la réalisation de dendrogramme (Wolfender et al. 2013).

#### **ACP :**

L'ACP est une méthode de réduction du jeu de données (multivariées) par projection de celui-ci sur deux axes afin de le rendre visualisable (2D). La méthode va chercher les deux axes (ou composantes) les « plus représentatifs », c'est-à-dire de façon à maximiser la variance, puis projeter le jeu de données sur ces deux axes. Elle fournit deux graphiques principaux :

- Le « score plot » ou graphique des individus représentant les échantillons
- Et le « loading plot » représentant les variables, dans notre cas les ions

Elle fournit également un « biplot » qui est la superposition des deux graphiques précédents.

Il est courant de trouver un histogramme appelé éboulis des valeurs propres où chaque barre représente un axe et la hauteur des barres est proportionnelle à la variance portée par l'axe en question. Cet histogramme permet de choisir le nombre d'axes à prendre en compte. En pratique, l'étude est généralement restreinte à deux axes pour visualiser de façon simple les graphiques.

### iii. Les analyses multivariées discriminantes (supervisées)

Il existe de nombreuses méthodes d'analyses supervisées. Elles s'appuient sur le principe de la régression des moindres carrés. Plusieurs sont employées en métabolomique comme par exemple l'O-PLS ou l'O2-PLS mais la plus courante est la PLS-DA (*Partial Least Squares-Discriminant Analysis*). (Lê Cao and Le Gall 2011; Wolfender et al. 2013)

#### PLS-DA :

La PLS-DA est comme l'ACP une méthode de projection du jeu de données, c'est la manière dont les axes sont construits qui diffère. Dans le cas de la PLS-DA, les composantes sont calculées afin de maximiser la co-variance, c'est-à-dire de trouver le plan de projection qui maximisera la séparation des groupes de l'étude. Le test va permettre de discriminer les groupes. (Worley and Powers 2013; Gromski et al. 2015)

La PLS-DA a deux objectifs : discriminer les groupes et prédire l'appartenance à ces groupes. En écologie chimique seul l'aspect descriptif de la PLS-DA est pris en compte. L'aspect prédictif de l'analyse est, quant à lui, employé notamment en médecine, par exemple pour prédire si un patient (échantillon) est malade (groupe d'appartenance) en fonction d'une prise de sang (profil chimique). (Hervé et al. 2018) Comme l'analyse a deux aspects différents, les indicateurs de validation du modèle vont revêtir eux aussi plusieurs aspects. Nous allons trouver l'indicateur  $R^2$  qui va être caractéristique de l'erreur de description du modèle et l'indicateur  $Q^2$  qui va être caractéristique de l'erreur de prédiction du modèle. Comme pour la PLS-DA, nous avons la matrice des données (X) et la matrice de design (Y), nous allons retrouver quatre indicateurs différents ( $R^2X$ ,  $R^2Y$ ,  $Q^2X$  et  $Q^2Y$ ) correspondants aux indicateurs précédents relativement à chaque matrice. C'est pour cette raison que l'ACP qui ne dépend que de X et qui est uniquement descriptive ne possède que  $R^2X$ , noté  $R^2$ , comme indicateur de validation du modèle. (Tenenhaus 1998) Tous les indicateurs ont des valeurs comprises entre 0 et 1 et le modèle est d'autant plus valide que les indicateurs tendent vers 1. De plus,  $Q^2$  sera toujours inférieur à  $R^2$ . (Saint-Lary 2015) Aucune valeur seuil de ces indicateurs ne fait consensus, d'un point de vue statistique pur le seuil de validation est généralement de 0,5. (Worley and Powers 2013; Gromski et al. 2015)

Pour la PLS-DA, l'indicateur important est  $Q^2Y$  qui est généralement abrégé en  $Q^2$ . Il est calculé en réalisant une validation croisée. Le principe de la validation croisée est de réaliser une nouvelle PLS sur le jeu de données initial retranché d'un échantillon et de comparer la réponse prédite par rapport à la réponse réelle. L'opération est répétée pour chaque échantillon et jusqu'à passer l'ensemble du jeu de données. (Szymańska et al. 2012; Worley and Powers 2013)

La PLS-DA est sensible au problème de sur-apprentissage du modèle, c'est-à-dire lorsque le modèle va réussir à séparer deux catégories de manière inopinée. Le cas peut survenir lorsque du bruit de fond subsiste et est fortuitement corrélé au groupe de la matrice de design (Y). La PLS-DA va alors créer un espace dans lequel les catégories de l'étude seront distinctes. La validation croisée n'est pas toujours suffisante pour détecter un sur-apprentissage, des tests de permutation peuvent être menés dans ce sens. Ils consistent à classer les échantillons de façon aléatoire dans les différents groupes puis à réaliser la PLS-DA avec ces nouvelles métadonnées, et à réitérer l'opération généralement entre 100 et 1000 fois. Les résultats des PLS-DA obtenus sont comparés à la PLS-DA faite sur le jeu de données initial qui sert de référence. Si la qualité des PLS-DA est basse et éloignée de celle de référence alors la séparation des groupes reflète bien une différence biologique et n'a pas été obtenue de façon fortuite. Il n'y a donc pas de sur-apprentissage du modèle. (Xia and Wishart 2011)

Les notations employées ci-dessus sont celles de la théorie statistique. En pratique, la dénomination des indicateurs et leur seuil de validation dépendent souvent du logiciel employé, d'autres indicateurs sont parfois proposés. Par exemple, le package « mixomics » sous R propose le pourcentage d'erreur de classification ou d'assignation (CER) pour valider la qualité de prédiction du modèle (Szymańska et al. 2012).

Une fois le modèle validé, un indicateur nommé VIP (variable importance in the projection) permet de connaître les variables initiales (dans notre cas, les ions) qui sont les plus impactantes dans la discrimination des groupes. Le VIP est calculé pour chaque ion et prend en compte le poids de cet ion sur un axe donné et le poids de cette même composante dans le modèle. (Xia and Wishart 2011)

### c. Annotation, identifications chimiques des métabolites et réseaux moléculaires

L'élucidation structurale est l'une des étapes les plus ardues et chronophages en métabolomique. Cette remarque est particulièrement vraie pour les études conduites par LC-MS. En effet, la technique ne dispose pas de bases de données de grande envergure et largement accessibles, contrairement à la GC-MS qui dispose elle de vastes bases de données telles que les bibliothèques NIST ou WILEY. Quant à la RMN, elle est par nature une technique analytique de caractérisation totale des composés et bénéficie de base de données (Wolfender et al. 2013; Spicer et al. 2017). L'identification des biomarqueurs est d'autant plus ardue chez les plantes car les métabolites secondaires sont souvent spécifiques à chaque espèce et qu'il n'existe aucune base de données de grande envergure et accessible (Wolfender et al. 2013).

#### Différents degrés d'identification en MS :

L'identification d'un composé totalement inconnu peut être putative ou totale. Pour qu'un composé soit dit identifié en spectrométrie de masse, il doit être comparé à un standard de grade analytique dans les mêmes conditions de façon stricte (même laboratoire, même machine et même paramétrage). Toute autre situation conduit à une identification partielle ou putative de la molécule. Quatre niveaux d'identification ont été définis par le « Metabolomics Standards Initiative » (MSI) (Spicer et al. 2017). Le niveau 4 correspond à une molécule inconnue, le niveau 3 à l'attribution à une famille chimique, le niveau 2 est la formulation d'une hypothèse d'identification, enfin le niveau 1 correspond à l'élucidation structurale complète explicité précédemment. Plusieurs classifications existent, dont la classification de Schymanski qui se décompose en 5 niveaux plus détaillés et vise les études faites par HRMS (Schymanski et al. 2014).

#### Processus d'identification en LC-MS :

La première étape pour l'élucidation structurale en LC-MS est d'identifier l'ion moléculaire des spectres de masse. Dans le cas de l'ESI de nombreux adduits sont produits (tel que  $[M+H]^+$ ,  $[M+Na]^+$  ou  $[M+H-H_2O]^+$  dans le cas d'une ionisation en mode positif) et peuvent souvent aider à la recherche de l'ion moléculaire. De façon à identifier sans ambiguïté l'ion moléculaire, la meilleure solution reste encore de comparer les modes d'ionisation positif et négatif (Wolfender et al. 2015). Ensuite à partir de la masse de l'ion moléculaire et de la correspondance avec le massif isotopique, une formule brute est proposée. Une bonne connaissance de la matrice d'étude peut aider à réduire les atomes potentiellement présents et faciliter la détermination de la formule brute des molécules. Généralement les logiciels constructeurs possèdent des modules proposant des formules brutes en fonction du massif isotopique et de la masse de l'ion. Une autre facilité est fournie par la haute résolution (HRMS) avec l'obtention d'une masse plus précise. Une fois la formule brute obtenue, elle peut être confrontée à des bases de données (cf. Tableau 6) afin de générer des hypothèses de formule développée. Une fois encore la connaissance de la matrice est une aide précieuse afin d'écartier certaines structures. L'étude des spectres de fragmentation est également un outil pour réduire le nombre de structure potentielle, ainsi que l'utilisation de la règle de l'azote et le calcul du nombre d'insaturation à partir de la formule brute. Si une hypothèse structurale est formulée, elle doit être validée par l'injection du standard analytique commercial comme expliqué précédemment. Si l'hypothèse n'est pas valide ou que le standard analytique commercial n'existe pas ou encore qu'aucune structure putative n'a pu être émise, il faudra isoler le composé afin d'effectuer des analyses complémentaires (UV, IR, RMN, ...). (Wolfender et al. 2013; Courant et al. 2014)



## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

Tableau 6 : Bases de données adaptées aux données MS et aux produits naturels (d'après (Wolfender et al. 2015))

Base de données	Types de métabolites	Nombre de composés	Données complémentaires			Site internet	Références
			MS <sup>2</sup>	UV	RMN		
AntiBase	Métabolites secondaires	42000	oui	oui	<sup>13</sup> C		(Laatsch 2014)
ChEBI	Métabolites primaires et secondaires	32000	x	x	x	<a href="http://www.ebi.ac.uk/chebi">http://www.ebi.ac.uk/chebi</a>	(Degtyarenko et al. 2008; Hastings et al. 2013)
Dictionary of Marine Natural Products	Métabolites secondaires d'origine marine	30000	x	oui	x	<a href="http://dmnp.chemnetbase.com/">http://dmnp.chemnetbase.com/</a>	
Dictionary of Natural Products	Métabolites secondaires	260000	x	oui	x	<a href="http://dnp.chemnetbase.com/">http://dnp.chemnetbase.com/</a>	
<i>E. coli</i> Metabolome Database (ECMDB)	Métabolites primaires et secondaires	3000	oui	x	oui	<a href="http://www.ecmdb.ca">http://www.ecmdb.ca</a>	(Sajed et al. 2016)
Golm Metabolome	Métabolites primaires et secondaires	2200	x	x	x	<a href="http://gmd.mpimp-golm.mpg.de/">http://gmd.mpimp-golm.mpg.de/</a>	(Kopka et al. 2005)
Human Metabolome database (HMDB)	Métabolites primaires et lipides	42000	oui	x	x	<a href="http://www.hmdb.ca/">http://www.hmdb.ca/</a>	(Wishart et al. 2007)
KNAPSAcK	Métabolites secondaires	51000	x	x	x	<a href="http://kanaya.naist.jp/KNAPSAcK_Family/">http://kanaya.naist.jp/KNAPSAcK_Family/</a>	(Shinbo et al. 2006)
Kyoto Encyclopedia of Genes and Genomes (KEGG)	Métabolites primaires et secondaires	17000	x	x	x	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>	(Kanehisa and Goto 2000)
Lipid Maps	Lipides	38000	oui	x	x	<a href="http://www.lipidmaps.org">http://www.lipidmaps.org</a>	(Fahy et al. 2007)
Madison Metabolomics Consortium Database	Métabolites primaires et secondaires	20000	x	x	oui	<a href="http://mmcd.nmrfam.wisc.edu/">http://mmcd.nmrfam.wisc.edu/</a>	(Cui et al. 2008)

## Chapitre 1 : Synthèse bibliographique et positionnement du sujet de thèse

MarinLit	Métabolites secondaires	5000	x	oui	<sup>1</sup> H	<a href="http://pubs.rsc.org/marinlit">http://pubs.rsc.org/marinlit</a>	
MassBank	Métabolites primaires et secondaires	15000	oui	x	x	<a href="http://www.massbank.jp/">http://www.massbank.jp/</a>	(Horai et al. 2010)
MetaboLights	Métabolites primaires et secondaires	11000	oui	x	oui	<a href="http://www.ebi.ac.uk/metabolights">http://www.ebi.ac.uk/metabolights</a>	(Haug et al. 2013)
MetaCyc	Métabolites primaires et secondaires	11000	x	x	x	<a href="http://www.metacyc.org">http://www.metacyc.org</a>	(Caspi et al. 2020)
MetLin	Métabolites primaires et secondaires	240000	oui	x	x	<a href="http://metlin.scripps.edu/index.php">http://metlin.scripps.edu/index.php</a>	(Smith et al. 2005)
MFSearcher	Flavonoïdes	/	x	x	x	<a href="http://webs2.kazusa.or.jp/mfsearcher/">http://webs2.kazusa.or.jp/mfsearcher/</a>	(Sakurai et al. 2013)
MZedDB	Métabolites primaires et secondaires	/	x	x	x	<a href="http://maltese.dbs.aber.ac.uk:8888/hrmet/index.html">http://maltese.dbs.aber.ac.uk:8888/hrmet/index.html</a>	(Draper et al. 2009)
NIST	Métabolites primaires et secondaires	7000	oui	x	x	<a href="https://webbook.nist.gov/chemistry/">https://webbook.nist.gov/chemistry/</a>	
Plant Metabolic Network (PMN)	Métabolites primaires et secondaires	4000	x	x	x	<a href="https://plantcyc.org/">https://plantcyc.org/</a>	(Schlöpfer et al. 2017)
Seaweed Metabolite Database	Métabolites secondaires	1000	x	x	x	<a href="https://www.swmd.co.in/">https://www.swmd.co.in/</a>	(Davis and Vasanthi 2011)
SuperNatural II	Métabolites primaires et secondaires	325000	x	x	x	<a href="http://bioinf-applied.charite.de/supernatural_new/index.php?site=home">http://bioinf-applied.charite.de/supernatural_new/index.php?site=home</a>	(Banerjee et al. 2015)
RIKEN	Métabolites secondaires	3500	oui	x	x	<a href="http://spectra.psc.riken.jp/">http://spectra.psc.riken.jp/</a>	(Sawada et al. 2012)
Yeast Metabolome database (YMDB)	Métabolites primaires et secondaires	2000	oui	x	oui	<a href="http://www.ymdb.ca">http://www.ymdb.ca</a>	(Ramirez-Gaona et al. 2017)

### Les réseaux moléculaires :

Pour faciliter la recherche des structures, des systèmes de modélisation de réseaux sont en plein essor. Le principe de base est la modélisation graphique de la proximité structurale des composés, cela se traduit par un ensemble de nœuds et d'arêtes, représentant respectivement les composés et leurs liens, plus ces derniers sont courts plus les deux molécules sont proches.

Ces réseaux ont plusieurs finalités, ils permettent à la fois de détecter des ensembles d'analogues structuraux, de formuler des hypothèses d'identification à partir d'une molécule de référence, mais aussi de comparer des analyses MS<sup>2</sup> de différents échantillons. Il existe à l'heure actuelle plusieurs plateformes en ligne permettant la création de ces réseaux comme GNPS (Global Natural Products Social Molecular Networking) (Wang et al. 2016), MetGem (Olivon et al. 2018) ou MetExplore (Cottret et al. 2018). La première étape pour générer un réseau moléculaire est l'acquisition de données MS<sup>2</sup>. Les spectres peuvent ensuite être injectés sur une plateforme. La plateforme GNPS se base sur le fait que les analogues structuraux vont avoir des fragmentations proches et par conséquent donner soit des fragments, soit des pertes neutres similaires. La plateforme va aligner les spectres du jeu de données et comparer deux à deux tous les spectres. Elle va ensuite attribuer un indice de corrélation (« cosine score » ou CS). Cet indice est compris entre 0 et 1 sachant qu'une valeur de 1 correspond à des spectres identiques. Il est communément admis que les spectres ayant un CS > 0,7 sont considérés comme ayant de forte similarité, et est donc employée comme valeur seuil pour générer la cartographie du réseau. Le réseau peut ensuite être annoté à partir des banques de données connectées à la plateforme en ligne ou manuellement. (Nothias-Scaglia et al. 2015; Fox Ramos et al. 2019)

En métabolomique, les réseaux moléculaires sont utilisés afin de faciliter l'élucidation structurale des molécules car ils permettent de découper en famille de composés le métabolome étudié. Si dans un cluster une molécule est identifiée, elle va potentiellement permettre d'identifier les autres molécules du cluster et ainsi élucider un biomarqueur qui était au départ complètement inconnu. (Demarque et al. 2020)

### IV. Objectifs de la thèse

Cette thèse s'inscrit dans un large projet visant à donner des moyens de lutte aux agriculteurs de la filière lavandicole durement touchés par le dépérissement du Stolbur. Les études conduites par les partenaires ces vingt dernières années ont permis d'avancer sur la compréhension de cette maladie, notamment en identifiant le pathogène majeur (*Ca. Phytoplasma solani*), puis l'insecte vecteur la cicadelle (*H. obsoletus*) et son cycle de vie. Ils ont également mis en avant des variétés de lavande plus ou moins tolérantes au Stolbur et corrélé cette sensibilité avec le nombre d'insectes présents en champ. Malgré ces avancées, de nombreuses zones d'incompréhension persistent :

- Une méconnaissance du phytoplasme et de son influence sur la plante et l'insecte. Notamment la question de la manipulation de l'hôte (de la plante et/ou de l'insecte vecteur) se pose.
- Une méconnaissance du comportement de l'insecte
- Une méconnaissance de la chimie de la plante, de ces réactions face à l'infection. Est-ce qu'elle développe des mécanismes de défense et quels sont-ils ?


La résolution de ses problématiques se heurte à la difficulté de mettre en place un élevage de cicadelles viable et à la maîtrise de l'infection des plants.

Cette thèse s'inscrit dans ce large projet et vise à donner des éclaircissements sur les interactions tripartites entre la lavande, *H. obsoletus* et *Ca. Phytoplasma solani*. Elle prend en charge l'action n°2 consistant à réaliser une étude métabolomique des CONVs, dans le but de caractériser des molécules impliquées dans les phénomènes de défenses de la plante induits à la suite de l'infection par le phytoplasme. L'objectif est donc de mettre en place une étude métabolomique non ciblée adaptée à la matrice d'étude (la lavande et le lavandin), dans le but de :

- Identifier des biomarqueurs de l'état sanitaire des plants (symptomatique/asymptomatique)
- Identifier des biomarqueurs du caractère sensible ou tolérant des variétés face au dépérissement

Pour ce faire, nous avons développé chaque étape comprenant en premier lieu la préparation des échantillons et leur analyse par UPLC-HRMS (chapitre 2), le traitement des données avec l'extraction des données spectrales, le nettoyage et la correction des données avant la réalisation des tests statistiques multivariés (chapitre 3). Dans une dernière partie seront présentés et discutés les résultats obtenus (chapitre 4).





## Chapitre 2 : Conception et optimisation de l'étude métabolomique

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

L'objet de ce chapitre est la mise en place d'un protocole expérimental permettant de préparer et analyser les échantillons en satisfaisant les conditions nécessaires au bon déroulement de l'étude métabolomique non ciblée. Concrètement, ces exigences se traduisent par :

- L'obtention de profils chimiques les plus riches possible d'un point de vue quantitatif et qualitatif ;
- L'obtention d'un protocole robuste, fiable, répétable et juste. En pratique, la variabilité technique doit être idéalement absente, ou à défaut réduite au maximum.

Hormis les problèmes inhérents à la mise en place d'une étude métabolomique, deux difficultés plus particulières à cette étude se posent, à savoir : la méconnaissance des analytes dans le sujet d'étude (lavande/lavandin) et l'échantillonnage en milieu non contrôlé. En effet, la connaissance sur les CONVs dans le genre *Lavandula* au début de cette thèse était restreinte. Néanmoins, des travaux sont venus enrichir ces connaissances durant ces trois dernières années. A notre connaissance, aucune étude métabolomique n'a été conduite sur CONVs de la lavande ou du lavandin.

A l'heure actuelle, les élevages de cicadelles ne sont pas viables et les paramètres d'infection ne sont pas contrôlés, ce qui conduit à la deuxième difficulté de ce sujet à savoir l'obligation d'échantillonner en champs, c'est-à-dire dans un milieu ouvert et donc non contrôlé. Cette contrainte impose de prélever un grand nombre d'échantillons afin de « s'affranchir » de la variabilité biologique. Cette quantité d'échantillon importante induit des problèmes dans le traitement et l'analyse des échantillons (gestion de la dérive analytique), et impose d'adopter un protocole simple nécessitant peu d'étapes afin d'assurer la répétabilité du processus.

Pour ce faire, dans un premier temps des tests préliminaires ont été effectués avec des plants de lavandin grosso supposés sains achetés en jardinerie afin de définir le protocole de préparation des échantillons. A partir de ces premiers essais, le protocole et les conditions analytiques ont été optimisés cette fois-ci sur des échantillons issus de la campagne d'échantillonnage.

Ce chapitre va se décomposer en cinq parties. Dans un premier temps, nous expliciterons le design de l'échantillonnage et présenterons la campagne de récolte réalisée. Dans une seconde partie, nous aborderons la sélection du solvant d'extraction. Puis dans un troisième temps, nous allons nous intéresser à la détermination du protocole de préparation des échantillons. Ensuite, nous verrons dans une quatrième partie le développement analytique ainsi que le contrôle de la dérive analytique. Enfin, il sera présenté dans une dernière partie le design expérimental final de l'étude.

### I. Echantillonnage

#### I.A. Design de l'échantillonnage

Le design de l'échantillonnage a été conceptualisé en amont de l'étude avec les partenaires (biologistes et agronomes). Comme expliqué précédemment l'échantillonnage doit satisfaire la puissance statistique, pour cela des triplicats sont généralement admis.

Du fait de la non-maitrise de l'infection des plants par le phytoplasme et d'un élevage de *H. obsoletus* sur lavande non opérationnel, les prélèvements ont été conduits directement en champs. Les terrains employés pour la récolte sont cultivés et exploités par des agriculteurs. Le choix de la matrice s'est donc porté sur les parties aériennes, même s'il aurait été intéressant de conduire une étude sur les racines de la plante. De plus, des tests préliminaires nous ont permis de montrer que les parties boisées ne présentaient pas d'intérêt. Leur profil chimique est plus pauvre (quantitativement et qualitativement) et sans informations supplémentaires par rapport à celui obtenu sur les feuilles (voir partie : Détermination du solvant d'extraction). Les parties boisées ont donc été écartées de l'étude métabolomique.

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

Nous avons fait le choix de séparer les parties aériennes en deux échantillons : un de type « feuille » contenant les tiges chlorophylliennes et les feuilles, et un comprenant uniquement les inflorescences du prélèvement. A priori, ces deux matrices végétales ont des empreintes chimiques qui diffèrent. Leur séparation permet de s'affranchir de la part de variabilité liée à la différence d'organes au sein de la plante. D'autre part, il est probable que chacune joue des rôles distincts dans l'attractivité et l'appétence de *H. obsoletus*. De plus, si la plante met un mécanisme de défense en place il est intéressant d'étudier la composition des parties chlorophylliennes séparément car ce sont les lieux de ponction alimentaire de la cicadelle. Or si la plante met un mécanisme de défense en place ce sera probablement le premier lieu où les modifications s'effectueront. Quant aux inflorescences, il est possible qu'elles entrent dans l'attraction préférentielle des cicadelles entre deux plants ou deux variétés que ce soit au niveau visuel (pigments) ou olfactif (molécules volatiles).

Le choix des espèces étudiées s'est porté sur la lavande et le lavandin car ce sont celles qui ont la valeur économique la plus importante. Afin d'étudier la sensibilité face au phytoplasme, deux variétés dites sensibles et deux tolérantes ont été sélectionnées. Pour rappel, les variétés dites « tolérantes » sont pour la lavande et le lavandin respectivement maillette et grosso. De plus, la lavande est plus sensible au dépérissement à phytoplasme que le lavandin. Dans le but d'étudier les variations de composition chimique en fonction du statut sanitaire de la plante, 10 plants symptomatiques et 10 plants asymptomatiques ont été sélectionnés pour chaque variété.

Le choix a été fait de réaliser 10 réplicats biologiques inter plants, pour anticiper la variabilité biologique et se prémunir de nombreux faux positifs au phytoplasme et ainsi assurer un nombre de réplicats suffisamment élevé. En ce qui concerne les réplicats biologiques intra plants, trois échantillons ont été prélevés (a, b et c). Le prélèvement de ces triplicats au sein d'un même plant, a été réalisé à des points distincts de la plante car nous méconnaissions la circulation et la répartition du phytoplasme au sein de la plante. De plus, nous ignorons si la plante a une réponse chimique globale ou locale à l'infection. Quand cela était possible, un échantillon est prélevé de part et d'autre du rang et un échantillon au milieu du pied.

Le design d'échantillonnage représente un total de 480 échantillons collectés pour chaque campagne (2017 et 2018).

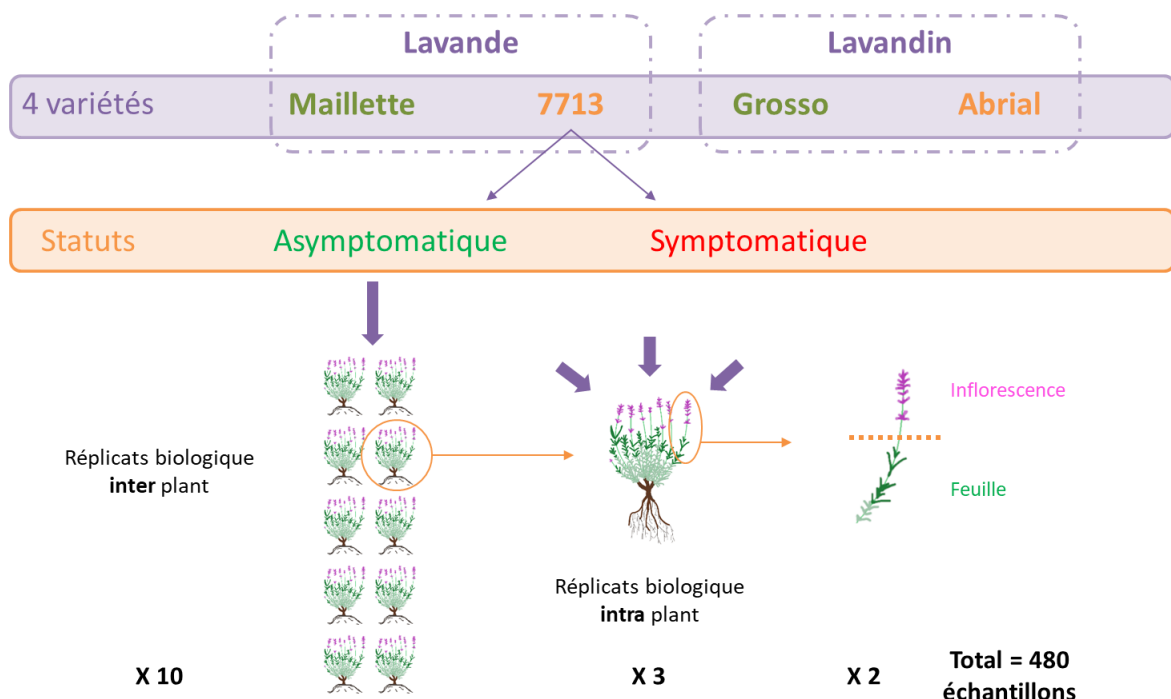


Figure 25 : Schématisation de la stratégie d'échantillonnage réalisée durant cette étude



### I.B. Campagne d'échantillonnage

Les parcelles de lavande et de lavandin utilisées sont situées sur la commune de Monieux dans le Vaucluse (84). Afin de ne pas induire de variabilité due à la zone géographique, les parcelles ont été sélectionnées les plus rapprochées possible. Les parcelles de lavande (7713 et Maillette) distantes de 100 m sont situées à 850 m d'altitude (coordonnées : 44.062, 5.34). Les deux variétés de lavandin étaient quant à elles sur la même parcelle à une altitude de 665 m (coordonnées : 44.08, 5.37). Les parcelles de lavande et celle de lavandin sont distantes d'approximativement 3,3 km.

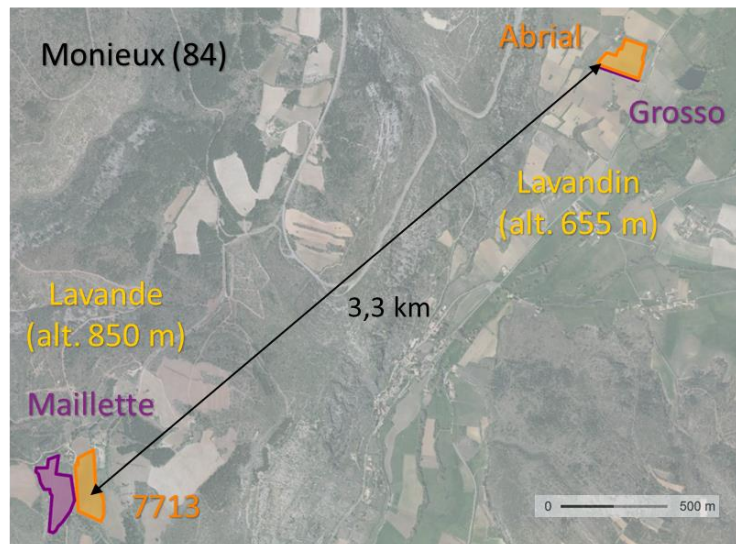


Figure 26 : Localisation des sites d'échantillonnage de lavande et de lavandin

Deux campagnes d'échantillonnage ont été réalisées en 2017 puis en 2018. Les prélèvements ont été effectués sur les mêmes parcelles et quand cela était possible (cf. mortalité des plants infectés) sur les mêmes plants. Les résultats présentés dans cette étude portent sur la campagne de 2017. Toutes les parcelles ont été prélevées sur deux jours (4 et 5 juillet 2017). Les deux jours de campagnes se sont déroulés dans des conditions météorologiques similaires, à savoir chaud et sec (Figure 27).

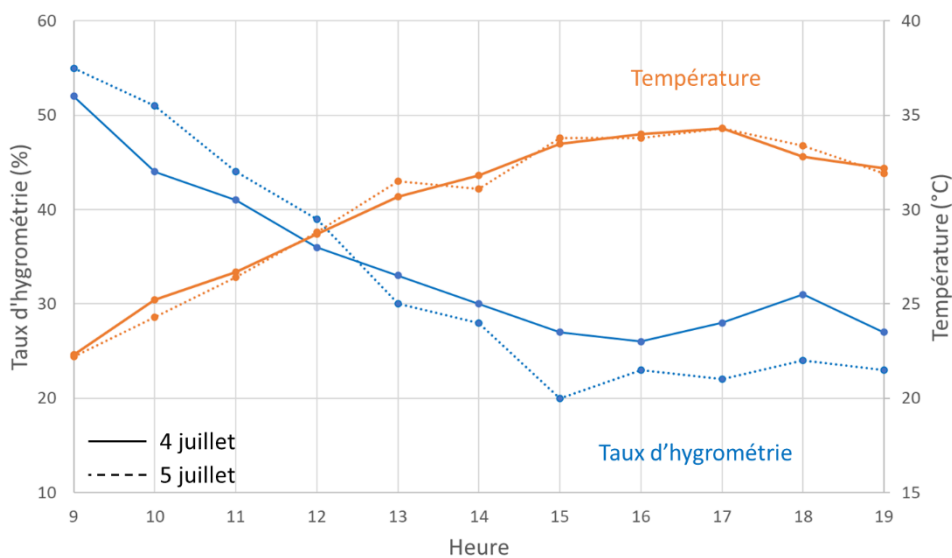


Figure 27 : Variation de la température (°C) en orange, et du taux d'hygrométrie (%) en bleu, durant les heures d'échantillonnages du 4 et 5 juillet 2017 relevés à la station de Carpentras (Vaucluse, France) situé à 25 km du lieu de prélèvement (Barbery 2017).

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

La campagne d'échantillonnage a été effectuée dans une période restreinte pour faire coïncider au maximum le pic de floraison de la lavande et du lavandin (le pic de floraison du lavandin précède celui de la lavande). A noter que cette contrainte est à la fois un avantage et un inconvénient pour l'étude. La courte période de récolte permet de limiter la variabilité dans le temps mais ne permet pas de respecter le cycle circadien de la plante. En d'autres termes, il est possible qu'une variabilité ait été induite car la chimie des plantes varie au cours de la journée. Cependant, les prélèvements étant effectués en milieu ouvert, donc non contrôlé, il n'était pas non plus judicieux d'étaler la campagne dans le temps. (Wolfender et al. 2013)

Lors de la récolte, le matériel végétal est manipulé avec des gants en nitrile pour éviter une éventuelle contamination par des éléments se trouvant sur les mains de l'opérateur. Les échantillons sont coupés sur la plante à l'aide d'un sécateur stérilisé à l'éthanol entre chaque plante. Les gants sont changés au minimum d'une variété à l'autre. Les plantes sont repérées par l'installation d'une étiquette avec un code associé à son statut symptomatique.

Une première coupe est réalisée à la limite parties chlorophylliennes-boisées, puis une deuxième coupe est réalisée à la base des inflorescences pour séparer directement le prélèvement en deux échantillons chacun : les inflorescences seules d'une part et le reste correspondant aux parties chlorophylliennes (tiges et feuilles). Dès qu'ils sont coupés, les échantillons sont immédiatement ensachés dans du papier d'aluminium (ce qui permet d'éviter les contaminations plastique et il résiste aux températures très basses) et plongés dans un bac contenant de l'azote liquide. L'azote liquide permet de réaliser une trempe des échantillons dans le but de figer le métabolisme de la plante. Il est important d'exécuter rapidement la trempe une fois l'échantillon coupé car les végétaux mettent très vite en place des réactions face au stress (ici coupure). (Wolfender et al. 2013; Courant et al. 2014)

Une fois les échantillons congelés, ils sont transférés dans des bacs en graphites remplis de glace carbonique pour le transport. Les échantillons sont ensuite stockés dans un congélateur à -80°C en attendant leur traitement.



Figure 28 : **A** échantillonnage en champ, **B** trempé des prélèvements dans l'azote liquide, **C** et **D** transport des échantillons dans de la carboglace et boîtes en graphite.

A noter, que les plants symptomatiques étant plus chétifs, des difficultés ont parfois été rencontrées pour obtenir des quantités suffisantes de matériel végétal et le nombre de prélèvements intra plants souhaités.

## II. Détermination du solvant d'extraction

Dans une étude métabolomique, le choix du solvant dépend en priorité de la matrice étudiée et des composés ciblés. Dans notre cas, nous avons adopté une approche non ciblée ce qui implique que le solvant ne doit pas être sélectif afin d'extraire la part la plus importante du métabolome (Kim and Verpoorte 2010; Wolfender et al. 2013; Mushtaq et al. 2014). Les solvants généralement employés en métabolomique peuvent être distingués en deux grands types :

- Les solvants acido-basiques
- Les solvants organiques

Les acides et les bases sont souvent utilisés pour des études métabolomiques portant sur une matrice d'origine animale, ils sont moins fréquents en phytochimie. Leur principal avantage est d'obtenir un arrêt immédiat des processus enzymatiques. En revanche, ils conduisent souvent à des réactions acido-basiques pouvant entraîner la production de sels néfastes pour les analyses en spectrométrie de masse. (Mushtaq et al. 2014)

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

La deuxième catégorie est les solvants organiques. Ils présentent plusieurs avantages. En effet, la plupart sont compatibles avec toutes les plateformes analytiques, sont souvent facile à évaporer, ne génèrent pas de sels.

Les solvants organiques les plus souvent employés sont les alcools tels que le méthanol et l'éthanol purs ou en mélange avec l'eau en proportions diverses. Les alcools ont un double avantage. Le premier est leur capacité à former des liaisons hydrogène qui leur confèrent une capacité à extraire des molécules hydrophiles. Le deuxième est de posséder des caractéristiques lipophiles, plus ou moins marquée en fonction de la longueur de la chaîne aliphatique de l'alcool. Ces deux propriétés combinées en font d'excellents candidats pour la métabolomique car elles leur permettent d'extraire une large gamme de métabolites. Les composés qu'ils extraient sont majoritairement des sucres, des acides aminés, des acides organiques, des alcaloïdes et des composés phénoliques. (Mushtaq et al. 2014)

Les autres solvants organiques usités sont les solvants apolaires ou faiblement polaires tels que le chloroforme, l'acétate d'éthyle ou l'hexane qui eux vont extraire préférentiellement des composés comme les terpènes ou les lipides. (Mushtaq et al. 2014)

Dans une étude métabolomique, lorsque la matrice est une plante, ce sont généralement les démarches classiques appliquées en phytochimie qui prévalent (Wolfender et al. 2013). Plusieurs paramètres sont à considérer mais le plus important est la polarité du solvant. Néanmoins, la valeur de polarité seule est insuffisante pour sélectionner le solvant le plus adapté ; il est donc recommandé de réaliser des tests, car en phytochimie les mécanismes de relargages des métabolites sont complexes et dépendants de la matrice étudiée. (Kim and Verpoorte 2010)

Dans cette optique, nous avons choisi de tester sept solvants (Tableau 7) de polarité croissante. Le critère de sélection principal sera sa capacité à extraire un très grand nombre de composés et en quantité suffisante pour permettre leur détection. Dans le cas où deux solvants auraient des performances similaires, le solvant le moins toxique et à moindre coût sera privilégié. Ces tests avaient comme deuxième objectif d'évaluer la complexité phytochimique du lavandin et de ses différents organes.

Tableau 7 : Relation entre les solvants envisagés et leur index de polarité (P') selon l'échelle de Snyder (Snyder 1974; Markom et al. 2007)

Solvant	Hexane	Ether diéthylique	Dichloro - méthane	Acétate d'éthyle	Ethanol	Acétone	Méthanol	Eau
Polarité (P')	0,0	2,9	3,4	4,3	5,2	5,4	6,6	9,0

### Matériel et méthodes :

Le matériel végétal employé est un plant de lavandin grosso asymptotique supposé sain, séché à température ambiante et broyé par organe (les feuilles et les tiges chlorophylliennes, les parties boisées et les racines). Une extraction solide-liquide assistée par ultra-sons est réalisée avec un ratio matière sèche-solvant de 10 mg/ml. Les extraits sont ensuite évaporés et les résidus sont dissouts avec le solvant de dilution correspondant (Tableau 41) à 10 mg/ml ou à défaut dilués dans 100 µL pour les résidus présents en quantité insuffisante. (Annexe 6)

Des conditions analytiques « généralistes » ont été utilisées pour ces premiers tests à savoir une colonne Luna C18 (150 x 4,6 mm, 5 µm) (pour plus de spécifications voir Annexe 4) et un gradient d'élution constitué d'eau et d'acétonitrile acidifiés à 0,1 % en acide formique, et d'isopropanol pur (voir gradient « généraliste » Annexe 5). La colonne est thermostatée à 25°C. Le volume d'injection est

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

de 10 µl. Plusieurs longueurs d'onde ont été acquises (254, 280, 330, 366, 445 nm). La longueur d'onde la plus informative pour les différents échantillons testés est 280 nm. Elle est celle pour laquelle les composés sont majoritairement plus intenses, et surtout elle est celle où le plus grand nombre de composés sont visibles. Néanmoins, pour certains composés (éluant en fin de gradient) peu visibles à 280 nm mais absorbant à 445 nm, il est préférable de vérifier cette longueur d'onde. Les profils DEDL ont été acquis pour obtenir une vision plus exhaustive et quantitative des extraits. En effet, le DEDL permet de détecter des composés sans groupement chromophore et vient donc compléter les profils UV acquis.

### Résultats :

Les premières extractions réalisées sur les feuilles (voir Figure 29), montrent une grande diversité chimique dans les extraits de lavandin. Nous pouvons constater que deux types de solvants se distinguent. D'un côté le méthanol et le mélange eau/éthanol avec une forte capacité d'extraction (voir Tableau 8) et une propension importante à l'extraction des composés de polarité intermédiaire (voir Figure 29). De l'autre côté, les solvants tels que l'hexane, le dichlorométhane, l'acétate d'éthyle et le diéthyle éther possèdent une capacité plus importante à extraire les composés à caractère lipophile (voir Figure 29) mais une capacité d'extraction inférieure (voir Tableau 8).

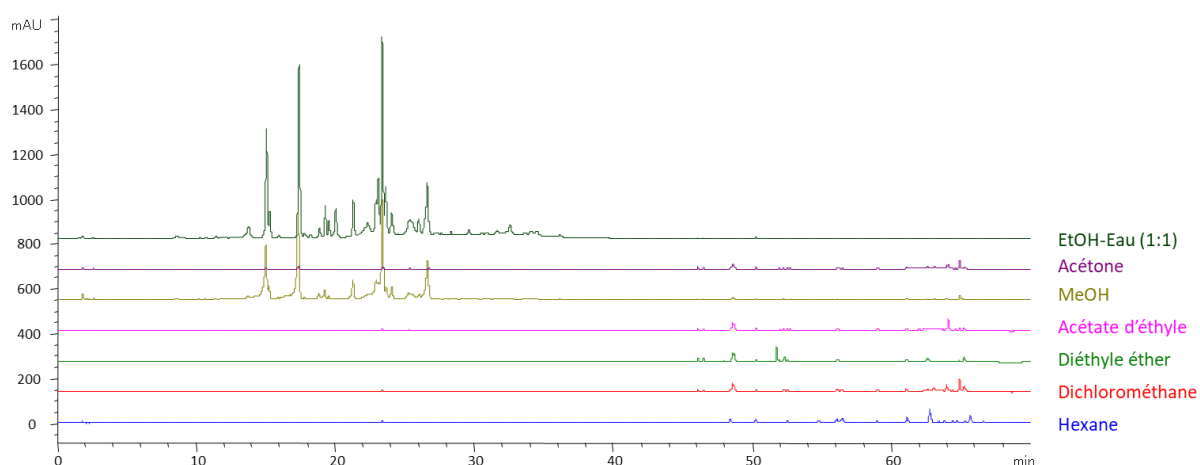


Figure 29 : Chromatogrammes HPLC-UV d'extraits de tiges chlorophylliennes et feuilles de lavandin correspondant aux sept solvants testés. Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; volume d'injection de 10 µL ; λ = 280 nm ; T = 25°C.

Pour le premier type de solvant (les alcools), le méthanol a été rapidement écarté au profit du mélange eau/éthanol (1/1) du fait de la similarité des profils obtenus (voir Figure 29), d'un rendement inférieur (voir Tableau 8) et de sa toxicité. L'acétone présente un type de profil intermédiaire, contenant à la fois des composés polaires et hydrophobes, cependant en quantité inférieure par rapport aux autres solvants (voir Tableau 8). Pour l'extraction des composés hydrophobes, l'acétate d'éthyle et le diéthyle éther présentaient des profils très similaires (voir Figure 29) et ce pour tous les organes. Cependant, l'acétate d'éthyle a été préféré du fait d'un rendement supérieur au diéthyle éther (voir Tableau 8). Le dichlorométhane a quant à lui été écarté car il n'apportait pas d'information supplémentaire (voir Figure 29) avec un rendement similaire à l'acétate d'éthyle (voir Tableau 8), et présentait une toxicité.

La formule utilisée pour calculer le rendement d'extraction en pourcentage est la suivante :

$$\text{Rendement (\%)} = 100 * \frac{\text{masse résidu sec}}{\text{masse matière première introduite}}$$

Tableau 8 : Rendements d'extraction obtenus en fonction des organes de la plante et du solvant d'extraction

Solvant d'extraction	Rendement (%)		
	Feuilles	Parties boisées	Racines
Hexane	0,54	0,49	NC
Dichlorométhane	7,65	NT	NT
Diéthyle éther	6,49	0,38	NC
Acétate d'éthyle	7,72	0,48	NC
Méthanol	19,32	9,59	11,00
Acétone	7,31	NT	NT
Eau/éthanol (1/1)	25,68	18,59	21,89

NC = non calculé (résidu présent en quantité insuffisante) ; NT = non testé

Comme le montre le Tableau 8, le rendement croît avec la polarité du solvant d'extraction. Les rendements diffèrent également selon l'organe de la plante, ce sont les parties chlorophylliennes qui possèdent le rendement le plus élevé. Les parties boisées ont, quant à elles, un rendement très faible, c'est entre autres pour cette raison et la similarité de leur profil avec les extraits de feuille (voir Figure 30) que cette partie de la plante n'a pas été retenue pour l'étude métabolomique.

### Composition des extraits et différences entre organes de la plante :

La Figure 30 présente les extraits hydroéthanoliques et la Figure 31 les extraits acétate d'éthyle pour les trois organes testés (les parties chlorophylliennes, les parties boisées et les racines). Les parties boisées présentent des profils proches de ceux obtenus avec des feuilles pour les extraits eau-éthanol. En revanche, les extraits acétate d'éthyle diffèrent quantitativement et qualitativement. En ce qui concerne les parties chlorophylliennes (feuille et tige), à savoir l'organe qui nous intéresse, les profils eau-éthanol sont plus riches que les profils obtenus avec l'acétate d'éthyle. De plus, de nombreux composés absorbent à une longueur d'onde de 280 nm et sont élués en début de chromatogramme sur les profils des extraits hydro-éthanoliques laissant fortement présager la présence de nombreux flavonoïdes et acides phénoliques (Stalikas 2007). Ces observations sont également en corrélation avec les composés décrits dans la littérature (voir chapitre 1 section II Les composés organiques non volatils (COnVs) dans le genre *Lavandula*).

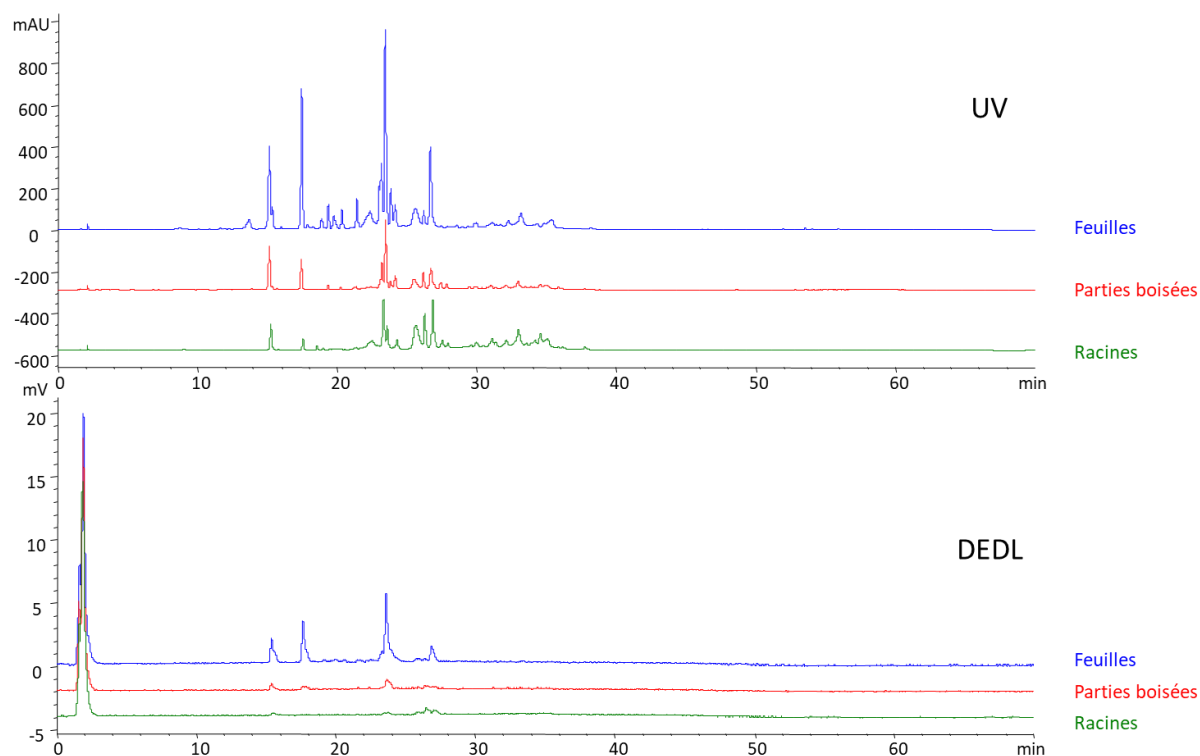


Figure 30 : Chromatogrammes HPLC-UV (haut) et DEDL (bas) des extraits eau/éthanol (1/1) des différents organes (tiges chlorophylliennes et feuilles, des parties boisées et des racines) de lavandin grosso. Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C. Paramètres DEDL : pression du gaz de nébulisation (azote) : 3,5 bars ; température du tube d'évaporation 40°C et gain 4.

Pour les extraits acétate d'éthyle (Figure 31) quelques composés répondent aux DEDL mais n'absorbent pas en UV, ce qui laisse penser à la présence de triterpènes (Oleszek and Stochmal 2010). Plusieurs études mentionnent d'ailleurs la présence de triterpènes dans le genre *Lavandula* (voir chapitre 1 section II.A) notamment dans les racines, qui rappelons-le, ont très peu été investiguées. Pour ces mêmes extraits, les composés éluant en fin de gradient, c'est à dire durant la phase organique, sont visibles à 280 nm mais ont une absorbance supérieure à 455 nm. Il est fort probable que nous soyons en présence de pigments tels que des composés chlorophylliens et surtout des caroténoïdes qui possèdent un maximum d'absorbance vers 450 nm. (Khachik et al. 1986; Brotas and Plante-Cuny 1996)

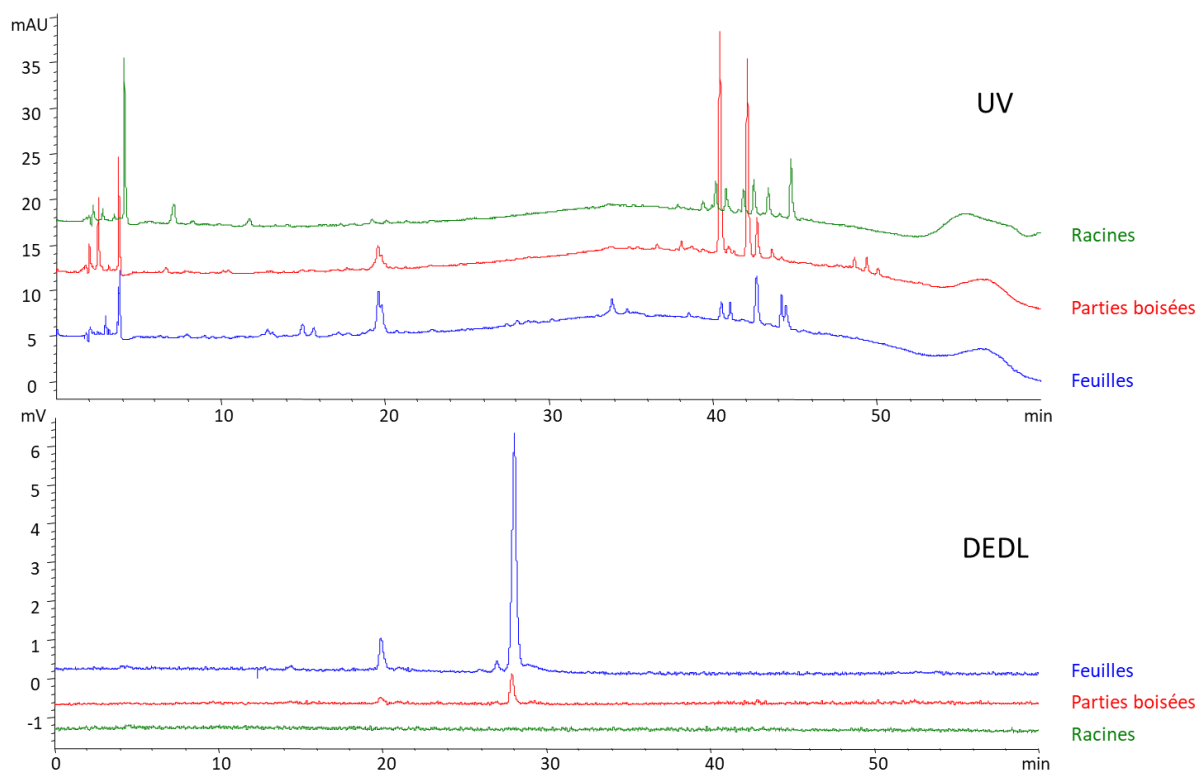


Figure 31 : Chromatogrammes HPLC-UV (haut) et DEDL (bas) des extraits acétate d'éthyle de différents organes (tiges chlorophylliennes et feuilles, des parties boisées et des racines) de lavandin grosso. Colonne Luna C18 (Annexe 4) ; gradient lipophile 2 (Annexe 5) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C. Paramètres DEDL : pression du gaz de nébulisation (azote) : 3,5 bars ; température du tube d'évaporation 40°C et gain 4.

Comparés aux chromatogrammes UV, les profils DEDL ont montré un nombre restreint d'analyte dont plusieurs étaient déjà visibles en UV. De plus, les intensités basses obtenues laissent présager de faibles quantités dans les extraits.

### Conclusion :

A l'issue de ces tests, les solvants retenus sont le mélange eau/éthanol (1/1) pour les composés à caractère hydrophile et l'acétate d'éthyle pour les composés à caractère lipophile. Néanmoins, l'étude métabolomique a été conduite uniquement sur les extraits hydroéthanoliques. L'étude des composés à caractère lipophile avait été envisagée du fait de leur potentielle interaction avec les insectes phytophages (Müller and Riederer 2005). De plus, les caroténoïdes et les molécules chlorophylliennes, potentiellement présentes dans les extraits acétate d'éthyle et absentes des extraits hydroéthanoliques, peuvent être impactées par les stress biotiques et abiotiques (Despinasse 2015). Il aurait été intéressant de voir l'impact du phytoplasme du Stolbur sur ces molécules en particulier. Cette étude n'a pas été menée faute de moyens et de temps. L'étude métabolomique réalisée durant cette thèse a donc porté sur les extraits eau/éthanol (1/1). A noter que les extraits hexaniques ont été réalisés et traités dans le cadre d'une autre thèse. Les composés majoritairement détectés dans ces extraits sont des monoterpènes oxygénés. D'autres analytes tels que des monoterpènes hydrocarbonés et des sesquiterpènes ont également été identifiés (Stierlin 2020).



### III. Préparation d'échantillon

Dans cette partie, nous allons de traiter la mise en place des différentes étapes du protocole de préparation des échantillons. La préparation d'échantillon comprend l'ensemble des étapes permettant de rendre les échantillons analysables. Elle désigne ici le broyage, le séchage, la conservation et la méthode d'extraction des échantillons. Le design de l'échantillonnage a été pensé en amont de cette étude et est présenté dans la partie I.

La méthodologie a été pensée afin d'éviter l'introduction de tout biais technique et de rendre le protocole simple pour assurer une répétabilité et une rapidité d'exécution de celui-ci.

#### III.A. Séchage

##### III.A.1. Mise en place

Une recherche bibliographique portée sur les méthodes de préparation d'échantillon sur les composés non-volatils de la lavande, et plus largement sur la famille des Lamiaceae a permis de mettre en avant qu'aucune méthode consensuelle n'existe. Le protocole employé découle de manière générale, de l'expérience, des habitudes acquises par les auteurs, et ne relève d'aucune standardisation. Les différents traitements qui ont pu être appliqués dans la littérature sur la famille des Lamiaceae (Borrás Linares et al. 2011; Ghasemi Pirbalouti et al. 2013; Silva and Câmara 2013; Cabana et al. 2013; Bogdanovic et al. 2016; Bernatoniene et al. 2016; Fotakis et al. 2016; Jovanović et al. 2017), sont les suivants :

- Utilisation du matériel frais (aucun traitement)
- Stockage dans le réfrigérateur ou dans le congélateur
- Séchage : à température ambiante (durée variable), à l'étuve (durée et température variable), au soleil, sous ventilation ou sous atmosphère contrôlée (température et hygrométrie variables)
- Lyophilisation

La méthode la plus souvent employée est le séchage et fréquemment à l'air libre (Borrás Linares et al. 2011; Ghasemi Pirbalouti et al. 2013; Cabana et al. 2013; Bogdanovic et al. 2016; Bernatoniene et al. 2016; Fotakis et al. 2016; Jovanović et al. 2017). Les différents traitements sont généralement suivis d'un broyage (au mortier ou la méthode n'est pas précisée) et parfois d'un tamisage.

D'après N. Mulinacci et al., dans le cas des feuilles de romarin, les méthodes de réfrigération et de congélation du matériel frais comme moyen de stockage n'étaient pas appropriées car conduisaient à la dégradation rapide de l'acide rosmarinique (Mulinacci et al. 2011). L'utilisation du matériel frais apparait comme la méthode la plus optimale cependant des réactions enzymatiques peuvent se produire (Bijttebier et al. 2016). La lyophilisation est le procédé qui représente souvent le meilleur moyen de conservation des échantillons avant analyse. Enfin, le séchage ne permet pas d'arrêter des réactions d'oxydation ou enzymatiques pouvant engendrer une surexpression de certains composés, et dans le cas d'un chauffage, il peut induire des dégradations thermiques. Néanmoins, dans le cas du romarin (*Rosmarinus officinalis* L.), le séchage limite les pertes en diterpènes comparativement à la lyophilisation et la congélation. (Mulinacci et al. 2011)

Dans le cadre d'une étude métabolomique adoptant une approche non-ciblée, la préparation des échantillons est une étape critique et souvent dépourvue d'investigations préliminaires. Comme évoqué dans le premier chapitre, la méthode de préparation des échantillons est dépendante de la matrice employée et des objectifs de l'étude (approche ciblée ou globale) (Royer et al. 2017). Par ailleurs, dans le cadre de cette étude l'échantillonnage sera réalisé sur le terrain. Il est donc primordial de réaliser une étude préalable pour déterminer l'impact des méthodes de conservation et de

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

préparation d'échantillon de lavande sur leur composition chimique. Six conditions, présentées ci-dessous, ont été testées :

- Condition n°1 : matrice fraîche

La matrice végétale est prélevée et immédiatement plongée dans l'azote liquide et broyée. Les échantillons ont été extraits et analysés sans délai.

Cette condition sur matériel végétal frais servira de référence afin de visualiser des potentielles dégradations chimiques pouvant survenir dans les autres conditions testées (Wolfender et al. 2013; Bijttebier et al. 2016). Cependant, ce prétraitement ne peut être mis en place pour cette étude à cause des contraintes techniques telles que l'échantillonnage en plein champ ou la durée de broyage au mortier incompatible avec le nombre d'échantillon à traiter.

- Condition n°2 : la lyophilisation

Pour cette condition, la matière végétale fraîche est immédiatement congelée à l'aide d'azote liquide puis lyophilisée.

- Condition n°3 : séchage à l'air libre pendant 13 jours

Un lot de végétal est prélevé puis déposé sur un papier dans un carton ouvert à température ambiante, sans lumière directe. Le lot est laissé à sécher pendant 13 jours.

- Condition n°4 : congélation (5 jours) suivi d'une lyophilisation

La matrice fraîche est prélevée puis directement placée dans le congélateur (-18°C) pendant 5 jours. Le lot est ensuite séché par lyophilisation.

- Condition n°5 : congélation (15 jours)

Afin d'estimer la qualité de la conservation de la matrice végétale, la condition n°1 a été conservée 15 jours dans le congélateur à -18°C. Passé ce délai, le matériel végétal a été extrait et analysé.

- Condition n°6 : matrice fraîche avec ajustement de la masse de la matière végétale introduite pour l'extraction

Le protocole utilisé est identique à celui employé pour la condition n°1, hormis le fait que la masse de matière végétale introduite pour l'extraction a été réévaluée à 86,3 mg. A noter que cette condition a été rajoutée a posteriori et a donc été prélevée sur un nouveau plant.

La quantité de matière végétale introduite lors de l'étape d'extraction a également été augmentée pour la condition n°5. En effet, le séchage induit une perte en eau et donc une perte de masse. Pour que la quantité d'analytes extraits soit équivalente entre les conditions avec et sans séchage, la masse introduite pour les conditions sans séchage a été augmentée. Pour déterminer, la masse supplémentaire à ajouter, un suivi de la perte massique sur les conditions n°2, 3 et 4 (c'est-à-dire avec séchage) a été effectué. A noter que ce suivi avait un deuxième objectif, à savoir déterminer la durée du séchage.

La matière végétale correspond à des feuilles et des tiges chlorophylliennes de plants de lavandin grosso asymptotique supposés sains. Une fois constitués comme explicité ci-dessus, tous les lots ont été broyés puis extraits et analysés. Les conditions n°2, 3 et 4 ont été broyées au mixeur de cuisine (broyeur couteau). Les conditions 1 et 6 ont été broyées au mortier dans de l'azote liquide jusqu'à l'obtention d'une poudre, en prenant garde de conserver l'échantillon congelé en permanence. Les extractions ont été réalisées avec un mélange eau/éthanol (1/1) et six réplicats d'extraction ont été réalisés par condition. Les chromatogrammes ont été réalisés par HPLC-UV équipée d'une colonne Kinetex PFP (100 x 4,60 mm, 2,6 µm) (détails sur la colonne en Annexe 4). Seuls les chromatogrammes UV à 280 nm des extraits eau/éthanol (1/1) seront présentés. A noter que les restes des lots des conditions n°2, 3 et 4 ont été conservés à température ambiante et à l'abri de la lumière. Pour les autres conditions, le lot est conservé dans le congélateur à -18°C. (Voir Annexe 7 pour plus de détails)

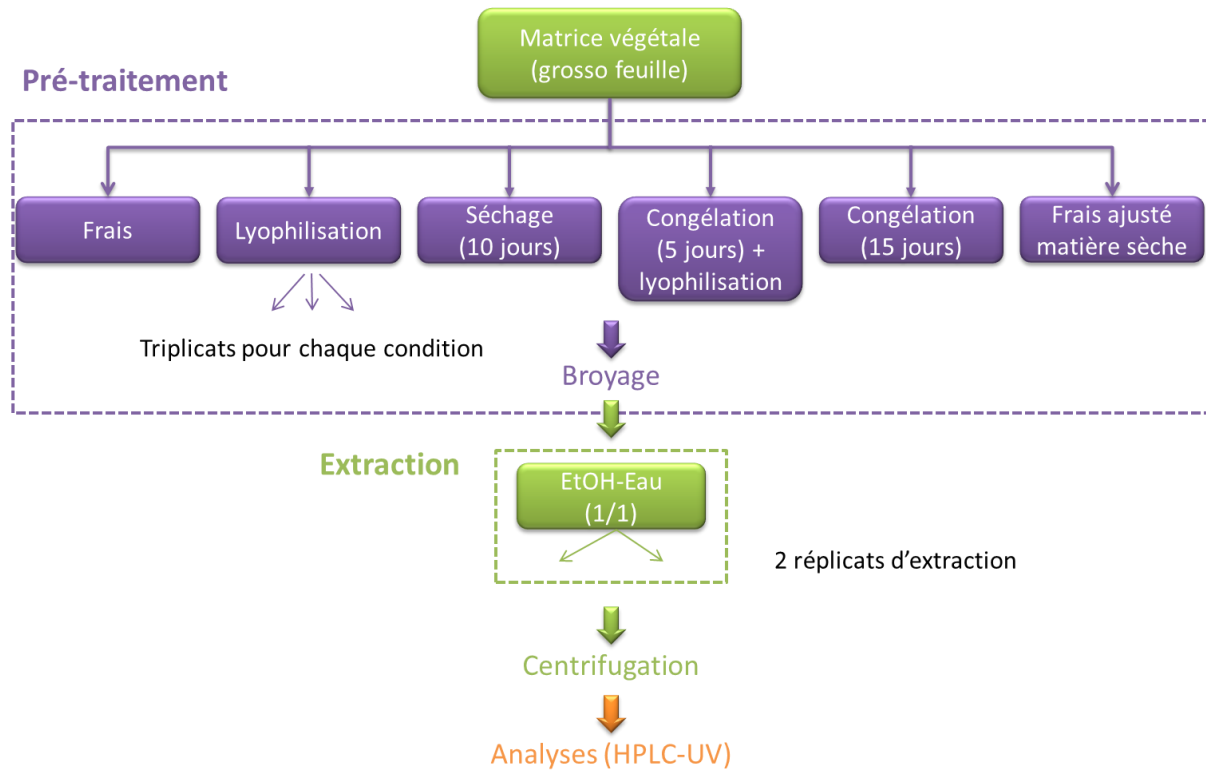


Figure 32 : Schématisation du protocole des tests préliminaires pour déterminer le prétraitement des échantillons

Le choix du processus de prétraitement se fera à partir du cahier des charges suivant :

- La faisabilité de mise en place du protocole pour l'étude métabolomique
- L'obtention d'une méthode répétable pour ne pas introduire de variabilité technique (ce qui va se traduire par une variation chimique entre les réplicats techniques tendant vers 0)
- L'informativité la plus grande, se traduisant par le nombre de pic le plus élevé à mettre cependant en parallèle avec le profil sur matériel végétal frais pour vérifier l'absence d'apparition d'artefacts.

### III.A.2. Résultats

#### Evaluation de la perte massique :

Pour comparer les différents tests préliminaires, la perte massique due au séchage a été évaluée et la masse de matière végétale employée pour l'extraction (50 mg sur matière sèche) a été ajustée pour les conditions sur « matériel frais ». Le pourcentage d'eau contenu dans un végétal est très variable et représente généralement 80 % de la masse (Despinasse 2015).

Comme attendu, les résultats HPLC (voir Figure 33 ci-dessous) montrent une intensité moindre pour la matrice fraîche (condition n°1). Les conditions 2, 3 et 4 faisant appel à un séchage, la perte d'eau induite « fausse » la concentration en analytes. La moyenne (conditions 2, 3 et 4) de la perte massique induite par le séchage est de 72,7 %. La masse de matrice fraîche qu'il faut introduire pour obtenir une concentration en analytes comparables à celle des conditions avec séchage est de 86,3 mg. Deux autres conditions (n°5 et 6) ont été alors réalisées en ajustant la quantité de matière introduite. La première pour tester l'impact de la congélation sans lyophilisation (condition n°5), et la seconde pour obtenir une analyse sur du matériel frais (condition n°6) « comparable » aux conditions avec séchage.

Tableau 9 : Suivi de la perte massique pour les conditions contenant une étape de séchage

Condition n°2 : lyophilisation			
Temps (h)	0	24	36
Masse matrice (g)	19,14	5,43	5,42
Condition n°3 : séchage			
Temps (jrs)	0	13	
Masse matrice (g)	19,60	5,08	
Condition n°4 : congélation et lyophilisation			
Temps (h)	0	23	30
Masse matrice (g)	19,51	5,88	5,78

**Comparaison des différentes conditions testées :**

La Figure 33 présente les chromatogrammes obtenus pour l'ensemble des conditions testées. Les conditions sur matières séchées c'est-à-dire lyophilisation (n°2), séchage à température ambiante (n°3) et congélation suivie d'une lyophilisation (n°4) ont des intensités supérieures aux autres conditions qui sont exemptes de séchage. Des différences de composition sont observables selon les conditions, le séchage à température ambiante (n°3) et la congélation suivie d'une lyophilisation (n°4) sont les deux conditions qui apparaissent comme présentant le plus grand nombre d'analytes.

La condition frais « concentré » (n°6) est assez décevante, l'ajout de matière végétale a permis d'augmenter la concentration en analytes sans pour autant égaler les concentrations obtenues pour les extraits sur matière sèche. De plus, des différences de composition ont été constatées avec la condition n°1 (matrice fraîche). (Figure 33) Cette différence a été imputée à l'utilisation d'un nouveau plant contraint car la condition a été ajoutée a posteriori.

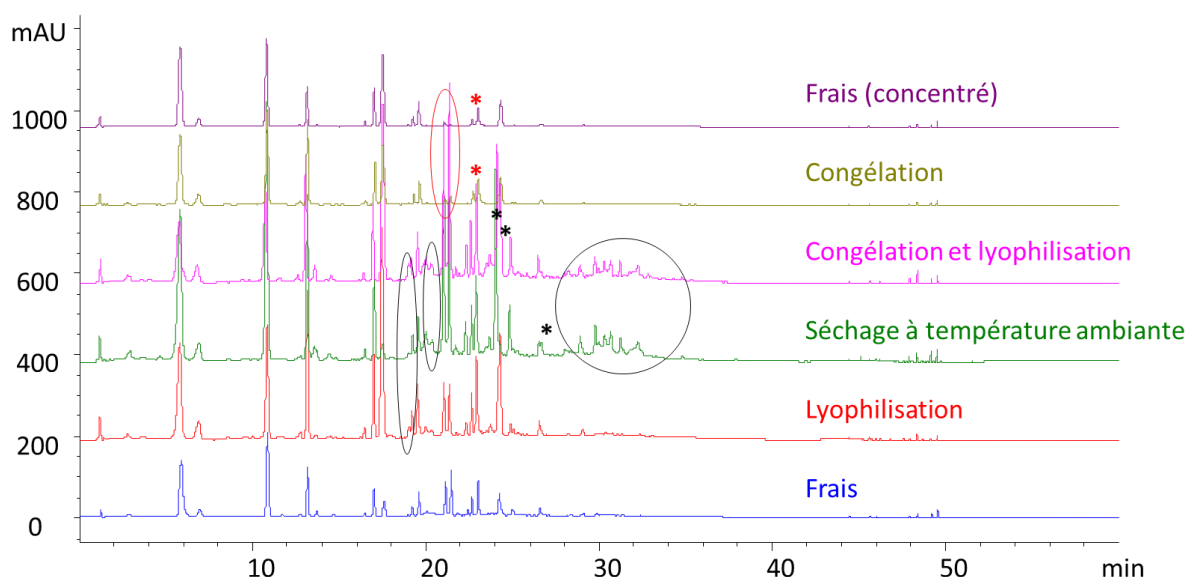


Figure 33 : Chromatogrammes HPLC-UV des extraits d'eau/éthanol (1/1) de feuilles de lavandin grosso obtenus pour les six conditions testées ; colonne Kinetex PFP (Annexe 4) ; méthode Polaire 2 constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

La présence de pics supplémentaires sur les chromatogrammes a été mis en avant soit par une étoile pour un pic isolé, soit par des cercles. La couleur noire signale la présence de pics supplémentaires, à l'inverse le rouge signifie une absence.

Afin de mieux comparer les différentes conditions, une ACP a été réalisée à partir des aires obtenues en UV à 280 nm sur le logiciel xIstat (logiciel de statistique additionnel d'excel). L'alignement a été effectué sur les temps de rétention à partir d'une macro (suite d'instructions automatisées sous excel), puis contrôlé et ajusté manuellement. Dans ce cas, l'ACP permet de visualiser la répétabilité et les similitudes chimiques des conditions testées. Chaque point sur l'ACP correspondant à un extrait, plus les points sont distants plus ces points ont une composition chimique différente. La répétabilité au sein d'une condition est donc visible par la proximité de ses répliqués.

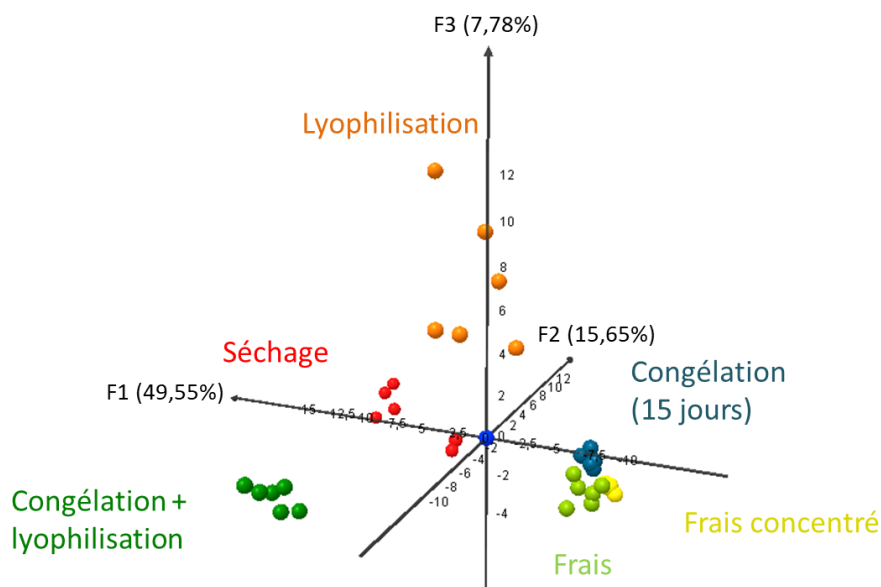


Figure 34 : Graphique de l'analyse en composante principale (ACP) des six conditions de prétraitement testées représentée avec les trois premières composantes (72,98%)

Le graphique des individus (Figure 34) obtenu montre que les conditions dites sur « frais » (n° 1, 5 et 6) présentent des similitudes chimiques ainsi qu'une très bonne répétabilité. Les conditions avec séchage (n° 2, 3 et 4) ont des différences de composition entre elles ainsi qu'avec les autres conditions, ce qui est en accord avec les observations faites sur les profils chromatographiques (Figure 33). En prenant la condition sur matrice fraîche comme référence, chaque prétraitement a bien une influence sur la composition chimique des échantillons. Les conditions avec séchage ne présentent pas toutes la même répétabilité. En effet, la lyophilisation a conduit à une variabilité entre les réplicats. Le séchage à température ambiante a une variabilité bien moindre mais supérieure à la condition avec congélation suivie d'une lyophilisation dont la répétabilité se rapproche de celle sur matrice fraîche.

Tableau 10 : Moyenne et écart-type du nombre de pics et des aires des réplicats des six conditions de prétraitement testées

Conditions		N°1 Frais	N°2 Lyoph.	N°3 Séchage	N°4 Congél. + Lyoph.	N°5 Congél.	N°6 Frais concentré
Nombre de pics	Moyenne	71	96	114	121	61	55
	Ecart- type	17	10	11	23	4	3
Aire	Moyenne	100	273	382	212	190	158
	Ecart- type	47	48	65	29	18	42

Le nombre de pics moyen par condition a été calculé (voir Tableau 10). Les conditions sur « frais » (n° 1, 5 et 6) présentent le nombre de pics le moins élevé et les aires les plus faibles des six conditions testées. Les écart-types sur les aires sont quant à eux faibles, ce qui corrobore les résultats de l'ACP. Les conditions comportant un séchage possèdent un plus grand nombre de pics, des aires plus élevées

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

mais aussi des écart-types plus importants. Le séchage à l'air libre est la condition qui a les aires les plus élevées et le nombre de pics le plus important loin devant les autres conditions. Là encore, les observations faites sont en accord avec les observations précédentes.

### III.A.3. Conclusion

La condition n°4, correspondant à la congélation suivie d'une lyophilisation, apparaît comme un bon compromis. Elle présente une bonne répétabilité (cf. ACP, écart-type). Elle compte parmi les conditions qui ont le plus grand nombre de pics et les aires les plus élevées ; et d'un point de vue pratique, elle est l'une des plus faciles à mettre en place pour un échantillonnage en champs.

De plus, la condition n°5, à savoir la conservation de la matrice fraîche dans le congélateur à -20°C pendant 15 jours, a été réalisée dans le but d'évaluer le vieillissement des échantillons. Les résultats obtenus sont très similaires à ceux de la matrice de départ, montrant qu'il est possible de stocker sur un délai raisonnable les échantillons de lavande dans le congélateur sans craindre des modifications chimiques.

Finalement, nous avons adopté le prétraitement n°4 pour l'étude métabolomique. Par conséquent, les échantillons ont été congelés dans l'azote liquide lors de la récolte puis maintenus à température très basse jusqu'à leur lyophilisation. Les échantillons ont été lyophilisés sur une durée minimale de 30h dans le lyophilisateur LYOVAC GT2 Finn-aqua de la marque G. Beyneix S.A.

A noter que la durée de la lyophilisation pour l'étude métabolomique a été déterminée en fonction des résultats obtenus pour les conditions 2 et 4 et en effectuant un suivi de la perte massique sur la première lyophilisation de l'étude métabolomique. Aucune perte massique n'a été constatée après 28 h.

### III.B. Broyage

Les broyages réalisés lors de la mise en place de l'étude ont été effectués sur un broyeur à marteau pour les parties boisées et un broyeur de type couteau (mixeur de cuisine Kenwood) pour les échantillons de feuille et de tige chlorophyllienne. Cependant, la quantité de matière disponible par échantillon pour l'étude métabolomique est très en deçà de celle employée pour les tests préliminaires, rendant obsolète l'utilisation des deux broyeurs précédents. De plus, il a été constaté que la qualité du broyage était insuffisante (hétérogénéité et granulométrie élevée). Or, un broyage de mauvaise qualité est gênant car il entrainera la présence de particules en suspension lors de l'étape de centrifugation des échantillons. D'autres types de broyage ont alors été envisagés : un broyeur à billes et un montage « artisanal » avec une tête de broyage créée par le CRIEPPAM.

La méthode de broyage pour cette étude devra répondre à plusieurs critères :

- Être adaptée à la quantité de matière végétale des échantillons (environ la moitié des échantillons ont une masse < 1 g et la moyenne se situe à 1,4 g)
- Éviter au maximum les possibilités de contamination croisée entre les échantillons. Ce critère est en lien étroit avec la facilité de nettoyage du broyeur, voire l'absence d'étape de nettoyage
- La qualité du broyage, à savoir l'obtention d'une granulométrie fine et homogène
- La faisabilité de la mise en place de la méthode de broyage dans une étude métabolomique : c'est-à-dire la rapidité d'exécution de l'étape qui dépend principalement du nettoyage afin d'être adapté au grand nombre d'échantillon à réaliser, la maîtrise des paramètres de broyage (durée, vitesse ou fréquence) ou degrés d'automatisation de la méthode pour augmenter la répétabilité
- Le coût de mise en place de la méthode.

### Description du broyeur du CRIEPPAM :

Le broyeur du CRIEPPAM est une tête de broyage, constituée de billes, montée sur une perceuse-visseuse à colonne. Les échantillons sont placés dans des sacs de broyage spécifique employables également pour l'échantillonnage. Les sachets sont constitués d'un plastique conçu pour résister aux frottements. En revanche, ils résistent difficilement au passage à l'azote liquide suivi d'un broyage.



Figure 35 : Broyeur fabriqué par le CRIEPPAM dans son intégralité à gauche, et à droite la tête de broyage

L'avantage de cette technique est principalement dû à l'usage des sacs qui évitent toute contamination croisée et suppriment le nettoyage et le transvasement ce qui représente un gain de temps considérable pour cette étape. Cependant cette méthode n'a pas été retenue car le broyat des échantillons de feuilles et de tiges chlorophylliennes est grossier et hétérogène (Figure 36).



Figure 36 : A gauche échantillons obtenus avec un broyeur de type couteau ; au centre échantillons broyés avec le broyeur du CRIEPPAM, à droite échantillon de feuille broyé avec un broyeur à billes

### Description du broyeur à billes :

Le deuxième instrument testé est un broyeur à billes MM 301 de chez Retsch (Figure 37). Il fait partie des vibro-broyeurs car il utilise les vibrations pour broyer les échantillons. Les échantillons sont placés dans une cellule avec une ou plusieurs billes en acier. Les cellules sont mises en vibrations et ce sont les chocs des billes contre l'échantillon qui vont le réduire en poudre. Il permet d'obtenir des broyats d'une granulométrie très fine et homogène (allant jusqu'à 5  $\mu\text{m}$ , donnée constructeur). Ce sont des broyeurs qui permettent des broyages à sec ou humide (dans notre cas le broyage a été effectué à sec) et adaptés à des échantillons semi-durs et fibreux comme les tiges de lavande. C'est un broyeur de laboratoire qui convient à de faible quantité d'échantillons.

Le broyeur retenu et utilisé pour l'étude métabolomique est ce broyeur à billes car il présente les avantages suivants :

- l'appareil est adapté à des échantillons de faibles quantités
- le broyage est rapide (voir Tableau 11 ci-dessous)



## Chapitre 2 : Conception et optimisation de l'étude métabolomique

- les paramètres de broyage sont précisément définis (durée et fréquence) (voir Tableau 11)
- les cellules sont faciles à nettoyer
- la qualité du broyage est excellente (granulométrie homogène et faible) (voir Figure 36)
- la méthode est répétable



Figure 37 : Broyeur à boules MM 301 de chez Retsch en fonctionnement

### Etude métabolomique :

Lors de l'étude métabolomique, les échantillons récoltés lors de la campagne de 2017 ont été broyés par variété, organes et état sanitaire séparés. Hormis ces critères, les échantillons ont été broyés au hasard. Les cellules ont été nettoyées et séchées entre chaque échantillon (voir protocole en Annexe 8). Les paramètres de broyages dépendent de l'organe de la plante et sont répertoriés dans le tableau ci-dessous :

Tableau 11 : Durées et fréquences de broyage employées pour l'étude métabolomique selon le type d'échantillon

Type échantillon	Temps (min)	Fréquence (1/s)
Feuille-tige (F)	0,40	30 (max)
Inflorescence	0,30	30 (max)

### III.C. Mise en place du protocole d'extraction

L'étape suivante dans la préparation des échantillons est la réalisation d'une extraction solide-liquide. L'objectif de cette étape est de rendre les échantillons compatibles avec la méthode analytique, pour ce faire les analytes doivent être extraits de la matrice. Les méthodes d'extraction conventionnelles en phytochimie telles que la macération ou la percolation sont lentes et peu adaptées pour réaliser un nombre élevé d'échantillons comme l'exige une étude métabolomique. Habituellement, des techniques d'extraction plus rapides (< 5 min) car accélérées par un apport d'énergie leur sont préférées. Les méthodes généralement employées en métabolomique sont l'extraction assistée par micro-ondes (MAE), par ultra-sons (UAE), par fluide pressurisé (PLE : *Pressurized Liquid Extraction*) et par fluide supercritique (SFE). (Kim and Verpoorte 2010; Wolfender et al. 2013; Mushtaq et al. 2014)

L'extraction assistée par ultra-sons est la plus souvent employée en métabolomique et celle que nous avons choisi d'employer. Elle a l'avantage d'être simple, rapide, adaptée à des échantillons de taille réduite, et n'est pas restrictive vis-à-vis du solvant (Toma et al. 2001; Vinatoru 2001; Kim and Verpoorte 2010). Les ultra-sons sont des ondes à hautes fréquences (> 16 kHz). Lors d'une extraction

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

solide-liquide, ils induisent des microbulles de gaz dans le liquide (phénomène de cavitation). L'amplitude des ondes vont générer une succession de compression-dépression au sein des microbulles, jusqu'à leur implosion qui va libérer une énergie élevée. Ce phénomène induit une hausse du rendement d'extraction en augmentant le transfert de masse et en cassant les cellules (Lavilla and Bendicho 2017).

Le protocole de départ est celui employé lors de la détermination du solvant d'extraction (voir partie I). Il a cependant fait l'objet de simplification afin d'être plus adapté aux exigences d'une étude métabolomique.

### III.C.1. Adaptation du protocole : suppression de l'étape d'évaporation et remplacement par une étape de centrifugation

L'étape d'évaporation peut être chronophage, notamment lorsqu'une solution hydroalcoolique est employée. Une telle perte de temps n'est pas envisageable au vu du nombre d'échantillons à réaliser. Une simplification du protocole de préparation des échantillons a été envisagée ; en substituant la filtration et l'étape d'évaporation à l'évaporateur-rotatif par une centrifugation et par la collecte du surnageant.

L'influence de ce nouveau protocole sur la nature et la quantité de composés extraits a été déterminée ainsi que la quantité de matière à introduire. Pour ce faire, cinq ratios plante sèche / solvant ont été testés : 10, 20, 30, 40 et 50 mg/ml. Des extraits eau/éthanol (1/1) ont été préparés à partir de feuille de lavandin grosso saine séchée et broyée. Les analyses effectuées par HPLC-UV ont été réalisées avec la méthode dite « généraliste » (détails en Annexe 5).

Les résultats, montrent que l'intensité des pics augmentent avec le ratio plante sèche / solvant (voir Figure 38). Le premier ratio à 10 mg/ml s'est avéré très insuffisant, la majorité des pics chromatographiques n'étant pas différentiable du bruit de fond. Les profils commencent à être visibles à partir de 20 mg/ml. Le ratio à 50 mg/ml confère les pics les plus intenses tout en conservant une bonne séparation. Il a donc été retenu pour la suite de l'étude.

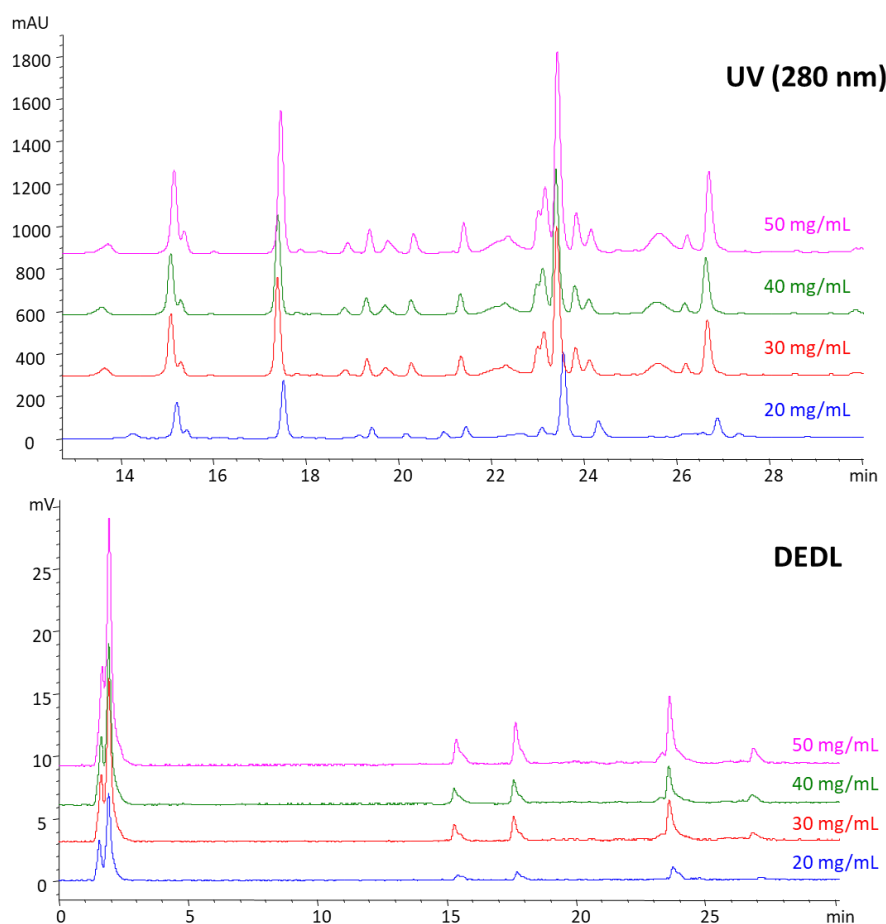


Figure 38 : Agrandissements des chromatogrammes HPLC-UV (haut) et DEDL (bas) des extraits eau/éthanol (1/1) des feuilles de lavandin grosso pour différents ratios plante sèche / solvant (20, 30, 40 et 50 mg/mL). Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5), constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

Ce nouveau protocole est beaucoup plus rapide, simple et efficace. Les profils sont très similaires à ceux obtenus avec le protocole précédent avec évaporation (voir Figure 39 ci-dessous). Aucune différence d'un point de vue qualitatif n'a été constatée, en revanche, un léger appauvrissement quantitatif a été noté.

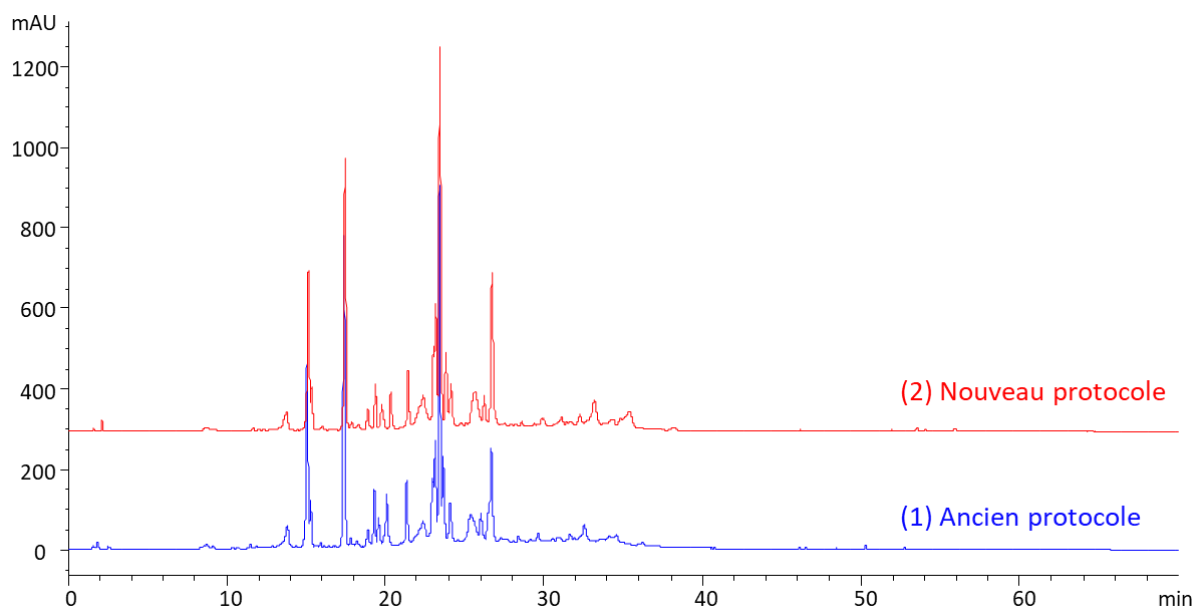


Figure 39 : Chromatogrammes HPLC-UV des extraits eau/éthanol (1/1) des feuilles de lavandin grosso obtenus avec le protocole d'extraction initiale (1) et sa version simplifiée (2). Colonne Luna C18 (Annexe 4) ; gradient généraliste (Annexe 5), constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

Finalement, le protocole d'extraction retenu est celui avec un ratio de 50 mg/ml et une centrifugation à 4000 rpm pendant 10 min.

### III.C.2. Suppression filtration et ajout deuxième centrifugation et précipitation à froid

Les tests précédents ont été conduits sur une centrifugeuse Sigma Fisher BioBlock Scientific 2-6. La centrifugeuse WiseSpin CF10 (Serflam) permet l'obtention de vitesse de rotation bien supérieure avec un maximum à 13000 rpm au lieu de 4000 rpm précédemment. De plus, l'utilisation du broyeur à billes a permis d'obtenir une granulométrie plus fine et plus homogène que celles obtenues avec les broyeurs couteux employés lors des premiers tests. Avec ces deux améliorations, nous avons pu envisager la suppression de la filtration (sur filtre PVDF à 0,2  $\mu$ m) et le remplacement par une étape de centrifugation supplémentaires. La filtration était réalisée jusqu'à présent, uniquement dans le but de protéger la colonne du système chromatographique des reliquats de particule provenant d'une mauvaise agglomération lors de la centrifugation. Ces deux techniques sont employées en métabolomique et il a été constaté dans certains cas que la filtration provoque la rétention de molécule. (Álvarez-Sánchez et al. 2010; Dietmair et al. 2010; Fan 2012)

Le protocole d'extraction finalisée est détaillé en Annexe 9.

## IV. Développement analytique

En amont, du travail effectué par UPLC-HRMS (Xevo G2 qTOF), des tests préliminaires ont été effectués par HPLC-UV dans le but d'identifier la phase stationnaire la plus adaptée aux extraits ainsi que d'évaluer la différence de composition entre les deux organes de la plante qui seront employés dans l'étude métabolomique (inflorescences et parties chlorophylliennes). Une étude par HPLC-UV a été également menée afin d'évaluer la variabilité biologique existante au sein d'un même plant. Nous verrons par la suite l'optimisation des conditions analytiques de l'UPLC-HRMS, ainsi que la détermination du standard interne, l'anticipation et le suivi de la dérive analytique au cours des séquences.

### IV.A. Développement préliminaire par HPLC-UV

Au vu des résultats d'analyse obtenus lors de la détermination du protocole d'extraction des co-élutions semblent exister. La séparation chromatographique est à optimiser ; pour ce faire, un criblage de colonne et une amélioration des paramètres analytiques ont été réalisés sur les extraits de feuille. Par la suite, des extraits d'inflorescences ont été analysés afin d'apprécier les différences de profils chimiques entre les deux organes et les potentiels ajustements analytiques à réaliser afin d'obtenir un paramétrage optimum et commun.

#### IV.A.1. Optimisation conditions analytiques par HPLC pour les échantillons de feuilles

##### Choix de la phase stationnaire

Les résultats précédents ont été obtenus avec une colonne Luna C18 (150 x 4,6 mm, 5  $\mu$ m). Afin d'optimiser la séparation des extraits eau-éthanol, d'autres types de phase stationnaire ont été testés. Pour effectuer ces essais, quatre colonnes supplémentaires ont été sélectionnées :

- Kinetex C18 (100 x 4,60 mm, 2,6  $\mu$ m)
- Kinetex PFP (100 x 4,60 mm, 2,6  $\mu$ m)
- Synergi C12 (250 x 4,60 mm, 4  $\mu$ m) max
- Synergi C18 (250 x 4,60 mm, 4  $\mu$ m) hydro

Ces colonnes ont été choisies pour leur différence de sélectivité dépendant à la fois du type de support solide que de la phase greffée (spécificités complètes et sélectivités de toutes les colonnes en Annexe 4). Trois phases greffées ont été testées comprenant deux greffons de type alkyle linéaire de taille différente (C12 et C18) et une phase de nature chimique différente (PFP). L'intérêt de l'allongement de la chaîne alkyle est qu'il confère une rétention supérieure des composés hydrophobes (Le Mapihan 2004). Quant à la phase PFP, elle est intéressante car elle possède une sélectivité différente des phases alkyles (Chirita 2009). Deux types de support solide ont également été envisagés. Les colonnes de type Kinetex ont une phase qui possède la technologie « *core-shell* ». Les particules de silice possèdent un noyau dur et une surface poreuse, contrairement aux phases classiques qui ont des particules de silice entièrement poreuse. Cette structure « *core-shell* » confère une réduction du chemin de diffusion, donc une réduction de la largeur des pics et une meilleure résolution (Fekete et al. 2012).

Les colonnes ont toutes été testées avec la même méthode d'analyse « généraliste » constituée d'eau et d'acétonitrile acidifiés à 0,1 % en acide formique, et d'isopropanol pur (paramètres détaillés en Annexe 5), hormis pour les colonnes de type Kinetex. En effet, la faible porosité des colonnes de type Kinetex (du fait de la structure *core-shell*) et la forte viscosité de l'isopropanol sont incompatibles. Afin d'éviter une élévation importante de la pression dans la colonne, l'étape d'élution avec de l'isopropanol en fin de gradient a été supprimée. Hormis la modification de la fin du gradient d'élution (voir Annexe 5 pour le descriptif complet), les deux méthodes analytiques sont identiques en tout point afin d'être comparables. Le volume d'injection est de 10  $\mu$ l, la longueur d'onde d'observation est toujours de 280 nm. L'extrait injecté est un extrait eau-éthanol (1/1) à 50 mg/ml réalisé précédemment avec le protocole comprenant la centrifugation (cf. partie III.C.1).

Les résultats sont présentés sur la Figure 40, les profils obtenus à 280 nm diffèrent en fonction des colonnes mais tous possèdent une ligne de base rectiligne et aucun bruit de fond n'est à noter.

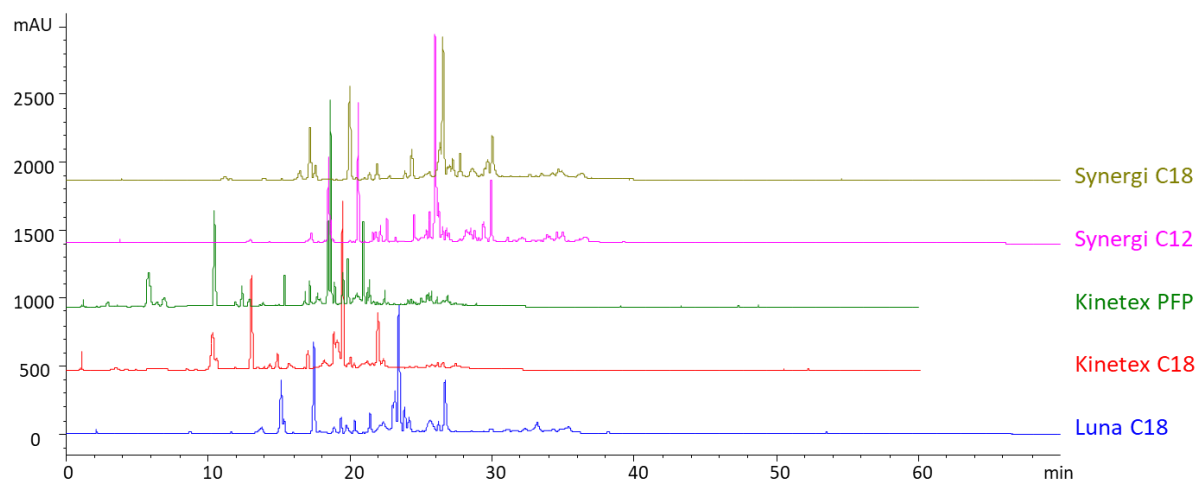


Figure 40 : Chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso réalisés avec les cinq colonnes testées ; gradients généraliste (Luna C18, Synergi C18 et Synergi C12) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure, et Polaire 1 (Kinetex C18 et Kinetex PFP) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

Parmi les colonnes de type C18 testées, la Synergi C18 apparaît comme celle possédant la meilleure séparation (voir la Figure 41).

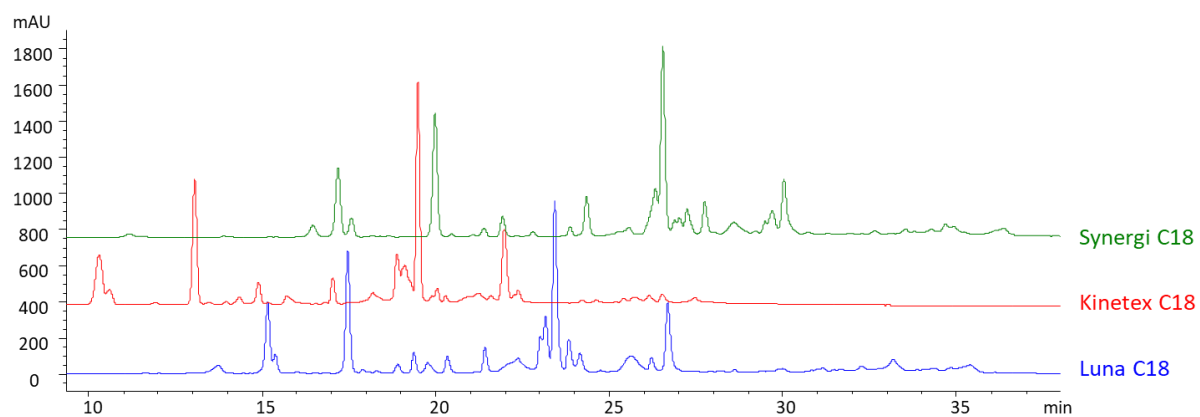


Figure 41 : Comparaison des colonnes de type C18 – agrandissement des chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso. Colonne Luna C18, Kinetex C18 et Synergie C18 ; gradients Généraliste (Luna C18, Synergie C18) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure, et Polaire 1 (Kinetex C18) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

Il est plus complexe de comparer la différence de sélectivité des colonnes avec des phases stationnaires de nature chimique différentes (voir la Figure 42). La Kinetex PFP notamment présente de nombreuses inversions dans l'ordre d'éluion. En effet, la phase PFP contient des greffons pentafluorophényl qui permettent des interactions de type  $\pi$ - $\pi$ , la rendant plus sélective aux composés aromatiques que les phases de type C18 ou C12. Les interactions de type  $\pi$ - $\pi$  se font entre le cycle phényl des greffons et les insaturations des analytes et plus particulièrement avec les cycles aromatiques très souvent présents chez les composés d'origine naturelle (Chirita 2009). De plus, la présence de fluor induit des liaisons hydrogène et des interactions dipôle-dipôle, facilitant la rétention

des composés polaires. Enfin l'encombrement stérique diffère par rapport aux greffons de type alkyle linéaire.

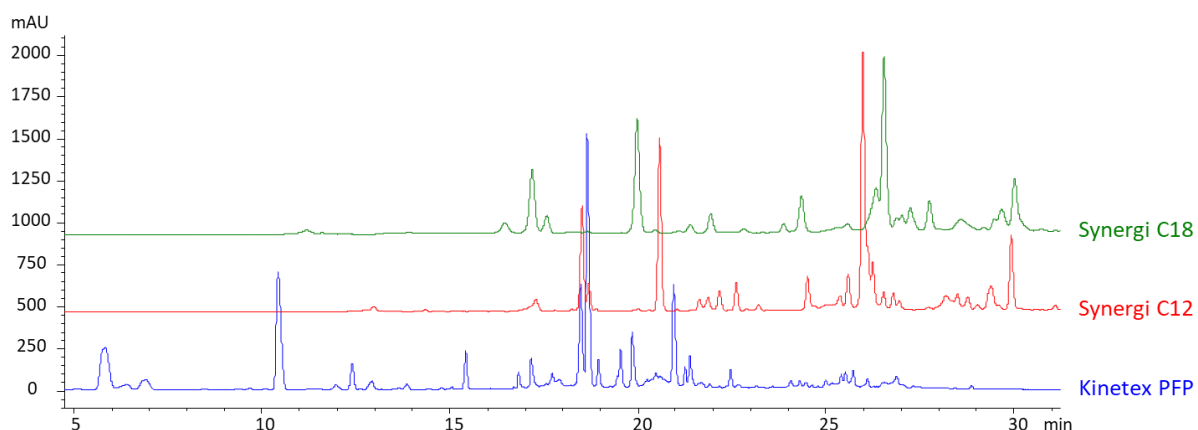


Figure 42 : Comparaison des colonnes ayant des phases stationnaires de nature chimique différentes - agrandissement des chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso. Colonne Synergi C18, Synergi C12 et Kinetex PFP (Annexe 4) ; gradients Généraliste (pour Synergi C18 et Synergi C12) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique, et d'isopropanol pure, et Polaire 1 (pour Kinetex PFP) constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

A partir des chromatogrammes, la colonne Kinetex PFP semble apparaître comme celle ayant la meilleure résolution pour les extraits eau-éthanol. Afin de comparer plus facilement les colonnes, le nombre de pics et la capacité des pics ont été calculés pour chacune.

### Capacité des pics pour les différentes colonnes :

La capacité totale des pics  $n_c$  permet de déterminer le nombre maximal de pics pouvant être séparés pour une fenêtre de temps et une résolution données. Elle permet d'apprécier la qualité d'une séparation, et dans notre cas de comparer les colonnes chromatographiques testées précédemment.

Dans le cas où un gradient est appliqué et que la largeur des pics est supposée constante  $W_{moy}$  tout au long du chromatogramme, l'équation est la suivante :

$$n_c = 1 + \frac{t_G}{W_{moy}}$$

Avec  $t_G$  la durée totale du gradient. (Attoma 2013)

Tableau 12 : Capacité totale et nombre de pics pour les différentes colonnes testées pour la séparation des extraits polaires

Colonne	Temps gradient (min)	Moyenne des largeurs des pics	Capacité totale	Nombre de pics <sup>(a)</sup>
Luna C18	70	0,191	367,49	77
Kinetex C18	70	0,167	420,16	85
Kinetex PFP	60	0,112	536,71	138
Synergi C12	70	0,140	501,00	112
Synergi C18	70	0,190	369,4f2	81

<sup>(a)</sup> le nombre de pics a été calculé par intégration automatique sur les chromatogrammes UV à 280 nm.

Le calcul de la capacité totale des pics conforte les observations faites jusqu'à présent, à savoir que la colonne Kinetex PFP est celle qui a la capacité de séparer le plus de composés. Nous avons donc choisi de conserver cette phase pour la suite des tests.

### Optimisation du gradient d'élution

Comme observé précédemment, la majorité des composés élue entre 5 et 40 min, une nouvelle méthode Polaire 2 (voir détails en Annexe 5), a été testée afin d'optimiser la séparation sur cette zone. Par rapport à la méthode Polaire 1, la pente du gradient a été adoucie pour passer de 114 à 86 %, ce qui a permis d'augmenter la séparation des composés. Les autres paramètres sont inchangés. L'extrait injecté est toujours le même que celui employé pour la détermination de la phase stationnaire, et la colonne est la Kinetex PFP (Annexe 4). Les profils obtenus sont présentés ci-dessous (Figure 43).

Ce nouveau gradient a été conservé pour la suite de l'étude.

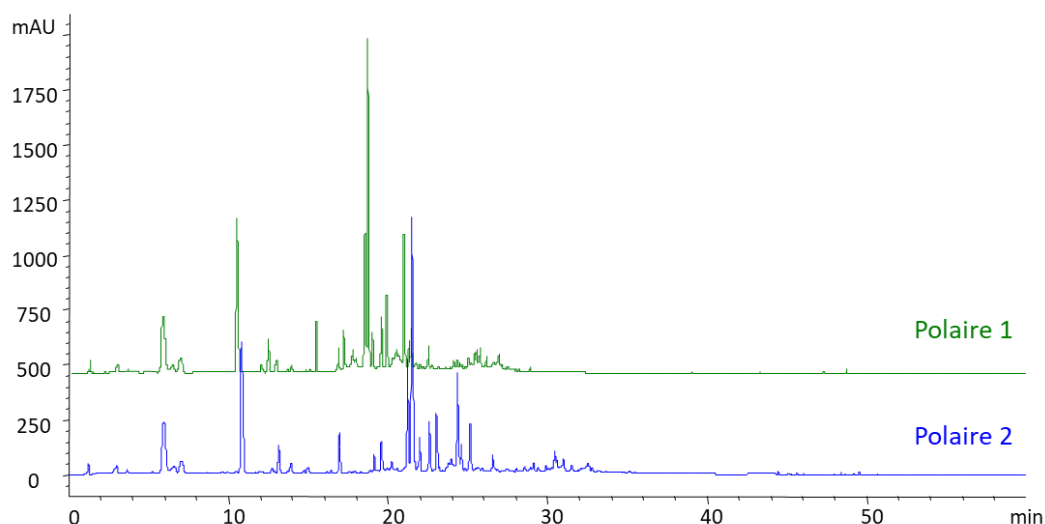


Figure 43 : Comparaison des gradients pour la séparation des extraits hydroalcoolique - chromatogrammes HPLC-UV des extraits eau-éthanol (1/1) de feuilles de lavandin grosso. Colonne Kinetex PFP (Annexe 4) ; gradients Polaire 1 et Polaire 2 constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; débit de 1 ml/min ; injection de 10 µL ; λ = 280 nm ; T = 25°C.



### IV.A.2. Optimisation des conditions analytiques HPLC pour les échantillons d'inflorescence

Pour des raisons techniques, l'optimisation des conditions analytiques a été effectuée intégralement sur des extraits de feuille. Cependant, les paramètres finaux doivent être optimaux pour toutes les conditions (variété, état sanitaire des plants, organes de la plante) car tous les extraits seront analysés avec des paramètres identiques pour permettre leur comparaison. Or, nous pouvons nous attendre à des profils chimiques différents pour les autres conditions, notamment avec le deuxième organe étudié : les inflorescences. Afin d'apprécier l'optimisation des conditions analytiques, un extrait d'inflorescences a été réalisé à partir d'un échantillon asymptotique de grosso (Gr9Sb) issu de la campagne d'échantillonnage 2017, et analysé selon les paramètres optimisés définis au préalable. Ainsi, un extrait eau-éthanol (1/1) a été réalisé selon le protocole de centrifugation établi précédemment. Les analyses ont été effectuées par HPLC-UV et la longueur d'onde exploitée est 280 nm.

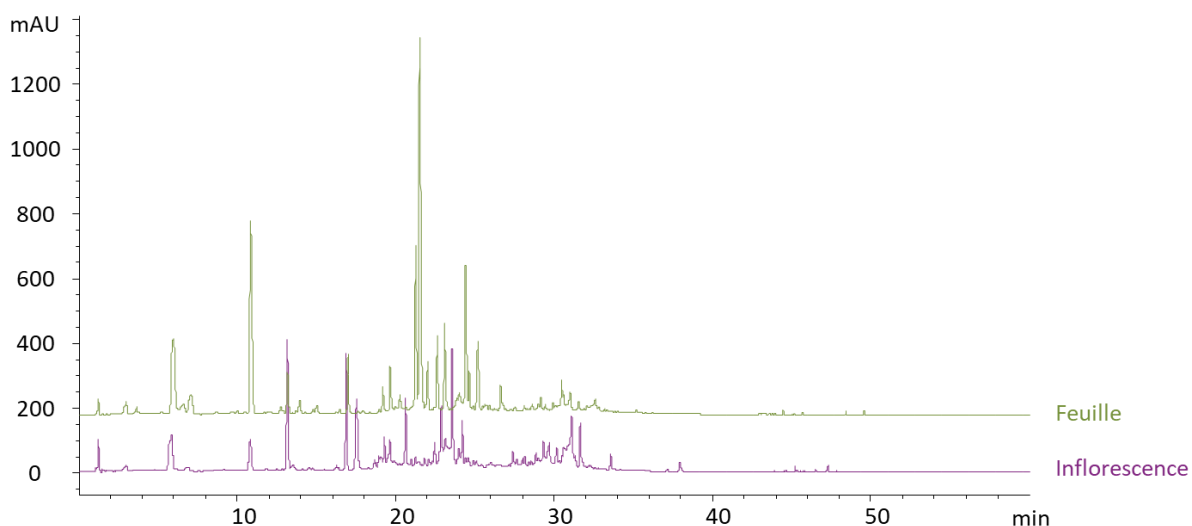


Figure 44 : Chromatogrammes HPLC-UV des extraits d'eau/éthanol (1/1) de feuilles et de fleurs de lavandin grosso asymptotique ; gradient d'élution Polaire 2 constitué d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; colonne Kinetex PFP (Annexe 4) ; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

Les extraits d'inflorescences se sont avérés très riches en composés et de nombreuses co-élutions apparaissent sur la zone d'élution 15 à 35 min (voir Figure 44). Un nouveau gradient Inflo 1 avec une pente inférieure a alors été testé. Il est constitué d'eau et d'acétonitrile acidifiés à 0,1 % en acide formique (détails en Annexe 5). Sur l'ensemble du chromatogramme, ce gradient est celui qui apporte la meilleure séparation (voir Figure 45). Un autre gradient d'élution Inflo 2 avait été testé, mais n'a apporté aucune amélioration significative (détails gradient en Annexe 5). Une diminution de la pente du gradient d'élution n'a pas été envisagée car aurait conduit à un allongement du temps d'analyse. D'autres paramètres tels que le volume d'injection, la température de la colonne ou le débit peuvent être utilisés et apparaissent plus appropriés afin d'améliorer la résolution. Malgré la présence de co-élution, l'optimisation des paramètres finaux se fera directement par UPLC-HRMS.

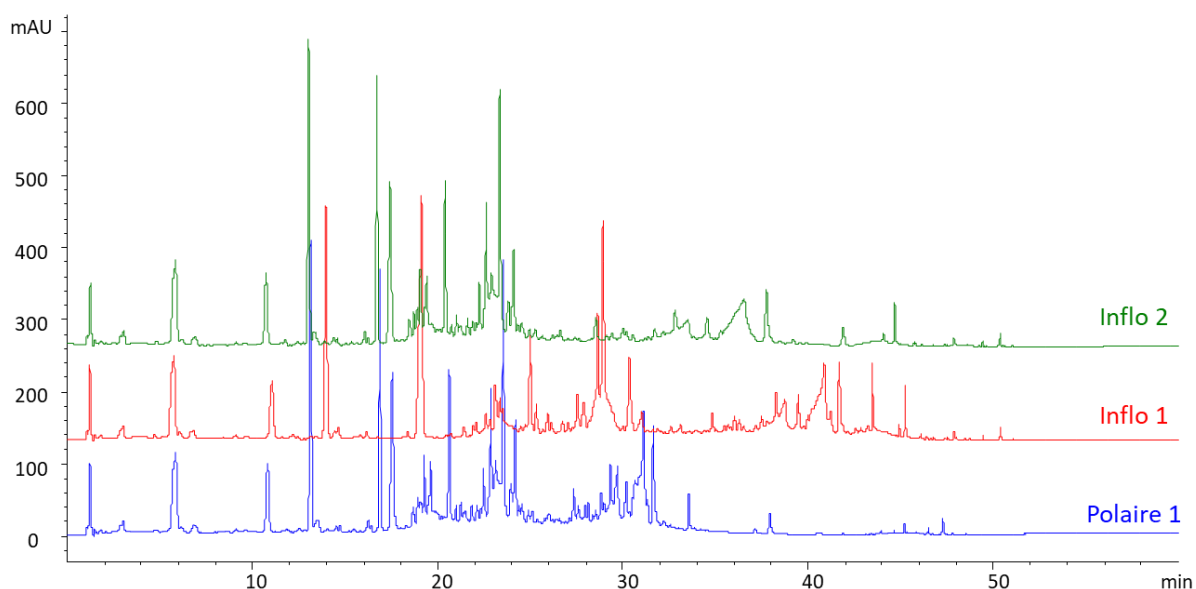


Figure 45 : Chromatogrammes HPLC-UV des extraits d'eau/éthanol (1/1) de fleurs de lavandin grosso asymptotique obtenus pour les trois gradient d'élution Polaire 1, Inflo 1 et 2 ; constitués d'eau et d'acétonitrile acidifiés à 0,1% en acide formique (Annexe 5) ; colonne Kinetex PFP; débit de 1 ml/min ; injection de 10  $\mu$ L ;  $\lambda$  = 280 nm ; T = 25°C.

### IV.B. Evaluation de la variabilité intra plant par HPLC-UV sur des extraits hydro-éthanoliques

Cette expérience avait pour but d'évaluer la variabilité intra plant dans l'optique de réduire le plan d'expérience. En d'autres termes, l'objectif était de regarder si la composition chimique était assez homogène pour permettre le regroupement des échantillons a, b et c d'un même plant (I.A). Cette réduction du plan d'expérience a plusieurs avantages dont celui de réduire le nombre de séquences d'analyse et ainsi éviter au maximum les effets de lot ; ou encore celui de diminuer la longueur des séquences afin de réduire la dérive analytique.

Afin que l'évaluation statistique soit convenable, nous avons choisi de conduire l'étude sur cinq plants. Au vu des disponibilités de la matière végétale, cette étude a été effectuée sur les échantillons a, b et c de feuilles et d'inflorescences des plants n°2, 7, 8, 9 et 10 de grosso asymptotique. Les extraits eau-éthanol (1/1) ont été réalisés en triplicats, selon le protocole d'extraction avec centrifugation établi précédemment (partie III.C). Deux pools, un pour chaque organe (feuille et inflorescence), ont été constitués en prélevant 25  $\mu$ L de chaque extrait (45). Les analyses ont été conduites par HPLC-UV avec la colonne Kinetex PFP (100 x 2,1 mm, 1,7  $\mu$ m) (spécificité Annexe 4) thermostatée à 25°C, et la méthode Polaire 2 (détails en Annexe 5). Une seule injection a été réalisée par échantillon. La longueur d'onde d'étude est 280 nm. Deux séquences analytiques ont été nécessaires pour chaque type d'extrait (feuille et inflorescence). Un blanc est injecté en début et fin de séquence ainsi que tous les 10 extraits. Le pool est injecté tous les 5 extraits. L'ordre d'injection suit celui d'extraction qui a été randomisé.

Le traitement des données comprend l'alignement des données brutes selon les temps de rétention qui a été effectué à l'aide d'une macro puis contrôlé voire ajusté manuellement, et les statistiques (ACP) qui ont été faites sous Xlstat.

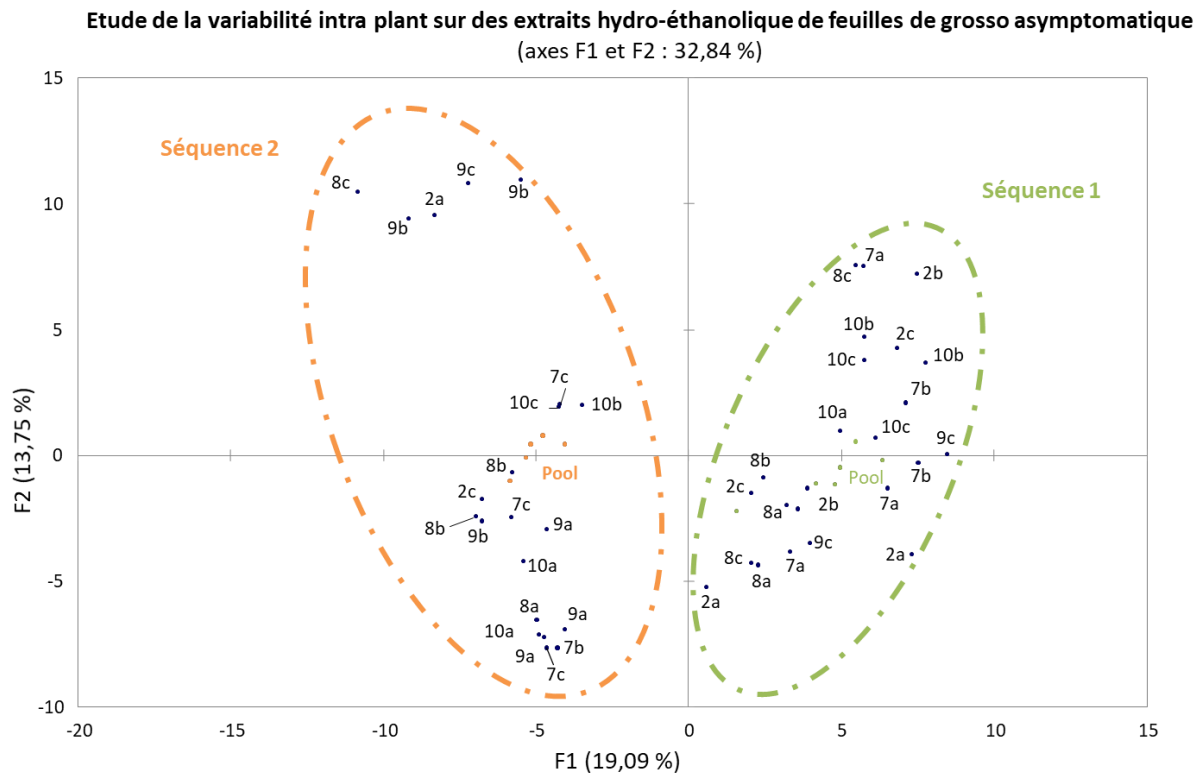


Figure 46 : ACP des extraits eau/éthanol (1/1) de feuilles de lavandin grosso, résultats obtenus avec Xlstat à partir des résultats HPLC-UV à 280 nm

Remarque : Les codes sur les graphiques des individus correspondent au numéro de plant et les réplicats biologiques sont différenciés par des lettres. Les triplicats d'extraction ne sont pas signalés et ont donc un code identique.

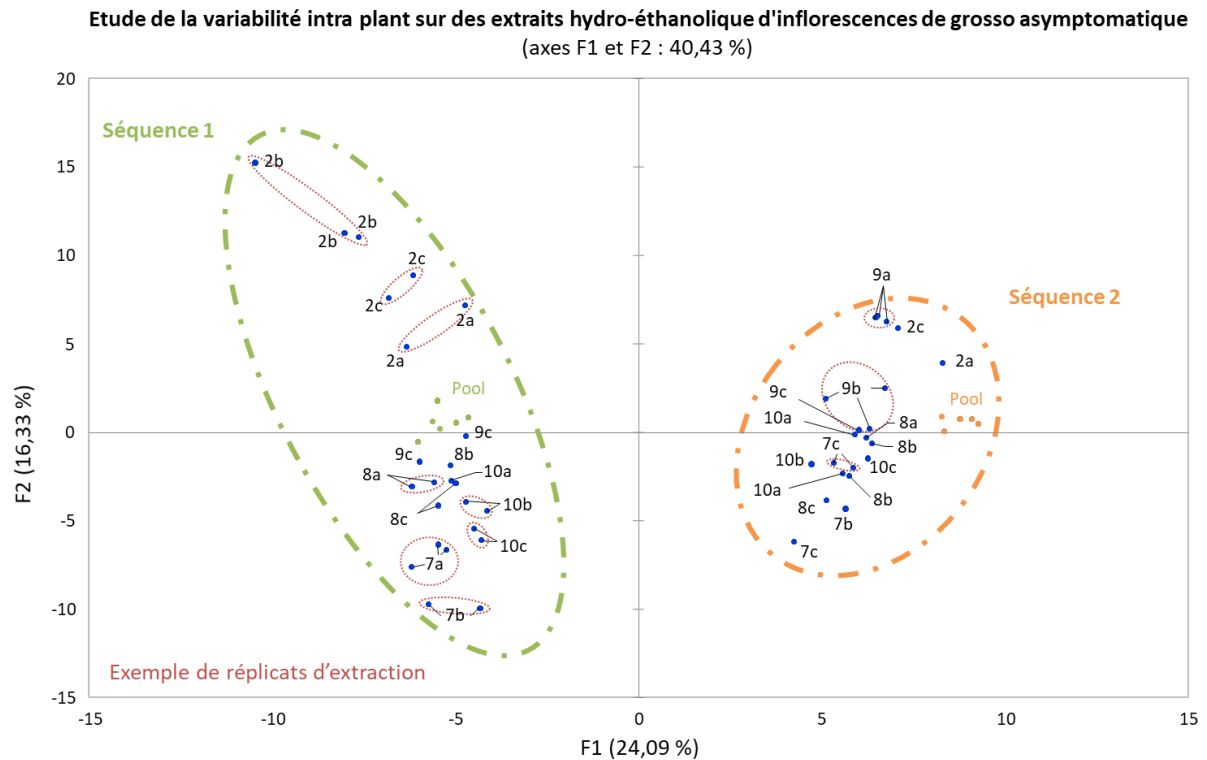


Figure 47 : ACP des extraits eau/éthanol (1/1) d'inflorescence de lavandin grosso, résultats obtenus avec Xlstat à partir des résultats HPLC-UV à 280 nm

Afin de visualiser le comportement du jeu de données, deux ACP ont été réalisées, une sur les extraits de feuilles (Figure 46) et la deuxième sur ceux d'inflorescences (Figure 47). Malheureusement la variabilité inter séquence l'emporte sur toute autre variabilité. Cependant, pour l'ACP des inflorescences, il semble se dégager des tendances prometteuses au sein d'une même séquence, à savoir que la variabilité biologique inter plant est supérieure à la variabilité intra plant, elle-même plus importante que la variabilité liée au protocole d'extraction (cf. proximité des répliquats d'extraction).

Afin de se prémunir de la variabilité inter séquence et ainsi obtenir une meilleure visualisation des variabilités intra et inter plant, nous avons réalisé l'ACP uniquement sur la première séquence (voir Figure 48 ci-dessous).

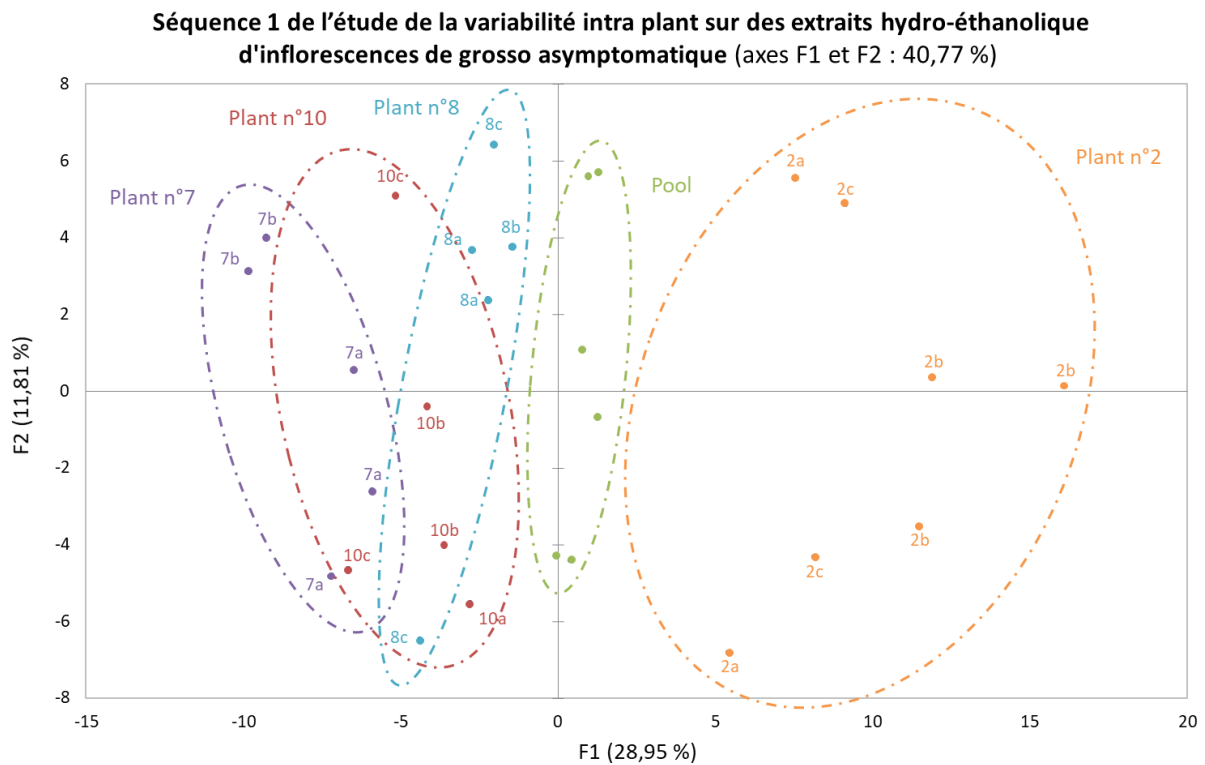


Figure 48 : ACP des extraits eau/éthanol (1/1) d'inflorescence de lavandin grosso de la séquence n°1, résultats obtenus avec Xlstat à partir des résultats HPLC-UV à 280 nm

L'interprétation de ce graphique (Figure 48), se fera essentiellement en fonction de la première composante car l'axe F2 porte essentiellement la dérive analytique au vu de l'éclatement du groupement des pools selon cette composante. Il est assez aisé de visualiser les groupements d'extraits appartenant à un même plant sur ce graphique et de constater la présence de variabilité intra et inter plant. Cependant, il est difficile de conclure que la variabilité intra est plus importante que la variabilité inter plant, notamment à cause de la superposition des plants 7, 8 et 10.

Même si les résultats semblent prometteurs, la variabilité inter-séquence l'emporte et ne nous permet pas de conclure en faveur de la réduction du plan d'expérience.

#### IV.C. Optimisation UPLC-HRMS

En se basant sur l'optimisation effectuée sur l'HPLC-UV, les conditions chromatographiques et d'ionisation ont été optimisées en UPLC-HRMS. Cette optimisation a également tenu compte des différentes catégories d'échantillons (feuille/inflorescences, variétés, état sanitaire) présents dans l'étude.

##### IV.C.1. Optimisation conditions chromatographiques

###### a. Modification colonne

A l'issu des tests préliminaires en HPLC, la colonne Kinetex PFP s'était avérée comme étant le meilleur compromis. La même phase stationnaire adaptée pour l'UPLC a été testée. Malgré de nombreux lavages, un échange de colonne et des tests sur la dernière version de la phase, à savoir la colonne Kinetex PFP F5 (plus résistante selon le constructeur), un bruit de fond non négligeable et attribué à la perte de phase stationnaire, est présent sur les analyses (voir Figure 49).

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

La colonne C18 HSS T3 (100 x 2,1 mm) de chez Waters (détails des spécificités en Annexe 4) a alors été testée et a montré une ligne de base de faible intensité, une absence de dérive et une bonne séparation des analytes. De plus, un nombre d'ions légèrement supérieur a été détecté sur cette colonne. Elle a donc fait l'objet d'une optimisation des paramètres analytiques et a été employée pour la suite de l'étude.

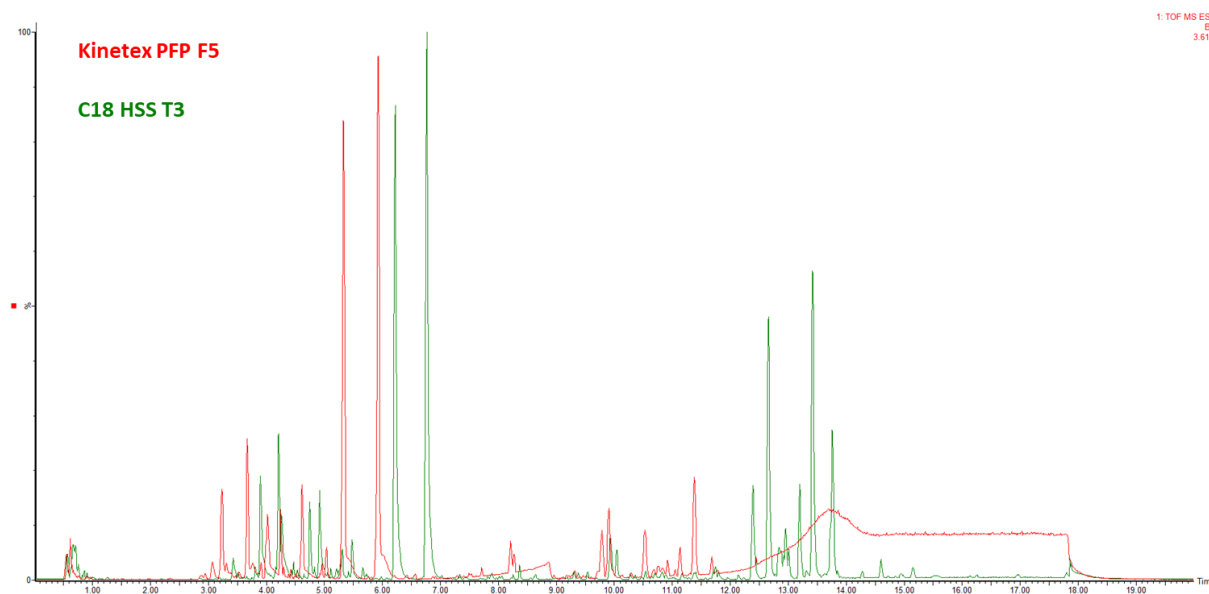


Figure 49 : Chromatogrammes UPLC-HRMS d'extraits d'eau/éthanol (1/1) de feuilles de lavande 7713 asymptomatique sur les colonnes Kinetex PFP F5 et C18 HSS T3 (Annexe 4) ; gradient d'élution eau et acétonitrile à 0,1% d'FA Annexe 5 ; débit de 0,35 ml/min ; injection de 1  $\mu$ L ; T = 40°C. ESI en mode positif

### Spécificité de la colonne : ACQUITY UPLC HSS T3 (C18)

Cette colonne dispose d'un greffage C18. Elle a été conçue pour être adaptée même à des composés très polaires et pour une utilisation avec une phase mobile aqueuse grâce à la technologie HSS (*High Strength Silica*). En effet, les groupements siloxanes non greffés par les groupes fonctionnels (C18) sont greffés avec des petites molécules afin de les rendre « inactifs ». Cet « *endcapping* » confère à la phase stationnaire une très bonne résistance et permet d'éviter des rétentions de composés très polaires de façon irréversible ou trop importante.

Tableau 13 : Spécificités de la colonne ACQUITY UPLC HSS T3

Gamme pH (à température ambiante)	Température limite	Taille des particules (sphérique)	Taille des pores/surface spécifique	Taux carbone
2-8	45°C	1,8 $\mu$ m	100 Å / 230 m <sup>2</sup> .g <sup>-1</sup>	11%

### b. Paramètres chromatographiques

L'optimisation du gradient d'élution a consisté à réduire au maximum la durée d'une analyse tout en conservant une bonne séparation. La réduction même minime du temps d'une analyse n'est pas négligeable au vu du nombre d'analyses à réaliser durant l'étude métabolomique. Ceci est possible grâce à l'UPLC qui assure une séparation accrue des composés.

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

Dans un premier temps, le palier en début d'élution a été ramené de 1,5 à 1 min. Dans un deuxième temps, des tentatives ont été menées afin de réduire la zone d'élution des composés hydrophiles. Cette zone d'élution est critique au niveau de la séparation du fait d'un nombre élevé d'analytes et n'a pu être raccourcie. Par la suite, la durée du gradient a ensuite été progressivement réduite de 20, puis 19, jusqu'à 18,5 min en rognant sur la zone d'élution de la phase organique. Plusieurs débits ont également été testés, à savoir 0,35 ; 0,4 et 0,45 ml/min. Afin de diminuer le temps d'analyse, un gradient basé sur le gradient n°4 (UPLC\_HSST3\_Grad04 voir Annexe 5) mais d'une durée inférieure (17 min) et avec un débit plus élevé de 0,6 ml/min a été testé. Cette diminution était un peu trop ambitieuse et des co-élutions sont apparues. Un gradient (UPLC\_HSST3\_Grad\_final) plus raisonnable de 18,5 min avec un débit de 0,4 ml/min a finalement été conservé (voir Annexe 5). La composition des gradients est inchangée (eau et acétonitrile acidifiés à 0,1% en acide formique).

D'autres paramètres sont généralement optimisés tel que le volume d'injection ou la température de la colonne. Le volume d'injection n'a pas fait l'objet d'optimisation (1 µl). Une augmentation du volume d'injection aurait été bénéfique pour les pics de faibles intensités mais n'aurait pas été judicieux car les pics majoritaires ont tendance à trainer.

Afin de pallier les variations de la température de la pièce, il est conseillé de thermostatier à minima la colonne à 30 voire 35 °C. Les températures de colonne de 35, 40 et 45 °C ont fait l'objet de tests. L'augmentation de la température a amélioré la qualité de la séparation des profils chimiques. Cependant pour éviter toute dégradation thermique, nous avons fait le choix de conserver une température de 40 °C.

### c. Vérification de la séparation des profils en fonction de la variété, de l'état sanitaire du plant et de l'organe

Cette étude regroupe plusieurs conditions biologiques différentes (variétés, état sanitaire et organes de la plante) pour lesquels nous pouvons nous attendre à rencontrer des profils chimiques dissemblables. Il est probable que la plus grande hétérogénéité chimique soit imputable aux deux conditions : variétés et organes de la plante. Il est important que les conditions analytiques optimisées soient adaptées et assurent une bonne séparation pour toutes ces catégories, afin de permettre leur comparaison.

Tableau 14 : Echantillons utilisés pour la mise en place des conditions analytiques en UPLC-HRMS

Espèces	Lavande				Lavandin		
Variétés	7713		Maillette		Abrial	Grosso	
Organes	Feuilles						Inflo.
Etat sanitaire du plant	Asy.	Sym.	Asy.	Sym.	Asy.	Sym.	Asy.
Code échantillon	77 10SaF	77 2IaF	Ma 2ScF	/	Ab 1SaF	Ab 1IaF	Pool <sup>(a)</sup>

<sup>(a)</sup> Utilisation des pools réalisés lors de l'étude de la variabilité intra plant par HPLC (IV.B)

Inflo. = Inflorescence ; Asy. = Asymptomatique ; Sym. = Symptomatique

La vérification de la qualité de la séparation selon les variétés a été effectuée en fonction de la disponibilité du matériel végétal à savoir sur un extrait asymptotique de feuille de chaque variété. Les profils obtenus présentent beaucoup de similarité et les conditions chromatographiques sont adaptées à toutes les variétés de l'étude. Les profils optimisés pour les différentes variétés sont représentés sur la Figure 50.

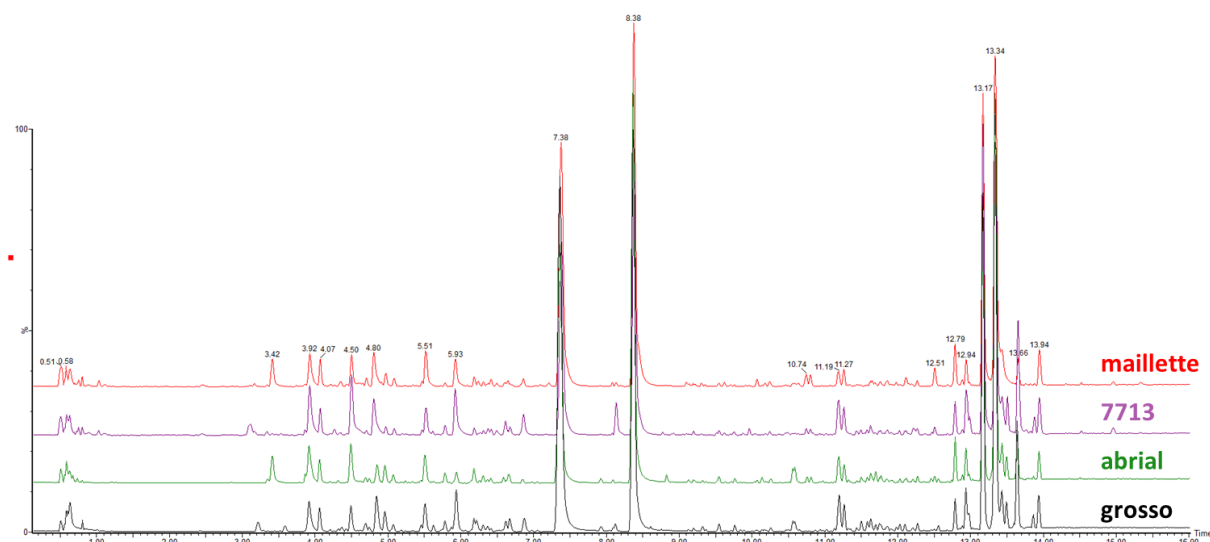


Figure 50 : Chromatogrammes UPLC-HRMS (BPI = Base Peak Ion) d'extraits d'eau/éthanol (1/1) de feuilles des quatre variétés étudiées (7713, maillette, abrial et grosso) ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; débit de 0,4 ml/min ; colonne C18 HSS T3 (Annexe 4) ; injection de 1  $\mu$ L ; T = 40°C. ESI en mode positif

En ce qui concerne l'état sanitaire des plantes, un extrait de feuille de chaque condition a été analysé pour la lavande (7713) et pour le lavandin (abrial). La comparaison des chromatogrammes n'a révélé que peu de dissimilarité sur le plan qualitatif entre les deux états des plants (voir Figure 51). Les profils asymptomatiques semblent plus riches quantitativement et qualitativement. La Figure 51 présente également les chromatogrammes obtenus pour les pools de grosso de feuilles et d'inflorescences (voir Tableau 14). Comme attendu, les profils présentent des différences non négligeables, notamment sur la zone d'élution comprise entre 8,5 et 13 min où les extraits de fleurs sont plus riches.

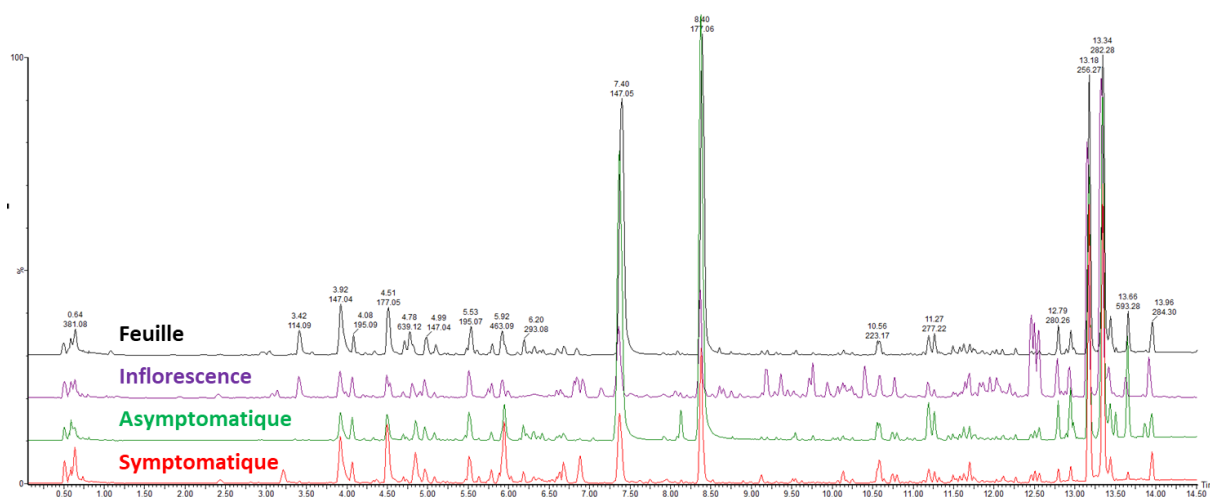


Figure 51 : Agrandissement des chromatogrammes UPLC-HRMS (BPI) d'extraits d'eau/éthanol (1/1) de feuilles d'abrial asymptomatique et symptomatique, et de feuilles et inflorescences de grosso asymptomatique ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; débit de 0,4 ml/min ; colonne C18 HSS T3 (Annexe 4) ; injection de 1  $\mu$ L ; T = 40°C. ESI en mode positif

Les conditions obtenues sont compatibles avec l'ensemble des catégories d'échantillons présents dans l'étude métabolomique.



### IV.C.2. Optimisation conditions d'ionisation

Le capillaire est généralement soumis à un fort potentiel positif (de 0,5 à 3,2 kV) ou négatif (de 0,5 à 2,6 kV). Le cône d'échantillonnage est quant à lui soumis à un potentiel compris entre 20 et 100 V. Ces deux paramètres sont à optimiser pour gérer la fragmentation en source, en sachant qu'une trop forte fragmentation serait néfaste pour la visibilité de l'ion moléculaire.

La diminution de la tension du capillaire de 1,85 kV à 1 kV, n'a eu que peu d'impact sur les résultats observés et nous avons fait le choix de conserver une tension de 1,85 kV.

Trois tensions appliquées au cône ont été testées : 20, 25 et 30 V. La tension de 20 V a fourni de meilleurs résultats en début d'élution alors qu'à l'inverse la tension de 30 V donnait de meilleurs résultats en fin d'élution. La tension de 25 V donnant un compromis a été conservée.

Les deux modes d'ionisation (positif et négatif) ont été testés afin de déterminer le mode le plus approprié pour cette étude. Même si l'ionisation est plus efficace sur la zone d'élution de la phase aqueuse en mode négatif, il a été constaté une perte d'information lors de l'élution de la phase organique. Le mode d'ionisation négatif ionise un nombre inférieur d'analyte que le mode positif, avec 792 ions détectés contre 1382 ions en positif. Pour ces raisons, le mode positif a été conservé pour la suite de l'étude.

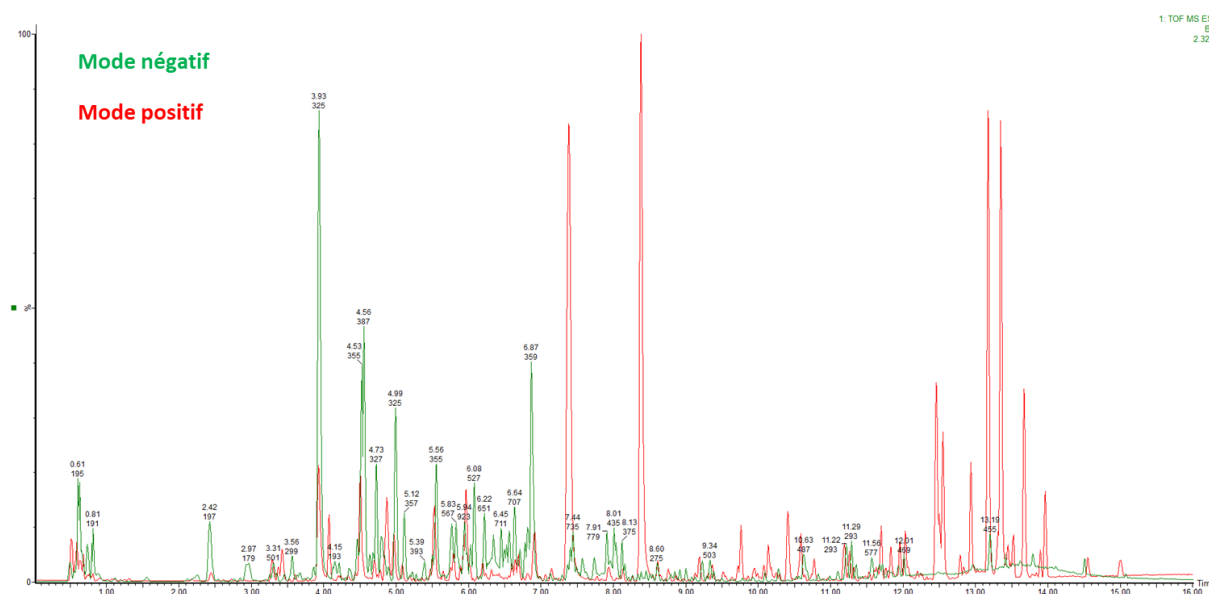


Figure 52 : Chromatogrammes UPLC-HRMS (BPI) d'extraits eau/éthanol (1/1) de feuilles de lavandin grosso asymptotique en mode positif et négatif ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; colonne C18 HSS T3 (Annexe 4) ; débit de 0,4 ml/min ; injection de 1 µL ; T = 40°C.

### IV.C.3. Séquence analytique

Chaque lot d'extraction a été divisé en deux séquences d'analyses. Par conséquent, chaque séquence comprend 180 extraits, ainsi que l'injection à intervalle régulier du pool et l'injection de blancs en début et fin de séquence. Une séquence correspond à la répétition (20 fois) d'un motif comprenant 9 injections d'extrait suivi d'une injection de pool. La présence du pool et de standard interne permet de surveiller et d'évaluer le biais analytique. Avant de lancer une séquence analytique, un pré-lancement contenant 2 blancs et l'injection de 10 pools successifs est réalisé afin de vérifier le système. En début de séquence, 10 injections du pool sont réalisées pour conditionner la colonne (Saint-Lary 2015; Favre 2017).

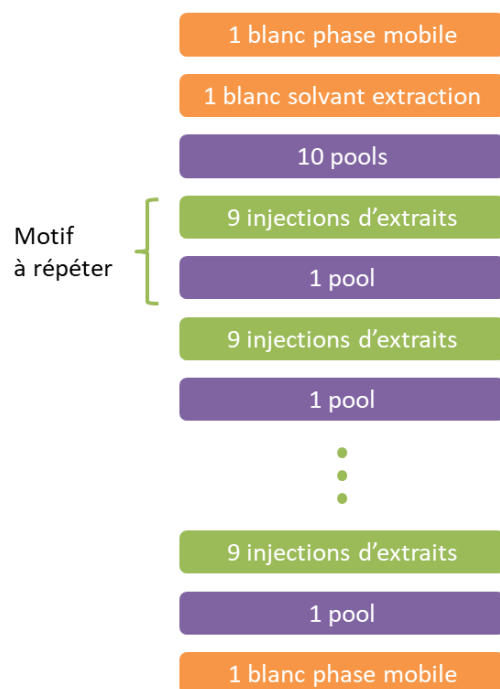


Figure 53 : Schématisation d'une séquence d'analyse type

Blanc phase mobile (acétonitrile) : permet de contrôler l'absence de contaminant dans le système analytique (*i.e.* appareil, phases mobiles), éventuellement de supprimer ces contaminants et dans le cas où le système n'en serait pas exempt de prendre les mesures adéquates (nettoyage de l'appareil, renouvellement des phases mobiles, ...), et afin de permettre d'identifier les ions à supprimer lors du retraitement des données.

Blanc solvant d'extraction (eau/éthanol (1/1) avec 5 ppm de caféine qui constitue le standard interne) : il permet de mettre en évidence la présence de contaminant dans le solvant d'extraction ou le standard interne, et d'établir une liste de ces contaminants en vue de les supprimer lors du retraitement des données.

Blanc phase mobile en fin de séquence : permet de vérifier la présence d'ions résiduels mal élués provenant des extraits analysés.

#### IV.D. Détermination standard interne

L'objectif est de trouver un standard interne utilisable pour l'ensemble de l'étude métabolomique. Pour ce faire, le composé doit répondre à plusieurs critères.

- Il doit être soluble à température ambiante dans le solvant d'extraction à savoir un mélange eau-éthanol (1/1)
- Il doit être ionisable par un électronebuliseur en mode positif
- Il doit être absent de tous les extraits de l'étude
- Le composé doit éluer dans une zone où il y a peu d'élution afin de ne pas interférer avec les analytes des extraits
- Il doit être disponible en quantité suffisante et d'une très grande pureté afin de ne pas introduire de contaminants
- Idéalement, il doit être de faible coût

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

De nombreux standards se sont révélés insolubles ou faiblement solubles dans l'eau à température ambiante. Selon ce premier critère et l'absence du composé dans la base de données bibliographique établie sur les composés non volatils dans le genre *Lavandula* (Annexe 1), nous avons fait le choix de tester deux molécules : la *o*-vanilline ( $m = 152,15$  g/mol) et la caféine ( $m = 194,19$  g/mol). Toutes les deux ayant une solubilité dans l'eau de 10 mg/ml et une solubilité dans l'éthanol de 50 mg/ml.

La caféine a été testée dans un extrait de feuille de 7713 à une concentration de 50 µg/ml. La molécule est éluée vers 4,10 min sans co-élution (voir Figure 54 ci-dessous) avec un rapport  $m/z$  de 195,089 pour l'ion  $[M+H]^+$ . Malgré une concentration élevée, la caféine ne présente pas d'effet rémanent ou mémoire (aucune trace n'a été observée dans le blanc). Après avoir testé différentes concentrations (5, 10 et 50 ppm), la concentration à 5 µg/ml a été finalement conservée (voir Figure 54 ci-dessous). Les concentrations supérieures étaient trop élevées, la phase stationnaire saturait provoquant un pic trainant dans une zone de forte élution ainsi qu'un pic majoritaire supérieur aux analytes des extraits.

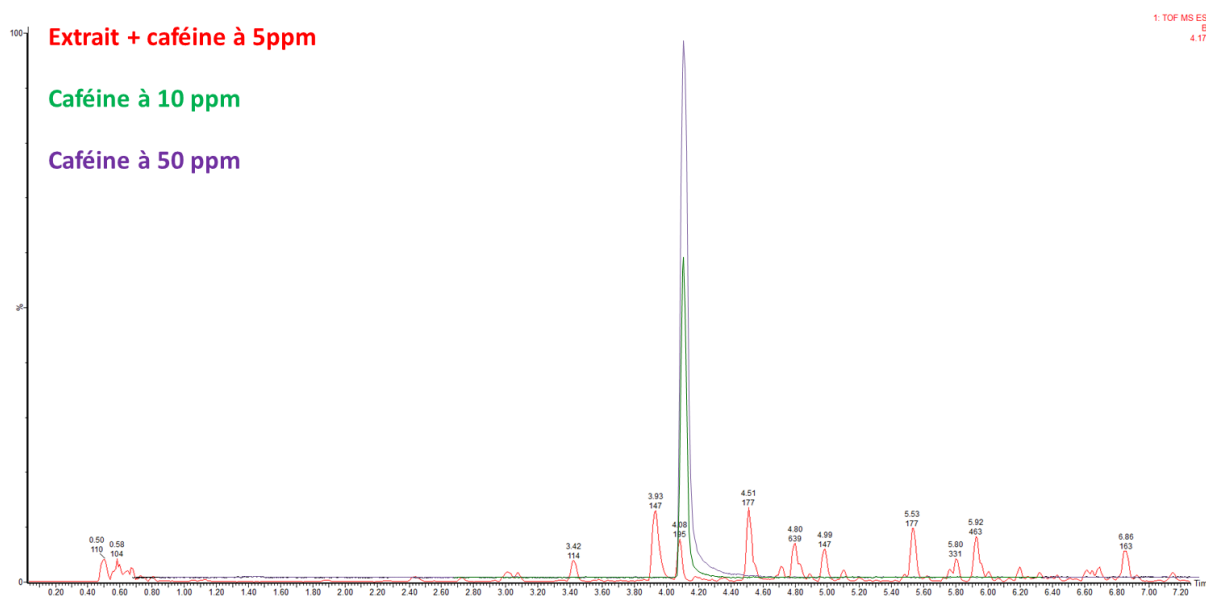


Figure 54 : Chromatogrammes UPLC-HRMS (BPI = Base Peak Ion) de la caféine à 10 et 50 ppm et d'un extrait eau/éthanol (1/1) de feuilles de grosso asymptotique contenant de la caféine à 5ppm ; gradient d'élution eau et acétonitrile à 0,1% d'FA ; débit de 0,4 ml/min ; colonne C18 HSS T3 (Annexe 4) ; injection de 1 µL ; T = 40°C. ESI en mode positif.

La deuxième molécule, la vanilline a été testée dans un extrait de feuille d'abrial à une concentration de 25 µg/ml. La molécule éluée vers 3,55 min avec un rapport  $m/z$  de 153,056 pour l'ion  $[M+H]^+$ . La zone d'élution de la vanilline est trop proche d'un pic de faible intensité éluant à 3,58 min. Il y a risque de co-élution.

La caféine a donc été conservée comme standard interne pour l'étude.

### IV.E. Contrôle de la dérive analytique

En amont du lancement de l'étude et en complément de l'optimisation des paramètres, un test de robustesse a été conduit afin d'évaluer la dérive analytique pouvant être engendrée par une longue séquence d'analyse. Par la suite, cette dérive a été suivie durant l'acquisition des données lors de l'étude métabolomique.

#### IV.E.1. Test de robustesse du système analytique

Comme explicité dans le chapitre 1, il est important d'essayer dans une étude métabolomique non ciblée, de réduire et de contrôler toutes les sources de variabilités indésirables. Dans ce cas nous

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

cherchons à anticiper et à minimiser la variabilité technique liée au système analytique ; notamment celle engendrée par une dérive du système à cause d'un nombre d'injection élevé. Rappelons, que nous avons 1429 extraits à analyser. (van der Kloet et al. 2009; Wolfender et al. 2013; Dudzik et al. 2018; Kovacevic and Simpson 2020)

Pour ce faire, un test de robustesse du système analytique a été mis en place. Ce test a pour objectif de visualiser une potentielle dérive analytique et d'évaluer la répétabilité du système lorsque celui-ci est soumis à de longues séquences d'analyse. Il a consisté à effectuer un nombre important d'injections successives (minimum 1/10ème de la séquence d'analyse prévue) d'un même échantillon avec la même méthode analytique. Dans le cadre de cette étude, 27 injections successives ont été conduites pour évaluer la robustesse de l'UPLC-HRMS. Afin de recréer des conditions proches des futures séquences d'analyse, le test a utilisé un pool d'extrait de feuille de grosso asymptomatique réalisé précédemment (IV.B), dans lequel de la caféine a été dissoute à 5 µg/ml.

### Evaluation de la dérive qualitative : temporelle et massique

Le test n'a montré aucun décalage sur les temps de rétention sur l'intégralité des profils. L'écart-type de la dérive de la caféine pour les temps de rétention sur l'ensemble du test est de 0,01 min et la dérive en masse de la caféine est de  $10^{-3}$  Da.

### Evaluation de la dérive quantitative : aire

Avec un grand nombre d'injections répétées, le système peut s'encrasser et la capacité d'ionisation diminuer (Furey et al. 2013). La déviation de l'aire du standard interne a été évaluée en regardant les variations en fonction du nombre d'injections. Sur ce test, aucune diminution n'a été observée. L'écart-type obtenu sur l'ensemble des analyses est de 627, soit 5,8 % de l'aire moyenne.

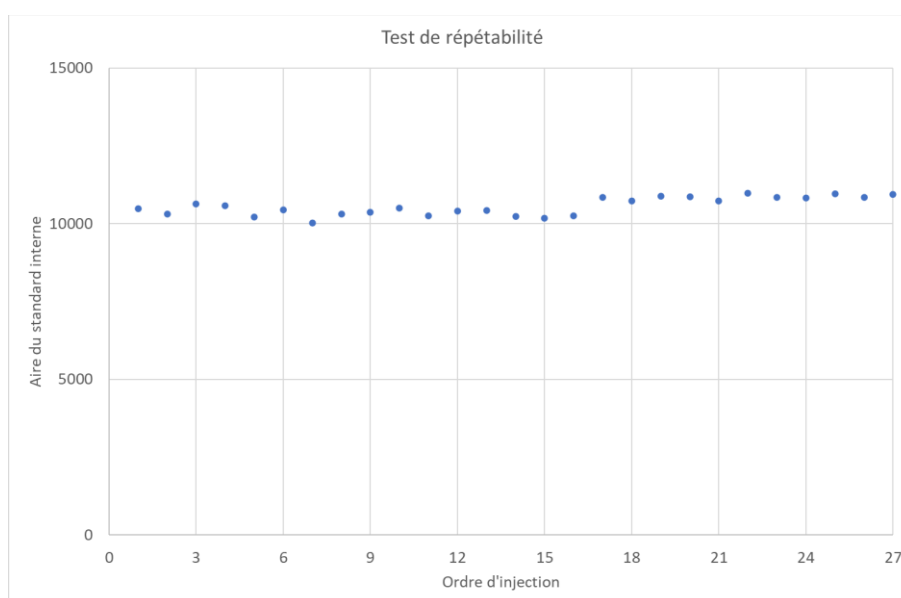


Figure 55 : Suivi de l'aire du standard interne (caféine) en fonction du nombre d'injection

Le seul point négatif constaté pendant ce test est un petit effet mémoire qui a été corrigé par une augmentation du lavage de la boucle d'injection de 300 µl à 500 µl. Le système analytique a très bien réagi à ce test. Au vu de ces résultats, l'étude métabolomique a été lancée.

IV.E.2. Contrôle de la dérive analytique au cours des séquences

a. Suivi stabilité du signal

Les longues séquences d'analyses peuvent causer une perte de sensibilité. Cette décroissance du signal est provoquée par l'encrassement du spectromètre de masse. Au fur et à mesure des injections, des particules s'accumulent dans la source induisant une diminution de l'ionisation. Ce phénomène est aussi nommé effet matrice (Furey et al. 2013; Dudzik et al. 2018). Son impact dans cette étude a été surveillé en suivant l'aire du standard interne contenu dans le pool injecté à intervalle régulier au cours de la séquence.

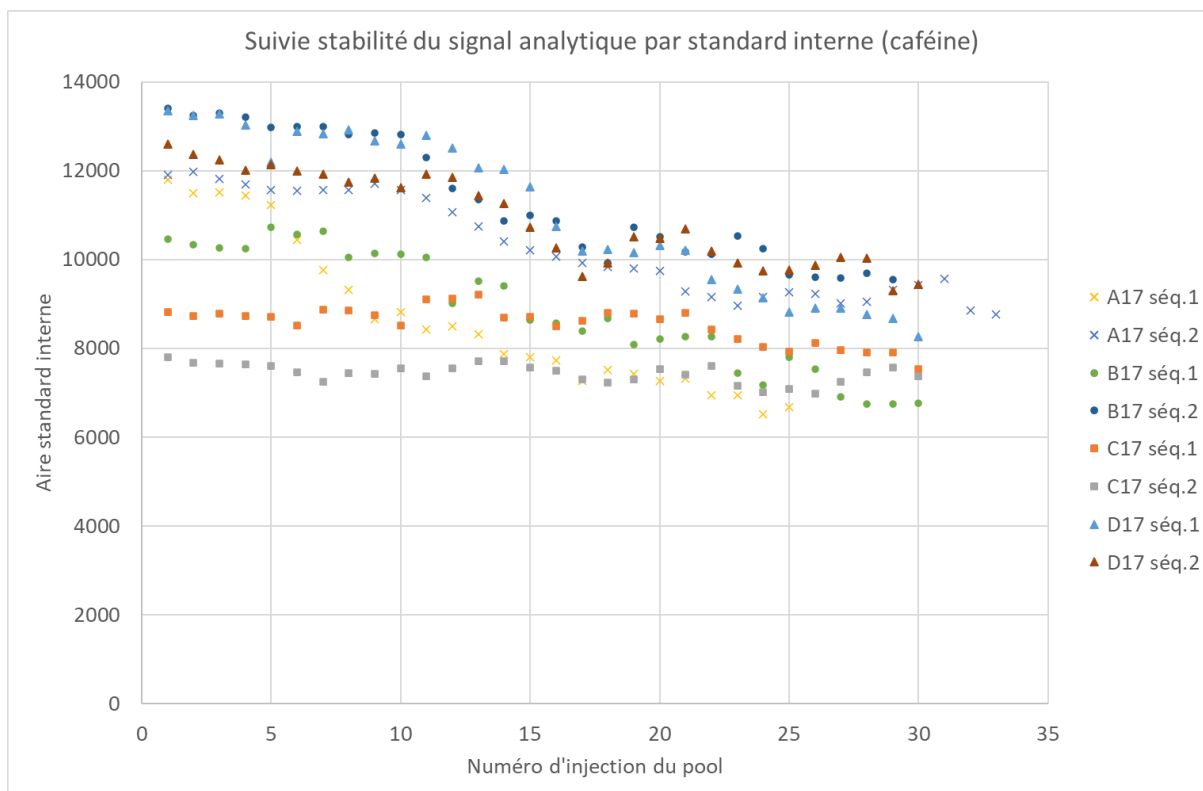


Figure 56 : Suivi de l'aire du standard interne dans les pools au cours de chaque séquence d'analyse

Comme nous pouvons le constater sur la Figure 56, de manière générale les premières séquences présentent une décroissance plus importante. Ce phénomène n'est pas surprenant car la source a été nettoyée entre chaque lot. Quant à la perte de signal des deux séquences du lot C17 très inférieure à celle des autres séquences, elle a été attribuée au changement de la pré-colonne.

Tableau 15 : Le classement des séquences en fonction de la différence maximale d'aire de la caféine dans les pools sur l'intégralité de chaque séquence

Séquence	A17 séq.1	D17 séq.1	B17 séq.1	B17 séq.2	D17 séq.2	A17 séq.2	C17 séq.1	C17 séq.2
$\Delta(\text{aire}) \text{ max}$	5280	5086	3986	3848	3297	3207	1666	813

Lorsque nous observons la variation de l'aire de la caféine sur l'intégralité des séquences, il est intéressant de noter que pour plusieurs d'entre elles la perte du signal suit une décroissance exponentielle et semble se stabiliser au fur et à mesure des injections. Le comportement des séquences a été modélisé grâce à un polynôme de degré 2 dans la Figure 57 ci-dessous.

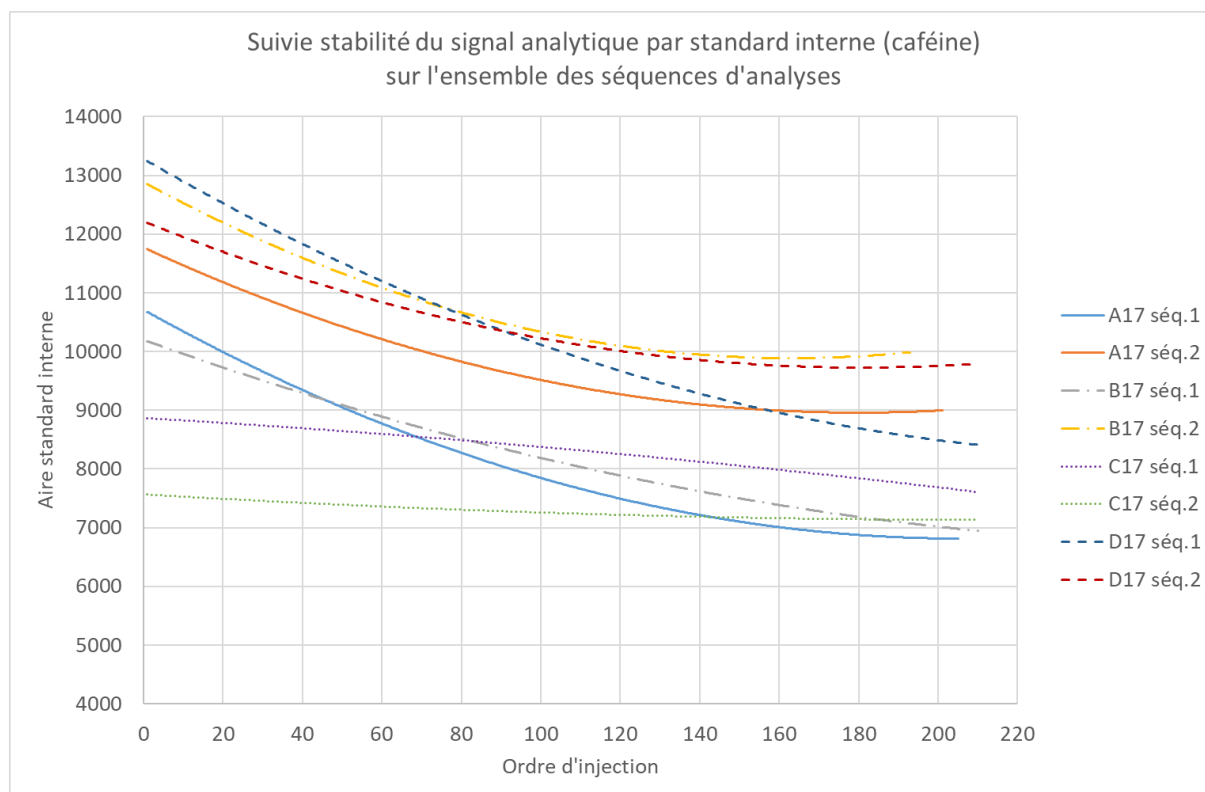


Figure 57 : Modélisation suivant un polynôme de degrés 2 de l'évolution de l'aire du standard interne suivant l'ordre d'injection pour l'ensemble des séquences d'analyse

Malgré les efforts déployés (changement de pré-colonne, nettoyage de la source, test de répétabilité, calibration du détecteur), une perte de signal importante est à déplorer. Ce cas illustre parfaitement l'importance de l'introduction d'un standard interne et de l'injection régulière des pools afin de permettre de palier à la dérive analytique lors du retraitement des données.

### b. Dérive temporelle et massique

La dérive analytique n'est pas seulement une perte de sensibilité, les analyses peuvent être sujettes à une dérive en masse ou sur les temps de rétention. Dans cette étude, ces dérives ont été surveillées en superposant les profils obtenus pour les injections des pools en début et fin de séquence.

En ce qui concerne la dérive massique, la répétabilité du TOF est excellente. La dérive en masse de la caféine est inférieure à 5 ppm pour toutes les analyses.

La dérive temporelle quant à elle, pour le standard interne est très faible et ne pose aucune difficulté sur l'ensemble des séquences. La dérive maximale de la caféine est de 0,06 min et l'écart-type de 0,008 min sur toutes les injections. Cependant, tous les analytes n'ont pas réagi de la même manière et une dérive temporelle a été observée pour les deuxième séquences des lots B17 et D17. Les composés affectés sont situés dans la zone d'élution de la phase aqueuse généralement entre 3 et 8 min (voir Figure 58). Cette dérive temporelle est irrégulière et a été attribuée à l'encrassement de la pré-colonne. Des précautions seront à prendre lors du retraitement des données afin de corriger ce décalage des temps de rétention.

## Chapitre 2 : Conception et optimisation de l'étude métabolomique

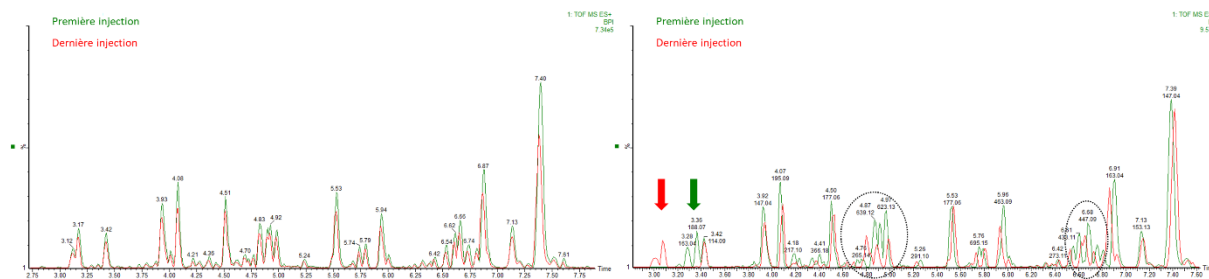


Figure 58 : Agrandissement des chromatogrammes UPLC-HRMS de la première injection et de la dernière injection du pool du lot B17 séquence 1 (à gauche) n'ayant aucune dérive temporelle et de la séquence 2 (à droite) dont les anomalies ont été entourées ou repérées par des flèches ; gradient d'éluion eau et acétonitrile à 0,1% d'FA ; colonne C18 HSS T3 (Annexe 4) ; débit de 0,4 ml/min ; injection de 1  $\mu$ L ; T = 40°C. ESI en mode positif

### V. Design expérimental

A l'issu des tests préliminaires, le design adopté pour la partie expérimentale comprend : l'échantillonnage en champ de quatre variétés avec trempé à l'azote liquide, la conservation à  $-80^{\circ}\text{C}$  jusqu'à lyophilisation, le stockage à l'abri de la lumière à température ambiante, le broyage de tous les échantillons, l'extraction hydro-éthanoïque assistée par ultra-sons avec la réalisation de triplicats, la centrifugation puis l'analyse des surnageants par UPLC-HRMS.

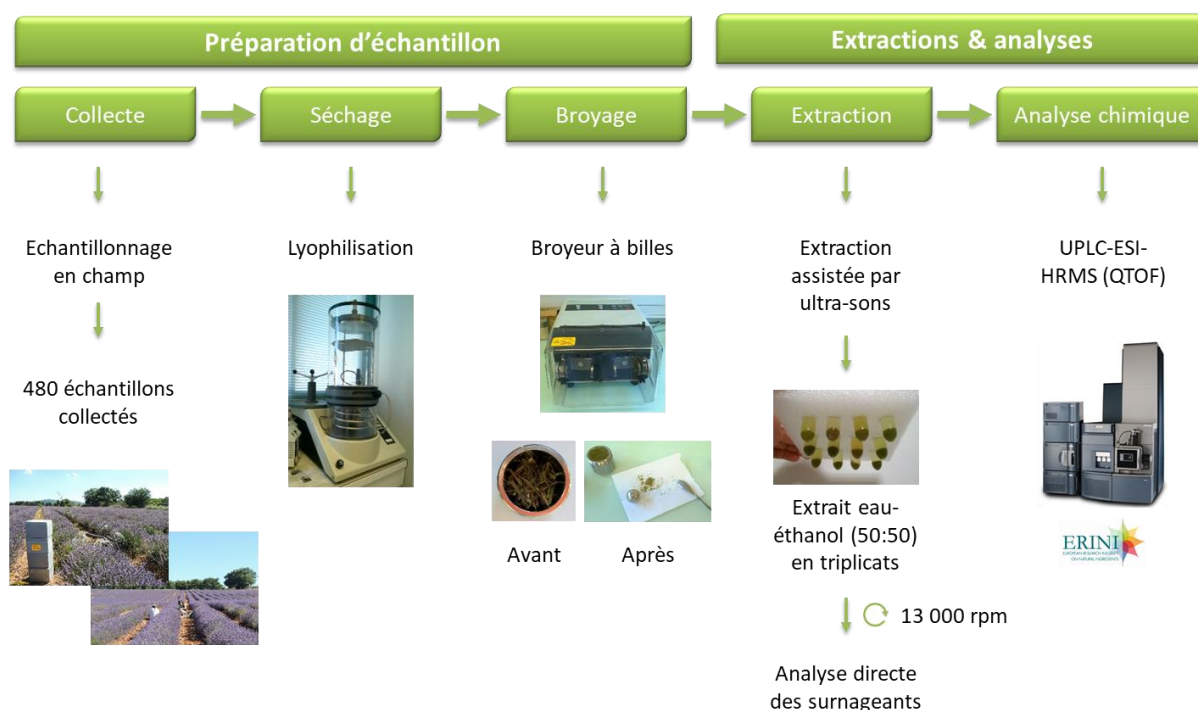


Figure 59 : Schématisation du design expérimental utilisé pour l'étude métabolomique non ciblée







## Chapitre 3 : Traitement des données

Ce chapitre traite du développement méthodologique et de la réalisation du traitement des données effectués au cours de cette étude. Dans une première partie, nous verrons l'extraction des données spectrales. Elle a été réalisée sur la plateforme en ligne W4M-Galaxy, et consiste à transformer les données brutes en matrice à deux dimensions. La deuxième partie de ce chapitre abordera la correction et la préparation des données aux analyses statistiques. Cette étape a été mise au point et effectuée sur une autre plateforme en ligne : MetaboAnalyst. Nous aborderons ensuite le nettoyage des jeux de données. Deux types de nettoyages ont été réalisés : le nettoyage sur les échantillons et le nettoyage sur les variables qui a pour but de supprimer les ions n'apportant aucune information utile pour les analyses statistiques (ex : adduits, contaminants, ...). Enfin, dans une dernière partie, nous aborderons la réalisation des analyses multivariées (ACP et PLS-DA) et l'obtention des composés discriminants.

### I. Extraction des données spectrales

#### I.A. Choix du logiciel de retraitement

Il existe de nombreux logiciels pour réaliser le prétraitement des données LC-MS en métabolomiques (voir Tableau 5). Ils peuvent être en libre accès ou commerciaux, permettre uniquement de réaliser le prétraitement des données spectrales ou l'intégralité du traitement des données. Ils peuvent soit être en local tels que MZmine 2, MarkerLynx XS (commercial) ou R ; soit hébergés sur des plateformes en ligne utilisant des serveurs délocalisés, c'est le cas de Xcms Online et de Galaxy - W4M (*Workflow4Metabolomics 3.0*). Dans un premier temps, des logiciels installés localement ont été testés pour traiter les données, notamment le module MarkerLynx XS du logiciel MassLynx de chez Waters. Malheureusement, la capacité de calcul s'est avérée insuffisante et nous nous sommes tournés vers des plateformes hébergées sur des serveurs. Néanmoins, la plupart des plateformes en libre accès possèdent un quota inférieur à la taille du jeu de données (ex : Xcms online limité à 25 Go pour une taille du jeu de données au format net.cdf de 161 Go). Nous nous sommes alors tournés vers la plateforme Galaxy – W4M (Giacomoni et al. 2015; Guitton et al. 2017) qui a l'avantage de posséder une capacité de stockage « illimitée », qui est gratuite, de plus elle a historiquement été développée pour la LC-MS et elle met à disposition un large choix de modules permettant la réalisation du traitement de données dans son intégralité et la création de *workflow*. Notre idée première était d'effectuer l'intégralité du traitement des données sous Galaxy. Finalement, seule l'extraction des données spectrales a été réalisée sur cette plateforme. La version employée durant cette thèse est la 3.2.0.

#### I.B. Stratégie adoptée pour la mise en place du *workflow*

##### **Workflow général :**

Le processus d'extraction des données spectrales réalisé sous Galaxy comprend plusieurs étapes effectuées dans l'ordre suivant :

- La détection des ions
- L'alignement des ions
- La correction de la dérive analytique temporelle
- Le remplissage des données manquantes
- La correction de la perte de signal due à la dérive analytique

### Jeux de données employés pour la mise en place du *workflow* :

La démarche adoptée a été de constituer un jeu de données réduit mais représentatif de l'ensemble des données quel que soit l'état sanitaire des plants, l'organe ou la variété ; afin de construire le *workflow* et de développer le paramétrage. Un jeu de données réduit (T03) a alors été créé contenant l'intégralité des pools de chaque séquence (soit 11 par séquence) et environ deux fois plus d'échantillons (soit environ 20 par séquence) répartis régulièrement sur toute la séquence. Soit un total non négligeable de 246 analyses. Le jeu de données est conséquent, pourtant cette quantité est nécessaire pour assurer la représentativité de l'ensemble des données et pour satisfaire aux conditions de certaines fonctionnalités de Galaxy. La majorité du développement et de l'optimisation a été réalisé sur ce jeu de données. Néanmoins, afin de faciliter la manipulation et de diminuer les temps de calcul, deux autres jeux de données très réduits ont été créés (T01 et T02) et utilisés en premier lorsque cela était possible. Le premier jeu (T01) réduit contient uniquement deux pools de chaque séquence et le deuxième (T02) contient l'intégralité des pools et deux échantillons pour toutes les séquences.

Une fois le processus créé et optimisé sur les jeux de données réduits, il a été testé et appliqué sur les quatre jeux de données correspondant aux lots d'analyse et d'extraction (voir Tableau 16) :

- Feuille de lavande
- Inflorescence de lavande
- Feuille de lavandin
- Inflorescence de lavandin

Tableau 16 : Correspondances entre les lots d'analyse et d'extraction et les caractéristiques biologiques (organes de la plante et espèces)

Espèce	Lavande		Lavandin	
Organe	Feuille	Inflorescence	Feuille	Inflorescence
Lot d'extraction et d'analyse	A17	B17	C17	D17

### Conversion des données :

Au préalable, il est nécessaire de convertir les données en un format compatible avec le logiciel de retraitement. Dans le cas de la plateforme Galaxy, elle accepte trois formats de fichiers : net.cdf, mzXML et mzData. Le logiciel de conversion Databridge a été utilisé afin de convertir les fichiers d'origine au format .raw au format net.cdf. A noter que la conversion des fichiers en net.cdf crée quatre fichiers correspondants respectivement à la MS, la MSe, la lockmass et l'UV. Seul le fichier MS, à savoir le 01, a été employé.

## I.C. Description et mise en place du *workflow* de prétraitement des données sous Galaxy

### I.C.1. Préparation et entrée des données et de la matrice de design dans le *workflow*

La matrice de design est le tableau regroupant les métadonnées. Les métadonnées sont les données relatives aux échantillons (voir définition p68), dans notre cas cela peut-être la variété, l'état symptomatique, l'organe de la plante, la séquence d'analyse, le code du plant, ...

### Préparation de la matrice de design :

Les matrices de design ont été créées sous excel. La première colonne doit contenir le nom des échantillons. La dénomination et l'ordre des analyses doit être strictement identique à celles de la matrice de données. Les autres colonnes contiennent les métadonnées de chaque échantillon, notamment les facteurs d'études (variété et état symptomatique) obligatoires pour la réalisation des analyses statistiques afin de définir les groupes biologiques. Les colonnes des matrices de design réalisées correspondent dans l'ordre à la dénomination des échantillons, la séquence d'analyse, l'ordre d'injection de l'extrait, le type d'échantillon (i.e. extrait, pool ou blanc), le premier facteur d'étude à savoir le statut symptomatique, puis le deuxième facteur d'étude c'est-à-dire la variété et enfin l'organe de la plante. Le type d'échantillon et la séquence d'analyse sont obligatoires pour réaliser une correction inter et intra séquence.

	A	B	C	D	E	F	G	H
1	Samples ID	batch	injectionOrder	sampleType	factor1_SymptomaticStatut	factor2_Varieties	factor3_PlantOrganes	
2	P1708002_SEQ1_A17_POOL_1_0501	A17_Seq1	5	pool	pool	pool	Feuille	
3	P1708002_SEQ1_A17_00101	A17_Seq1	6	sample	sain	7713	Feuille	
4	P1708002_SEQ1_A17_00201	A17_Seq1	7	sample	sain	Ma	Feuille	
5	P1708002_SEQ1_A17_00301	A17_Seq1	8	sample	sain	7713	Feuille	
6	P1708002_SEQ1_A17_00401	A17_Seq1	9	sample	infecte	7713	Feuille	
7	P1708002_SEQ1_A17_00501	A17_Seq1	10	sample	sain	7713	Feuille	
8	P1708002_SEQ1_A17_00601	A17_Seq1	11	sample	infecte	Ma	Feuille	
9	P1708002_SEQ1_A17_00701	A17_Seq1	12	sample	sain	7713	Feuille	
10	P1708002_SEQ1_A17_00801	A17_Seq1	13	sample	sain	Ma	Feuille	
11	P1708002_SEQ1_A17_00901	A17_Seq1	14	sample	sain	Ma	Feuille	
12	P1708002_SEQ1_A17_POOL_1_0601	A17_Seq1	15	pool	pool	pool	Feuille	
13	P1708002_SEQ1_A17_01001	A17_Seq1	16	sample	sain	Ma	Feuille	
14	P1708002_SEQ1_A17_01101	A17_Seq1	17	sample	infecte	Ma	Feuille	
15	P1708002_SEQ1_A17_01201	A17_Seq1	18	sample	infecte	Ma	Feuille	
16	P1708002_SEQ1_A17_01301	A17_Seq1	19	sample	sain	Ma	Feuille	

Figure 60 : Exemple de matrice de design réalisée sous excel

### Entrée dans le workflow sous Galaxy :

Le principe est le même pour les données et pour la matrice de design. Ils doivent être soit téléchargés dans l'historique, et dans le cas où un workflow est employé, ils doivent faire l'objet d'une entrée ce qui est possible grâce au module « Input dataset » pour la matrice de design et « Input dataset collection » pour les données.

#### I.C.2. Lecture des données (module MSnbase readMSData)

Une fois les données introduites dans le workflow, il est nécessaire d'utiliser le module « MSnbase readMSData » pour permettre au logiciel de lire les données brutes. Ce module permet également de créer une matrice de design à partir des données d'entrée. Dans notre cas, cette matrice nous a permis de vérifier l'ordre et la dénomination des échantillons, afin d'introduire la matrice de design créée en amont. En effet, il faut une stricte coïncidence entre la matrice de design et de données pour les modules se servant des deux sous peine d'échec du traitement.

#### I.C.3. Détection des ions (module findChromPeak ou xcmsSet)

L'étape suivante est celle de la détection des pics pour tous les échantillons. C'est une étape importante qui doit permettre l'intégration des analytes de faibles intensités tout en évitant d'intégrer du bruit de fond (Tautenhahn et al. 2008). Elle comprend trois étapes : la filtration, la détection et l'intégration de tous les ions. Ces trois étapes sont réalisées grâce à l'outil « Find » qui propose trois algorithmes différents pour y parvenir : *MatchedFilter*, *Centwave* et *CentWaveWithPredIsoROIs*. L'algorithme le plus performant dans le cas de données acquises en HRMS en mode centroïde est *Centwave* et est donc l'algorithme que nous avons employé.

Dans un premier temps, un « pré-filtre » est appliqué afin de retirer les ions de faible intensité, ce qui facilite et accélère la suite du processus. La filtration des ions est basée sur trois paramètres à définir,

à savoir le rapport signal/bruit « snthresh » et les deux paramètres « k,l ». Le rapport signal sur bruit et le paramètre k, correspondant au nombre minimum de scans consécutifs avec une intensité supérieure ou égale à l que doit contenir un pic pour être pris en compte, n'ont pas été modifiés. En revanche, le paramètre l correspondant à l'intensité minimale que doit avoir un pic pour être détecté par l'algorithme, a été déterminé à 1000 en fonction de la valeur minimale observée pour la ligne de base sur les chromatogrammes.

Dans un deuxième temps, tous les ions sont extraits pour chaque échantillon indépendamment. Les algorithmes de détection de type « binning » balayent les rapports  $m/z$  (en segmentant l'axe en portions) et détectent des augmentations d'intensités. L'approche développée avec l'algorithme *Centwave* est différente et consiste à chercher les plages de temps de rétention où un nombre défini de scans consécutifs possèdent une intensité minimum définie et un rapport  $m/z$  identique. En prenant en compte la constance de la valeur du rapport  $m/z$  comme paramètre de détection d'un pic, cet algorithme permet d'être plus précis et de détecter des pics co-élués. De plus, le fait de cibler des zones plutôt que de balayer l'intégralité des temps de rétention permet de diminuer le temps de calcul nécessaire. (Tautenhahn et al. 2008)

Les deux paramètres principaux à gérer pour la détection sont « ppm » et « peakwidth ». Le paramètre ppm correspond à la différence de  $m/z$  tolérer entre deux scans d'acquisition consécutifs (pour un même échantillon pour un même  $m/z$ ) et permet de créer la tranche de  $m/z$ . Il dépend de la précision du spectromètre de masse. La précision du TOF étant excellente (< 1 ppm, selon les données constructeurs pour le XEVO G2-qTOF) et la dérive en masse étant quasi-nulle, le paramètre a pu être abaissé à 10 ppm. Le paramètre *peakwidth* (min, max) correspond aux largeurs minimale et maximale autorisées pour un pic en seconde. En-dessous et au-dessus de ces deux valeurs, un pic ne sera pas détecté. Les pics les plus fins observés sur des chromatogrammes de pool ont une largeur de 3,6 s et les pics les plus larges de 18 s. Le *peakwidth* a alors été fixé à 2,20.

Enfin, dans un troisième temps, les pics restants sont intégrés. L'algorithme *Centwave* propose deux méthodes d'intégration, celle du « chapeau mexicain » et une deuxième méthode se basant sur les points des données réelles. Nous avons fait le choix d'utiliser la dernière méthode car plus juste même si le risque d'intégrer du bruit est accru. Le paramètre permettant de gérer l'intégration est « mzdifff ». Il correspond à la différence de  $m/z$  limite permettant de distinguer deux analytes co-élués (i.e. ayant des temps de rétention similaire). La différence maximale observée sur les chromatogrammes sous un pic est de 0,007 Da. Le signe négatif traduit le possible chevauchement de deux pics.

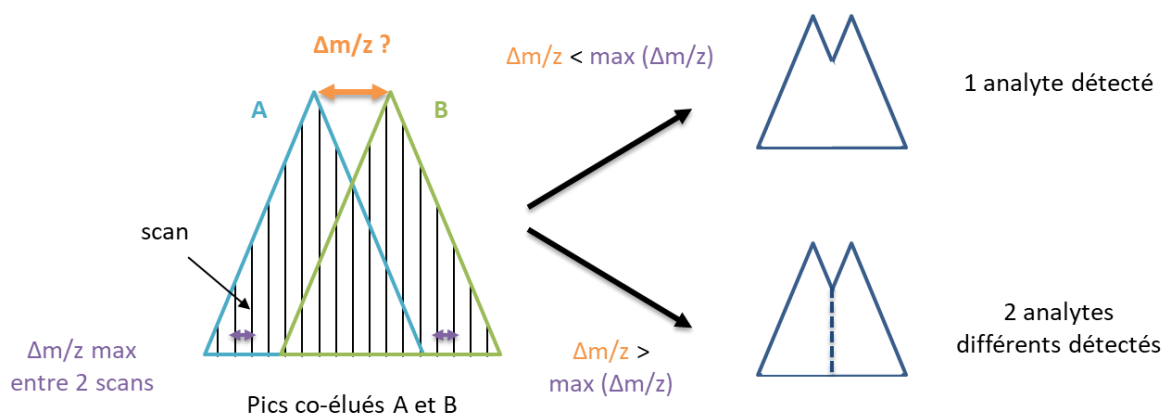


Figure 61 : Illustration de l'influence du paramètre « mzdifff » sur l'intégration des données

### I.C.4. Groupement des fichiers (module Merger)

Le module « Merger » regroupe l'ensemble des données en un seul fichier. Son application est obligatoire pour la suite du traitement. Il est utilisé après l'étape de détection des ions, dans le cas où les données ont été introduites sous forme d'une collection de fichiers séparés.

I.C.5. Alignement des ions (module group)

L'alignement des ions se fait grâce à l'outil « group » qui va rassembler les ions de tous les échantillons correspondant au même analyte en fonction du  $m/z$  dans les différents échantillons. Cette étape permet de passer de plusieurs matrices (une par échantillon) à une seule matrice (contenant tous les échantillons). Galaxy propose trois méthodes de regroupement : *peakdensity* (Smith et al. 2006), *MzClust* (conçu pour l'infusion directe) et *NearestPeaks* (Katajamaa et al. 2006). La méthode *peakdensity* a été choisie pour réaliser l'alignement des ions. Elle se base en priorité sur les rapports  $m/z$  pour regrouper les ions plutôt que sur les temps de rétention. Cette approche est plus fiable du fait de la grande précision des spectromètres de masses alors que la chromatographie en phase liquide est généralement sujette à une dérive sur les temps de rétention. (Smith et al. 2006)

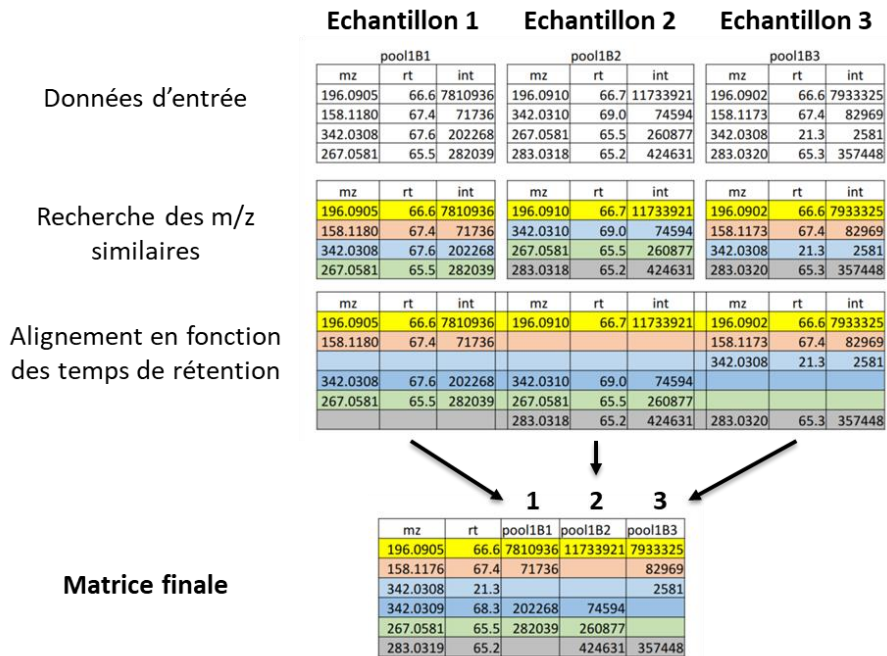


Figure 62 : Schématisation du fonctionnement de l'alignement des ions (figure adaptée d'après (Martin 2017))

Dans un premier temps, l'algorithme *peakdensity* va rassembler les ions selon leur rapport  $m/z$  (quel que soit leur temps de rétention et leur échantillon d'appartenance). Le paramètre « binsize » permet de définir la fenêtre de  $m/z$  utilisée pour réaliser ce premier regroupement (plus précisément, il définit la largeur de la superposition des tranches de  $m/z$  autorisée).

En second lieu, les ions d'une même fenêtre de  $m/z$  sont regroupés par temps de rétention. La largeur de la fenêtre autorisée sur les temps de rétention est gérée par le paramètre « bandwidth » (ou bw). Il correspond exactement à la largeur à mi-hauteur que l'on autorise pour le pic simulé soit la déviation standard (sur les temps de rétention en seconde) de l'enveloppe après lissage. L'enveloppe des pics est simulée par un modèle gaussien.

Les autres paramètres importants dans cette étape sont « minfract » et « minsamples ». Le paramètre « minfract » est la fraction minimale d'échantillon dans un groupe biologique pour qu'un groupe d'ions soit considéré comme un composé (et non un artefact). Il est donc compris entre 0 et 1, et se détermine en fonction des groupes biologiques (et non sur l'ensemble des échantillons) définis dans la matrice de design. En d'autres termes, si le rapport entre le nombre d'échantillons dans lequel le pic a été détecté et le nombre d'échantillons total dans une catégorie biologique est inférieur au *minfract*, alors l'ion n'est pas considéré comme un analyte. Quant au paramètre « minsamples », il correspond au

## Chapitre 3 : Traitement des données

nombre minimal d'échantillons dans lequel le groupe de pic doit être présent pour être considéré comme valide (i.e. comme un analyte).

Finalement, l'algorithme regroupe pour chaque ion le rapport  $m/z$  et le temps de rétention pour créer les « features » définis par un code MxxxTxxx. Par exemple, la caféine est définie par M195T245 avec un rapport  $m/z$  de 195 Da et un temps de rétention de 245 s.

L'outil « group » possède une sortie graphique nommée « chromdensity plot » permettant de visualiser chaque fenêtre de  $m/z$ , et ainsi évaluer la qualité de l'alignement des ions. Dans chaque fenêtre de  $m/z$  sont représentés les échantillons (en ordonnées) en fonction du temps de rétention. Chaque point représente un échantillon et la hauteur des gaussiennes est liée au nombre d'échantillons présents. Lorsqu'un carré gris est présent, le pic est validé comme un ion.

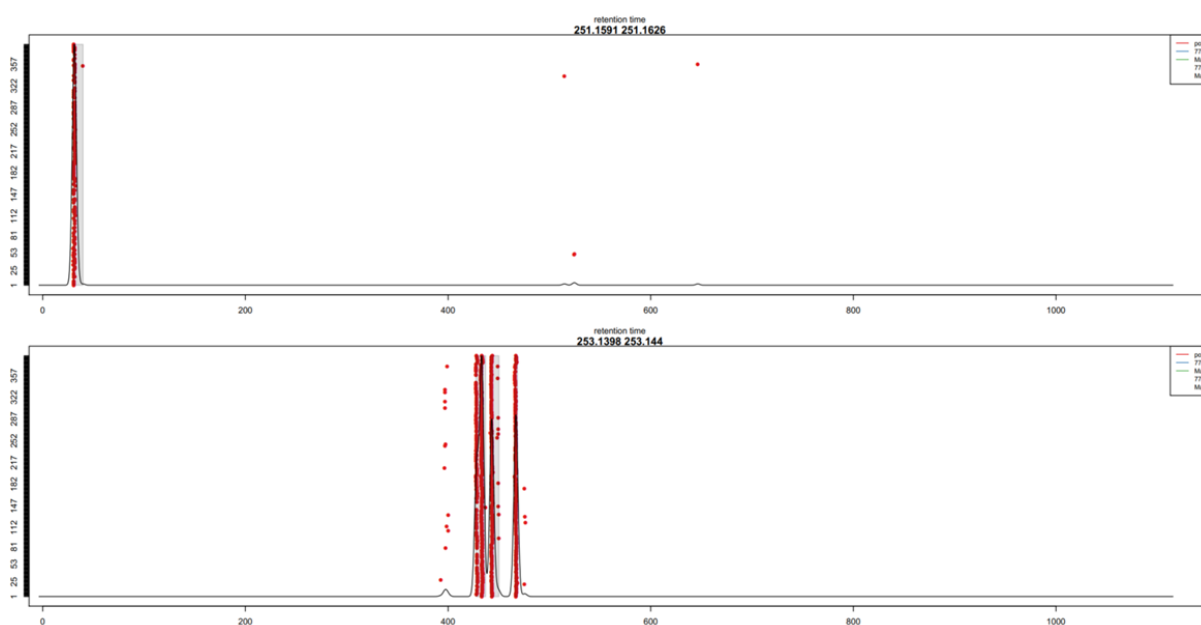


Figure 63 : Exemple de deux fenêtres de  $m/z$  consécutives pour la sortie graphique de l'outil « group » réalisé sur le jeu de données des feuilles de lavande (A17).

### I.C.6. Corrections des temps de rétention (module adjustRtime)

#### Correction des temps de rétention (module adjustRtime) :

Galaxy offre la possibilité de corriger d'éventuelle dérive sur les temps de rétention à l'aide du module « adjustRtime ». Cette étape est optionnelle et répétable autant de fois que nécessaire.

L'outil *adjustRtime* propose deux méthodes pour réaliser la correction à savoir « peak groups » ou « obiwrap ». Le premier réalise la correction sur les « features » en se basant uniquement sur les temps de rétention. La deuxième méthode utilise à la fois les temps de rétention et les rapports  $m/z$  et se base sur les données brutes. Les deux méthodes ont été testées et la qualité de la superposition des chromatogrammes en sortie a été comparée. *Peak groups* est légèrement plus efficient mais aucune des deux méthodes ne se distingue vraiment par rapport à l'autre.

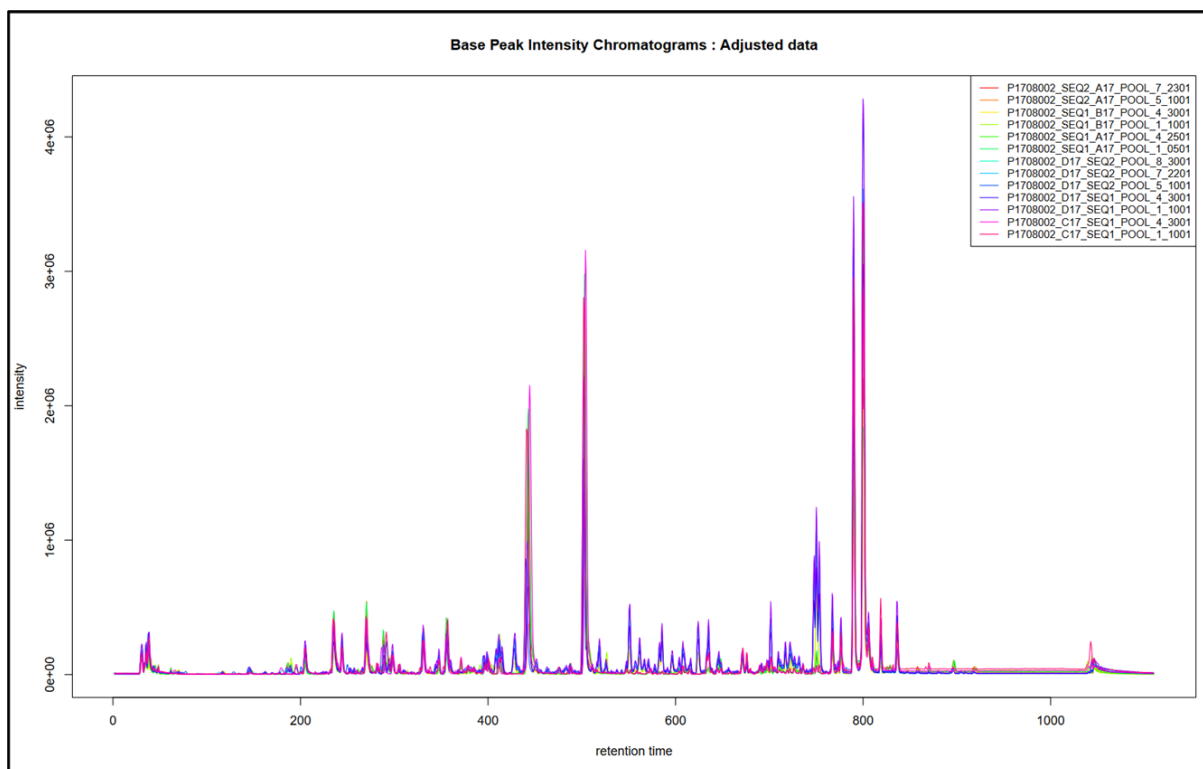


Figure 64 : Chromatogrammes obtenus après la correction des temps de rétention avec la méthode *peak groups* sur un jeu de données test contenant uniquement des pools des différentes séquences (T01).

La méthode *peak groups* a été conservée. Le paramétrage contient un « minfract » similaire à celui défini pour l'outil *group* (voir partie I.C.5 Alignement des ions (module *group*)), et un paramètre nommé « *extrapeaks* » qui correspond au nombre de pics à ajouter au nombre total d'échantillons pour définir le nombre maximum de pics pouvant être pris en compte pour la correction. Enfin, il est possible de choisir la méthode de lissage, à savoir linéaire ou polynomiale, utilisée pour le tracé des chromatogrammes à partir des points.

Le module *adjustRtime* possède également une sortie graphique qui permet de visualiser l'amplitude de la correction réalisée, en représentant la différence entre le temps de rétention initial et celui corrigé en fonction du temps.



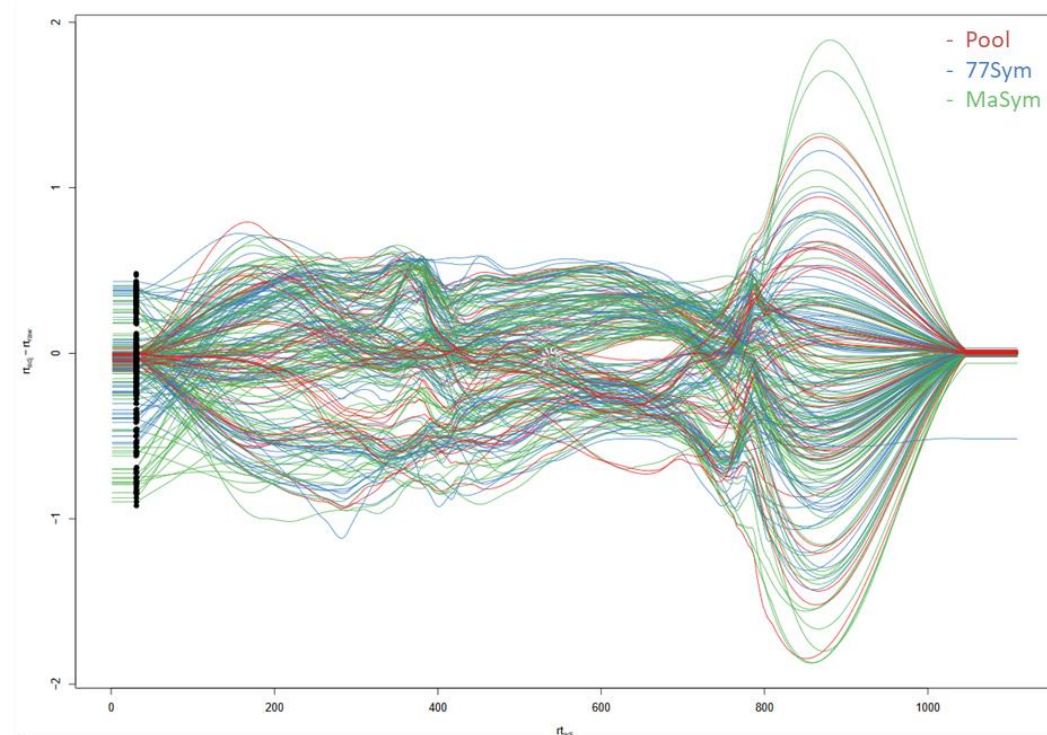


Figure 65 : Déviation des temps de rétention en fonction du temps pour le jeu de données des feuilles de lavande (lot A17)

### Ré-alignement des ions (module groupChromPeaks) :

Après avoir effectué un ajustement des temps de rétention avec la méthode *peakgroup*, il est indispensable de réaligner les ions, pour ce faire l'outil « group » est à nouveau employé. Nous avons fait le choix d'utiliser les mêmes paramètres que précédemment mais il est possible de changer certains paramètres (*bw*) afin d'être moins contraignant et de diminuer les temps de calcul.

#### I.C.7. Attribution des aires aux pics manquants (module fillPeaks)

L'alignement des ions provoque des données manquantes au sein de la matrice. Ces absences sont liées au fait que tous les analytes ne sont pas détectés dans l'ensemble des échantillons. Afin de combler la matrice, l'outil « fill » est employé. L'algorithme détermine la valeur de la ligne de base autour du pic fictif et attribue à celui-ci une intensité égale au bruit de fond qui l'entoure. La zone d'intégration est définie par la valeur du *m/z* médian et les temps de rétention de début et de fin pour l'ion considéré.

#### I.C.8. Corrections de la dérive analytique (modules *determine batch correction* et *batch correction*)

Le suivi de la décroissance du signal du standard interne effectué simultanément à l'acquisition des données (voir partie IV.E du chapitre 2), a montré une perte de signal au cours de chaque séquence d'analyse. Une différence de signal entre chaque séquence a également été observée et est imputable aux procédures de nettoyage et/ou de calibration effectuées entre les séquences d'acquisition. Au vu de l'importance de la dérive constatée, il apparaît essentiel d'y remédier. Galaxy offre la possibilité de corriger la dérive analytique intra et inter séquence grâce à l'outil « batch correction » qui s'appuie sur l'algorithme de Van der Kloet (van der Kloet et al. 2009). Cette méthode est plus juste (que celles utilisant un standard interne) car elle se base sur les pools pour effectuer l'ajustement des données. En effet, rien n'assure que la décroissance du signal affecte de manière homogène tous les analytes, donc qu'un seul standard interne permette de modéliser le comportement de l'ensemble des

## Chapitre 3 : Traitement des données

composés contenu dans un mélange complexe. La correction avec ce module va s'effectuer en plusieurs étapes. Dans un premier temps, l'algorithme va modéliser le comportement des pools dans le temps (ce qui revient à considérer l'ordre d'injection). A partir de la régression, un facteur de correction va être calculé pour chaque pic. Enfin, ces facteurs vont être appliqués aux intensités de chaque ion afin de les corriger indépendamment. (van der Kloet et al. 2009)

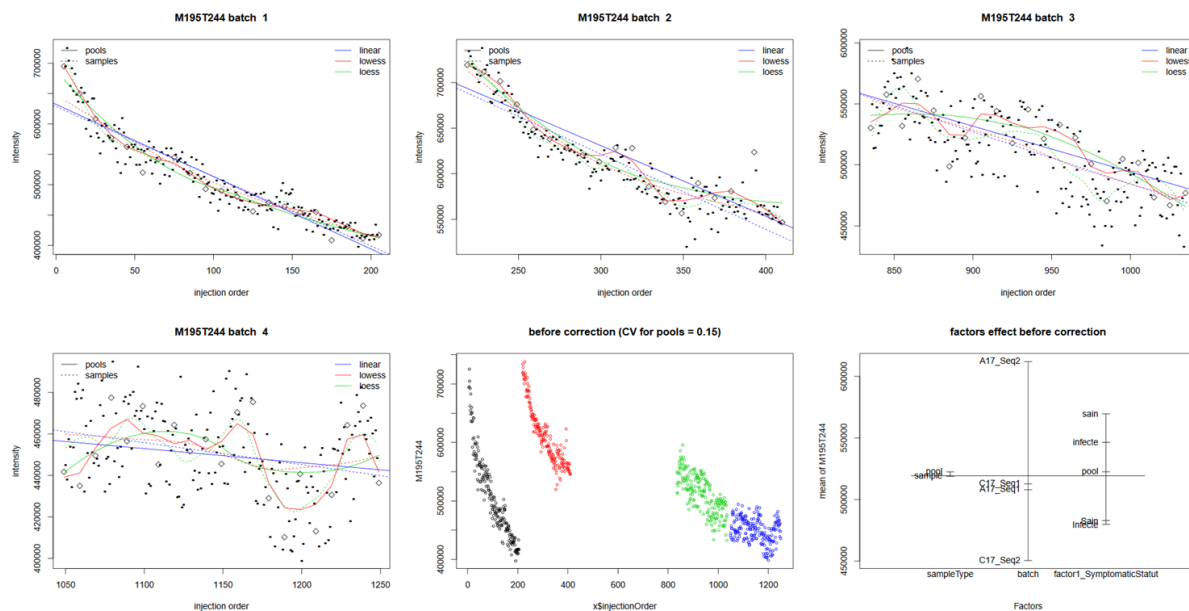
### Détermination du modèle de régression à appliquer :

La plateforme W4M met également à disposition un outil pratique « determine batch correction » afin de visualiser et de déterminer le modèle de régression le plus approprié au jeu de données étudié. Trois modèles de régression sont disponibles :

- Un modèle linéaire,
- Et deux modèles basés sur le principe de la régression locale (« lowess » et « loess »). La régression locale consiste à calculer des régressions simples sur des sous-ensembles locaux de données (au lieu de chercher à décrire l'ensemble par une unique fonction globale). Le modèle *lowess* est basé sur un modèle de régression linéaire alors que *loess* s'ajuste avec une fonction polynomiale de faible degré.

Le paramètre le plus important est « span ». Il permet de déterminer le nombre de points voisins à prendre en considération pour la modélisation de la régression locale. Pour un *span* de 1, l'ensemble du jeu de données est pris en compte. Plus le *span* est élevé, plus le lissage est important. Il est cependant conseillé d'opter pour des valeurs élevées afin d'éviter le sur-ajustement du modèle.

Dans cette thèse, la détermination du modèle de régression le plus adapté pour chaque jeu de données utilisé a été réalisée sur le standard interne. De manière générale, le modèle linéaire ne s'ajustait pas convenablement aux jeux de données étudiés et le modèle *lowess* ne moyennait pas suffisamment les résultats. Le choix s'est porté sur le modèle *loess* qui s'adaptait au mieux aux données.



Les quatre premiers graphiques (un par séquence d'analyse) représentent l'aire de la caféine pour chaque analyse (pool et échantillon) et la simulation de chaque méthode de régression. Le cinquième représente l'aire du standard interne en fonction de l'ordre d'injection, ce qui permet de visualiser la

décroissance du signal inter et intra séquence. Ainsi, il est facilement possible de constater que la première séquence du jeu de données considéré est la plus impactée par la dérive intra séquence ; et que la différence inter séquence est très marquée entre les deux premières alors que les deux dernières présentent une dérive beaucoup plus faible. Le dernier graphique permet d'identifier le facteur qui influence le plus la dérive analytique pour l'ion considéré, c'est-à-dire dans le cas présenté les séquences d'acquisition.

### Correction de la dérive intra et inter séquence :

Une fois le modèle de régression déterminé, il est facile d'utiliser l'outil « batch correction ». Les paramètres principaux sont les mêmes que pour le module d'aide à la détermination du modèle de régression. A noter que pour le modèle *loess*, il est possible de choisir si la modélisation de la régression s'appuie sur les pools ou les échantillons. Le paramètre « factor of interest » permet de visualiser le facteur biologique souhaité sur les sorties graphiques. La dénomination doit correspondre parfaitement à celle introduite dans la matrice de design. Ce paramètre n'influence en rien le calcul de la régression et de la correction.

La sortie graphique du module permet de visualiser l'impact de la rectification sur le jeu de données pour chaque ion (Figure 67) mais aussi dans son ensemble (Figure 68). Ci-dessous les résultats de la correction intra et inter séquence pour le jeu de données des feuilles de lavande (lot A17). La première illustration est la correction sur le standard interne pour lequel nous constatons la normalisation de l'aire. La différence inter séquence a été supprimée. La deuxième figure permet de visualiser l'impact de la correction sur l'ensemble du jeu de données. Nous observons également la disparition de la différence d'intensité entre les séquences d'acquisition mais aussi au sein des séquences. Les graphiques des individus et des variables des ACP sont également tracés sur les données avant et après corrections. Dans le cas de ce jeu de données, le graphique des individus des données brutes comporte deux groupes d'échantillons correspondant aux deux séquences d'acquisition. La dérive analytique est portée par l'axe F2 qui est facilement visible sur les pools. Après corrections, nous pouvons constater que la variabilité contenue par les échantillons n'est plus due à la déviation intra et inter séquence.

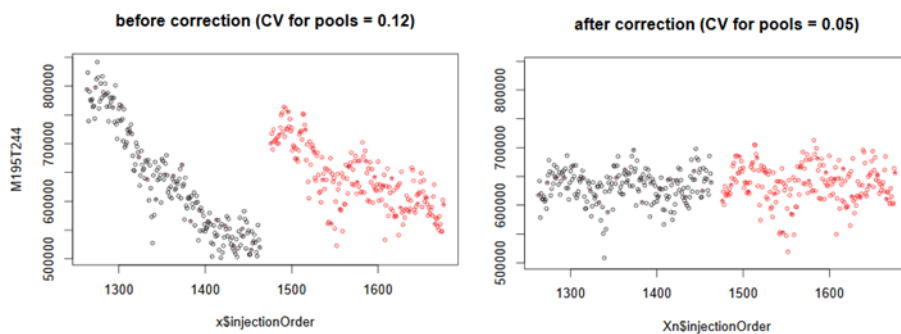


Figure 67 : Aire du standard interne en fonction de l'ordre d'injection avant et après correction, sur le jeu de données des feuilles de lavande (lot A17)

## Chapitre 3 : Traitement des données

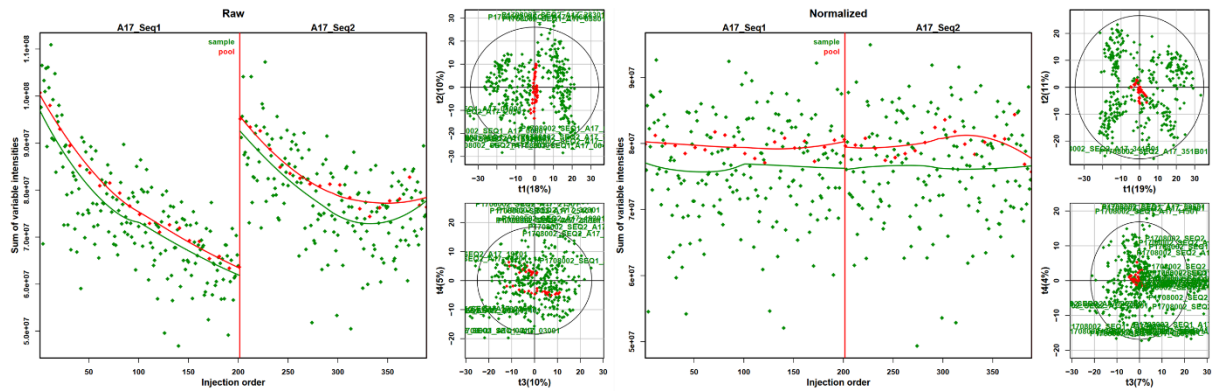


Figure 68 : A gauche les données avant rectification, à droite les résultats de la correction de la décroissance du signal intra et inter séquence pour le jeu de données des feuilles de lavande (lot A17) par l'outil « batch correction » de Galaxy

### I.D. Workflow et paramétrage finals

Le workflow utilisé sous Galaxy durant cette étude est présenté ci-dessous :

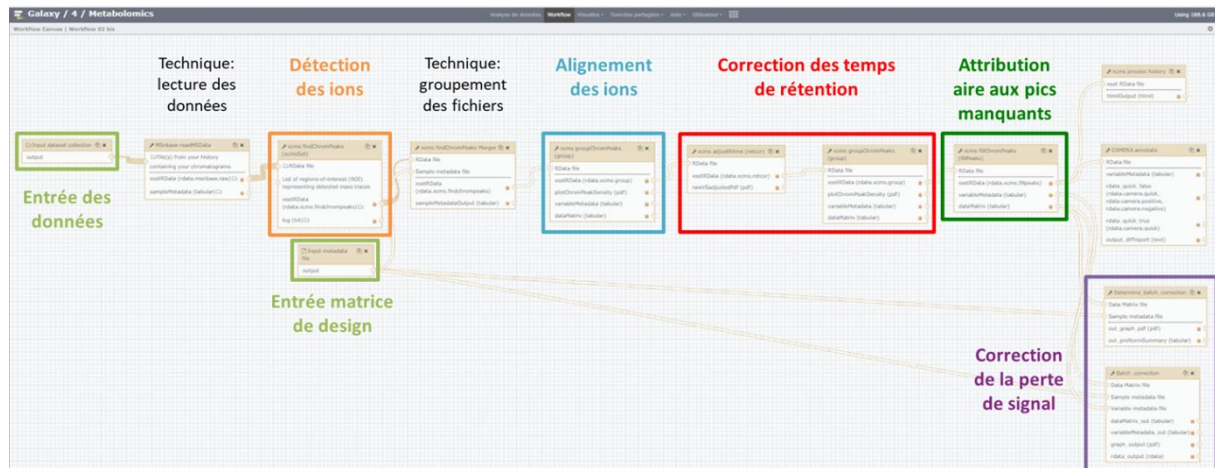


Figure 69 : Schématisation du workflow du prétraitement des données réalisé sous Galaxy-W4M

Le paramétrage optimisé obtenu sous Galaxy pour l'ensemble des outils utilisés est synthétisé dans le tableau ci-dessous. Le même paramétrage a pu être employé pour tous les jeux de données étudiés.

Tableau 17 : Paramètres finaux obtenus pour le traitement spectral des données sous Galaxy-W4M

Modules / Outils	Paramètres	Valeur paramètres
<b>ReadMSData</b>		
<b>Find</b>	Spectra filters	Hide
	Extraction method for peaks detection	CentWave
	Max tolerated ppm $m/z$ deviation in consecutive scans in ppm (ppm)	10
	Min,Max peak width in seconds (peakwidth)	2,20
	Advanced options (centwave)	
	Signal to Noise ratio cutoff (snthresh)	10
Pre-filter step for the first analysis step (prefilter) k,l	3,1000	

		Name of the function to calculate the $m/z$ center of the chromatographic peak (mzCenterFun)	Intensity weighted mean of the peak's $m/z$ values
		Integration method (integrate)	Peak limits based on real data (more accurate but prone to noise)
		Minimum difference in $m/z$ for peaks with overlapping retention times (mzdiff)	-0,007
		Fitgauss	No
		Noise filter (noise)	0
		Verbose Columns	No
		List of regions-of-interest (ROI)	
<b>Merger</b>			
<b>Group</b>		Method to use for grouping	Peak density
		Bandwith (bw)	2
		Minimum fraction of samples necessary (minfract)	0,75
		Minimum number of samples (minSamples)	2
		Width of overlapping $m/z$ slices (binSize)	0,01
		Advanced options	
		Maximum number of groups to identify in a single $m/z$ slice (maxFeatures)	1000
		Get the Peak List	Yes
		Convert retention time (seconds) into minutes	Yes
		Number of decimal places for mass values reported in ions identifiers	4
		Number of decimal places for retention time values reported in ions identifiers	1
		Reported intensity values	intb
		Replace the remain NA by 0 in the dataMatrix	Yes
<b>AdjustRtime</b>		Method to use for retention time correction	Peak groups
		Minimum required fraction of samples in which peaks for the peak group were identified (minFraction)	0,75
		Maximal number of additional peaks for all samples to be assigned to a peak group for retention time correction (extraPeaks)	1
		Smooth method (smooth)	Loess (non-linear alignment)
		Advanced options	hide
		Degree of smoothing for the loess fitting (span)	0,2
		Family	Gaussian
		Bin size (in $m/z$ dimension) to be used for the profile matrix generation	
<b>Group</b>		Method to use for grouping	Peak density
		Bandwith (bw)	2
		Minimum fraction of samples necessary (minfract)	0,75

	Minimum number of samples (minSamples)	2
	Width of overlapping $m/z$ slices (binSize)	0,01
	Advanced options	
	Maximum number of groups to identify in a single $m/z$ slice (maxFeatures)	1000
<b>Fill</b>	Advanced options	Hide
	Value by which the $mz$ width of peaks should be expanded (expandMz)	0
	Value by which the RT width of peaks should be expanded (expandRt)	0
	Specifying a ppm by which the $mz$ width of the peak region should be expanded (ppm)	0
	Peak List	
	Convert retention time (seconds) into minutes	Yes
	Number of decimal places for mass values reported in ions identifiers	4
	Number of decimal places for retention time values reported in ions identifiers	1
	Reported intensity values	intb
	Replace the remain NA by 0 in the dataMatrix	Yes
<b>Determine batch correction</b>	Factor of interest	Batch
	Advanced options	Hide
	Span	0,85
<b>Batch correction</b>	Type of regression model	All loess sample
	Span	1
	Null values	consider it as a null intensity
	Factor of interest	Symptomatic Statut
	Level of details for plots	Complete

## II. Corrections et préparation aux analyses statistiques

MetaboAnalyst est une plateforme en ligne d'accès libre basée sur le logiciel R permettant le retraitement des données de type « omics », issues de RMN, GC ou LC-MS. La version utilisée dans cette étude est la 4.0 (Xia and Wishart 2011; Chong et al. 2019). La plateforme met à disposition de nombreux modules mais seul celui pour les analyses statistiques a été employé. Ce module permet à la fois de réaliser des analyses statistiques uni- et multivariées mais aussi la correction et la transformation des données. Pour des jeux de données de taille importante, l'extraction des données spectrales (détection, alignement, ...) doit être effectuée au préalable.

### II.A. Préparation, téléchargement, vérification et filtration des données

#### Préparation de la matrice de données et formats acceptés :

MetaboAnalyst prend en charge les formats de fichier de type .txt (séparateur espace) et .csv (séparateur virgule) pour les données prétraitées et déjà sous forme de tableau ; et le format .zip pour les données brutes dans le cas de jeux de données restreints. Nous avons fait le choix de mettre les échantillons en ligne mais la plateforme autorise également la matrice transposée (échantillons en

## Chapitre 3 : Traitement des données

colonne). MetaboAnalyst ne permet de rentrer qu'une seule colonne de métadonnées. Pour chaque jeu de données, deux matrices ont donc été créées correspondant aux deux facteurs biologiques de l'étude : les variétés et l'état symptomatique des plants.

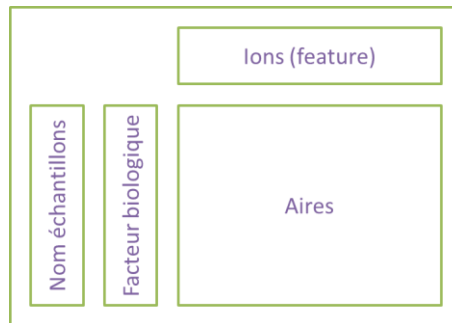


Figure 70 : Agencement de la matrice de données adaptée à MetaboAnalyst

### Téléchargement et vérification des données :

Une fois les données téléchargées en précisant ses caractéristiques et l'agencement de la matrice, MetaboAnalyst génère un document de vérification des données, spécifiant entre autres le nombre d'échantillons, d'ion détectés ou de groupes biologiques, et d'éventuelles données manquantes.

### Filtration des données :

MetaboAnalyst propose une première étape de filtration des données. Cette étape est fortement recommandée pour les études de métabolomiques non-ciblées. Elle permet de supprimer les ions provenant du bruit de fond et/ou proche de la limite de détection. Dans notre cas cette étape n'est pas nécessaire car un seuil de détection élevé a déjà été appliqué lors de la détection des ions dans Galaxy. De plus, le nombre de *features* total est bien inférieur à 5000 (limite imposée par la plateforme pour limiter les temps de calculer), ce qui nous autorise à ne pas employer cette étape.

## II.B. Normalisation

Comme explicité dans le chapitre 1 (voir partie III.B.5.a paragraphe Transformations des données), il existe plusieurs types de normalisation :

- La normalisation par rapport aux échantillons
- La normalisation par rapport aux ions (« features »)

L'étape de normalisation sous MetaboAnalyst comprend trois catégories : la normalisation sur les échantillons, la transformation puis la réduction des variables (Xia and Wishart 2011).

### II.B.1. Normalisation par rapport aux échantillons (« Sample Normalization »)

Cette normalisation a pour but de réduire la variabilité technique ayant pu être introduite lors de la préparation des échantillons ou de leurs analyses (Xia and Wishart 2011). La correction de la dérive analytique est également assurée par le module « batch correction » de Galaxy. Les deux outils seront comparés dans cette partie.

#### a. Type de normalisation sur les échantillons proposé

Sept normalisations différentes sont réalisables sous MetaboAnalyst, à savoir :

- Par une valeur spécifique à chaque échantillon, comme par exemple le poids ou le volume
- Par la totalité des aires

## Chapitre 3 : Traitement des données

- Par la médiane des aires
- Par un échantillon de référence
- Par un échantillon regroupant l'ensemble des échantillons du groupe analysé (pool)
- Par un ion de référence, comme un standard interne
- Par normalisation quantile. Cette normalisation permet de faire coïncider les distributions des échantillons entre eux (Bolstad et al. 2003).

MetaboAnalyst propose une sortie graphique permettant de visualiser l'impact de la normalisation sur les échantillons et sur les ions du jeu de données (voir exemple ci-dessous). Elle a pour but de permettre de choisir la normalisation la plus adaptée à l'étude. Chaque vue (échantillons et ions) contient deux types de graphiques déclinés avant et après normalisation. Le premier type de graphique modélise la distribution générale des échantillons ou des ions. Le deuxième graphique représente la distribution individuelle soit des échantillons, soit de l'aire des ions sous forme de boîte à moustaches. La normalisation idéale fournira une distribution générale proche d'une gaussienne.

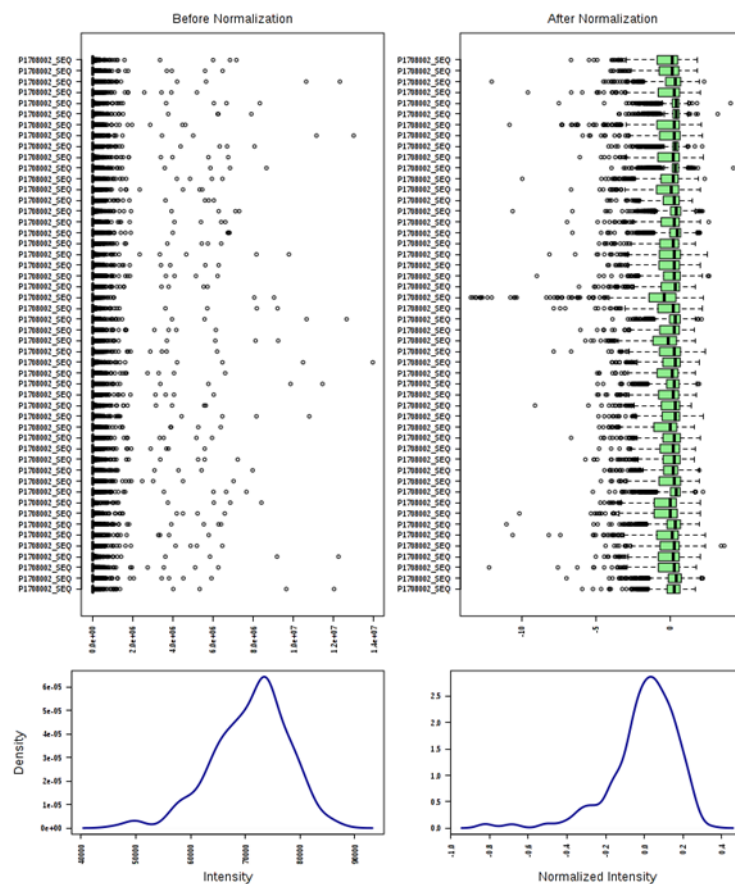


Figure 71 : Graphiques des échantillons obtenus après l'étape de normalisation sous MetaboAnalyst représentant à gauche le jeu de données avant, et à droite après normalisation. En bas, la modélisation de la répartition générale des échantillons en fonction de leurs aires ; et en haut, la distribution individuelle des plusieurs échantillons (matrice feuille-lavande et normalisation avec le standard interne)

De prime abord, la normalisation par les pools semble la plus adaptée à cette étude. Néanmoins, d'autres normalisations ont été testées telles que la normalisation par standard interne ou quantile (voir Figure 72 ci-dessous). Les essais ont été réalisés sur la matrice des variétés pour les extraits de feuille de lavande (A17), une transformation logarithmique et un « autoscaling » sont également



## Chapitre 3 : Traitement des données

appliqués, et les ACP ont été réalisées. La normalisation quantile et par standard interne ne fournissent pas une distribution normale, et sur l'ACP de la normalisation par standard interne les injections de pool suivent un axe montrant que la dérive analytique n'a pas été complètement corrigée. Comme attendu, les meilleurs résultats ont été obtenus avec la normalisation par les pools, qui a été conservée.

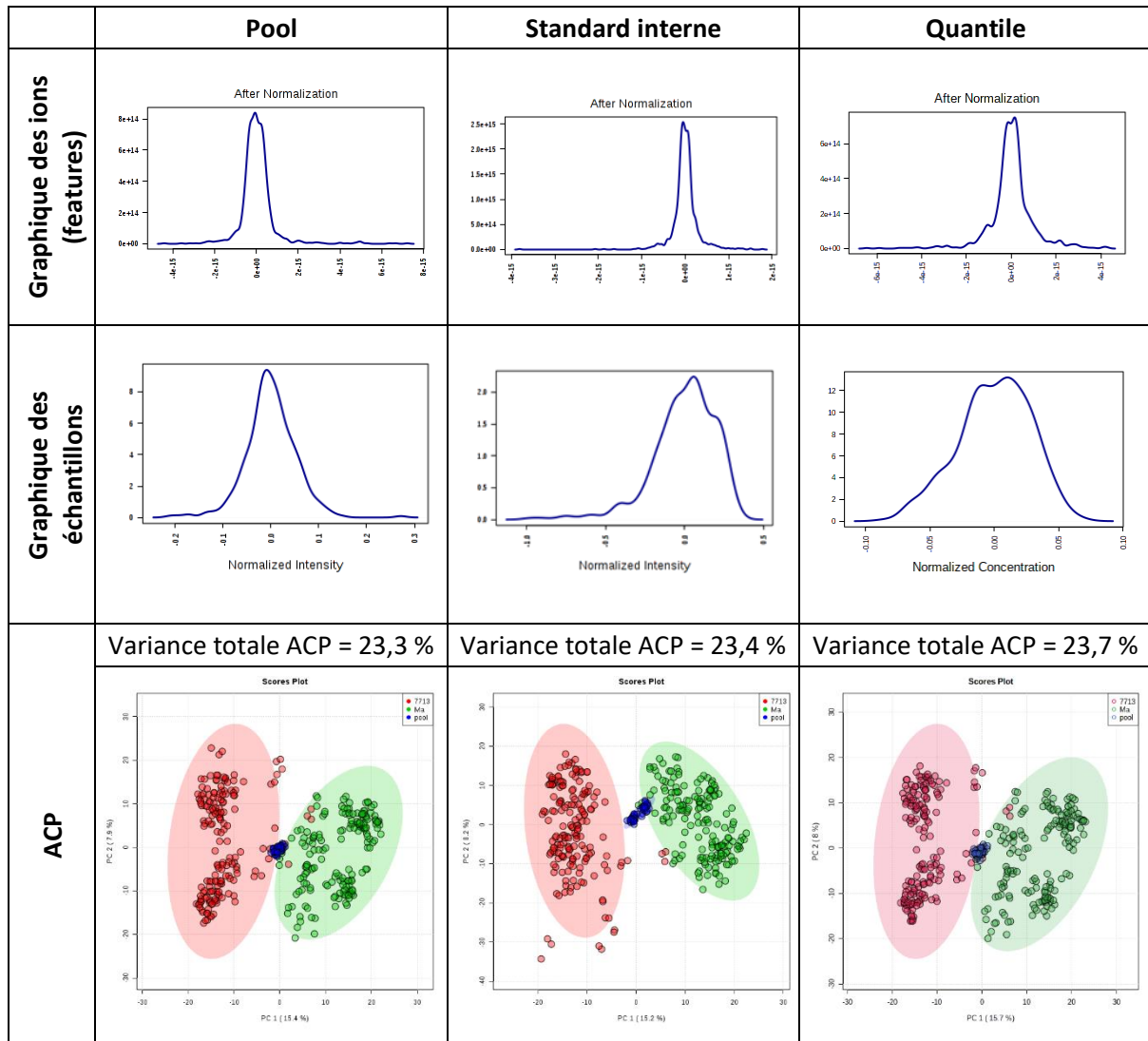


Figure 72 : En haut, les graphiques représentant la distribution générale des données (*features*) ; au milieu, les graphiques représentant la distribution générale des échantillons ; et en bas, les graphiques des individus des ACP obtenus pour chacune des trois méthodes testées (normalisation par les pools, par le standard interne ou normalisation quantile) de MetaboAnalyst (matrice des variétés des feuilles de lavande, autres paramètres : normalisation automatique et transformation logarithmique)

### b. Comparaison avec « batch correction » proposé par Galaxy

Les deux plateformes employées pour le traitement des données à savoir Galaxy et MetaboAnalyst, proposent toutes les deux un système de correction de la dérive analytique intra et inter séquence : par l'intermédiaire de l'outil « batch correction » pour Galaxy et par la normalisation avec les pools pour MetaboAnalyst. Nous avons cherché à déterminer quel outil était le plus performant pour les données étudiées. Pour ce faire, la matrice du jeu de données des extraits de feuille de lavande (A17) a été sortie avant et après le module « batch correction ». Pour les deux matrices, la transformation logarithmique et un « autoscaling » ont été appliqués. Afin de les comparer, les ACP et les PLS-DA ont également été réalisés.

Tableau 18 : Correction et paramétrages appliqués pour effectuer la comparaison entre les deux plateformes sur la correction de la dérive analytique

Nom test	Galaxy	MetaboAnalyst		
	Batch correction	Normalisation sur les échantillons	Transformation	Réduction des données
<b>BC</b>	oui	non	log	autoscaling
<b>Pool</b>	non	oui (pool)	log	autoscaling

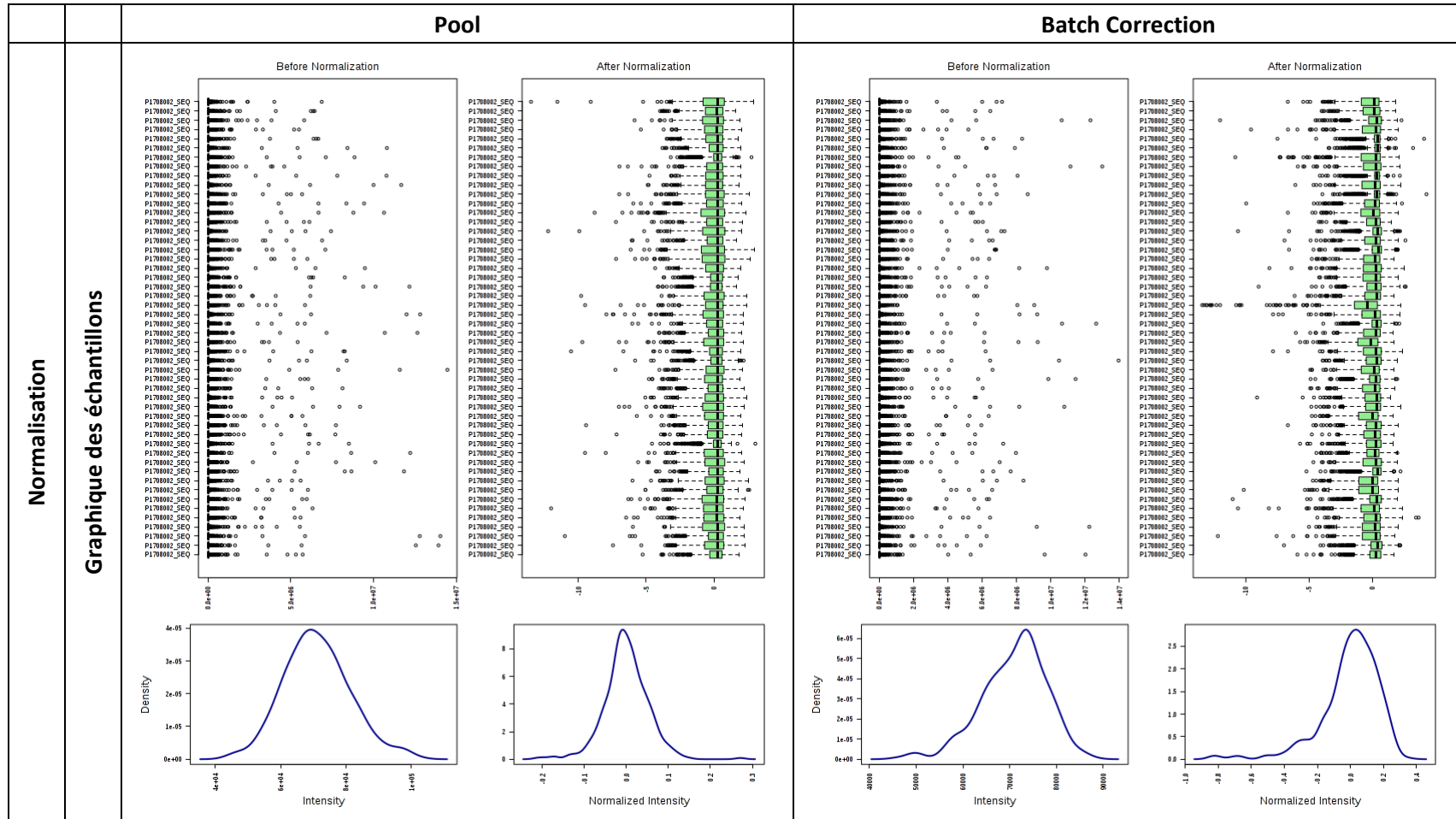


Figure 73 : Graphiques des échantillons, à gauche après une normalisation sur les pools sous MetaboAnalyst et à droite après l'application de l'outil « Batch correction » de Galaxy pour le jeu de données feuille-lavande (A17). Dans les deux cas, une transformation logarithmique et une normalisation automatique ont été appliquées

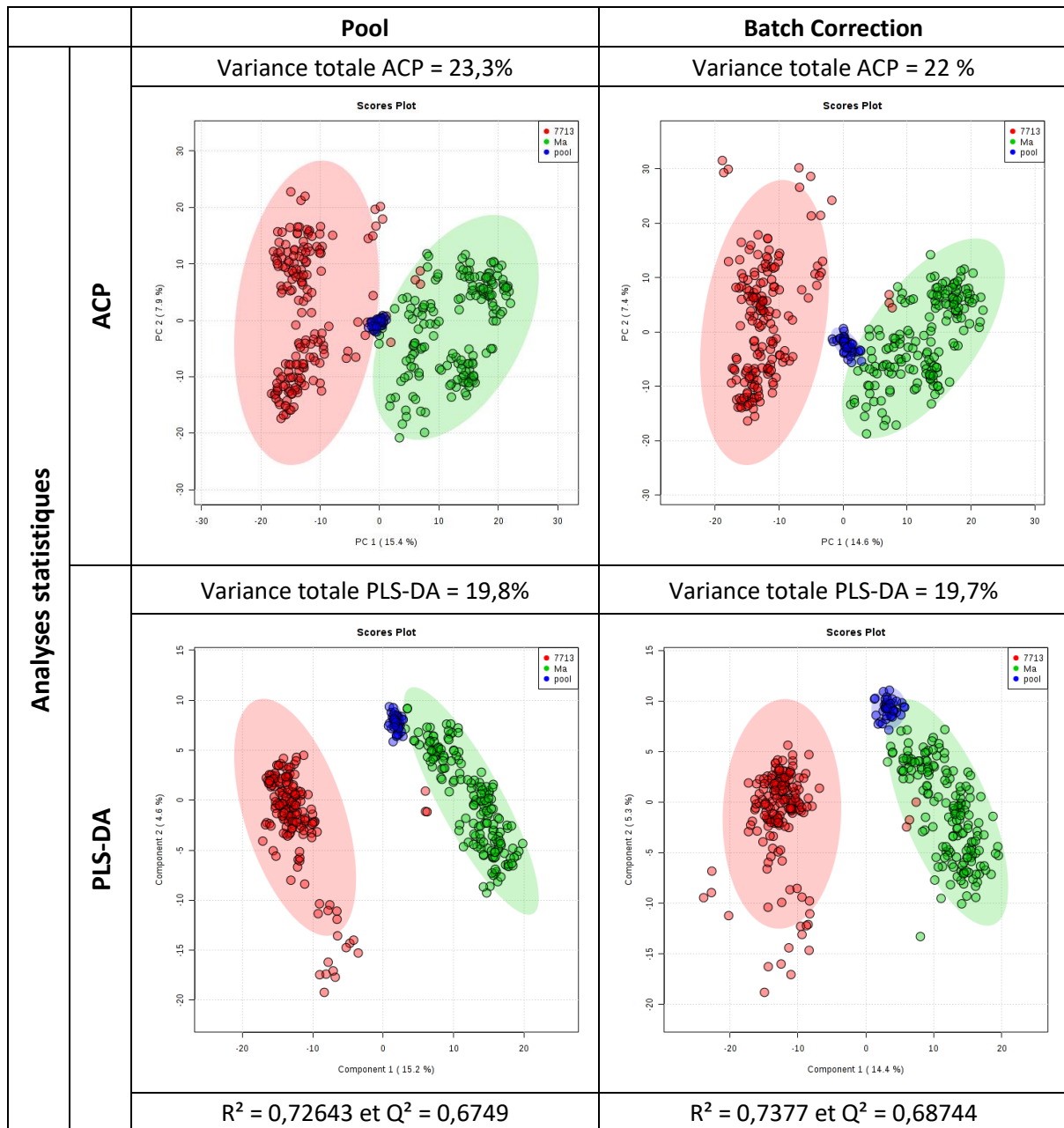


Figure 74 : Les analyses statistiques (ACP et PLS-DA) obtenues après correction de la dérive analytique sous MetaboAnalyst à gauche et sous Galaxy à droite.

Les résultats des analyses statistiques sont très proches. L'influence sur les trente premiers VIP des deux méthodes a également été regardée, hormis de légers changements de rang dans la classification, les différences entre les deux méthodes sont ici aussi minimales. En revanche, la distribution générale des échantillons est mieux normalisée par MetaboAnalyst que par Galaxy. Nous avons donc choisi de conserver la normalisation des pools sous MetaboAnalyst afin de palier la dérive analytique intra et inter séquence d'acquisition.

## Chapitre 3 : Traitement des données

### II.B.2. Normalisation par rapport aux ions

La normalisation sur les échantillons est découpée en trois catégories : le centrage, la réduction et la transformation des données. Ces normalisations sont indispensables et ont pour but de rendre les analytes comparables en leur conférant un poids similaire et ce quel que soit leur ordre de grandeur initial. (Van Den Berg et al. 2006; Xia and Wishart 2011; Wolfender et al. 2013)

#### a. Réduction des données (« Data scaling »)

Le centrage et la réduction des données sont vivement recommandés en métabolomique. La réduction a pour but d'équilibrer le poids des différents composés. En effet, les statistiques multivariées ont tendance à sur-exprimer les composés ayant une concentration élevée et ne pas refléter la variance due au problème biologique. « L'auto-scaling » est recommandé lorsqu'aucune hypothèse n'est faite entre la variation de composition et le problème biologique. Cependant, l'inconvénient de cette méthode est d'augmenter la variance, ce qui peut conduire à une surexpression de constituants non informatifs (Hervé et al. 2018). Une alternative à « l'auto-scaling » a été développée pour palier à ce phénomène, c'est la méthode Pareto qui est très similaire mais permet de limiter l'inflation du bruit de fond. Une autre méthode est proposée par MetaboAnalyst : « range ». Elle est également sujette à la sur-expression des ions de faible intensité, mais aussi sensible à la présence de données aberrantes. (Van Den Berg et al. 2006)

Les quatre méthodes de réduction des données proposées par MetaboAnalyst sont décrites dans le tableau suivant :

Tableau 19 : Méthodes de réduction des données proposées par MetaboAnalyst

Méthode	Action	Centre	Réduit
None	Rien		
Mean centering	Centre selon la moyenne uniquement	x	
Auto scaling	Centre et réduit en divisant par la déviation standard pour chaque variable	x	x
Pareto scaling	Centre et réduit en divisant par la racine carrée de la déviation standard pour chaque variable	x	x
Range scaling	Centre et réduit en divisant par l'amplitude de chaque variable pour chaque variable	x	x

Les quatre méthodes de réduction ont été testées sur la matrice des variétés des feuilles de lavande (lot A17). Toutes les méthodes ont été essayées avec la normalisation par les pools et avec soit aucune transformation soit avec le logarithme. Les analyses statistiques ACP et PLS-DA ont été réalisées pour chacune des conditions.

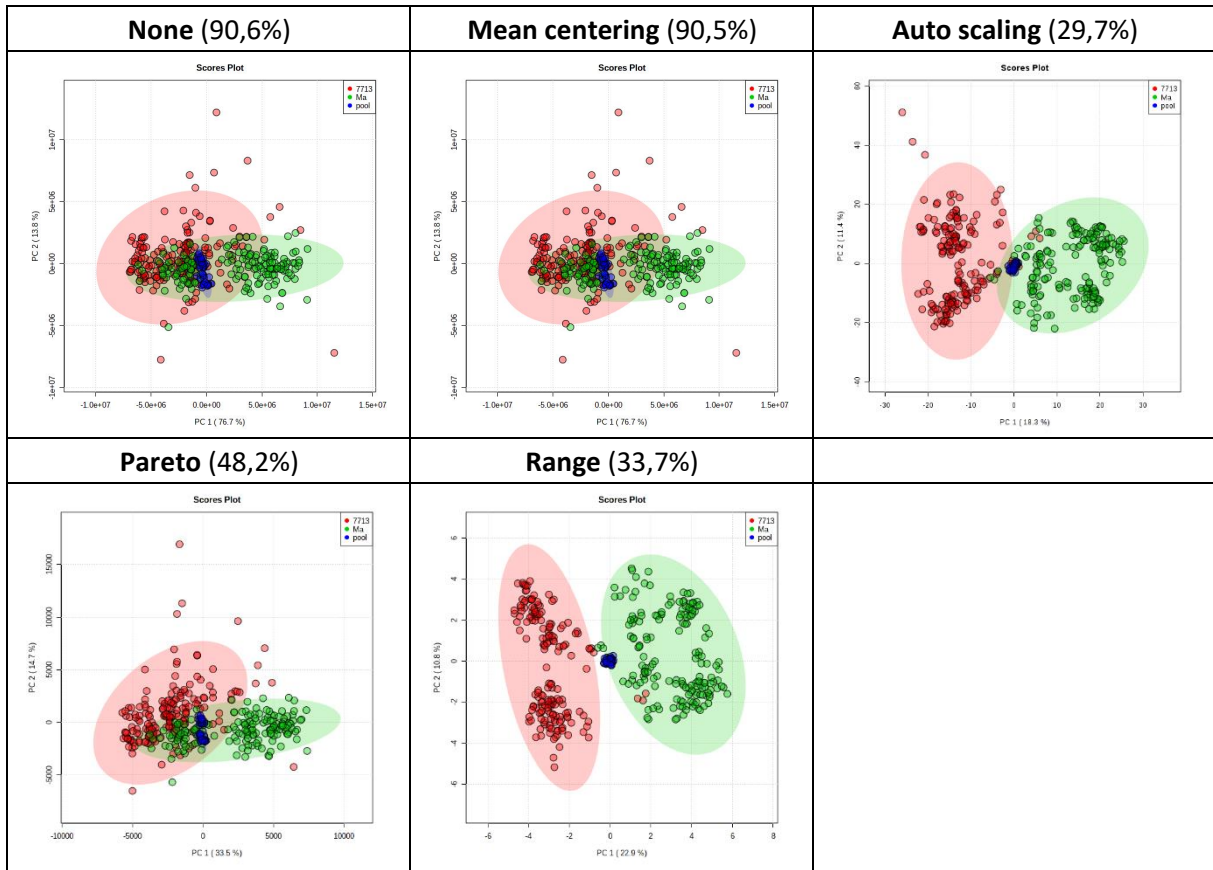


Figure 75 : Graphiques des échantillons des ACP pour les différentes méthodes de réduction des données de MetaboAnalyst sans transformation (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les pools). Le pourcentage entre parenthèse correspond à la variance totale exprimée par les deux premiers axes.

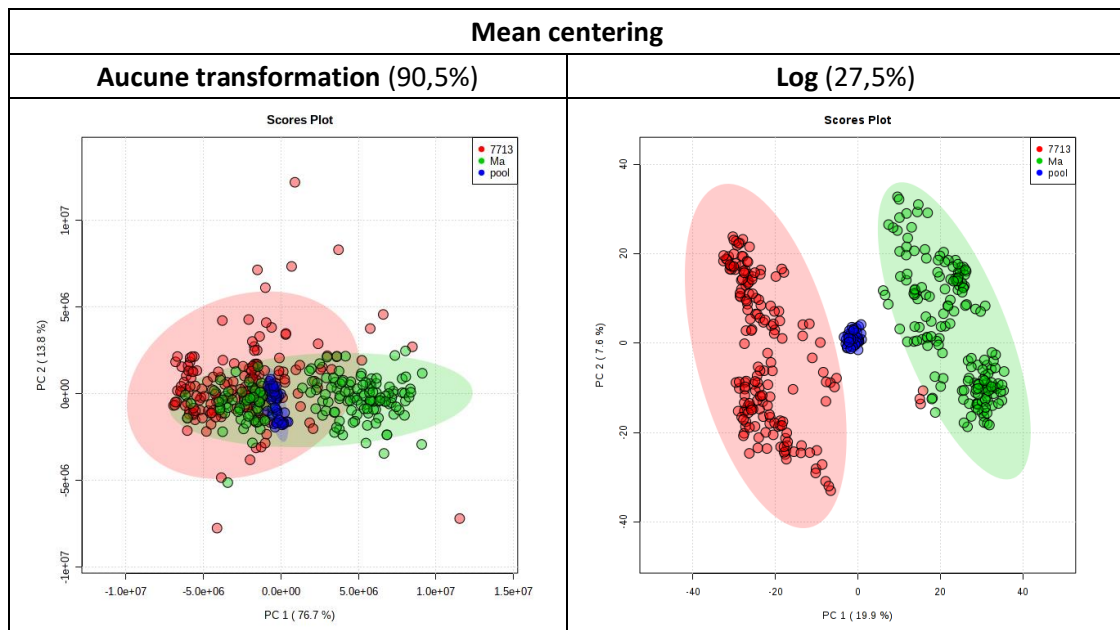


Figure 76 : Graphiques des échantillons des ACP avec et sans transformation logarithmique avec centrage des données de MetaboAnalyst sans transformation (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les pools). Le pourcentage entre parenthèse correspond à la variance totale exprimée par les deux premiers axes.

La comparaison de la méthode « mean » avec les autres fonctions illustre l'importance de réaliser une réduction sur ce jeu de données. Les groupes biologiques sur les ACP ne sont pas différenciés si la transformation logarithmique n'est pas appliquée. De plus, la fonction Pareto (sans log) ne permet pas non plus de différencier les groupes sur les ACP contrairement aux résultats obtenus pour les méthodes *autoscaling* et *range*. La fonction Pareto diminue l'importance des variables de faible intensité. De plus, la transformation logarithmique apporte une grande importance aux petites variables. Ces deux constatations laissent à penser que les marqueurs les plus importants du jeu de données sont des analytes de faible intensité. Par conséquent, les méthodes *autoscaling* et *range* sont à privilégier. *L'autoscaling* a été privilégié car la distribution des données obtenue est celle qui se rapproche le plus de la loi normale.

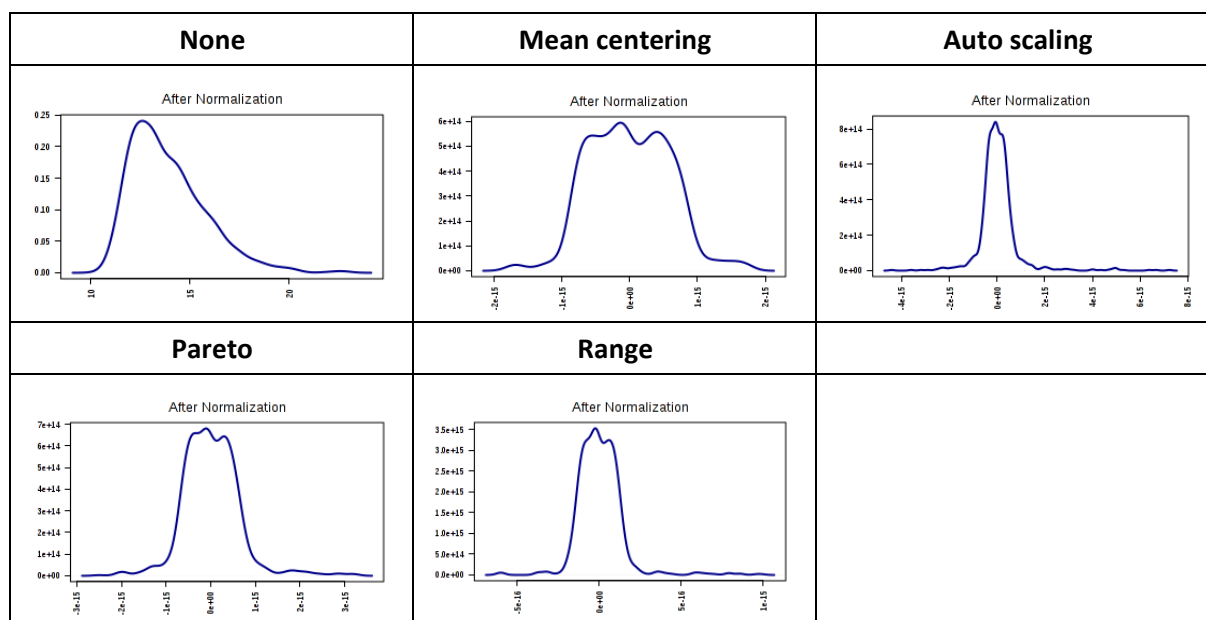


Figure 77 : Graphique représentant la distribution générale des données (features) obtenue pour chacune des méthodes de réduction des données de MetaboAnalyst (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les pools et transformation logarithmique).

### b. Transformation des données (« Data transformation »)

La transformation des données n'est pas obligatoire, elle vient en complément de la réduction et du centrage des données mais ne peut se substituer à eux. L'interaction entre les méthodes de transformation et de centrage et réduction ne sont pas bien comprises. Cependant, l'utilisation de transformation améliore les résultats des analyses statistiques multivariés. La transformation logarithmique est la plus usitée et la racine carrée est à favoriser si le jeu de données contient des valeurs nulles. (Van Den Berg et al. 2006; Hervé et al. 2018)

MetaboAnalyst permet soit de ne pas réaliser de transformation (None), soit d'appliquer la fonction logarithme (Log), soit d'utiliser la racine carrée (Racine). Nous avons testé les trois transformations proposées sur la matrice des variétés des feuilles de lavande (lot A17) avec pour autres paramètres la normalisation par les pools et « l'auto-scaling ». Les analyses statistiques (ACP et PLS-DA) ont été réalisées.

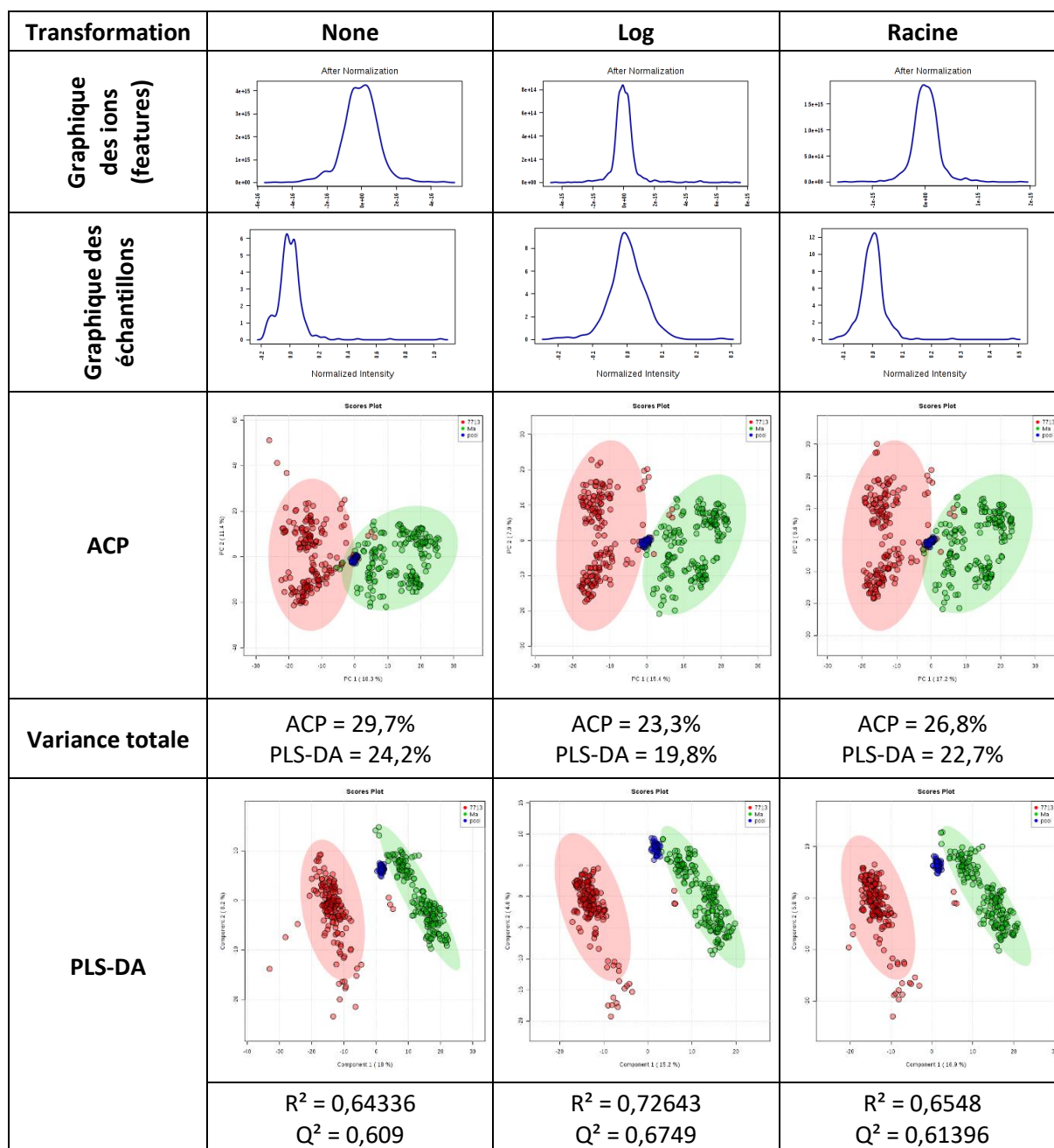


Figure 78 : Tests des différentes transformations disponibles sous MetaboAnalyst (matrice des variétés des feuilles de lavande, autres paramètres : normalisation par les pools et l'auto-scaling)

L'utilisation d'une transformation améliore la distribution générale des données. Le logarithme apparaît comme la transformation qui confère la distribution la plus normale. Les résultats des analyses statistiques sont également légèrement améliorés par l'application d'une transformation. Les analyses statistiques obtenues avec le logarithme et la racine carrée sont très proches. Néanmoins, les indicateurs de validation de la PLS-DA  $R^2$  et  $Q^2$  (voir partie III.B.5.b.iii du chapitre1) sont plus élevés pour la transformation log.

Au vu des résultats le logarithme est la transformation la plus adaptée à cette étude, et celle qui a été conservée pour la suite.



### II.C. Paramétrage final

Les paramètres employés pour la correction et la préparation des jeux de données aux analyses statistiques dans MetaboAnalyst sont synthétisés dans le tableau ci-dessous. Ce paramétrage a été employé pour toutes les matrices.

Tableau 20 : Paramètres finaux obtenus pour la correction et la préparation des données aux analyses statistiques sous MetaboAnalyst

Module	Etape	Intitulé paramètre	Paramètre appliqué	
Statistical analysis	Upload	Data type	Peak density table	
	Data Filtering		None	
	Normalization	Sample normalization		Pool group
		Data transformation		Log
Data scaling			Auto scaling	

### III. Nettoyage des données

Comme explicité dans le chapitre 1, il existe deux types de nettoyage : le nettoyage sur les variables et le nettoyage sur les échantillons. L'objectif du nettoyage est principalement d'enlever des données qui n'apportent pas d'informations voire faussent les résultats (Worley and Powers 2013). Dans cette étude, nous avons appliqué aux jeux de données quatre nettoyages différents dont trois correspondants à un nettoyage sur les ions. Le premier nettoyage a consisté à supprimer les ions fournissant une information redondante. Dans le deuxième nettoyage, nous avons identifié la présence de contaminants et les avons supprimés des jeux de données. Dans un troisième temps et à l'aide de tests statistiques, nous avons identifié et supprimé tous les ions qui n'apportent aucune information vis-à-vis de la problématique biologique (*i.e.* soit par rapport au caractère sensible/tolérant, soit par rapport à l'état symptomatique des plants). Enfin, nous avons effectué un nettoyage sur les échantillons consistant à supprimer ceux identifiés comme aberrants.

#### III.A. Nettoyage sur les ions

##### III.A.1. Suppressions des données redondantes : isotopes, fragments et adduits

Les jeux de données contiennent l'intégralité des ions, or une même molécule peut être corrélée à plusieurs ions tels que des ions isotopes, des ions fragments ou des adduits. Ces ions n'apportent pas d'informations supplémentaires et ne constituent qu'une information redondante pour les analyses statistiques. Ils peuvent même être nuisibles, en introduisant une variabilité résiduelle qui va parasiter les analyses. Il est donc préférable de les supprimer. L'objectif de ce nettoyage est de ne conserver que l'ion majoritaire de chaque molécule. La première étape consiste à obtenir une annotation des isotopes, fragments et des adduits par molécules. Dans un second temps, les isotopes sont supprimés manuellement puis les fragments et des adduits.

#### Obtention d'une matrice annotée (outil CAMERA.annotate de Galaxy) :

Afin de faciliter la suppression des données redondantes, le premier objectif est d'obtenir une matrice annotée. Pour ce faire l'outil CAMERA de Galaxy a été utilisé, il permet d'attribuer les isotopes, fragments et adduits pour chaque molécule.

L'outil a été appliqué aux données brutes obtenues en sortie du module « fill » et ceux pour les quatre jeux de données. Ensuite, la matrice des variables est récupérée en sortie. Les informations utiles de cette matrice sont ensuite copiées dans la matrice de données. A savoir, à l'aide du logiciel excel, les

## Chapitre 3 : Traitement des données

colonnes n°3, 6, 15 et 16 de la matrice des variables correspondant respectivement à « mz », « rt », « isotopes » et « adduct » sont collées dans la matrice des données. La correspondance entre les lignes (i.e. les *features*) des matrices de variables et de données correspondantes a été vérifiée pour chaque jeu de données.

Tableau 21 : Paramétrage de l'outil « CAMERA.annotate »

CAMERA	Group co-eluted peaks based on RT (groupFWHM)	
	Multiplier of the standard deviation [sigma]	6
	Percentage of FWHM width [perfwhm]	0,6
	Annotation general options	
	General ppm error [ppm]	5
	General absolu error in $m/z$ [mzabs]	0,015
	Annotate isotopes (indisotopes)	
	Max. ion charge [maxcharge]	3
	Max. number of expected isotopes [maxiso]	4
	The percentage number of samples, which must satisfy the C12/C13 rule for isotope annotation [minfrac]	0,5
	Mode	All functions
	Verifying grouping co-eluted peaks (groupCorr)	
	Correlation threshold (0..1) [cor_eic_th]	0,75
	Method selection for grouping peaks after correlation analysis into pseudospectra [graphMethod]	hcs
	Significant correlation threshold [pval]	0,05
	Use correlation inside samples for peak grouping [calcCiS]	Yes
	Use isotopic relationship for peak grouping [calcIso]	No
	Use correlation across samples for peak grouping [calcCaS]	No
	Annotate Adducts (finAdducts)	
	Which polarity mode was used for measuring of the ms sample (polarity)	positive
	How much peaks will be calculated in every thread using the parallel mode ([max_peaks])	100
	Use a personal ruleset file	False
	If no ruleset is provided, calculate ruleset with max. number n of [nM+x] clusterions [multiplier]	3
	Statistics and results export (doffreport)	
	Number of condition	One condition
	Export options	Hide

### Suppression des isotopes :

Ensuite, sous excel les lignes correspondant à des isotopes sont supprimées. Les isotopes sont annotés par  $[M+1]^+$ ,  $[M+2]^+$ , et  $[M+3]^+$ .

Néanmoins avant de supprimer un isotope, les temps de rétention et la différence de  $m/z$  sont systématiquement vérifiés. De plus des erreurs d'annotation sont possibles. Des points d'alertes sont la présence d'isotope  $[M+3]^+$  et  $[M+4]^+$ . La présence d'isotopes  $[M+3]^+$  a été attestée mais uniquement lorsque l'ion majoritaire  $[M]^+$  est très intense. Outre la vérification des rapports  $m/z$  et des temps de rétention entre les isotopes, ce sont les variations d'intensités entre les isotopes qui doivent alerter.

## Chapitre 3 : Traitement des données

En effet, nous n'avons pas de composés bromés ou chlorés dans les extraits (cf. partie Les composés organiques non volatils (CONVs) dans le genre *Lavandula*). Par conséquent, la variation d'intensité doit être décroissante au sein du massif isotopique.

1012			631.20025035613	M631T422	17573.181434	14415.24372
1013	[138][M]+	[M+H]+ 634.28	635.286961561055	M635T836	14416.153076	15668.63406
1014	[139][M]+		635.394707077442	M635T699	103116.28997	99239.74368
1015	[138][M+1]+		636.290213286883	M636T836	0	6463.980019
1016	[139][M+1]+		636.398114096207	M636T699	41365.905846	45697.95395
1017	[140][M]+		639.120236950284	M639T289	1090262.9709	1109573.583
1018	[140][M+1]+		640.123519306142	M640T289	366520.14230	368421.4248
1019	[140][M+2]+		641.125542346028	M641T289	102732.45066	100075.1577
1020	[140][M+3]+		642.128229518334	M642T289	14359.446911	19721.21414
1021	[141][M]+	[M+Na]+ 620.3	643.289611922254	M643T871	27519.325163	30806.88356
1022			643.367138996089	M643T774	39574.586719	42950.47768
1023	[141][M+1]+		644.29280745007	M644T871	11180.150253	13488.09640
1024			651.459377476951	M651T876	14495.126988	0
1025		[M+Na]+ 634.28	657.26885842025	M657T836	0	0
1026			657.376579104459	M657T702	23365.196201	29007.28401

Figure 79 : Matrice de données prête pour la suppression des isotopes

### Suppression des fragments et des adduits :

La deuxième étape est la suppression des adduits et des fragments. Le principe est le même que pour la suppression des isotopes. Tous les ions correspondant à des fragments ou adduits du type : +Na, +K, -H<sub>2</sub>O, +HCOOH et -HCOOH. La présence d'adduits sodium et potassium est fréquente en LC-MS (DeHaven et al. 2010; Ernst et al. 2014; Wolfender et al. 2015) et a été observée dans les analyses. Les pertes en eau sont également très fréquentes et l'ion HCOOH correspond à l'acide formique qui a servi de tampon dans les solvants d'élution.

Un adduit n'est supprimé que lorsque l'ion [M+H]<sup>+</sup> correspondant a été retrouvé dans la liste des ions. Lorsque plusieurs annotations sont proposées, l'ion n'est supprimé que si tous les ions [M+H]<sup>+</sup> correspondants à tous les adduits annotés sont retrouvés. Si une des annotations est [M+H]<sup>+</sup>, l'ion est automatiquement conservé. Pour que la correspondance soit acquise, il est nécessaire que le rapport masse sur charge (mz) et le temps de rétention (rt) de l'ion [M+H]<sup>+</sup> correspondant soit exactement les mêmes.

Afin de rendre la matrice de donnée conforme à la suite du traitement les colonnes rajoutées ont été supprimées. Ensuite, la matrice finale est copiée-collée dans un nouveau classeur et enregistrée au format approprié. Cette étape permet d'éviter tout déboire avec la comptabilisation des lignes précédemment supprimées, qui est entre autres mal gérée par Galaxy.

### III.A.2. Suppression des contaminants

La présence de contaminants (plastique) dans les extraits a été suspectée et confirmée (voir ci-dessous). Une extraction à blanc a été conduite afin de pouvoir déterminer l'origine de cette pollution et élaborer une liste de contaminants. Cette liste a ainsi permis d'effectuer une annotation des contaminants pour chaque matrice de données, puis de les supprimer.

L'outil « CAMERA » n'acceptant pas de matrice de données en entrée, l'étape de suppression des contaminants, initialement prévue en premier, a finalement été réalisée après l'étape de suppression des données redondantes.

### a. Extraction à blanc

Lors des premières expérimentations, la présence de contaminants plastiques dans les analyses a été suspectée. Afin de déterminer l'origine de cette contamination, nous avons vérifié que celle-ci ne provienne ni du système analytique, ni du solvant utilisé pour réaliser les extractions. Pour ce faire, un blanc contenant le solvant d'éluion (acétonitrile) et un blanc avec le solvant d'extraction (mélange eau-éthanol contenant de la caféine à 5 ppm) ont été injectés et n'ont révélé aucune trace de contamination. Ces deux possibilités écartées, l'hypothèse formulée est la suivante : le matériel utilisé (tubes Eppendorf et cônes pour propipette en plastique) relargue des composés lors de l'extraction. Afin de vérifier cette hypothèse, trois extractions à blanc c'est-à-dire sans matière première (Dudzik et al. 2018), ont été réalisées. L'analyse du mélange des trois extraits a révélé la présence de plusieurs contaminants (voir Figure 80 ci-dessous).

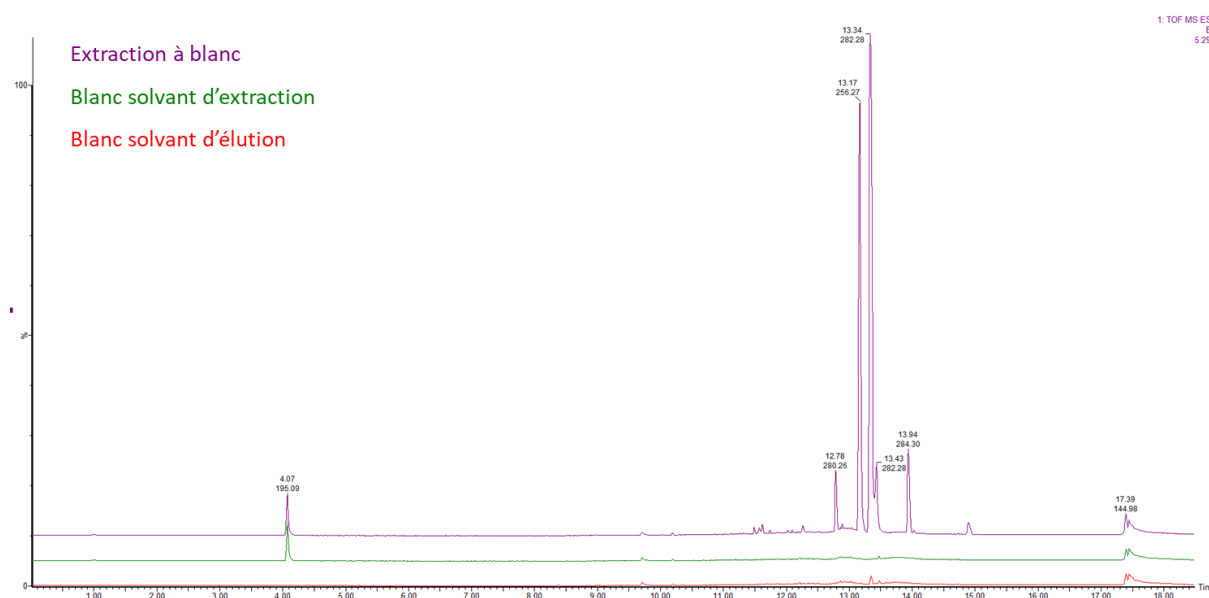


Figure 80 : Chromatogrammes UPLC-HRMS de l'extraction à blanc, des solvant d'éluion et d'extraction ; gradient d'éluion eau et acétonitrile à 0,1% d'FA ; colonne C18 HSS T3 (Annexe 4) ; débit de 0,4 ml/min ; injection de 1  $\mu$ L ; T = 40°C. ESI en mode positif

### b. Suppression des contaminants des matrices de données

L'objectif de cette procédure est d'enlever des matrices de données les contaminants plastiques relargués par le matériel expérimental durant l'étape d'extraction.

#### Elaboration de la liste des contaminants :

Tout d'abord, une liste des contaminants a été élaborée à partir de l'analyse de l'extraction à blanc explicité ci-dessus. Pour ce faire un *workflow* basé sur celui élaboré précédemment sous Galaxy pour l'extraction des données a été simplifié. Plus précisément, les étapes de corrections (des temps de rétention et de la dérive analytique) ont été supprimées. Le paramétrage est identique. Il est nécessaire d'avoir un nombre minimum d'analyses pour utiliser certains modules. Par conséquent, deux échantillons pris au hasard ont été utilisés pour créer un jeu de données fictif afin de lancer le *workflow*. Une fois la matrice obtenue, toutes les données relatives aux échantillons ont été supprimées pour ne conserver que la liste des ions appartenant aux contaminants.

#### Banque interne et annotation des matrices :

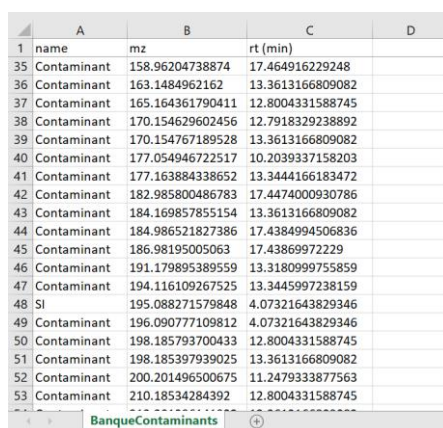
L'objectif est d'annoter automatiquement les matrices de données grâce à l'outil « bank\_inhouse » de Galaxy. Ce module permet d'annoter une matrice à partir d'une banque de données conçue par

## Chapitre 3 : Traitement des données

l'utilisateur. La correspondance s'appuie sur les rapports  $m/z$ , et éventuellement, selon le souhait de l'utilisateur, sur les temps de rétention. Pour utiliser ce module, il est donc nécessaire de concevoir au préalable une banque de données interne. Les matrices de données doivent également faire l'objet d'ajustement car cet outil travaille sur les matrices de variables (les matrices de variables sont des matrices utilisées par Galaxy contenant de nombreux paramètres calculés par la plateforme tels que le nombre de pics par échantillons, les rapports masse/charge et les temps de rétention minimum, maximum et moyen). Les matrices de données ayant fait l'objet d'un premier nettoyage, nous avons fait le choix de récupérer les informations nécessaires sur les matrices des variables (rapports  $m/z$  et temps de rétention) et de les intégrer aux matrices de données.

Création de la banque interne :

En premier lieu un fichier de type banque interne a été créé sous excel. La première colonne doit correspondre au nom de l'analyte, ici tous les ions ont été dénommés contaminants hormis l'ion majoritaire du standard interne. La deuxième et troisième colonnes doivent respectivement correspondre au rapport  $m/z$  et aux temps de rétention en minutes des ions. Ces données ont été récupérées sur la matrice des variables obtenue précédemment lors de l'élaboration de la liste des contaminants (cf. ci-dessus).



	A	B	C	D
1	name	mz	rt (min)	
35	Contaminant	158.96204738874	17.464916229248	
36	Contaminant	163.1484962162	13.3613166809082	
37	Contaminant	165.164361790411	12.8004331588745	
38	Contaminant	170.154629602456	12.7918329238892	
39	Contaminant	170.154767189528	13.3613166809082	
40	Contaminant	177.054946722517	10.2039337158203	
41	Contaminant	177.163884338652	13.3444166183472	
42	Contaminant	182.985800486783	17.4474000930786	
43	Contaminant	184.169857855154	13.3613166809082	
44	Contaminant	184.986521827386	17.4384994506836	
45	Contaminant	186.98195005063	17.43869972229	
46	Contaminant	191.179895389559	13.3180999755859	
47	Contaminant	194.116109267525	13.3445997238159	
48	SI	195.088271579848	4.07321643829346	
49	Contaminant	196.090777109812	4.07321643829346	
50	Contaminant	198.185793700433	12.8004331588745	
51	Contaminant	198.185397939025	13.3613166809082	
52	Contaminant	200.201496500675	11.2479333877563	
53	Contaminant	210.18534284392	12.8004331588745	

Figure 81 : Illustration de la banque de contaminant élaborée

Ajustement des matrices :

L'outil « bank\_inhouse » s'appuyant sur les rapports  $m/z$  et les temps de rétention en minutes pour effectuer l'annotation, ces deux paramètres doivent être introduits dans les matrices de données nettoyées. Ces données sont accessibles dans la matrice de variables correspondante. Cependant, les matrices de données ont été nettoyées, de ce fait elles comportent un nombre inférieur de données. La solution adoptée a été de réaliser une jointure entre les deux matrices à l'aide de la fonction RECHERCHEV ou RECHERCHEH selon l'orientation de la matrice.

Annotation :

Lorsque la matrice et la banque de contaminants sont prêtes, la matrice est annotée avec l'outil « bank\_inhouse » de Galaxy.

Tableau 22 : Paramétrage de l’outil d’annotation banque de données interne de Galaxy

<b>Bank inhouse</b>	Would you use a file	yes
	File of masses	MatriceA17_Annotable
	Do you have a header?	yes
	Number of header lines	1
	Would use retention time (RT)?	yes
	Column of retention time (RT)	Column: 3
	Delta of retention time (RT) (in minutes)	0,5
	Column of masses (MZ)	Column: 2
	Delta type	ppm
	Delta of mass (MZ) (in ppm)	5
	Use personal data bank?	yes
	bank file	BanqueContaminants
	Column of masses (MZ) in bank file	Column: 2
	Have you used the retention time (RT) for your input data?	yes
	Column of retention time (RT) in bank file	Column: 3
Would you more output?	yes	

**Suppression des contaminants :**

Une fois la matrice annotée, les lignes correspondant aux contaminants sont supprimées sous excel. Le pic majoritaire du standard interne a été conservé pour servir d’indicateur sur la qualité des corrections des jeux de données.

The screenshot shows a spreadsheet with columns labeled NR, NS, NT, NU, NV, NW, NX, NY, NZ, OA, OB, OC, OD, OE, OF. The rows contain numerical data, with some cells containing text like 'No\_match\_ir NA' or 'Contaminant189.50744812'. The spreadsheet is titled 'Galaxy10-(bank\_inhouse\_MatriceB)'.

Figure 82 : Exemple d’annotation réalisée par Galaxy en utilisant la banque de contaminant (matrice inflorescence-lavande)

L’annotation et la suppression ont été répétées pour chaque matrice (au nombre de 4).

**III.A.3. Suppression des variables non significatives**

L’objectif de ce nettoyage est de supprimer tous les ions qui n’apportent aucune information intéressante vis-à-vis de la problématique de l’étude. L’intérêt de cette procédure est dû au fait que les données redondantes ou non pertinentes perturbent les résultats des analyses supervisées (Wolfender et al. 2013; Worley and Powers 2013). Pour ce faire, des tests statistiques univariés ont été employés (t-test) afin de déterminer quels ions étaient significatifs en fonction de la problématique.

## Chapitre 3 : Traitement des données

Dans un premier temps, pour chaque jeu de données deux matrices ont été nettoyées par rapport aux métadonnées : une en fonction de la significativité par rapport aux variétés et une deuxième en fonction de la significativité en fonction de l'état symptomatique des plants, que nous dénomons respectivement par matrice « variété » et « statut ». Cependant, la variance due aux variétés est plus élevée que celle imputable au statut des plants. Afin de discerner la différence de composition chimique entre les états symptomatiques des plants, ceux-ci ont été considérés pour chaque variété séparément. Dès lors, les matrices de type statut ont été abandonnées au profit de huit matrices de type « statut par variété ». A l'issue de ce nettoyage, nous avons donc douze matrices différentes présentées dans le tableau ci-dessous.

Tableau 23 : Dénomination des douze matrices conservées et leurs caractéristiques

Nom matrice	Métadonnées	Jeux de données	Lots	Variétés	Organes plante
Lavande-feuille	Variété	Lot	A17	77-Ma	Feuille
Lavande-inflorescence			B17	77-Ma	Inflorescence
Lavandin-feuille			C17	Ab-Gr	Feuille
Lavandin-inflorescence			D17	Ab-Gr	Inflorescence
77F	Statut	Par variété	A17	77	Feuille
MaF			A17	Ma	Feuille
77Inflo			B17	77	Inflorescence
Malnflo			B17	Ma	Inflorescence
AbF			C17	Ab	Feuille
GrF			C17	Gr	Feuille
AbInflo			D17	Ab	Inflorescence
GrInflo			D17	Gr	Inflorescence

77 = 7713, Ma = maillette, Ab = abrial, Gr = grosso, Inflo = Inflorescence

La réalisation de la suppression des ions non significatifs comprend plusieurs grandes étapes. La procédure à appliquer est détaillée en Annexe 11.

La première étape consiste à normaliser la matrice. A noter que les nettoyages précédents (contaminants, isotopes, adduits et fragments) ont été réalisés sur les matrices « brutes ». Dans cette partie, le nettoyage faisant appel à des tests statistiques, les données doivent être corrigées et transformées. Pour ce faire, nous avons utilisé MetaboAnalyst avec le paramétrage défini précédemment (voir Tableau 20 : Paramètres finaux obtenus pour la correction et la préparation des données aux analyses statistiques sous MetaboAnalyst).

Une fois la matrice normalisée, les pools sont supprimés de la matrice. Les pools servent à normaliser et deviennent gênants car ils interfèrent avec l'analyse supervisée. En effet, ils seraient considérés comme un groupe à part entière par ces analyses.

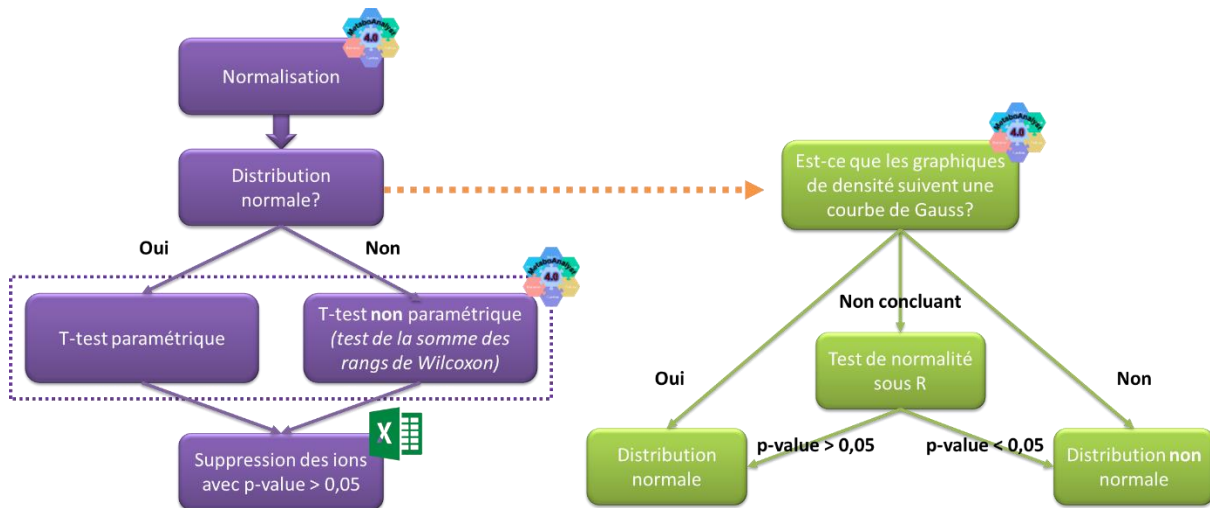


Figure 83 : Logigramme du processus utilisé pour la suppression des ions non significatifs

La troisième étape consiste à déterminer si la distribution générale des données suit une loi normale. En fonction du résultat, le choix de l'analyse statistique se portera sur un T-test paramétrique ou non. Pour ce faire et comme évoqué précédemment (voir partie II), MetaboAnalyst possède des sorties graphiques après normalisation permettant d'observer la distribution des données (ions et échantillons). Si la distribution générale suit une courbe de Gauss (ie en forme de cloche), elle est dite normale et un test paramétrique (T-test « classique ») est employé. Dans le cas contraire, un test non-paramétrique est appliqué. Dans ce cas, MetaboAnalyst utilise le test de la somme des rangs de Wilcoxon (Wilcoxon 1945; Kotz and Johnson 1992). Dans de nombreux cas, en ne prenant en compte uniquement le graphique fourni par MetaboAnalyst, le doute subsistait sur la normalité de la distribution. Pour y remédier, la fonction « tapply » sous R a été utilisée. Cette fonction permet d'affirmer qu'une distribution est normale si la p-value obtenue est supérieure à 0,05. Toutes les matrices de l'étude se sont avérées posséder des distributions non normales.

Par conséquent, la quatrième étape a consisté à appliquer un t-test non-paramétrique. MetaboAnalyst propose une sortie graphique interactive permettant de faire apparaître les boîtes à moustaches de chaque ion. L'absence de significativité de chaque ion supprimé a été vérifiée sur les graphiques.



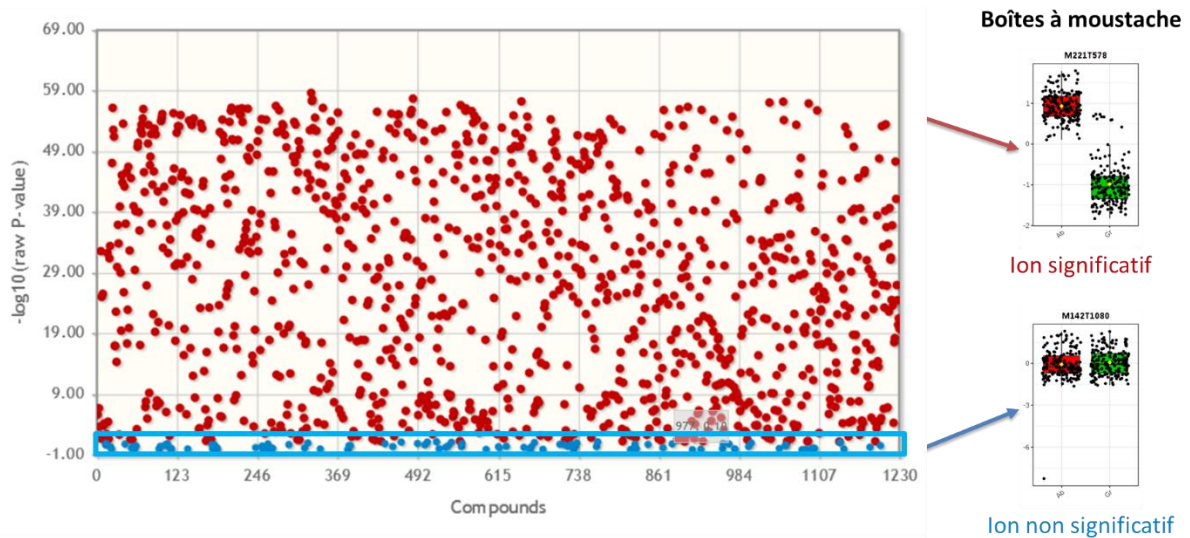


Figure 84 : Graphique des p-value obtenus pour chaque ion à l'issue du t-test pour la matrice des extraits d'inflorescence de lavandin et deux exemples de boîtes à moustaches pour un ion significatif et pour un ion non significatif.

La dernière étape consiste à effectuer une jointure entre la matrice obtenue à l'issue de l'étape de nettoyage précédente et la liste d'ions significatifs obtenus à l'issue du test. De cette manière, la matrice obtenue est « brute » (pas de normalisation, de mise à l'échelle ou transformation), contient les analyses de pool et uniquement les ions significatifs.

#### III.A.4. Impact du nettoyage sur les ions

Le nettoyage des données a permis d'éliminer les ions non informatifs voire redondants. Il a induit une réduction d'environ de moitié du nombre d'ions présents dans les matrices (voir Tableau 24). Comme il est visible sur les graphiques des individus et des variables ci-dessous, les différents nettoyages successifs sur les ions ont bien fonctionné. En effet, ils n'ont pas impliqué de modification majeure sur les graphiques des individus (donc la composition des extraits), tout en diminuant le nombre de ions (cf. graphique des variables) présents sans changer pour autant leur répartition. De plus, la variabilité totale expliquée a augmenté.

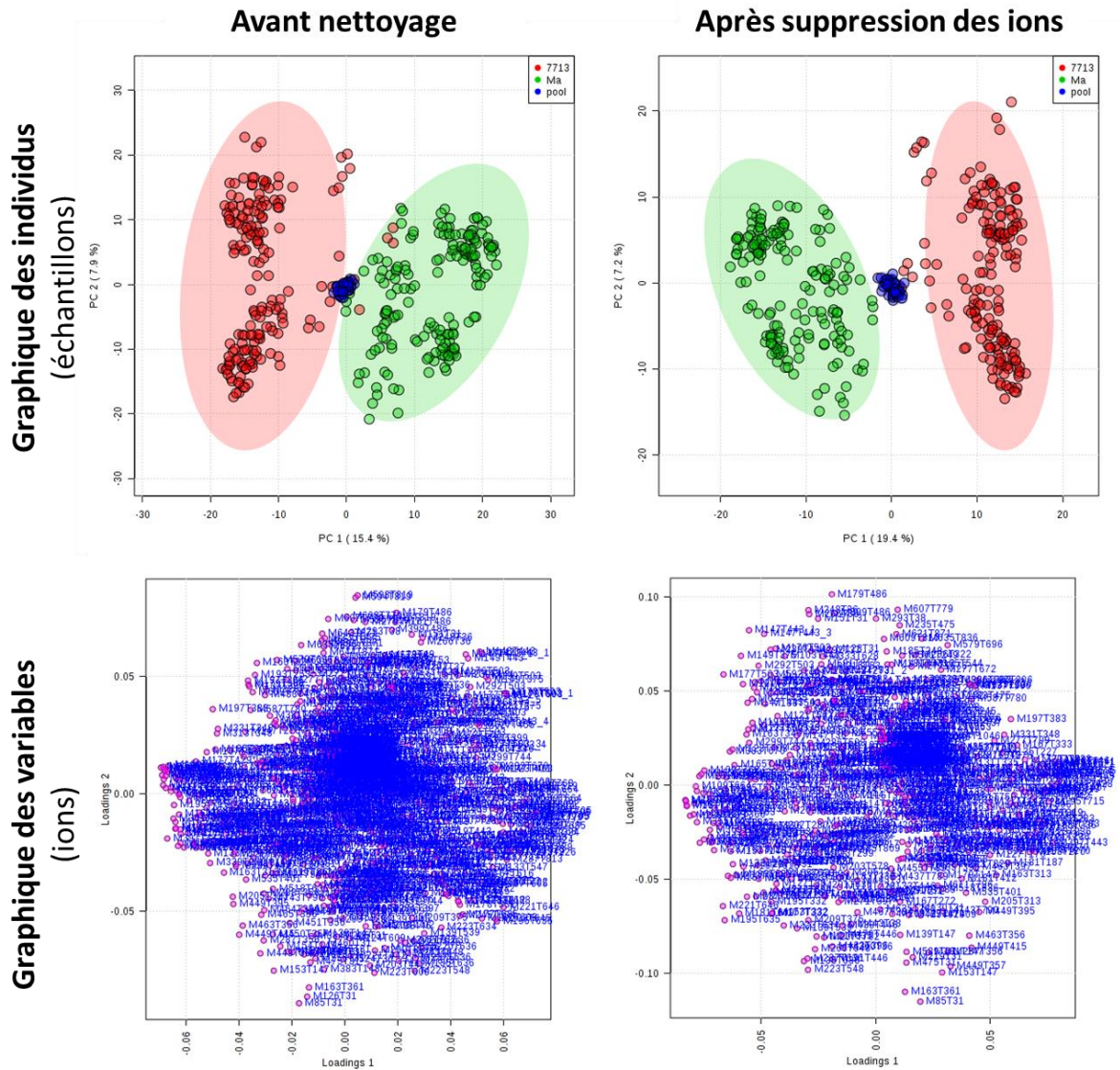


Figure 85 : Exemple de l'influence de la suppression des ions lors du nettoyage sur la matrice des extraits de feuille de lavande. En haut, les graphiques des échantillons (*score plot*) ; en bas, les graphiques des ions (*loading plot*) ; à gauche, avant nettoyage ; et à droite, après suppression des ions.

Tableau 24 : Nombre d'ions avant et après le nettoyage des données

Matrice	Nombre d'ions	
	Avant nettoyage	Après nettoyage
Lavande-feuille	1082	597
Lavande-inflorescence	1183	726
Lavandin-feuille	983	575
Lavandin-inflorescence	1646	1127
77F	1082	482
MaF	1082	507
77Inflo	1183	477
MaInflo	1183	473
AbF	983	435
GrF	983	421
AbInflo	1646	546
GrInflo	1646	686

### III.B. Nettoyage sur les échantillons

#### III.B.1. Suppression des données aberrantes

##### a. Procédure générale

Le but de cette partie est de supprimer les données aberrantes des différents jeux de données. Pour réaliser ce nettoyage sur les échantillons, nous utilisons l'ACP qui est une analyse multivariée descriptive. Cette analyse statistique permet de visualiser le jeu de données sans le forcer, et ainsi détecter d'éventuelles valeurs aberrantes et les exclure. Par exemple, si un échantillon apparaît comme complètement isolé ou s'il est présent dans un groupe autre que le sien, il sera considéré comme aberrant.

Pour rappel, à l'issu des étapes de nettoyage précédentes, nous avons obtenu douze matrices se décomposant en quatre matrices de variétés et huit matrices en fonction de l'état symptomatique des plants par variété (statut par variété). Afin de détecter les échantillons aberrants, une ACP est réalisée pour chaque matrice sur la plateforme MetaboAnalyst, sans filtration et avec l'application d'une normalisation par les pools, d'une transformation logarithmique, d'un centrage et d'une réduction automatique (voir Tableau 20). Pour chaque résultat, les extraits aberrants sont relevés, ainsi que leur type d'aberrations (ex : isolé, dans le groupe opposé). Ensuite, d'éventuelles erreurs lors des expérimentations sont recherchées afin d'expliquer ces aberrations. En fonction, des résultats les analyses sont supprimées de la matrice de données de départ. A noter que ces suppressions doivent conserver un caractère exceptionnel.

### b. Analyses supprimées

Plusieurs aberrations ont été retrouvées dans les différentes matrices de l'étude. La description de chaque erreur est explicitée ci-dessous. Une même erreur est généralement présente dans plusieurs matrices et un récapitulatif des suppressions par matrices est présenté à la fin de cette partie.

#### Matrices des extraits de lavande :

- Echantillons 7713 4SbF et 7713 4Sb :

L'ACP sur les extraits de feuille de lavande (voir Figure 86) a mis en évidence trois échantillons aberrants, à savoir les extraits A207, A252 et A326. Après vérification, il se trouve que ces trois extraits appartiennent tous au même échantillon (7713 4SbF). De surcroît, l'ACP réalisée sur la matrice des extraits d'inflorescences de lavande (voir Figure 86), a elle aussi fait apparaître trois extraits aberrants (B004, B013 et B149) correspondant à la partie inflorescence du même échantillon (7713 4Sb). Ces mêmes aberrations ont également été constatées sur les ACP effectuées sur les matrices de 7713 feuille et inflorescence (statut par variété). Aucun incident n'a été répertorié sur ces deux échantillons. Néanmoins, une erreur lors de l'échantillonnage n'est pas exclue (erreur de marquage, non-respect de la trempé, contamination lors du prélèvement), et nous avons fait le choix de supprimer ces échantillons des différents jeux de données impliqués.

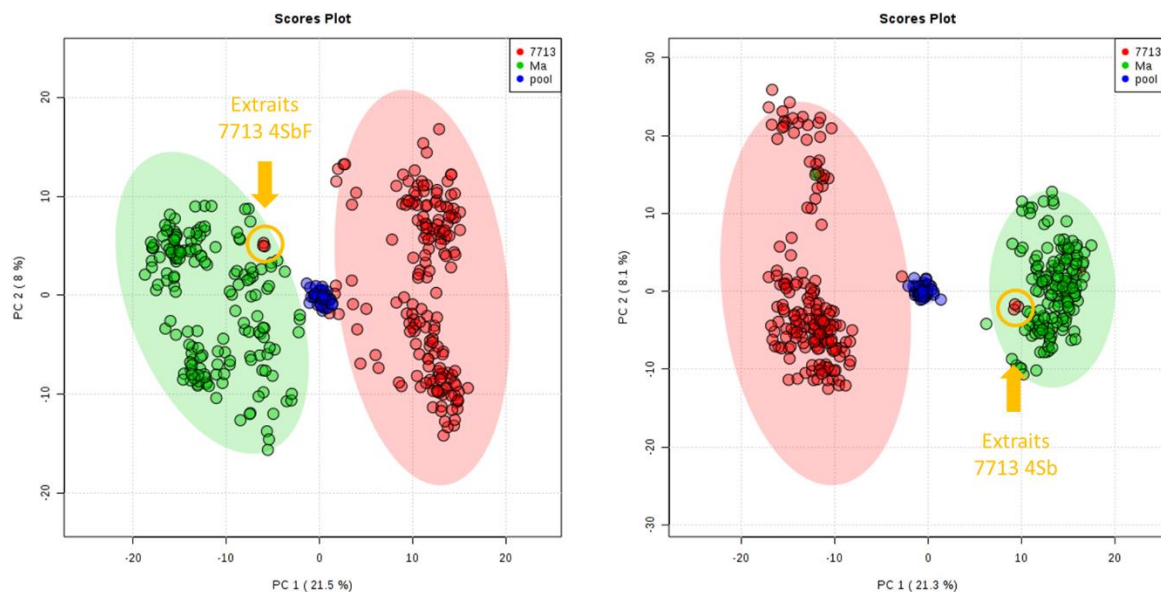


Figure 86 : ACP des matrices de feuilles (à gauche) et d'inflorescences (à droite) de lavande en fonction des variétés avant suppression des échantillons aberrants

- Mélange de deux extraits lors de la préparation :

Lors de l'étape d'extraction une erreur a été commise et deux extraits ont été mélangés (B320 et B321). L'extrait B320 est composé des extraits B320 et B321, ce qui correspond à un extrait des variétés 7713 et de Maillette. Ceci explique la présence de l'extrait B320 proche du groupe des pools sur les ACP (voir Figure 87). Malgré une très faible quantité restante, l'extrait B321 a été analysé. L'analyse apparaît comme aberrante uniquement sur l'ACP de la matrice des extraits d'inflorescences de maillette en fonction de l'état symptomatique du plant (voir figure ci-dessous). Néanmoins, il apparaît après vérification dans les matrices de données que cet extrait comporte de nombreux pics ayant une intensité nulle ou anormalement faible. Les résultats pour cet extrait dénotent clairement par rapport aux autres. Nous supposons que l'analyse a été faussée du fait de la trop faible quantité d'extrait restant. Nous avons préféré supprimer cet extrait.

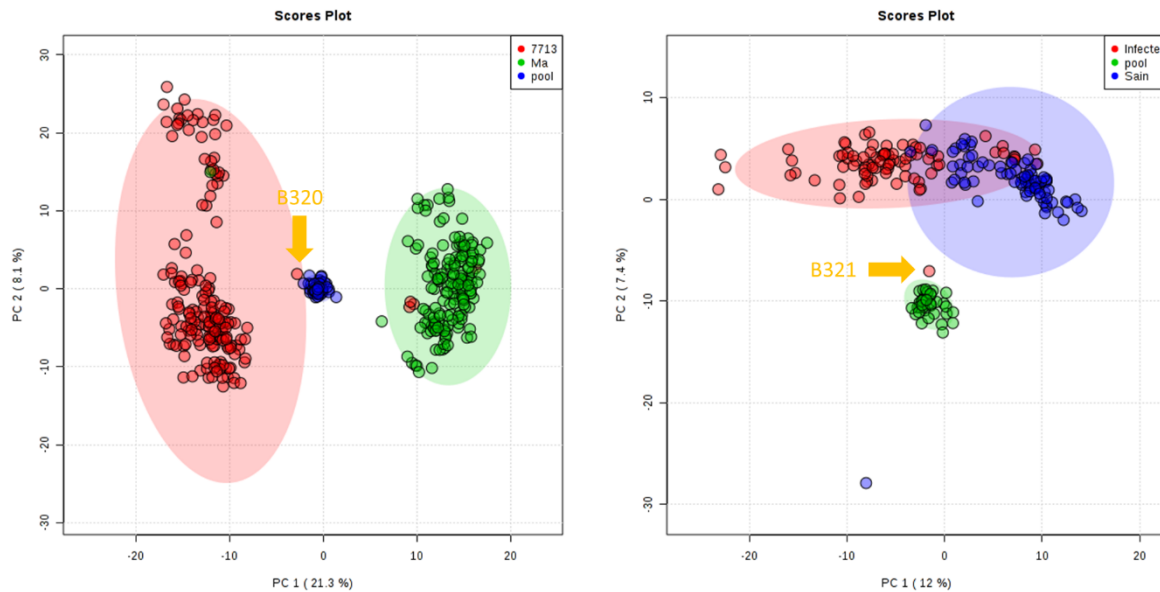


Figure 87 : ACP de la matrice d'inflorescence de lavande en fonction des variétés (à gauche) et de la matrice d'inflorescence de maillette en fonction de l'état symptomatique du plant (à droite) avant suppression des échantillons aberrants

Les extraits B320 et B321 ont été supprimés de toutes les matrices de données dans lesquels ils apparaissent (i.e. matrices des extraits d'inflorescence de lavande par variété et les deux matrices de statut par variété).

- Inversion de deux extraits lors de l'analyse :

A la fin de la séquence d'analyse n°2 du lot B17, une inversion de place sur le rack de l'UPLC-HRMS entre deux extraits (B234 et B235) a été relevée. L'ACP de la matrice sur les variétés des extraits d'inflorescence de lavande (voir Figure 88) a fait apparaître ces deux extraits dans le groupe opposé, ce qui est cohérent avec la réelle appartenance de chaque extrait avec sa variété de lavande (B234 = 7713 et B235 = maillette). Pour la matrice sur les variétés, une correction a été effectuée en inversant les lignes correspondantes aux analyses B234 et B235 dans la matrice de données. Cependant, cette correction est impossible pour les matrices de type statut par variété correspondante. Par conséquent, l'extrait a été supprimé dans chacune de ces matrices.

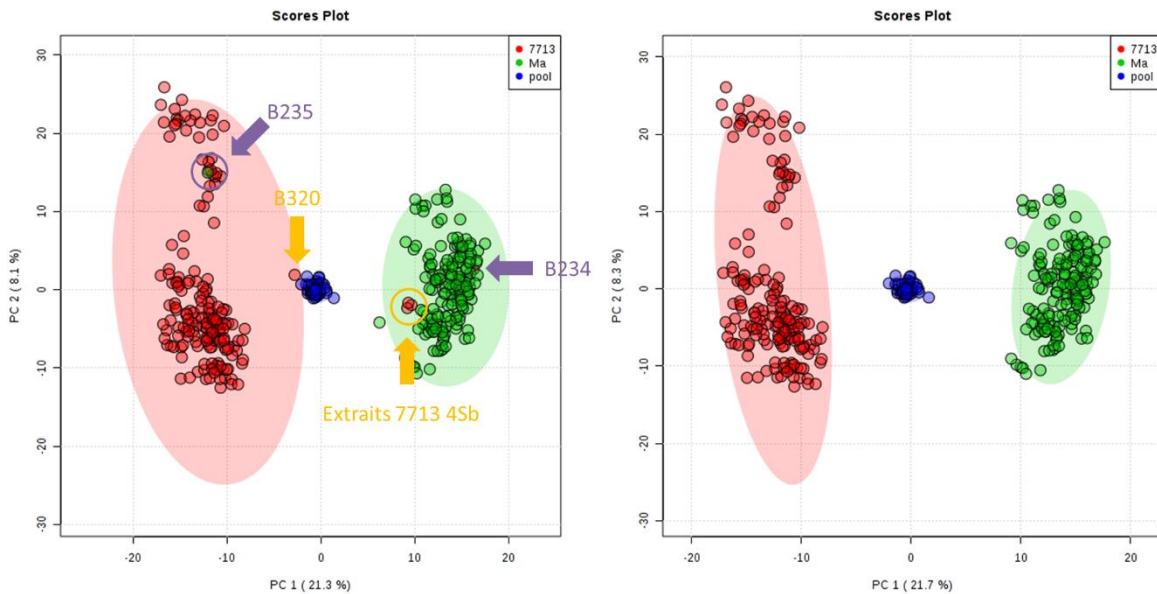


Figure 88 : ACP de la matrice d'inflorescence de lavande en fonction des variétés avant (à gauche) et après (à droite) suppression des échantillons aberrants

**Matrices des extraits de lavandin :**

- Plant grosso n°4 symptomatique :

Pour les extraits d'inflorescence de lavandin (lot D17), neuf analyses correspondant à des extraits de la variété grosso sont localisées sur les ACP dans le groupe abrial (voir Figure 89). Les codes extraits correspondant sont D002, D031, D087, D102, D140, D219, D247, D296 et D348. Ils appartiennent tous au plant n°4 de grosso symptomatique.

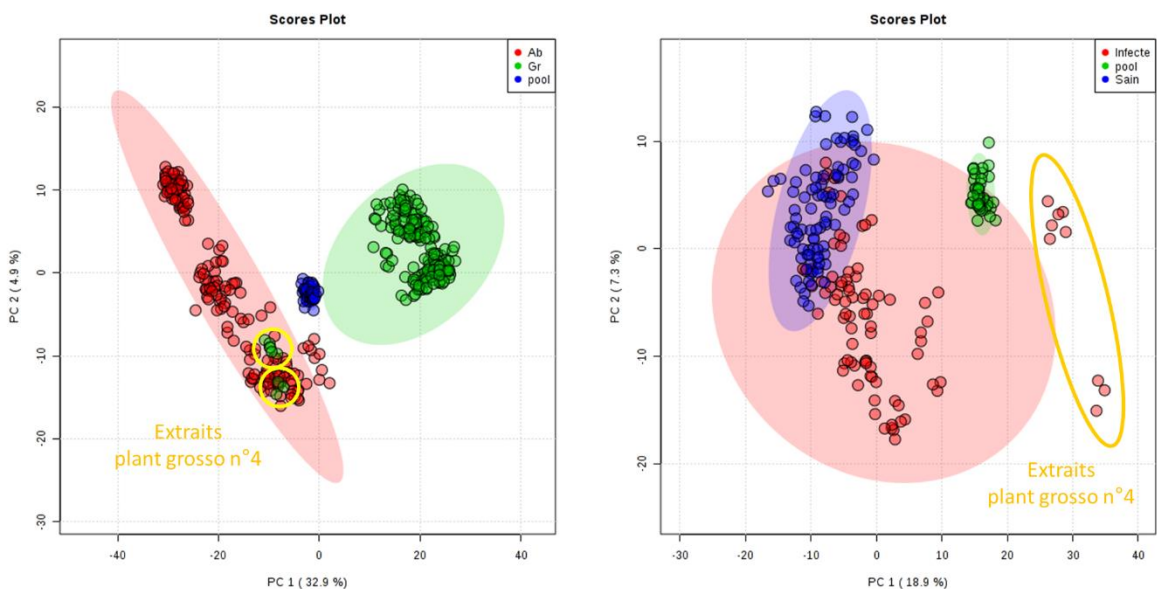


Figure 89 : ACP de la matrice d'inflorescence de lavandin en fonction des variétés (à gauche) et de la matrice d'inflorescence de grosso en fonction de l'état symptomatique du plant (à droite) avant suppression des échantillons aberrants

Concernant les extraits de feuille, les extraits du plant n°4 (C043, C048, C127, C138, C210, C216, C298, C310 et C331) apparaissent comme douteux sur les ACP mais de façon moins flagrante que pour les

extraits d'inflorescence (voir Figure 90). La visualisation 3D interactive permet de confirmer l'isolement de ces extraits notamment pour la matrice de feuille de grosso en fonction de l'état symptomatique du plant.

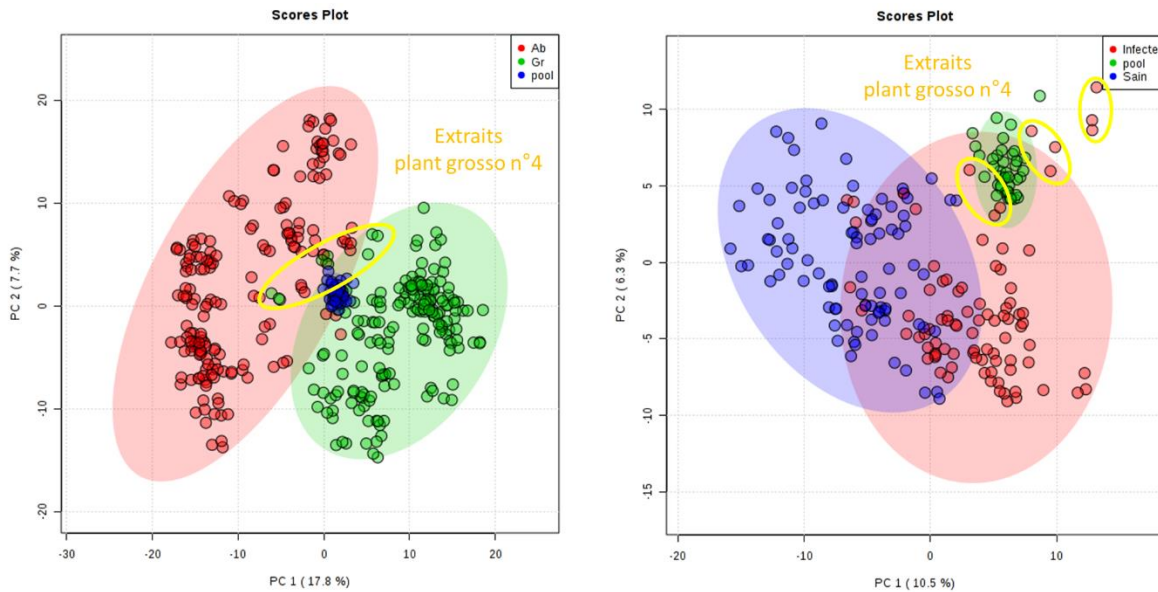


Figure 90 : ACP de la matrice de feuille de lavandin en fonction des variétés (à gauche) et de la matrice de feuille de grosso en fonction de l'état symptomatique du plant (à droite) avant suppression des échantillons aberrants

Aucune corrélation ou erreur n'a été relevée qui pourrait expliciter que les extraits de ce plant apparaissent comme aberrants. A noter, qu'aucune confusion avec le plant n°4 symptomatique abrial ou avec le plant de n°4 grosso asymptotique n'a pu avoir lieu, car ces plants ont leurs extraits et ne sont pas aberrants. L'hypothèse qui nous apparaît la plus plausible est une erreur sur la variété du plant. En effet, la parcelle d'échantillonnage de grosso et d'abrial est la même. Nous avons fait le choix de supprimer l'intégralité des extraits correspondant à ce plant sur les matrices concernées.

- Abrial :

Pour les matrices de la variété abrial en fonction de l'état symptomatique des plants, les extraits appartenant aux deux réplicats biologiques a et c du plant symptomatique d'abrial n°3 sont isolées (voir Figure 91). Cette constatation est valable pour les extraits de feuilles et d'inflorescences. Les extraits correspondants au troisième réplica biologique (b) du même plant n'ont montré aucun caractère aberrant. Aucun problème n'a été reporté sur ces extraits durant la préparation. Cependant, comme les extraits de feuilles et d'inflorescences sont tous aberrants pour ces deux échantillons (a et c), ils ont été supprimés.

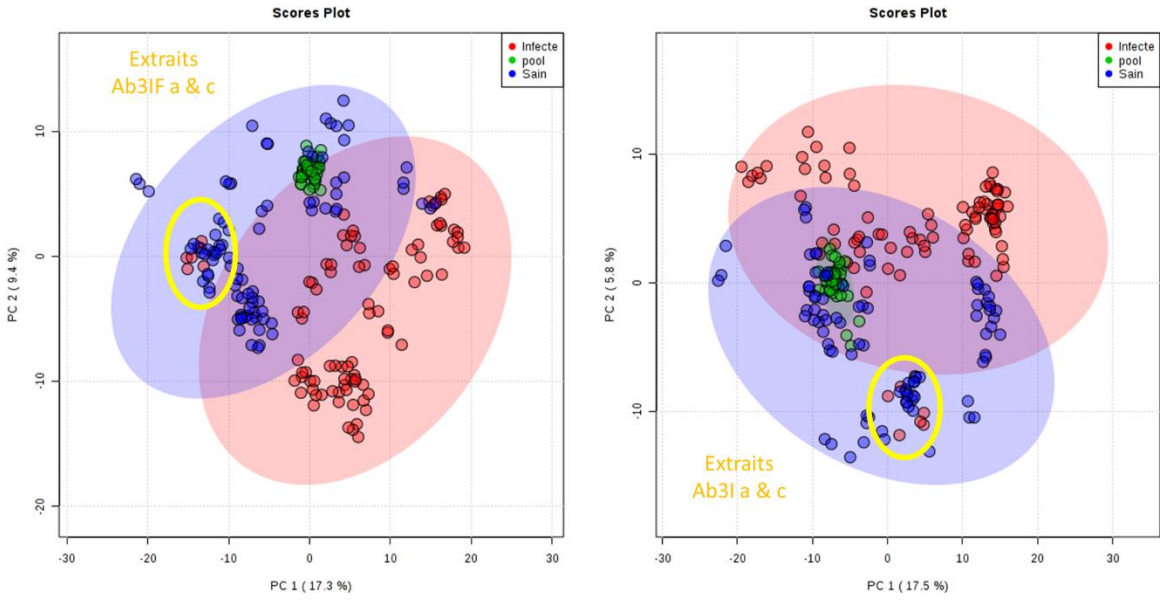


Figure 91 : ACP de la matrice de feuille (à gauche) et d’inflorescence d’abrial (à droite) en fonction de l’état symptomatique du plant avant suppression des échantillons aberrants



**Synthèse :**

Tableau 25 : Récapitulatif des échantillons aberrants supprimés par matrice sur l'ensemble des jeux de données

Lots	Variétés	Statut par variété	
		7713	Maillette
Lavande - Feuille (A17)	<b>Suppression :</b> 3 extraits (échantillon 7713 4SbF) = A207, A252 et A326	<b>Suppression :</b> 3 extraits (7713 4SbF)	∅
Lavande – Inflorescence (B17)	<b>Correction :</b> B234 et B235 (inversion sur le rack) <b>Suppression :</b> B320 et B321 (mélange des extraits lors de la préparation) + 3 extraits (7713 4Sb) = B004, B013, B149	<b>Correction :</b> B252 erreur d'annotation (sain) <b>Suppression :</b> B234 + B320+ 3 extraits (7713 4Sb)	<b>Suppression :</b> B235 et B321
		Statut par variété	
		Abrial	Grosso
Lavandin – Feuille (C17)	<b>Suppression :</b> 9 extraits plant Gr 4I = C043, C048, C127, C138, C210, C216, C298, C310 et C331	<b>Suppression :</b> 6 extraits (Ab 3IF a et c) = C052, C098, C197, C234, C266 et C275	<b>Suppression :</b> 9 extraits plant Gr 4I
Lavandin – Inflorescence (D17)	<b>Suppression :</b> 9 extraits plant Gr 4I = D002, D031, D087, D102, D140, D219, D247, D296 et D348	<b>Suppression :</b> 6 extraits (Ab 3I a et c) = D025, D027, D038, D074, D293 et D308	<b>Suppression :</b> 9 extraits plant Gr 4I

∅ = aucune suppression ou correction

#### IV. Analyses statistiques et obtention des composés discriminants

Les tests statistiques mis en place dans cette étude métabolomique sont des analyses multivariées, plus précisément, une ACP et une PLS-DA ont été réalisées pour chaque jeu de données. Nous verrons également dans cette partie comment ont été obtenus les biomarqueurs ainsi que leurs caractères discriminants respectifs. Les tests ont été effectués sur la plateforme Metaboanalyst et avec le logiciel R à l'aide du package « mixomics ». Les résultats obtenus pour les deux logiciels sont très similaires, c'est pour cette raison que nous ne présenterons que les résultats obtenus par Metaboanalyst (voir chapitre 4).

### IV.A. Statistiques multivariées

#### IV.A.1. Analyses descriptives ou exploratoires : ACP

Le premier test effectué est une ACP. Comme explicité dans le chapitre 1, les ACP sont des analyses multivariées descriptives créant un système de coordonnées sur lequel est projeté le jeu de données de façon à maximiser la variance. Elles ont été réalisées dans le but d'obtenir une première vision du jeu de données et évaluer la qualité des analyses avec les pools. Elles ont été générées à l'aide de l'outil « Statistical analysis » sous Metaboanalyst après nettoyage, correction, centrage-réduction et transformation des matrices (cf. parties ci-dessus).

#### IV.A.2. Analyses discriminantes : PLS-DA

Dans un second temps, des PLS-DA ont été effectuées dans le but de mettre en évidence des différences de compositions chimiques (quantitatives ou qualitatives) en fonction de facteurs biologiques, à savoir la variété (qui renvoie au caractère sensible/tolérant) ou l'état sanitaire des plants. Pour rappel (voir partie III.B.5.b.iii du chapitre 1), les PLS-DA sont des analyses discriminantes et vont créer un nouvel espace, de façon à maximiser les différences entre les groupes de métadonnées choisis par l'utilisateur. Comme précédemment pour les ACP, les PLS-DA ont été réalisées grâce à l'outil « Statistical analysis » de Metaboanalyst. Contrairement aux jeux de données employés pour les ACP, les analyses des pools ont été supprimées après normalisation. Les transformations (log et autoscaling) ont été effectuées après la suppression des pools.

#### Validation des PLS-DA :

Comme vu dans le chapitre 1 (partie III.B.5.b.iii), les indicateurs pour valider une PLS-DA sont  $R^2$  et son analogue  $Q^2$  (obtenu par validation croisée). Elles sont considérées comme valides lorsque les valeurs de  $R^2$  et  $Q^2$  sont supérieures à 0,5. Metaboanalyst génère automatiquement ces deux indicateurs lorsqu'une PLS-DA est réalisée. Il est possible de les retrouver dans l'onglet « Cross Validation ».

Néanmoins, la PLS-DA est sujette au risque de sur-apprentissage (ou « over-fitting ») du modèle, c'est-à-dire que l'algorithme ne modélise pas de façon suffisante les données et vient coller de façon excessive à ces dernières (voir Figure 92). Un sur-apprentissage va induire l'incorporation d'informations telles que du bruit de fond dans la construction du modèle. Dans le cas d'une utilisation prédictive du modèle, celui-ci étant trop proche du jeu de données qui a servi à sa création, il ne va pas être généralisable.

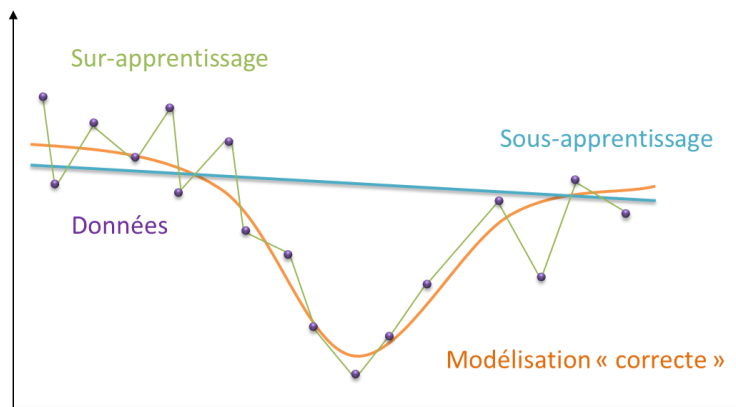


Figure 92 : Illustration du sous- et sur-apprentissage d'un modèle

Pour évaluer le risque de sur-apprentissage du modèle, la validation croisée est parfois insuffisante et les tests de permutations sont de plus en plus couramment employés (Xia and Wishart 2011). Le principe de ce test est d'échanger les classifications des échantillons de façon aléatoire entre les

## Chapitre 3 : Traitement des données

différents groupes biologiques, puis de générer une nouvelle PLS-DA. Le processus est ensuite réitéré un grand nombre de fois (Favre 2017).

Metaboanalyst permet de générer simplement des tests de permutations lors de la réalisation d'une PLS-DA grâce à l'onglet « Permutation ». Deux paramètres sont à définir : l'indicateur statistique de performance de la PLS-DA et le nombre de permutations à réaliser. Les paramètres de tests à disposition sont la justesse (« Prediction accuracy during training ») et le ratio entre la somme des carrés des distances intra- et inter- groupes (« Separation distance (B/W) »). L'indicateur conseillé et que nous avons choisi d'employer est la justesse (Xia and Wishart 2011). Le nombre de permutations à effectuer peut-être de 20, 100, 1000 ou 2000. Nous avons choisi d'appliquer 1000 permutations afin d'avoir un nombre de permutations supérieur au nombre d'extraits contenus dans un jeu de données (environ 200), tout en conservant une durée de calcul acceptable. Les résultats des tests de permutation se présentent sous la forme d'un histogramme (voir Figure 93 ci-dessous). (Chong et al. 2019)

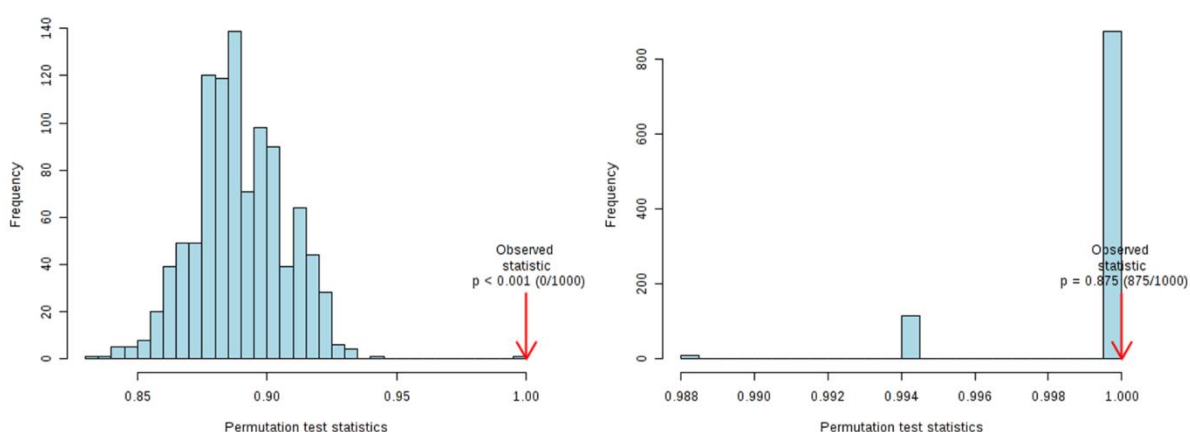


Figure 93 : Deux exemples d'histogrammes obtenus à l'issue de tests de permutation effectués sous Metaboanalyst. A gauche, un test valide ; et à droite, un test ayant un sur-apprentissage du modèle.

La PLS-DA initiale sert de référence et son résultat est modélisé par une flèche rouge sur le graphique. Si les résultats obtenus lors des permutations sont supérieurs à ceux de la PLS-DA (exemple test de droite) alors la séparation obtenue est due à un sur-apprentissage du modèle. Le rapport écrit au-dessus de la flèche rouge, indique que 875 permutations sur les 1000 effectuées ont obtenu une valeur de l'indicateur supérieure à la configuration de départ, la séparation des groupes est fortuite et ne correspond pas à une réalité biologique. A l'inverse, un écart marqué entre les permutations et la PLS-DA initiale sera la preuve que la séparation des groupes biologiques reflète une réalité (exemple test de gauche sur la figure).

### IV.B. Biomarqueurs

#### IV.B.1. Obtention des biomarqueurs

A l'issue de l'analyse PLS-DA effectuée pour chaque matrice sur la plateforme MetaboAnalyst, un tableur est disponible contenant l'intégralité des ions de la matrice avec leur VIP scores. Habituellement un composé est considéré comme discriminant lorsque son score est supérieur à 1 (Lê Cao and Le Gall 2011). Dans cette étude, le nombre de composés ayant un score supérieur à 1 est élevé. Nous avons donc fait le choix de travailler arbitrairement sur les 50 composés les plus discriminants pour chaque matrice, ce qui représente environ 10 % du nombre total des composés présents dans les matrices après nettoyage (voir Tableau 26).

Tableau 26 : Nombre d'ions contenus dans chaque matrice après nettoyage

Matrice en fonction de variétés (caractère sensible/tolérant)								
Nom	Lavande F		Lavande Inflo.		Lavandin F		Lavandin Inflo.	
Nombre d'ions	597		726		575		1127	
Matrice en fonction de l'état symptomatique								
Nom	7713 F	7713 Inflo.	Maillette F	Maillette Inflo.	Abrial F	Abrial Inflo.	Grosso F	Grosso Inflo.
Nombre d'ions	482	477	507	473	435	546	421	686

F = parties chlorophylliennes ; Inflo. = inflorescences

#### IV.B.2. Détermination du caractère discriminant de chaque marqueur

MetaboAnalyst offre une sortie graphique permettant de visualiser les composés les plus discriminants (Figure 94). Le graphique est généré lors de la réalisation de la PLS-DA et se situe dans l'onglet « Imp. Features ». Le graphique principal représente les composés les plus discriminants en ordonnées en fonction des valeurs des VIP scores. Plus le score du composé est élevé, plus celui-ci est discriminant. Par conséquent, les molécules les plus discriminantes sont celles qui apparaissent en haut de l'axe des abscisses. A droite du graphique principal, un petit graphique de couleur représente les variations de concentration de chaque métabolite au sein des différents groupes biologiques définies par l'utilisateur. Le code couleur est explicité par une échelle à sa droite.

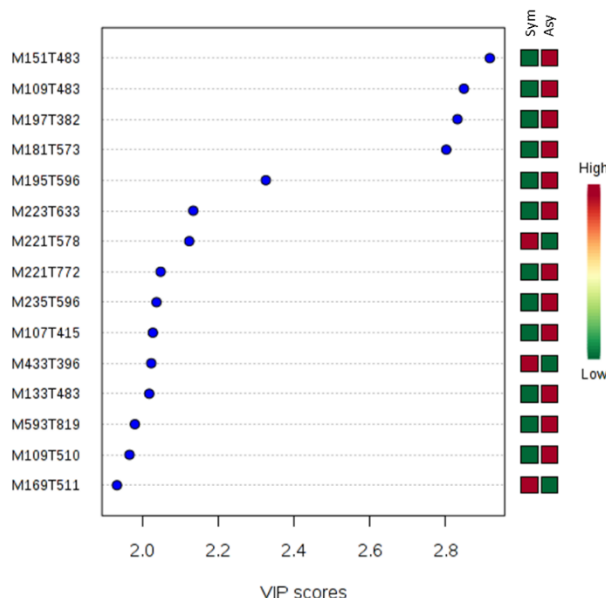


Figure 94 : Exemple d'un graphique contenant les ions les plus discriminants (en ordonnées) en fonction de leur VIP score (en abscisse) (réalisé avec Metaboanalyst)

La plateforme permet également d'accéder au boîte à moustache de chaque ion (Chong et al. 2019). La sur ou sous-expression de chaque composé retenu a été déterminée à partir de ces graphiques.

#### IV.B.3. Priorisation des biomarqueurs

Nous avons cherché à déterminer quels étaient les composés les plus pertinents vis-à-vis de la problématique. Ces molécules ont été nommées composés cibles. Nous verrons tout d'abord pourquoi et comment ces composés ont été choisis.

## Chapitre 3 : Traitement des données

### a. Détermination des composés cibles

L'objectif de cette partie est d'identifier les marqueurs (parmi les 50 premiers composés discriminants) communs à un maximum de matrice par catégorie biologique. Nous avons différencié l'intégralité des données générées en quatre catégories, nommées et définies de la façon suivante :

- **Statut F** : correspondant à l'ensemble des extraits des **parties chlorophylliennes** dont le traitement statistique a été conduit afin de mettre en avant le **caractère symptomatique** des plants,
- **Statut Inflo** : correspondant à l'ensemble des extraits **d'inflorescences** dont le traitement statistique a été conduit afin de mettre en avant le **caractère symptomatique** des plants,
- **Variété F** : correspondant à l'ensemble des extraits des **parties chlorophylliennes** dont le traitement statistique a été conduit afin de mettre en avant le **caractère sensible ou tolérant** des variétés,
- **Variété Inflo** : correspondant à l'ensemble des extraits **d'inflorescences** dont le traitement statistique a été conduit afin de mettre en avant le **caractère sensible ou tolérant** des variétés.

Ces catégories de matrice dépendent des facteurs biologiques de l'étude et du caractère que nous souhaitons mettre en évidence (sensibilité des variétés au phytoplasme ou état symptomatique des plants). Les matrices correspondant aux différentes catégories sont regroupées dans le tableau ci-dessous.

Tableau 27 : Catégories d'étude et matrices associées

Groupe	Statut F	Statut Inflo.	Variété F	Variété Inflo.
Matrices	77F, MaF, AbF & GrF	77Inflo, MaInflo, AbInflo & GrInflo	A17 & C17	B17 & D17

77 = 7713, Ma = maillette, Ab = abrial, Gr = grosso, F = parties chlorophylliennes, Inflo. = Inflorescences

Les catégories auxquelles nous portons le plus d'intérêt sont celles mettant en avant le caractère symptomatique des plants (statut). De plus, l'identification des biomarqueurs des jeux de données contenant les extraits des parties chlorophylliennes seront privilégiés parce qu'elles constituent les lieux de ponction alimentaire de *Hyaalsthes*.

### b. Composés cibles par catégorie

Pour chaque catégorie de matrices définies précédemment, nous avons cherché dans les 50 composés les plus discriminants ceux qui apparaissent plusieurs fois et leur nombre de répétition (voir Tableau 28 ci-dessous). Pour chaque ion une « note VIP » a été attribuée correspondant au nombre de fois où un même ion apparaît dans des matrices différentes. Nous avons également établi pour chaque ion un « classement » qui correspond à la somme des positions occupées par ledit ion dans les listes de composés discriminants auxquels il appartient. Cet indicateur permet de comparer et de classer les composés présentant le même nombre de récurrence. Plus la valeur de cet indicateur est faible, plus l'ion est discriminant et donc un marqueur d'intérêt.

Tableau 28 : Nombre de composés discriminants en fonction de leur nombre d'apparition pour chaque catégorie

Nombre de récurrence	Statut F	Statut Inflo.	Variété F	Variété Inflo.
4	2	0		
3	8	1		
2	16	19	6	0

F = parties chlorophylliennes, Inflo. = Inflorescences

La catégorie en fonction de l'état symptomatique pour les extraits des parties chlorophylliennes est celle pour laquelle un maximum de récurrence a été trouvé. De plus, elle possède deux ions très intéressants car présents dans toutes les matrices de la catégorie. A l'inverse, aucun composé n'a été retrouvé en double dans la catégorie variété inflorescence. A noter que pour les catégories en fonction de l'état symptomatique des plants, il a été constaté dans le cas où deux récurrences sont obtenues, que les ions appartiennent généralement à la même espèce (lavande ou lavandin). Les résultats des trois catégories ayant des récurrences sont présentés dans le chapitre 4.

### IV.C. Identification : confrontation aux banques de données

La dernière étape d'une étude métabolomique est l'identification des biomarqueurs. Cependant, elle est souvent la plus complexe, la plus longue et la plus limitante notamment en LC-MS (voir partie Annotation, identifications chimiques des métabolites et réseaux moléculaires du chapitre 1). La réussite de l'élucidation structurale dans une étude métabolomique conduite par LC-MS dépend souvent de la présence de banque interne au laboratoire (Wolfender et al. 2015), c'est dans cette optique que nous avons réalisé en amont une base de données des composés organiques non volatils décrits dans la littérature dans le genre *Lavandula* (voir Annexe 1).

#### Confrontation aux banques de données :

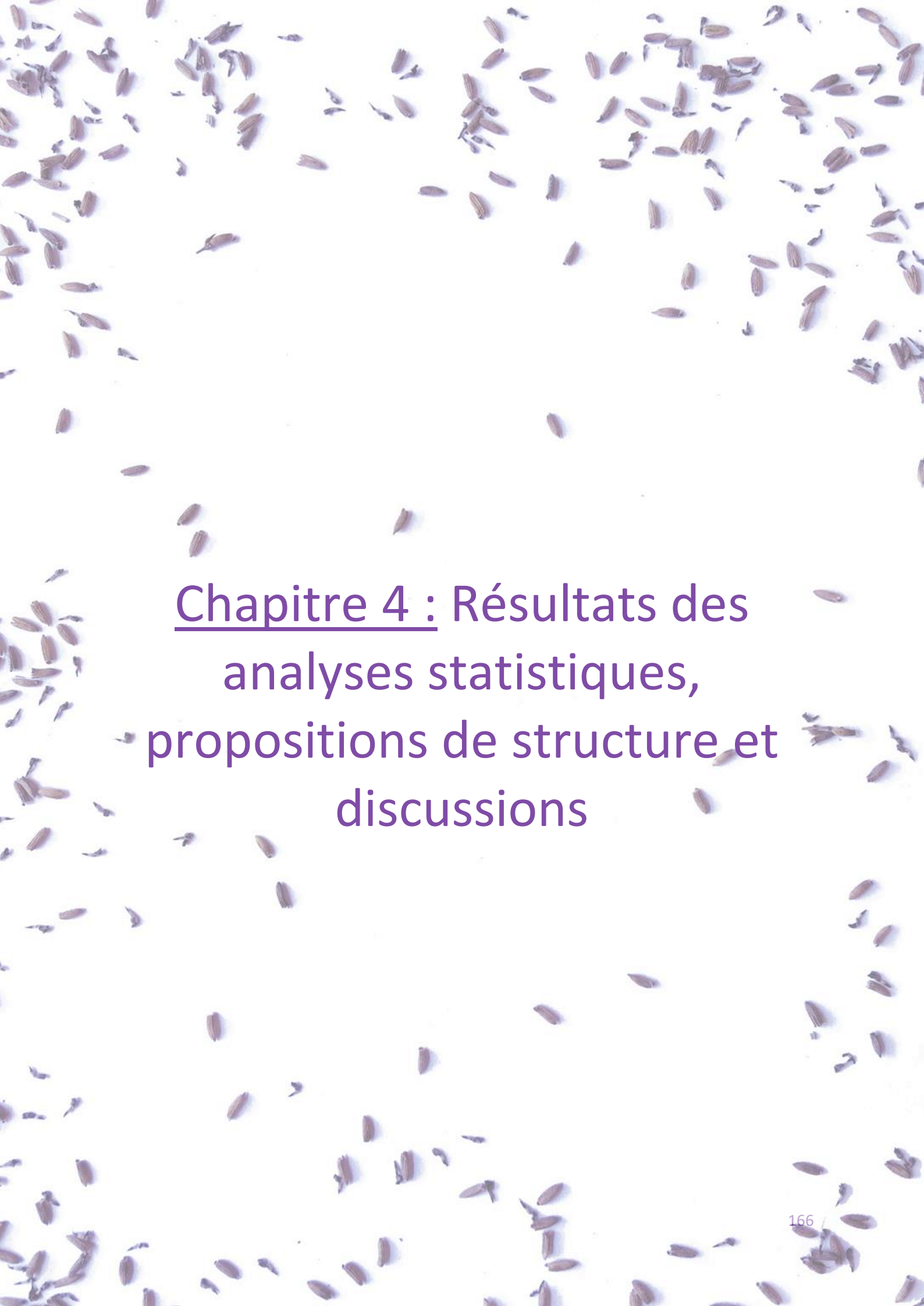
Afin d'automatiser la recherche de correspondance entre la banque et les ions obtenus, un script a été réalisé sous le logiciel Python 3 (voir Annexe 12). La recherche des correspondances par le script se base sur les masses mesurées. La précision du qTOF étant élevée, nous avons accepté une différence de masse de 0,05 Da pour valider une correspondance. A noter que le script corrige la différence de masse de 1,007825 (correspondant à un atome d'hydrogène), lié au fait que la banque de données a été conçue à partir des masses exactes des molécules trouvées dans la littérature alors que les listes de composés discriminants contiennent les ions moléculaires  $[M+H]^+$ .

Certaines molécules ayant un rapport  $m/z$  relativement bas, les résultats obtenus ont également été confrontés aux résultats obtenus durant la thèse d'Emilie Stierlin portant sur l'étude des composés volatils de lavande et de lavandin (Stierlin 2020). Une nouvelle banque (COVs) a alors été créée. Néanmoins, les données étant en basse résolution le delta de masse à dû être réévalué à 0,5 Da. De nombreuses correspondances ont été obtenues, cependant ces résultats sont à relativiser du fait de l'amplitude de masse élevée tolérée pour la recherche des correspondances.

Cette démarche a été effectuée sur les douze listes des 50 composés les plus discriminants obtenus pour chaque matrice étudiée. Les résultats sont présentés en Annexe 13.

Néanmoins, la confrontation aux banques de données a fourni peu de correspondances dont certaines sont douteuses ou redondantes. Il est donc nécessaire de réaliser *de novo* les identifications, en commençant par l'exploitation des spectres MS qui seront présentés dans le chapitre 4.





Chapitre 4 : Résultats des  
analyses statistiques,  
propositions de structure et  
discussions



## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Dans ce chapitre sera présenté l'intégralité des résultats des analyses statistiques obtenus. Les résultats seront découpés selon les questions correspondant à :

- La tolérance et la sensibilité des variétés testées face au dépérissement
- L'influence de l'état sanitaire des plants (caractère symptomatique/asymptomatique)

L'influence des différents facteurs de l'étude sur la variabilité chimique induite fera l'objet d'une première partie. Dans un deuxième temps, nous verrons les résultats des analyses et les composés discriminants cibles mis en avant par l'étude ainsi que les hypothèses d'identification qui ont pu être générées grâce à la confrontation aux banques de données internes. Enfin, dans une dernière partie nous discuterons de l'ensemble des résultats obtenus ainsi que des perspectives du projet.

### I. Etude de la tolérance des variétés au dépérissement

#### I.A. Résultats des analyses statistiques

L'un des deux objectifs de cette thèse est l'identification des biomarqueurs du caractère sensible ou tolérant des variétés face au dépérissement du Stolbur. Pour ce faire, deux variétés dites sensibles ont été sélectionnées, à savoir 7713 pour la lavande et abrial pour le lavandin, ainsi que deux variétés dites tolérantes au dépérissement correspondant à la variété maillette pour la lavande et grosso pour le lavandin. Les deux organes prélevés c'est-à-dire les parties chlorophylliennes simplement désignées sous le terme de « feuille » et les inflorescences, ont été étudiés séparément. Les graphiques correspondants aux extraits de feuilles seront toujours présentés dans la colonne de gauche et ceux des inflorescences dans la colonne de droite.

##### I.A.1. La lavande

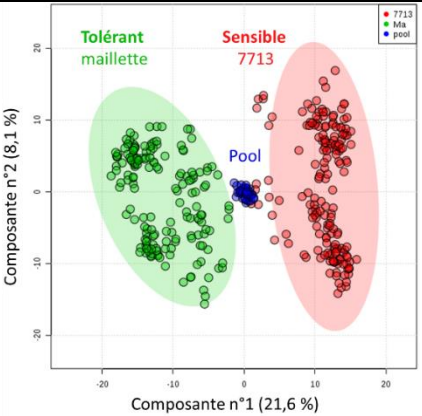
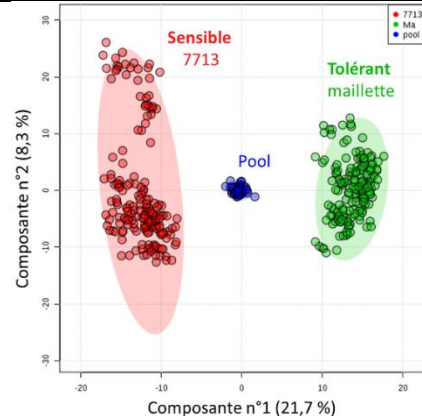
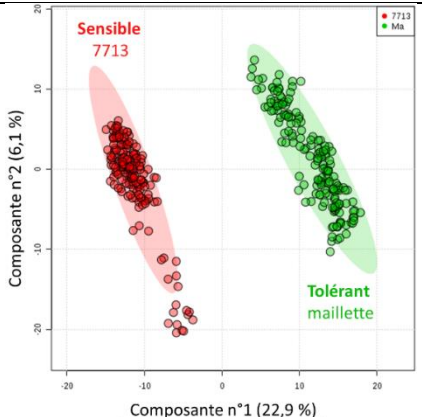
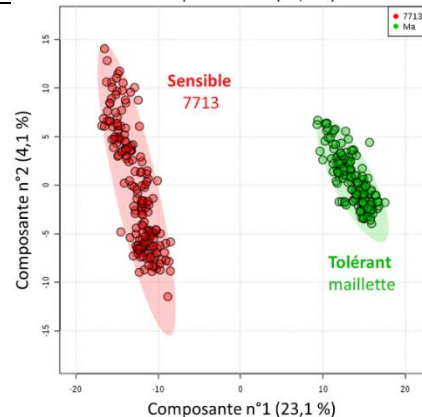
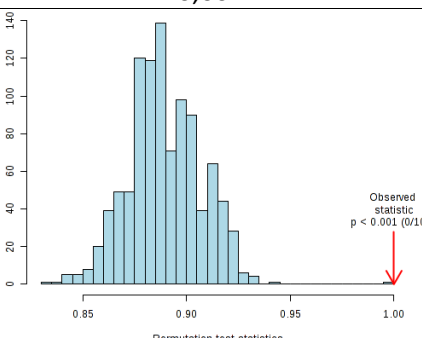
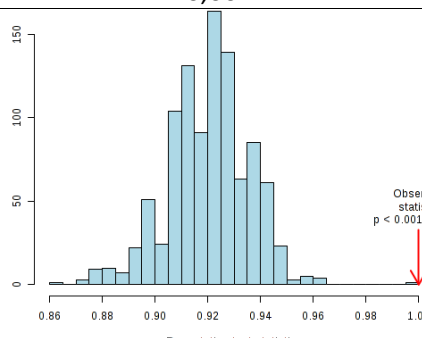
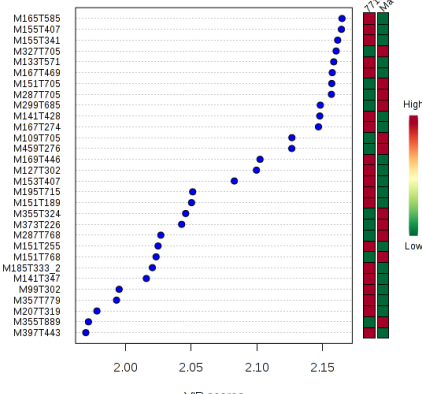
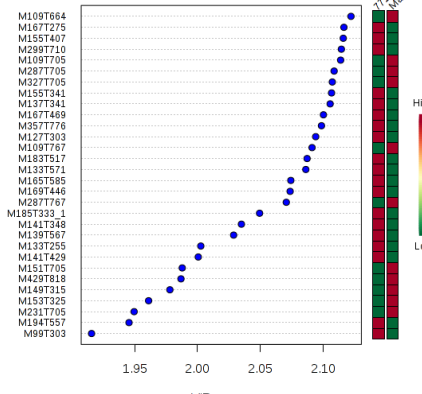
Les résultats des analyses statistiques obtenus avec MetaboAnalyst pour les extraits de lavande (feuille et fleur) sont regroupés dans le Tableau 29. A noter que pour l'ensemble des graphiques présentés dans cette partie, les variétés dites sensibles sont représentées en rouge et les variétés tolérantes en vert. Les analyses correspondantes aux pools seront quant à elles en bleu.

Les premiers graphiques présentés sont ceux des individus des ACP, ils permettent de donner une première visualisation du jeu de données sans a priori. Dans le cas de la lavande et ce pour les deux organes, les variétés sont distinctes et la variabilité cumulée sur les deux premières composantes avoisine les 30 %. Il n'était pas évident qu'une séparation soit visible sur ce test uniquement descriptif ce qui témoigne de la différence de signature chimique marquée entre les deux espèces.

Par leur principe de constitution même, à savoir un mélange de tous les extraits analysés, les pools contiennent toute la variabilité biologique et par conséquent sont le reflet de la variabilité technique induite au cours de l'étude (Wolfender et al. 2013). Ils permettent donc de vérifier la qualité des données acquises. Dans notre cas, les analyses des pools sont extrêmement bien centrées et regroupées sur les ACP ce qui implique que les analyses sont répétables, justes et que la dérive analytique constatée précédemment (voir chapitre 2, partie Contrôle de la dérive analytique au cours des séquences) a été remarquablement corrigée par la normalisation effectuée sous MetaboAnalyst. Le traitement des données est donc performant et permet de visualiser la variabilité biologique des échantillons.

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 29 : Analyses statistiques (ACP et PLS-DA) obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite de lavande.

Lavande		Feuille	Inflorescence
ACP	Graphique des individus		
		Composante n°2 (8,1 %)	Composante n°2 (8,3 %)
PLS-DA	Graphique des individus		
		Composante n°2 (6,1 %)	Composante n°2 (4,1 %)
	R <sup>2</sup>	0,98	0,99
	Q <sup>2</sup>	0,98	0,99
	CER (%)*	0 % ± 0	0 % ± 0
p-value*	0,001	0,001	
Test de permutation	Graphique		
		Permutation test statistics	Permutation test statistics
VIP	Graphique des 30 premiers VIP		
		VIP scores	VIP scores

\* Valeurs obtenues avec le logiciel R ; CER = Classification Error Rate

Comme nous pouvions nous y attendre au vu des ACP, les variétés sont parfaitement distinctes sur les graphiques des individus des PLS-DA. La variabilité est majoritairement portée par l'axe numéro 1 avec approximativement 23 % de la variabilité totale exprimée contre 6,1 % et 4,1 % pour respectivement les extraits de feuilles et d'inflorescences sur la composante numéro deux. De plus, le pourcentage de variabilité par la première composante est élevé, ce qui est intéressant puisque la séparation des groupes biologiques s'effectue selon cet axe.

Les  $R^2$  et  $Q^2$  sont très élevés avec 0,98 et 0,99 et valident donc les deux PLS-DA. De plus, les deux histogrammes des tests de permutations écartent tout risque de sur-apprentissage du modèle. Les résultats générés par les permutations sont tous inférieurs et même éloignés des résultats des PLS-DA pour les extraits de feuilles et d'inflorescences montrant que les modèles obtenus ne sont pas le fruit du hasard.

### I.A.2. Le lavandin

Les résultats des analyses statistiques obtenus avec MetaboAnalyst pour les extraits de lavandin (feuilles et fleurs) sont regroupés ci-dessous dans le Tableau 30. Comme pour la lavande, les variétés dites sensibles sont représentées en rouge, les variétés tolérantes en vert et les pools en bleu.

Les premiers graphiques correspondent toujours aux graphiques des individus des ACP. Pour les extraits d'inflorescences les groupes sont parfaitement distincts, la variabilité cumulée est de 38 % avec une composante n°1 forte à 33,3 %. En revanche pour le graphique des parties chlorophylliennes, les deux groupes se chevauchent partiellement. De plus, la variabilité cumulée est assez basse avec 21,5 %, et contrairement aux extraits de fleurs la première composante ne se démarque pas avec seulement 13,9 % de variabilité portée.

Les graphiques suivants sont les graphiques des individus des PLS-DA. Pour les deux organes de la plante les deux variétés sont parfaitement distinctes. Comme pour les ACP, le graphique des inflorescences a une variabilité cumulée bien supérieure avec plus de 40%, contre 26,3 % pour les parties chlorophylliennes. Les deux PLS-DA sont valides, avec des valeurs de  $R^2$  et  $Q^2$  de 0,95 et 0,99. De plus, les tests de permutations ne montrent pas de sur-apprentissage du modèle.

Les derniers graphiques représentent les 30 composés les plus discriminants obtenus. Les scores des VIP obtenus sont élevés pour les deux organes, et plusieurs composés sont caractéristiques pour chacun des caractères sensible et tolérant.

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 30 : Analyses statistiques (ACP et PLS-DA) obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite de lavandin

Lavandin		Feuille	Inflorescence
ACP	Graphique des individus		
	Graphique des individus		
PLS-DA	$R^2$	0,95	0,99
	$Q^2$	0,95	0,99
	CER (%)*	0,2 % ± 0,04	0 % ± 0
	p-value*	0,001	0,001
Test de permutation	Graphique		
	Graphique des 30 premiers VIP		

\* Valeurs obtenues avec le logiciel R ; CER = Classification Error Rate

Que ce soit pour la lavande ou pour le lavandin des signatures chimiques distinctes ont pu être mises en évidence en fonction du caractère sensible et tolérant des plants. Les résultats obtenus sont tous valides. L'étude des composés discriminants cibles est présentée ci-dessous.

### I.B. Composés discriminants cibles

L'ensemble des 50 premiers composés discriminants des quatre jeux de données présentés précédemment, figurent dans l'Annexe 13. Dans cette partie, seront exposés uniquement les composés discriminants cibles (voir définition dans la partie Priorisation des biomarqueurs du chapitre 3) en fonction du caractère sensible ou tolérant des variétés pour les extraits des parties chlorophylliennes (VarF). En effet, aucun doublon n'a pu être mis en évidence pour les échantillons d'inflorescences en fonction de la sensibilité des variétés au dépérissement.

Pour les parties chlorophylliennes, nous avons six composés discriminants communs qui ont été relevés dans les deux matrices (*i.e.* lavande feuille et lavandin feuille). Ces ions sont regroupés dans le Tableau 31 ci-dessous en fonction de leur classement. Le caractère discriminant de chaque ion est indiqué dans le tableau et correspond au caractère (sensible ou tolérant) pour lequel l'ion en question est sur-exprimé. Les boîtes à moustaches des ions cibles sont également présentées dans la Figure 95.

Tableau 31 : Composés discriminants en fonction du caractère sensible ou tolérant des variétés communs aux jeux de données correspondant aux extraits de feuille

VIP	Note	Classement	Sensible/Tolérant
M155T341	2	20	Sensible
M133T571	2	32	Sensible
M167T274	2	47	Sensible
M155T407	2	51	Sensible
M361T283	2	63	!
M369T768	2	68	Tolérant

! = discordance sur le caractère discriminant du composé entre les matrices

Parmi ces six ions cibles, quatre se sont avérés caractéristiques des variétés dites « sensibles », à savoir 7713 et abrial. Les boîtes à moustaches de ces quatre ions ont été extraites pour la matrice des feuilles de lavande et sont présentées dans la Figure 95. Pour chaque ion la différence entre la variété 7713 et maillette est marquée, et les ions sont sur-exprimés par la variété 7713. Parmi les ions cibles, seul l'ion M369T768 est caractéristique des variétés dites « tolérantes » (maillette et grosso), comme l'illustre sa boîte à moustache pour les variétés de lavandin. En revanche, le caractère sensible ou tolérant de l'ion M361T283 n'a pu être établi. En effet, cet ion est sur-exprimé par la variété de lavande 7713 qui est considérée comme sensible ; et sur-exprimé par la variété grosso qui est une variété de lavandin tolérante au dépérissement. Ces deux résultats sont en contradiction et ne permettent pas de définir le caractère discriminant de cet ion.

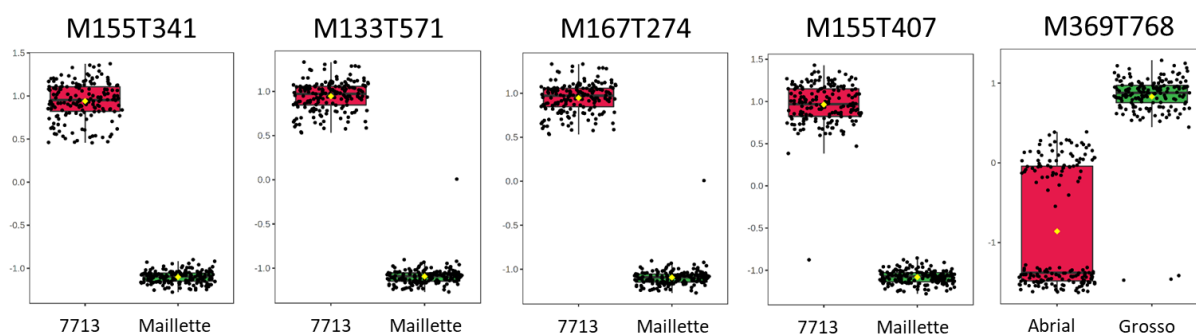


Figure 95 : Boîtes à moustaches des composés discriminants cibles en fonction du caractère sensible (en rouge) ou tolérant (en vert) des variétés

### Proposition de structure :

Pour plusieurs de ces composés cibles des correspondances ont été obtenues lors de la comparaison des masses expérimentales avec celles des banques de données réalisées sur le genre *Lavandula*. Ces hypothèses sont présentées dans le Tableau 32 et sont uniquement une base de travail, qu'il est nécessaire d'étudier plus en détail. Par exemple, les deux ions de rapport masse sur charge de 155,107 u ne peuvent tous les deux correspondre à l'acide protocatéchique. En revanche, ces deux composés peuvent avoir des schémas de fragmentation similaires et potentiellement appartenir à une même famille chimique. En effet, l'exploitation des spectres de masses de ces deux composés a révélé qu'un autre fragment à 133,0968 u était commun à ces deux molécules. En revanche, c'est le seul autre fragment commun aux deux VIP. Quant à la correspondance obtenue grâce à la banque de données, la fragmentation des deux ions ne correspond pas à celle relevée dans la littérature pour l'acide protocatéchique (Contreras et al. 2018). Aucune autre proposition de structure n'a pu être émise pour ces deux composés.

Les deux premiers ions M155T341 et M133T571, ont obtenu des valeurs de classement faibles ce qui fait d'eux deux marqueurs forts. L'ion M133T571 correspond potentiellement à la molécule de p-cymène ( $C_{10}H_{12}$ ) qui a été identifiée dans les composés volatils émis par les racines du lavandin (Stierlin 2020). Néanmoins, cette molécule étant volatile, il semble plus raisonnable de penser que l'ion M133T571 est un fragment d'une molécule de masse plus importante. Le spectre de masse de ce composé a révélé la présence d'un ion de rapport m/z 203,1802 u dont la formule brute correspondante est  $C_{15}H_{23}^+$ . La formule brute  $C_{15}H_{22}$  correspond à quatre molécules déjà répertoriées dans le genre *Lavandula*, à savoir : le cuparène (Sudria et al. 1999), l' $\alpha$ -curcumène (Guillén et al. 1996), les *cis*- et *trans*-calamenène (de Pascual-T et al. 1983). Les informations dont nous disposons à l'heure actuelle, ne nous permettent pas de trancher entre ces quatre hypothèses.

En revanche, une proposition de structure a pu être émise pour l'ion M361T283. L'exploitation de son spectre de masse montre un ion moléculaire  $[M+H]^+$  à 361,0923 u. Aucun adduit n'a pu être mis en évidence néanmoins le dimère  $[2M+H]^+$  à 721,1742 u a pu être observé. La première formule proposée pour l'ion moléculaire est  $C_{18}H_{17}O_8^+$  qui possède une correspondance parfaite (0 ppm) ainsi qu'une bonne corrélation entre le massif isotopique théorique et l'expérimental. La formule brute  $C_{18}H_{16}O_8$  a une correspondance dans la banque des composés organiques non volatils réalisée sur le genre *Lavandula* à savoir l'acide rosmarinique. Le spectre MSe expérimental a été comparé à celui de l'acide rosmarinique dans la banque MoNA (voir Figure 96) qui a été obtenu avec des paramètres expérimentaux proches (ESI-QToF de chez Waters en mode en positif, voltage du capillaire 3,0 kV, tension du cône d'échantillonnage 23 V et température de la source 120°C) (Matsuda). Les spectres présentent des similitudes importantes venant conforter cette hypothèse de structure.

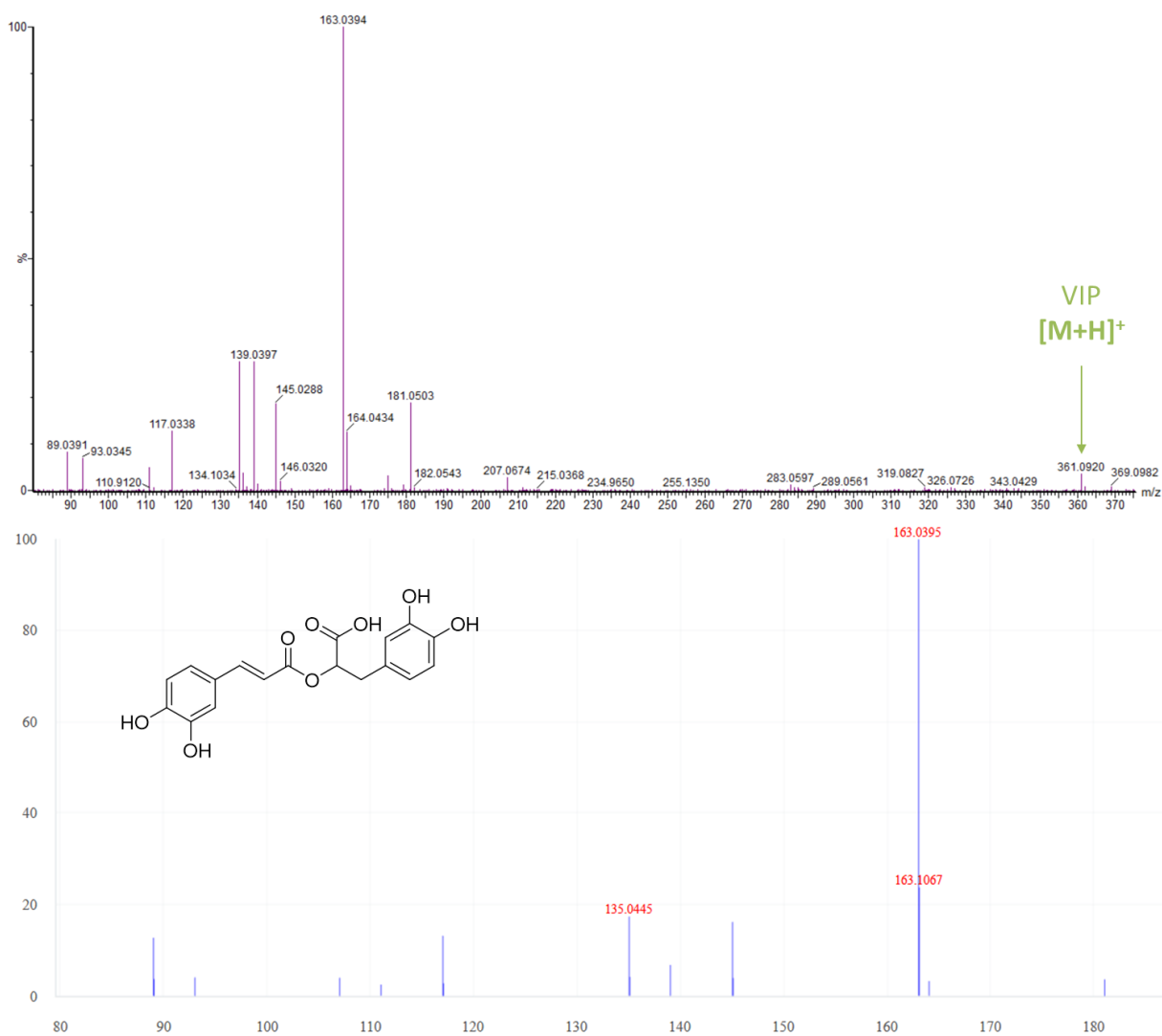


Figure 96 : Comparaison entre le spectre MSe expérimental obtenu pour l'ion M361T283, en haut, et le spectre MS/MS de l'acide rosmarinique (Matsuda), en bas.

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 32 : Hypothèses d'identification pour les composés cibles discriminants du caractère sensible ou tolérant des plants contenus dans les extraits des parties chlorophylliennes

VIP	Rt (min)	[M+H] <sup>+</sup> m/z expérimental	Formule brute [M]	Δm (ppm) expérimental et théorique	m/z fragments	m/z adduits	Hypothèse identification	Banques <i>Lavandula</i>
<b>M155T341</b>	5,679	155,1073*	C <sub>9</sub> H <sub>14</sub> O <sub>2</sub>	0,6	137,0965	177,0555[M+Na] <sup>+</sup>		Protocatechuic acid
<b>M133T571</b>	9,528	203,1802	C <sub>15</sub> H <sub>22</sub>	1,0	133,1019 105,0698	x	Cuparène, α-Curcumène, <i>cis</i> - ou <i>trans</i> -Calamenène	p-Cymenene
<b>M167T274</b>	4,573	167,071	C <sub>9</sub> H <sub>10</sub> O <sub>3</sub>	1,2				/
<b>M155T407</b>	6,792	155,1071	C <sub>9</sub> H <sub>14</sub> O <sub>2</sub>	-0,6				Protocatechuic acid
<b>M361T283</b>	4,749	361,0923	C <sub>18</sub> H <sub>16</sub> O <sub>8</sub>	0	181,0503 163,0394	721,1742[2M+H] <sup>+</sup>	Acide rosmarinique	3.4-Dihydroxyphenyllactic acid hexoside
<b>M369T768</b>	12,792	369,2407	C <sub>24</sub> H <sub>32</sub> O <sub>3</sub>	1,1		x		/

\* = hypothèse forte de l'ion moléculaire

/ = aucune correspondance avec les deux banques du genre *Lavandula*

A noter, qu'aucun doublon n'existe entre ces composés cibles en fonction de la tolérance des variétés et les autres listes de composés cibles à savoir celles correspondant à l'étude de l'état symptomatique des plants, qui est présentée ci-dessous.



## II. Etude de l'influence de l'état sanitaire des plants : caractère symptomatique ou asymptomatique

Du fait de l'importance de la variété sur la composition chimique des extraits, l'étude sur le caractère symptomatique des plants a été effectuée sur les variétés et les organes séparément. Les résultats obtenus sont présentés ci-dessous par variété. Pour chaque variété, les résultats des analyses statistiques sont présentés dans un tableau avec à gauche les analyses des extraits des parties chlorophylliennes (« feuilles ») ; et à droite, celles correspondant aux extraits d'inflorescence.

Dans un second temps, nous verrons les composés cibles discriminants en fonction de l'état sanitaire des plants, obtenus pour les extraits des parties chlorophylliennes et des inflorescences séparément, puis les composés communs.

### II.A. Résultats des analyses statistiques

#### II.A.1. Lavande

##### a. Variété 7713

7713 est une variété de lavande considérée comme sensible au dépérissement du Stolbur. La première analyse réalisée sur les deux jeux de données (feuilles et inflorescence) de cette variété est une ACP. Les graphiques des individus sont présentés en premier dans le Tableau 33 ci-dessous. Pour les extraits de feuilles, une distinction est visible entre les deux groupes symptomatique et asymptomatique. En revanche, il n'y a pas du tout de séparation dans le cas des inflorescences, même si les deux groupes sont visibles. La variabilité cumulée sur les deux premiers axes est de 23,1 % pour les parties chlorophylliennes et 23,8 % pour les inflorescences. Les analyses des pools sont bien regroupées dans les deux cas, ce qui prouve la bonne répétabilité des mesures et la bonne rectification de la dérive analytique grâce aux corrections apportées.

Plusieurs échantillons appartenant à la fois aux deux groupes biologiques (symptomatique et asymptomatique) et aux deux organes de la plante sont apparus comme isolés sur les ACP. L'explication de cet isolement est exposée ci-dessous.

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 33 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavande 7713

7713		Feuille	Inflorescence
ACP	Graphique des individus		
	Graphique des individus		
	R <sup>2</sup>	0,94	0,90
	Q <sup>2</sup>	0,92	0,87
	CER (%)*	0 % ± 0	0,45 % ± 0,1
	p-value*	0,001	0,001
Test de permutation	Graphique		
VIP	Graphique des 30 premiers VIP		

\* Valeurs obtenues avec le logiciel R ; CER = Classification Error Rate

### Echantillons isolés : cas de la lyophilisation n°5

Après la suppression des échantillons aberrants sur les deux matrices (feuille et inflorescence) en fonction de l'état symptomatique pour la variété de lavande 7713, de nombreux extraits sont apparus comme aberrants (voir Figure 97). Nous nous sommes rendu compte que tous ces extraits appartenaient à la même lyophilisation (n°5) au cours de laquelle un incident de pression est survenu, contraignant à arrêter et relancer la lyophilisation. Il est difficilement envisageable de supprimer une telle quantité d'analyses. En effet, pour la matrice des extraits de feuille cela représente la suppression de 17 extraits soit 9,6 % du jeu de données et pour la matrice des extraits d'inflorescence cela représente 35 extraits soit 19,4 % du jeu de données.

La conséquence sur les résultats des analyses statistiques de la suppression des extraits sur ces deux jeux de données a été testée. Les résultats des tests de permutations des deux jeux de données sont très mauvais après la suppression des échantillons appartenant à la lyophilisation n°5. La suppression a engendré un sur-apprentissage du modèle de la PLS-DA, certainement à cause du nombre d'extraits devenu insuffisant face au nombre de variable (resté constant). L'influence de la suppression des extraits sur les VIP est faible. La suppression a surtout engendré des différences dans le classement des VIP puisque la quasi-totalité des 30 premiers VIP avant suppression se retrouvent dans les 50 premiers VIP après suppression.

Finalement, nous avons fait le choix de ne pas supprimer les échantillons de la lyophilisation n°5. Premièrement, du fait du nombre élevé d'analyses concernées. De plus, les analyses correspondant à cette lyophilisation sont dissociées des autres sur la deuxième composante. Or, la discrimination entre les plants symptomatique et asymptomatique s'effectue sur la composante n°1, c'est donc cet axe qui nous intéresse. De plus, sur cette composante, les extraits des échantillons de la lyophilisation n°5 sont bien alignés avec le reste des échantillons. De surcroit, la comparaison entre les premiers VIP a montré que la majorité des composés les plus discriminants est inchangée avec ou sans la suppression.

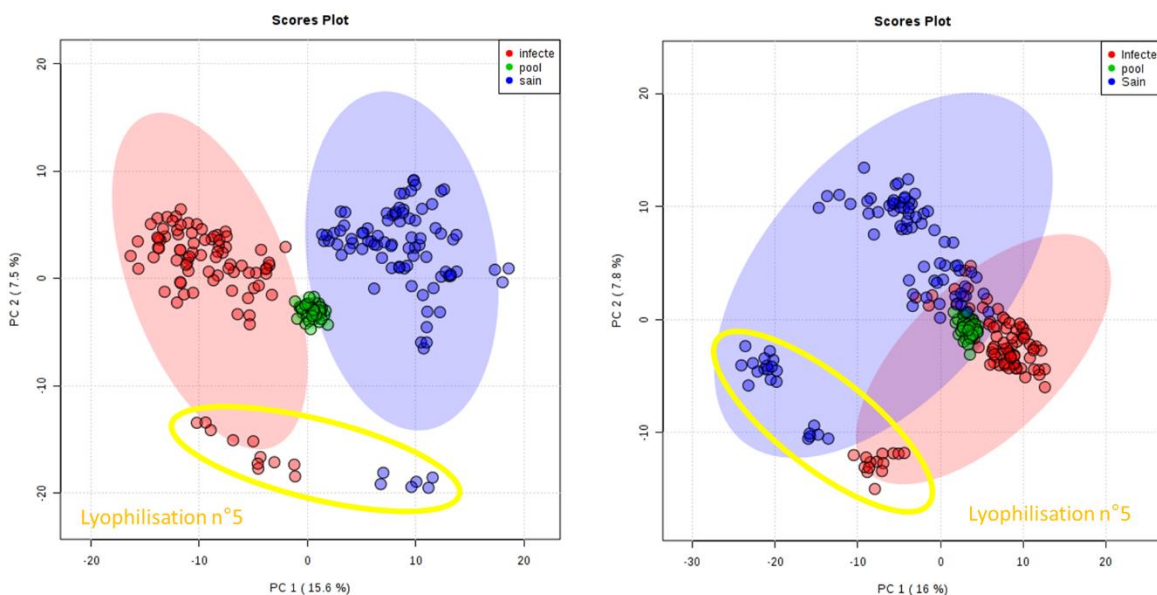


Figure 97 : Mise en évidence des analyses correspondant aux échantillons lyophilisés pendant le cycle n°5, sur les ACP de la matrice de feuille (à gauche) et d'inflorescence (à droite) de lavande 7713 en fonction de l'état symptomatique du plant après suppression des autres échantillons aberrants

### PLS-DA :

Que ce soit pour les extraits des parties chlorophylliennes ou pour les inflorescences, les groupes symptomatiques et asymptomatiques sont distincts sur les graphiques des individus des PLS-DA (voir Tableau 33 ci-dessus). A noter que la séparation est plus marquée pour les extraits de « feuilles » et survient principalement sur la première composante qui porte 17,5 % de variabilité. La variabilité totale est respectivement de 23,4 % et 25,7 % pour les extraits de feuilles et de fleurs.

La séparation des analyses liées aux échantillons appartenant à la lyophilisation n°5 est moins nette sur les PLS-DA que pour les ACP. Cependant, dans le cas des extraits asymptomatiques des inflorescences, un groupe, situé en bas à droite du graphique, est clairement disjoint. Cette coupure intervient surtout sur la composante numéro 2, à l'inverse de la séparation entre les deux groupes biologiques qui est principalement portée par l'axe numéro 1. Il est probable que la séparation entre les groupes plus marquée pour les extraits de feuilles soit imputable à la présence d'un nombre inférieur d'échantillons appartenant à la lyophilisation n°5.

Dans les deux cas, les PLS-DA sont valides avec des valeurs de  $R^2$  de 0,94 et 0,90, et des valeurs de  $Q^2$  de 0,92 et de 0,87, pour respectivement la matrice des parties chlorophylliennes et des inflorescences.

Les tests de permutations laissent penser qu'il n'y a pas de sur-apprentissage du modèle, même si plusieurs permutations ont obtenu des résultats supérieurs. Leur nombre reste néanmoins inférieur à 5% (23 et 34 permutations sur 1000) sur la totalité des permutations effectuées. Les résultats obtenus pour l'étude en fonction de l'état sanitaire des plants sont en-dessous de ceux obtenus pour le caractère sensible/tolérant des plants. Ce phénomène est certainement imputable à la réduction de moitié des jeux de données liée à leur segmentation par variété (dans le cas de l'étude sur l'état sanitaire).

Les derniers graphiques présentés dans le Tableau 33 sont les 30 ions les plus discriminants avec leur sur- ou sous-expression en fonction du groupe biologique. A noter que des ions caractéristiques de chacun des deux états sanitaires ont été mis en évidence aussi bien dans les parties chlorophylliennes que dans les inflorescences. Les tableaux complets des 50 ions les plus discriminants sont présents en Annexe 13.

### b. Variété maillette

La variété de lavande maillette a été choisie pour sa tolérance face au phytoplasme. En premier lieu ont été réalisées des ACP (voir Tableau 34 ci-dessous). Aucune séparation entre les groupes n'apparaît pour aucun des organes de la plante sur les ACP. Quant aux analyses de pools, elles sont bien différenciées du reste des extraits et forment un groupe compact témoignant d'une bonne répétabilité des analyses. Il n'est pas surprenant que les pools ne soient pas centrés, comme dans l'étude des caractères sensible et tolérant, puisqu'ils contiennent également les extraits de la variété 7713. Les variabilités cumulées des deux premières composantes atteignent 25,7 % et 20,6 %. Il est intéressant de relever que la variabilité chimique semble plus importante pour les échantillons symptomatiques et pour les extraits des parties chlorophylliennes (attention aux échelles des graphiques qui ne sont pas équivalentes et qui pourraient laisser penser le contraire de prime abord).

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 34 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavande maillette

Maillette		Feuille	Inflorescence
ACP	Graphique des individus		
	Graphique des individus		
PLS-DA	R <sup>2</sup>	<b>0,85</b>	<b>0,81</b>
	Q <sup>2</sup>	<b>0,79</b>	<b>0,73</b>
	CER (%)*	<b>0 % ± 0</b>	<b>5,6 % ± 0,3</b>
	p-value*	<b>0,001</b>	<b>0,001</b>
Test de permutation	Graphique		
	Graphique des 30 premiers VIP		

\* Valeurs obtenues avec le logiciel R ; CER = Classification Error Rate

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

La réalisation des PLS-DA a permis de séparer le groupe symptomatique des asymptotiques pour les extraits de feuilles (voir graphique des individus dans le Tableau 34 ci-dessus). La distinction entre les deux groupes biologiques est beaucoup moins évidente dans le cas des inflorescences avec un chevauchement des ellipses dû à quelques extraits symptomatiques. Les pourcentages de variabilité cumulée pour les deux axes, sont assez faibles avec un total de 24 % pour les extraits de feuilles et de 21,2 % pour les inflorescences. Les analyses sont néanmoins valides avec des valeurs de  $R^2$  et  $Q^2$  de 0,85 et 0,79 pour les parties chlorophylliennes et 0,81 et 0,73 pour les fleurs. Ces chiffres sont cependant en-deçà de ceux obtenus jusqu'à présent. Ce phénomène est probablement lié à la diminution du nombre d'échantillons dans les jeux de données. En revanche, pour les tests de permutation, nous avons clairement un sur-apprentissage du modèle dans le cas des inflorescences avec près de 20 % des permutations qui obtiennent un résultat égal ou supérieur à la PLS-DA d'origine. Pour les extraits de feuilles le nombre d'erreur reste inférieur à 5 % (26 permutations sur 1000), ce qui laisse penser qu'il n'y a pas de sur-apprentissage.

Comme précédemment, les derniers graphiques présentés dans le Tableau 34 sont les 30 ions les plus discriminants. Des ions caractéristiques des deux états sanitaires ont été mis en évidence pour chacun des organes de la plante. Les tableaux complets des 50 ions les plus discriminants sont présents en Annexe 13.

### II.A.2. Lavandin

#### a. Variété abrial

La variété abrial est une variété de lavandin caractérisée comme sensible au dépérissement du Stolbur. Les premières analyses statistiques faites sur les jeux de données de cette variété, sont les ACP (voir Tableau 35 ci-dessous). Les graphiques des individus montrent un chevauchement des groupes biologiques. Les données sont de manière générale très dispersées, hormis pour les analyses de pools qui forment, pour les deux organes, un groupe compact gage de la qualité des analyses. La variabilité totale portée par les deux premières composantes est légèrement supérieure dans le cas des parties chlorophylliennes avec 26,5 %, pour 23,4 % pour les inflorescences.

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 35 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavandin abrial

Abrial		Feuille	Inflorescence
ACP	Graphique des individus		
	Graphique des individus		
PLS-DA	R <sup>2</sup>	<b>0,86</b>	<b>0,80</b>
	Q <sup>2</sup>	<b>0,83</b>	<b>0,75</b>
	CER (%)*	<b>0,9 % ± 0,2</b>	<b>2,7 % ± 0,2</b>
	p-value*	<b>0,001</b>	<b>0,001</b>
Test de permutation	Graphique		
	Graphique des 30 premiers VIP		

\* Valeurs obtenues avec le logiciel R ; CER = Classification Error Rate

La réalisation des PLS-DA a permis de différencier les groupes symptomatique et asymptomatique dans le cas des extraits de feuilles. De plus, l'analyse est validée par des valeurs de  $R^2$  et  $Q^2$  de respectivement 0,86 et 0,83. Le test de permutation correspondant permet d'éliminer le risque de sur-apprentissage du modèle. Pour les extraits d'inflorescences, même s'il y a une tendance, les deux groupes ne sont pas clairement distincts. La variabilité cumulée est acceptable et s'élève à 26,2%. La variabilité est répartie de manière à peu près égale entre les deux axes. Néanmoins, c'est la deuxième composante qui porte le plus de variabilité avec 14,1 %, or la « séparation » des deux groupes aurait tendance à être portée par l'axe numéro un. En revanche, la PLS-DA est valide avec des valeurs de  $R^2$  et  $Q^2$  respectivement de 0,80 et 0,75. Les résultats du test de permutation montrent que 6 permutations sur 1000 ont eu des résultats supérieurs à la PLS-DA initiale, ce qui est négligeable.

A noter que les tableaux des 50 ions les plus discriminants pour cette variété sont présentés en Annexe 13.

### b. Variété grosso

La variété de lavandin grosso a été choisie pour réaliser cette étude du fait de sa tolérance au phytoplasme du Stolbur. Les ACP réalisées sur les deux organes ne montrent aucune séparation des groupes biologiques (voir Tableau 36 ci-dessous). Pour les extraits d'inflorescences, les échantillons symptomatiques paraissent plus épars que les asymptomatiques. Cette constatation ne se retrouve pas sur les extraits de feuilles. Il est à nouveau normal que les analyses du pool ne soient pas centrées car il a été constitué avec l'ensemble des extraits de lavandin contenant à la fois les extraits de grosso et ceux d'abrial. Les données sont néanmoins bien regroupées ce qui est un gage de la qualité des analyses. La variabilité cumulée est plus élevée pour les extraits d'inflorescences avec 24,2 % alors que les extraits des parties chlorophylliennes ont un pourcentage assez faible de 18,3 %. Dans le cas des inflorescences, les échantillons symptomatiques paraissent contenir une plus grande variabilité chimique.



## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 36 : Analyses statistiques (ACP et PLS-DA) en fonction de l'état symptomatique des plants obtenues avec MetaboAnalyst pour les extraits de feuille à gauche et d'inflorescence à droite pour la variété de lavandin grosso

Grosso		Feuille	Inflorescence
ACP	Graphique des individus		
	Graphique des individus		
PLS-DA	R <sup>2</sup>	<b>0,83</b>	<b>0,85</b>
	Q <sup>2</sup>	<b>0,68</b>	<b>0,74</b>
	CER (%)*	<b>7,25 % ± 0,43</b>	<b>1,9 % ± 0,25</b>
	p-value*	<b>0,001</b>	<b>0,001</b>
Test de permutation	Graphique		
	Graphique des 30 premiers VIP		

\* Valeurs obtenues avec le logiciel R ; CER = Classification Error Rate

## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Pour les analyses discriminantes (voir graphique des individus des PL-SDA du Tableau 36 ci-dessus), les ellipses des groupes biologique symptomatique et asymptotique ne sont pas distinctes, ni se chevauchent vraiment ; ce qui rend l'interprétation de la séparation des deux groupes tendancieuse. De plus, les pourcentages de variabilité cumulée sont faibles avec 14,9 % pour la matrice des feuilles et 15,8 % pour les inflorescences. Pour le jeu de données des parties chlorophylliennes, les valeurs de  $R^2$  et  $Q^2$  sont respectivement de 0,83 et 0,68 et valident la PLS-DA, néanmoins la valeur de  $Q^2$  est assez basse et est donc à nuancer. De plus, aucune permutation du test n'a obtenu un score supérieur à la PLS-DA initiale montrant l'absence de sur-apprentissage du modèle. Dans le cas des extraits d'inflorescences, la PLS-DA est validée par les valeurs de  $R^2$  et  $Q^2$  de 0,85 et 0,74 respectivement. En revanche, nous avons un sur-apprentissage du modèle avec un très grand nombre de permutations qui ont obtenu un score supérieur. Les résultats contrastés obtenus pour la variété *grosso* sont potentiellement imputables à la suppression des analyses aberrantes lors du nettoyage des données. En effet, cette catégorie est celle qui a subi le plus de corrections avec notamment neuf extraits appartenant à un même plant qui ont été supprimés des deux jeux de données (feuilles et inflorescences).

La liste des 50 ions les plus discriminants est présentée en Annexe 13. A noter que pour la matrice d'inflorescences (voir graphique des 30 premiers VIP correspondant dans le Tableau 36 ci-dessus), tous les ions discriminants parmi les 30 premiers sont sur-exprimés par les plants symptomatiques. Quant aux composés discriminants cibles, ils sont présentés ci-dessous.

### II.B. Composés discriminants cibles

Dans cette partie seront présentés les composés discriminants cibles en fonction de l'état sanitaire des plants pour chaque organe de la plante séparément, c'est-à-dire les composés ayant obtenu un VIP score élevé et ayant des récurrences entre les listes des 50 composés les plus discriminants des matrices contenues dans chaque catégorie (voir partie IV.B.3 du chapitre 3). Dans un premier temps seront présentés les ions cibles correspondants aux extraits des parties chlorophylliennes, puis ceux des inflorescences. Enfin dans une dernière partie, nous verrons les composés discriminants cibles qui ont particulièrement attiré notre attention car communs aux deux catégories précédentes.

#### II.B.1. Etude sur les parties chlorophylliennes (Statut F)

##### a. Composés cibles en fonction de l'état symptomatique des plants pour les extraits des parties chlorophylliennes

Les parties chlorophylliennes des plants comportent 26 composés discriminants cibles, c'est-à-dire ayant été identifiés comme discriminants (et faisant parti des 50 premiers) dans au moins deux matrices. Ces ions sont regroupés dans le Tableau 37 ci-dessous en fonction de leur note puis de leur classement. Le caractère discriminant est également indiqué et correspond, ici, à l'état symptomatique sur-exprimé.

Tableau 37 : Composés discriminants en fonction de l'état symptomatique des plants commun aux jeux de données correspondant aux extraits de feuilles

VIP	Note VIP	Classement	Caractère discriminant
M593T819	4	49	Asymptomatique
M607T779	4	114	Asymptomatique
M179T486	3	35	Asymptomatique
M223T548	3	57	Symptomatique
M399T486	3	62	Asymptomatique
M193T253	3	68	Asymptomatique
M221T442	3	87	Symptomatique
M139T539	3	90	Symptomatique
M209T278	3	92	Asymptomatique
M635T836	3	104	Asymptomatique
M181T573	2	7	Asymptomatique
M167T549	2	8	Symptomatique
M85T31	2	12	Symptomatique
M163T361	2	14	Symptomatique
M126T31	2	19	Symptomatique
M169T512	2	27	Symptomatique
M219T748	2	42	Symptomatique
M453T615	2	45	!
M207T702	2	47	Symptomatique
M153T609	2	53	Symptomatique
M153T147	2	54	Symptomatique
M235T475	2	59	Asymptomatique
M151T31	2	75	Asymptomatique
M248T36	2	88	Asymptomatique
M397T443	2	90	Symptomatique
M221T578	2	95	Symptomatique

! = discordance sur le caractère discriminant du composé entre les matrices

Parmi ces 26 ions cibles, 11 sont caractéristiques des plants asymptomatiques. La plupart font partie des ions qui ont obtenu le plus d'occurrence et les meilleurs classements. Nous avons en particulier les ions M593T819 et M607T779 qui sont marqueurs dans toutes les matrices des parties chlorophylliennes. Ces deux composés sont donc particulièrement intéressants pour spécifier le caractère asymptomatique des plants. L'ion M179T486 est également intéressant car il apparaît comme marqueur dans trois matrices sur les quatre et il a obtenu le meilleur classement pour cette récurrence (note VIP).

Parmi ces ions cibles, nous en avons 14 qui se sont révélés comme étant marqueur des symptômes du dépérissement. Même si aucun n'a pu être retrouvé dans l'ensemble des matrices, trois de ces ions sont marqueurs dans trois matrices sur quatre. Il s'agit des ions M223T548, M221T442 et M139T539

En revanche, le caractère discriminant de l'ion M453T615 n'a pu être défini puisqu'il est sur-exprimé par les plants symptomatiques de lavandin abrial et sur-exprimé par les plants asymptomatiques de grosso.



Figure 98 : Boîtes à moustaches des composés discriminants cibles en fonction de l'état sanitaire (symptomatique en rouge et asymptomatique en vert) des plants pour les parties chlorophylliennes

Nous pouvons constater que certains ions possèdent des temps de rétention similaires ou très proches. Nous pouvons notamment relever les ions suivants :

- M399T486 et M179T486
- M397T443 et M221T442
- M85T31, M126T31, M151T31 et M248T36

Il est envisageable que ces ions appartiennent à la même molécule. Par exemple, il est fort probable que l'ion M179T486 soit un fragment de l'ion M399T486. En effet, ils possèdent le même caractère discriminant et le même temps de rétention. A l'inverse, il serait surprenant que les quatre ions ayant un temps de rétention proche de 31s appartiennent tous au même analyte. De plus, ce temps de rétention correspond à une zone d'élution où de nombreux composés sont co-élus (voir Figure 99).

#### b. Proposition de structure

Afin de compléter la simple confrontation des masses expérimentales avec les deux banques de données (CONVs et COVs) dans le genre *Lavandula* dont le résultat est présenté dans le Tableau 38, nous avons cherché à émettre des hypothèses de structure pour les composés cibles à partir des spectres MS et, si cela était possible, des spectres MSe et UV.

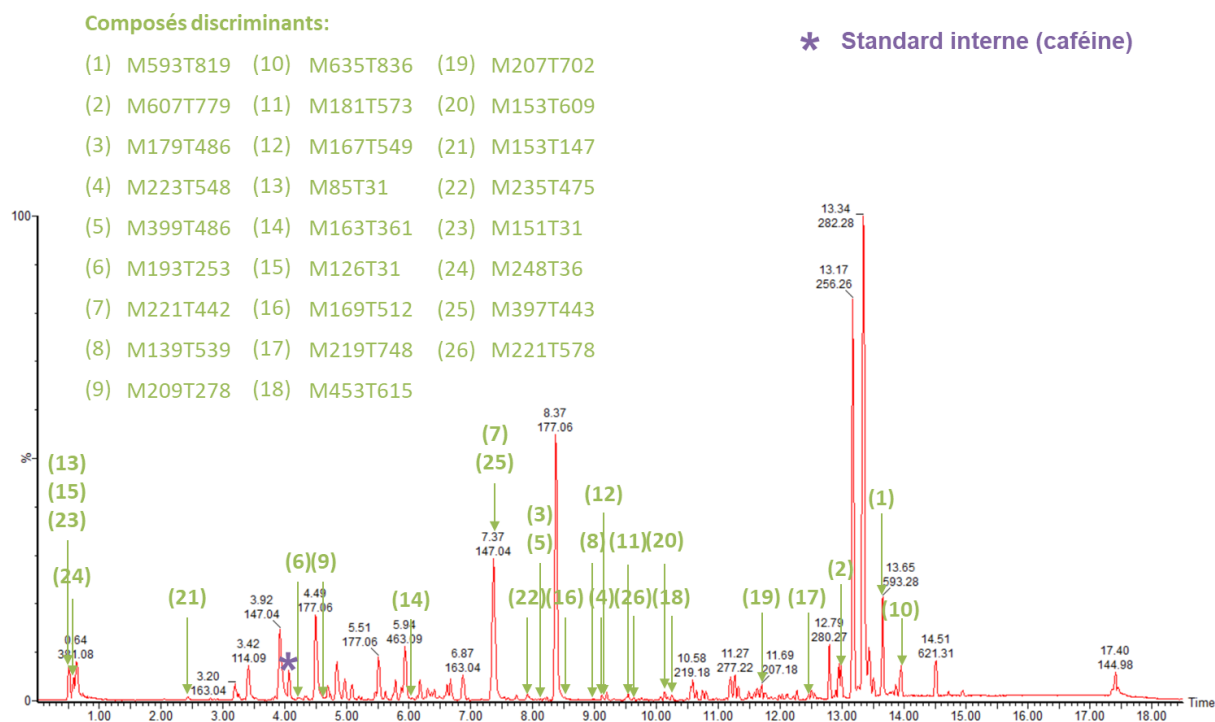


Figure 99 : Chromatogramme d'un extrait des organes chlorophylliens de lavandin annoté avec les composés cibles discriminants de l'état symptomatique des plants numérotés en fonction de leur pertinence (*i.e.* en fonction de leur note puis de leur classement)

Le premier composé discriminant cibles en fonction de l'état sanitaire des plants pour les extraits est l'ion M593T819. L'exploitation de son spectre de masse, montre un ion moléculaire avec un rapport  $m/z = 593,2756$  u. La première formule proposée est  $C_{34}H_{37}N_4O_5^+$  (-0,5 ppm) et possède une bonne corrélation entre le massif isotopique théorique et l'expérimental. La formule brute  $C_{34}H_{36}N_4O_5$  a été employée pour réaliser une recherche dans plusieurs banques de données. En premier lieu, des banques de données spécifiques telles que Knapsack pour les composés d'origine végétale, et la banque HMDB, ont été utilisées et n'ont pas fourni de résultat. Des banques généralistes telles que Reaxys ou Pubchem ont ensuite été testées. Parmi les nombreuses propositions de structure, celle du phéophorbide A (ou pheaphorbide A) a attiré notre attention car c'est un dérivé de chlorophylle (Hörtensteiner and Kräutler 2011). En effet, l'ion étudié absorbe en UV (voir Figure 100), avec un spectre spécifique comportant plusieurs bandes d'absorption (408, 503, 534, 605, 664 nm) qui se rapproche du spectre de la chlorophylle A. Après vérification dans la littérature, le spectre UV expérimental correspond à celui décrit pour le phéophorbide A (Mohn et al. 2009; Chen et al. 2015). De plus, la MSe révèle la présence d'un fragment à 533,2543 u, potentiellement attribuable à la perte d'un groupement -  $CH_3COOH$ , et cette fragmentation correspond avec celle trouvée dans la littérature (Juin et al. 2015; Chen et al. 2015). Les phéophorbides sont des dérivés naturels issus de la dégradation des chlorophylles et sont donc très répandus dans le règne végétal (Chen et al. 2015). Ils n'ont pas fait l'objet d'une étude particulière dans le genre *Lavandula*.

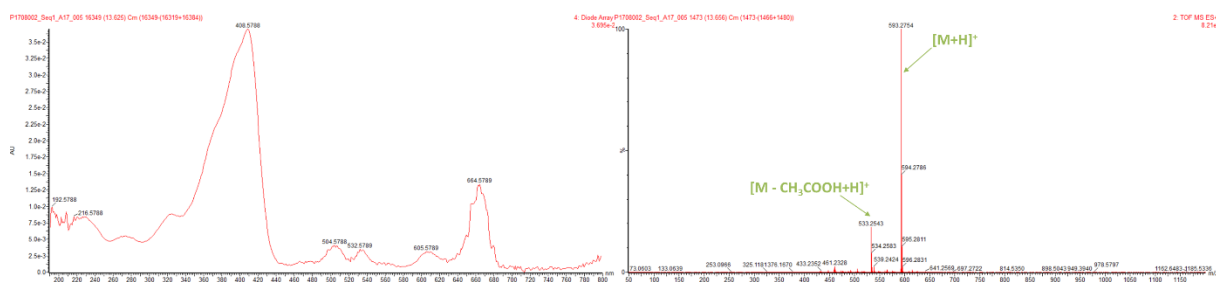


Figure 100 : Spectres UV nettoyés et MS de l'ion M593T819 dont la proposition de structure est celle du phéophorbide A

Le deuxième composé discriminant cible le plus intéressant pour les extraits de feuille est le VIP M607T779. Le spectre de masse montre la présence d'un ion moléculaire avec un rapport  $m/z$  de 607,2555 u. La première formule brute proposée pour l'ion est  $C_{35}H_{35}N_4O_6^+$  (-0,3 ppm). La formule brute ( $C_{35}H_{34}N_4O_6$ ) a été soumise aux banques de données HMDB, DNP et Knapsack, une seule structure potentielle a été obtenue, celle du phéophorbide B. Au vu de la proximité des temps de rétention (40 s) avec le composé potentiellement identifié comme le phéophorbide A, la proposition de structure est cohérente. De plus, le spectre MS a mis en évidence trois fragments de rapport  $m/z$  579,2595 u; 547,2325 u et 519,2375 u attribuables respectivement à des pertes de CO,  $CH_3COOH$  puis CO qui ont été décrits dans la littérature comme les fragments principaux du phéophorbide B (Chen et al. 2015). Le spectre UV vient renforcer cette hypothèse avec la présence de deux bandes d'absorption à 434 et 653 nm qui ont été décrites dans la littérature comme les deux longueurs d'onde d'absorption pour ce composé (Chen et al. 2015).

Un troisième VIP M635T836, moins pertinent (en 10ème position), a également fait l'objet d'une proposition de structure. Comme nous le montre la Figure 99 ; il a la particularité d'éluer dans la même zone que les deux ions décrits ci-dessus (classés en 1<sup>ère</sup> et 2<sup>ème</sup> positions). Les deux premières propositions de formule brute obtenues pour l'ion moléculaire sont  $C_{36}H_{43}O_{10}^+$  (0,8 ppm) et  $C_{37}H_{39}N_4O_6^+$  (-1,4 ppm). La première formule a une très mauvaise corrélation entre le massif isotopique théorique et expérimental, contrairement à la deuxième formule. La formule brute  $C_{37}H_{38}N_4O_6$  a été soumise à plusieurs banques de données (HMDB, DNP et Knapsack). Seule la banque DNP a fourni une seule proposition de structure, à savoir l'ester éthylique du phéophorbide B. Aucun spectre MS/MS n'a été trouvé dans la littérature pour ce composé. En revanche, la masse de l'ion  $[M+H]^+$  est en accord avec la littérature (Chee et al. 2005), ainsi que le spectre UV. Ce dernier possède un profil de type composé chlorophyllien et exhibe deux maxima à 433 et 650 nm. De plus, le temps de rétention du composé est cohérent avec la proposition de structure et les autres hypothèses d'identification formulées précédemment.

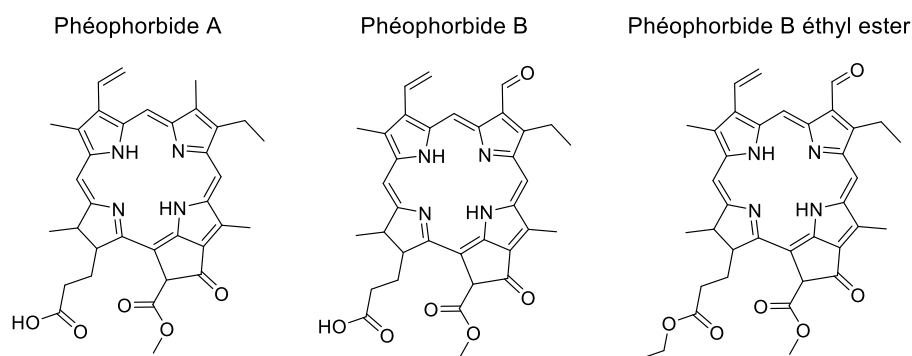
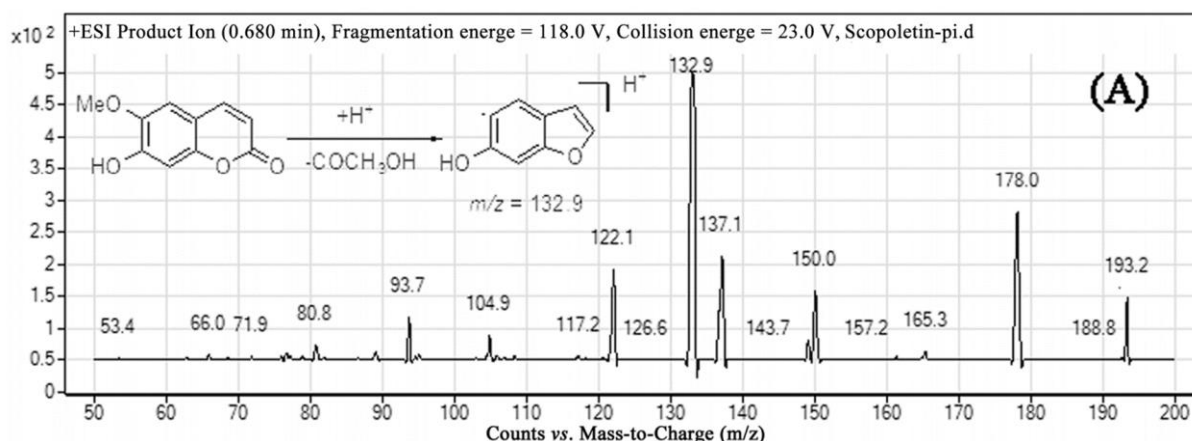


Figure 101: Structure des trois dérivés de chlorophylle potentiellement identifiés dans les extraits de feuilles

Nous avons également émis une hypothèse de structure pour l'ion M193T253 placé cette fois-ci en 6ème position. L'exploitation du spectre de masse a pu mettre en évidence la présence d'un adduit sodium  $[M+Na]^+$  à 215,0315 u et la présence d'un ion moléculaire  $[M+H]^+$  à 193,0501 u. Des fragments caractéristiques ont également été observés à 178,0251 u correspondant à la perte d'un groupement  $CH_3$  (perte de neutre de  $-15$  u) et à 150,0321 u correspondant à la déshydratation du fragment précédent (perte de neutre de  $-18$  u). Une corrélation parfaite a été obtenue avec la masse de l'ion  $[M+H]^+$  est la formule brute  $C_{10}H_9O_4^+$  (0 ppm). La formule brute  $C_{10}H_8O_4$  a alors été soumise à plusieurs banques de données (HMDB, DNP et Knapsack) et de nombreux résultats ont été obtenus. En excluant les structures ne comportant pas de méthoxy, les propositions restantes sont la scopolétine (7-hydroxy-6-methoxy-2H-chromen-2-one) une coumarine méthoxylée ou ses isomères de position. La fragmentation décrite dans la littérature (voir Figure 102) est proche de celle obtenue (Zeng et al. 2015).



Plusieurs coumarines et dérivés ont été reportés dans le genre *Lavandula* mais ce n'est pas le cas de la scopolétine. Pourtant, elle est fréquemment rencontrée dans le règne végétal et dans les différents organes (feuilles, fruits, graines, racines, ...). Il a été reporté dans la littérature que la production de scopolétine augmente chez les plantes infectées et que cette augmentation est corrélée avec la résistance à une attaque microbienne ou un autre stress subit par la plante tel qu'une coupure ou une déshydratation (Gnonlonfin et al. 2012). Cette remarque est en contradiction avec les résultats obtenus car ce composé a été relevé comme sur-exprimé dans les plants asymptomatiques. Par conséquent, il est possible que la bactérie arrive à inhiber les voies de défense de la plante.

Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 38 : Hypothèses d'identification pour les composés cibles contenus dans les extraits des parties chlorophylliennes

VIP	Rt (min)	[M+H] <sup>+</sup> m/z expérimental	Formule brute [M]	Δm (ppm) expérimental et théorique	m/z fragments	m/z adduits	Proposition structure	Banques <i>Lavandula</i>
<b>M593T819</b>	13,669	593,2756*	C <sub>35</sub> H <sub>36</sub> N <sub>4</sub> O <sub>5</sub>	-0,5	533,2543	615,2592 [M+Na] <sup>+</sup> 1185,5468[2M+H] <sup>+</sup>	Phéophorbide A	/
<b>M607T779</b>	12,98	607,2555*	C <sub>35</sub> H <sub>34</sub> N <sub>4</sub> O <sub>6</sub>	-0,3	579,2595; 547,2325; 519,2375	629,2399 [M+Na] <sup>+</sup>	Phéophorbide B	/
<b>M179T486</b>	8,09	197,1544	C <sub>12</sub> H <sub>20</sub> O <sub>2</sub>	1,0	179,1433; 161,1329	x		Hydrocinnamyl acetate
<b>M223T548</b>	9,125	241,1790*	C <sub>14</sub> H <sub>24</sub> O <sub>3</sub>	-5,8	223,1700; 205,1596; 177,1639; 147,1179	263,1621 [M+Na] <sup>+</sup>		Multi.
<b>M399T486</b>	8,09	399,1987	C <sub>16</sub> H <sub>26</sub> N <sub>6</sub> O <sub>6</sub>	-1,3		x		/
<b>M193T253</b>	4,227	193,0501*	C <sub>10</sub> H <sub>8</sub> O <sub>4</sub>	0,0	178,0251; 150,0321	215,0315 [M+Na] <sup>+</sup>	Dérivé scopolatin	Citric acid/isocitric acid
<b>M221T442</b>	7,345	249,1489	C <sub>15</sub> H <sub>20</sub> O <sub>3</sub>	-0,8	221,1907 ; 203,1801	x		Multi.
<b>M139T539</b>	8,976	139,0759	C <sub>8</sub> H <sub>10</sub> O <sub>2</sub>	0,0				Hydroxybenzoic acid isomer 1
<b>M209T278</b>	4,626	209,1543	C <sub>13</sub> H <sub>20</sub> O <sub>2</sub>	0,5	191,1442; 167,1433; 149,1347 ; 125,0961	x		Multi.
<b>M635T836</b>	13,932	635,2861*	C <sub>37</sub> H <sub>38</sub> N <sub>4</sub> O <sub>6</sub>	-1,4	607,2933; 575,2642; 547,2625	657,2699 [M+Na] <sup>+</sup>	Phéophorbide B éthyle ester	/
<b>M181T573</b>	9,546	181,1224*	C <sub>11</sub> H <sub>16</sub> O <sub>2</sub>	-3,3		203,1039 [M+Na] <sup>+</sup>		Caffeic acid



## Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

<b>M167T549</b>	9,152	167,1071*	C <sub>10</sub> H <sub>14</sub> O <sub>2</sub>	-0,6	121,1018	189,0879 [M+Na] <sup>+</sup>		/
<b>M85T31</b>	0,529	84,9602	X	X				/
<b>M163T361</b>	6,003	181,0504	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>	1,7	163,0398	X		/
<b>M126T31</b>	0,529	125,9869	X	X				/
<b>M169T512</b>	8,528	563,1391	C <sub>26</sub> H <sub>26</sub> O <sub>14</sub>	-1,8		X		Multi.
<b>M219T748</b>	12,458	219,1754	C <sub>15</sub> H <sub>22</sub> O	2,3				14-Oxy-a-murolene
<b>M453T615</b>	10,257	453,3366	C <sub>30</sub> H <sub>44</sub> O <sub>3</sub>	-0,7	407,3328			/
<b>M207T702</b>	11,704	207,1751	C <sub>14</sub> H <sub>22</sub> O	0,5				/
<b>M153T609</b>	10,152	153,1278	C <sub>10</sub> H <sub>16</sub> O	-0,7				Multi.
<b>M153T147</b>	2,451	153,0552	C <sub>8</sub> H <sub>8</sub> O <sub>3</sub>	0,0				Multi.
<b>M235T475</b>	7,924	235,1339	C <sub>14</sub> H <sub>18</sub> O <sub>3</sub>	2,1				/
<b>M151T31</b>	0,529	151,0356	C <sub>3</sub> H <sub>6</sub> N <sub>2</sub> O <sub>5</sub>	0,7				Multi.
<b>M248T36</b>	0,609	266,1253	C <sub>11</sub> H <sub>15</sub> N <sub>5</sub> O <sub>3</sub>	0	248,1146			/
<b>M397T443</b>	7,388	397,1468	C <sub>15</sub> H <sub>20</sub> N <sub>6</sub> O <sub>7</sub>	-1,0				/
<b>M221T578</b>	9,634	221,1913	C <sub>15</sub> H <sub>24</sub> O	3,6	203,1806	X		Multi.

\* = hypothèse forte de l'ion moléculaire

Multi. = correspondance multiple avec la banque des composés organiques volatils

/ = aucune correspondance avec les deux banques du genre *Lavandula*

II.B.2. Etude sur les inflorescences (StatutInflo)

a. Composés cibles en fonction de l'état symptomatique des plants pour les extraits d'inflorescences

Pour cette catégorie sur les extraits d'inflorescence, nous avons 20 composés discriminants communs. En revanche, aucun ion n'a pu être relevé dans les quatre matrices et seulement, un composé, l'ion M223T548 a été retrouvé dans trois matrices sur quatre. Les résultats sont présentés dans le Tableau 39 ci-dessous.

Tableau 39 : Composés discriminants en fonction de l'état symptomatique des plants commun aux jeux de données correspondant aux extraits d'inflorescence

VIP	Note VIP	Classement	Caractère discriminant
M223T548	3	66	Symptomatique
M169T587	2	6	Asymptomatique
M273T386	2	10	Symptomatique
M353T386	2	10	Symptomatique
M109T484	2	19	Asymptomatique
M593T819	2	32	Asymptomatique
M125T587	2	34	Asymptomatique
M107T584	2	35	Asymptomatique
M223T633	2	38	!
M111T587	2	41	Asymptomatique
M109T510	2	42	!
M207T702	2	47	Symptomatique
M277T676	2	48	Asymptomatique
M169T511	2	49	Symptomatique
M109T511	2	53	Symptomatique
M139T539	2	58	Symptomatique
M221T442	2	62	Symptomatique
M151T610	2	63	Asymptomatique
M463T357	2	66	Symptomatique
M262T571	2	74	Asymptomatique

! = discordance sur le caractère discriminant du composé entre les matrices

Pour les inflorescences, nous constatons que 9 ions sont caractéristiques de l'état symptomatique des plants. Parmi lesquels l'ion M223T548 qui est présent dans trois matrices sur quatre.

Nous avons également 9 ions caractéristiques de l'état asymptomatique des plants. Cependant, aucun n'a été retrouvé dans plus de deux matrices. L'ion M169T587 présente cependant un excellent classement, ce qui le rend intéressant vis-à-vis de la problématique biologique.

Les boîtes à moustaches des composés cibles des inflorescences sont exposées dans la Figure 103. Les résultats pour les échantillons symptomatiques sont représentés en rouge et les résultats pour les extraits asymptomatiques sont présentés en vert. A noter que la différence de significativité est souvent moins marquée pour les inflorescences que pour celles des parties chlorophylliennes observées sur la Figure 98.

Dans cette catégorie des inflorescences, le caractère discriminant de deux ions n'a pu être identifié. Le premier ion M223T633 est surexprimé dans les inflorescences des plants asymptotique de lavandin abrial alors qu'il est sous-exprimé pour la variété grosso. Le deuxième est l'ion M109T510 qui est lui aussi sur-exprimé chez abrial et sous-exprimé pour grosso pour les plants asymptotiques.

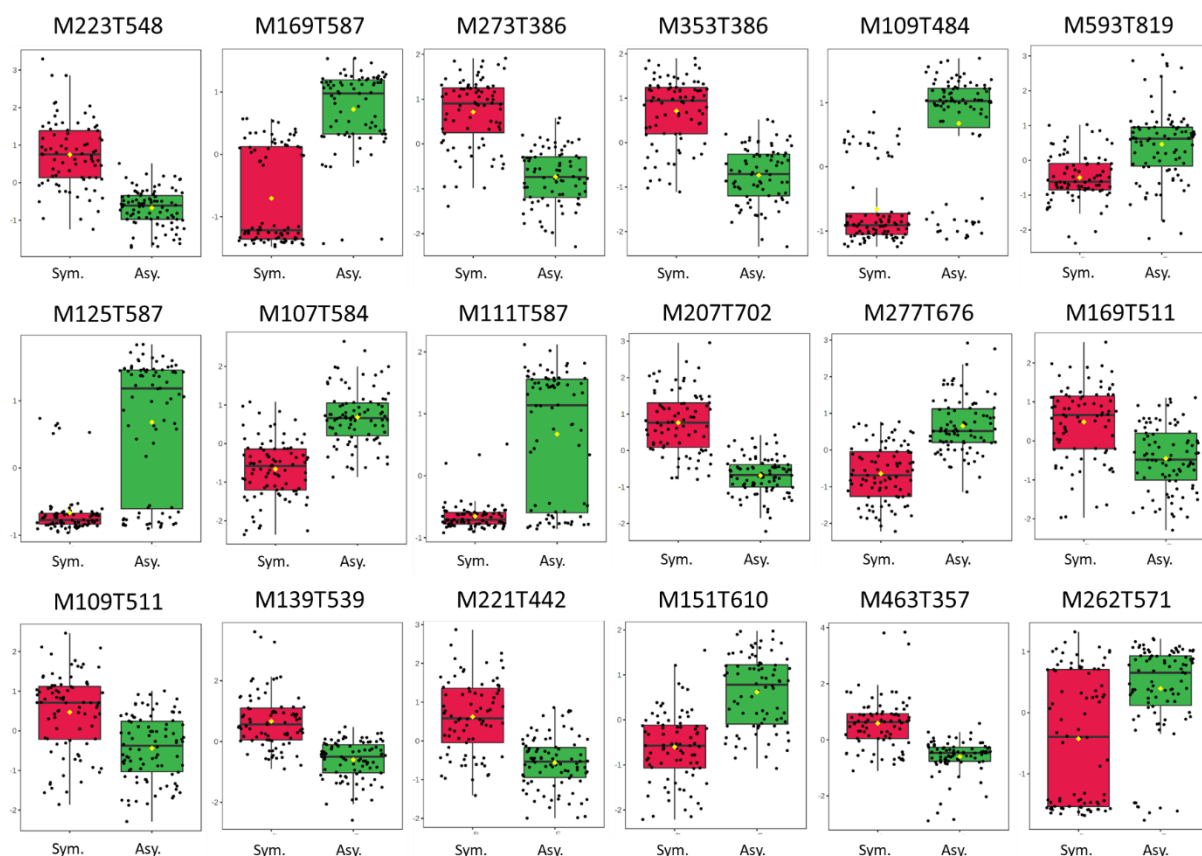


Figure 103 : Boîtes à moustaches des composés discriminants cibles en fonction de l'état sanitaire des plants (symptomatique en rouge et asymptotique en vert) pour les inflorescences de lavande et de lavandin

De même que précédemment pour les extraits des organes chlorophylliens, plusieurs ions ont des temps de rétentions identiques ou très proches pouvant faire penser qu'ils appartiennent à la même molécule. Nous avons pu relever les ions :

- M169T587, M125T587, M111T587 et M107T584 tous les quatre sont surexprimés par les plants asymptotiques
- M273T386 et M353T386 sont caractéristiques des plants symptomatiques
- M169T511 et M109T511 surexprimés par les plants symptomatiques

Les ions M273T386 et M353T386 sont particulièrement intéressants puisqu'ils ont une valeur de classement très faible révélant qu'ils sont fortement discriminants.

#### b. Proposition de structure

Comme pour les extraits des organes chlorophylliens de la plante, nous avons cherché à émettre des propositions de structure à partir des spectres MS, MSe et UV des composés discriminants cibles. Comme le montre la Figure 104, les composés discriminants sont généralement de faible intensité voire co-élus, ce qui rend la tâche plus ardue.

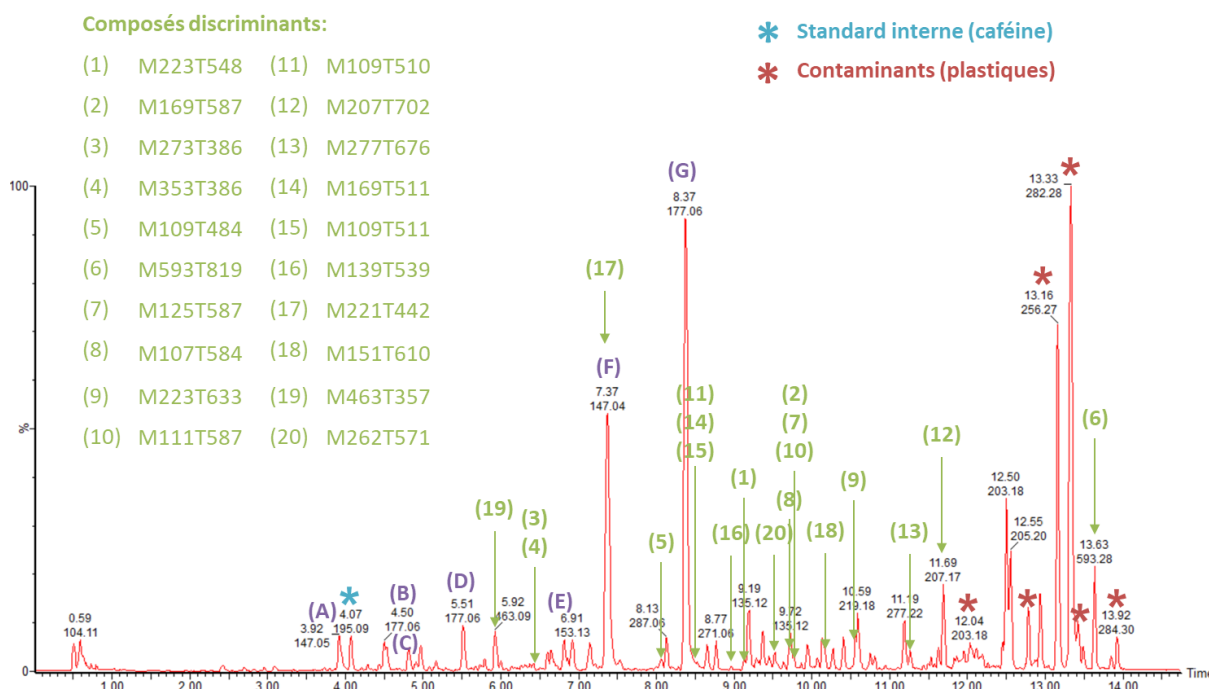


Figure 104 : Chromatogramme des inflorescences de lavandin annoté avec les composés cibles discriminants de l'état symptomatique des plants. (A) O-coumaric acid; (B) Methyl caffeate; (C) Luteolin di-glucuronide; (D) Ferulic acid; (E) Caffeic acid; (F) Coumarin; (G) Herniarin

Dans les composés discriminants cibles en fonction de l'état sanitaire des plants pour les extraits d'inflorescences, le VIP M463T357 a obtenu une correspondance avec la banque des composés organiques non volatils dans le genre *Lavandula*. La différence de masse entre la masse expérimentale du VIP et celle de la base de données est extrêmement faible ( $\Delta m = 0,00095$  u), ce qui fait de cette correspondance une hypothèse solide. La proposition de structure fournie par la banque est la lutéoline 7-O-glucuronide (Contreras et al. 2018). Le spectre de masse contient l'adduit sodium (485,0676 u) et le dimère (925,1655 u) du VIP, ce qui nous conforte dans l'idée que le pic de rapport  $m/z = 463,0875$  u est l'ion moléculaire. La formule brute la plus probable pour cet ion est  $C_{21}H_{19}O_{12}^+$  (-0,4 ppm). De plus, le profil isotopique théorique a une bonne correspondance avec le massif expérimental. Un autre pic à  $m/z = 287,0564$  u est présent et correspond à la perte d'un motif glucuronide. Le spectre de la MSe apporte peu d'information supplémentaire par rapport à celui du spectre de masse, la partie aglycone n'a pas ou presque pas fragmenté puisque seul un ion de faible intensité est visible à  $m/z = 153,0180$  u. Cet ion est cependant caractéristique de la perte d'un fragment  $^{1,3}A^+$  d'un flavonoïde (Cuyckens and Claeys 2004) comme présenté dans la Figure 105. La proposition de structure de la lutéoline 7-O-glucuronide semble conforme. Cette hypothèse est d'autant plus confortée par le spectre UV (longueurs maximales d'absorbance : 253 et 347 nm) correspond parfaitement avec celui de la lutéoline (De Rijke et al. 2006).

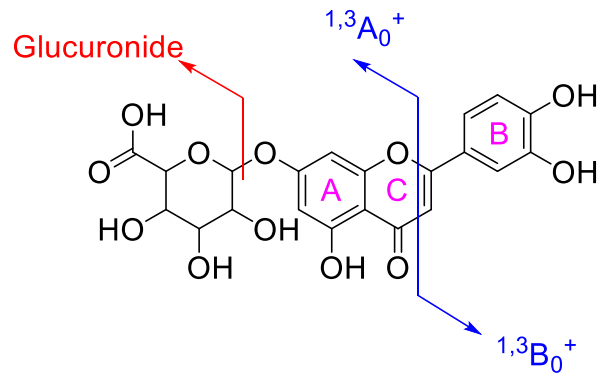


Figure 105 : Proposition de fragmentation de la lutéoline 7-O-glucuronide.

Le travail effectué sur les autres composés cibles des inflorescences est présenté dans le Tableau 40. Le phéophorbide A précédemment identifié dans les extraits des parties chlorophylliennes est également présent dans les composés cibles des inflorescences. D'autres hypothèses de structure forte n'ont pu être émises pour cette catégorie. Des propositions ont cependant pu être formulées pour certains ions, c'est par exemple le cas pour le composé correspondant à l'ion M262T571. Nous avons pu observer sur le spectre de ce composé la présence d'un adduit sodium  $[M+Na]^+$  à 302,2085 u et un ion moléculaire  $[M+H]^+$  à 280,2266 u. Par conséquent, l'ion discriminant (d'intensité supérieure, voir Figure 106) correspondrait à l'ion moléculaire déshydraté (perte de neutre de  $-18$  u). La formule brute pour l'ion moléculaire ayant une différence de masse la plus basse est  $C_{21}H_{19}O_{12}^+$  ( $-3,9$  ppm), malgré une correspondance faible entre le profil isotopique théorique et le massif expérimental, nous avons privilégié cette formule à celle en  $C_{12}H_{30}N_3O_4^+$  dont la différence de masse est bien supérieure (10,7 ppm). La formule brute  $C_{17}H_{29}NO_2$  a alors été soumise à plusieurs banques de données (Knapsack, HMDB, DNP). Parmi les propositions de structures seule la stemaphylline a été retenue. Les autres molécules ont été écartées car elles comportaient toutes un N-méthyl aromatique, or aucune perte de méthyl n'a pu être observée sur les spectres MS ou MSe. Concernant l'hypothèse de la stemaphylline, aucune information sur ses spectres de masse ou UV, n'a pu être trouvée dans la littérature permettant d'infirmer ou confirmer cette hypothèse.

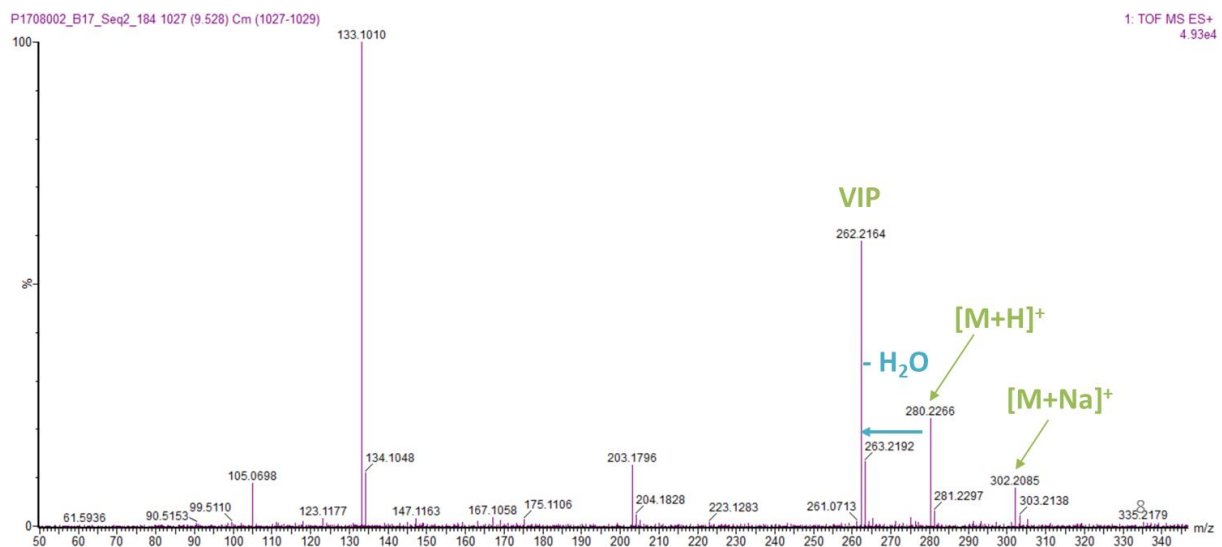


Figure 106 : Agrandissement du spectre de masse du pic correspondant à l'ion cible M262T571

Parmi les autres composés discriminants de l'état symptomatique des plants pour les extraits d'inflorescences, deux ions éluent au même temps de rétention le M169T511 et le M109T511. Le

Le spectre UV est celui d'une flavone voire plus spécifiquement d'une 5-Hydroxyflavone (De Rijke et al. 2006). Nous pouvons observer sur le spectre MS, un premier ion à 585,1129 u correspondant à l'adduit sodium  $[M+Na]^+$  de l'ion moléculaire  $[M+H]^+$  à 563,1387 u. Le spectre MSe, quant à lui, montre un fragment de cet ion à 315,0869 u correspondant à la perte d'une partie glycosylée acylée à savoir celle d'un malonyl glucuronide (perte de neutre de  $-248$  u). La partie aglycone se fragmente également en perdant un groupement  $-CH_3$  (perte de neutre de  $-15$  u) qui révèle la présence d'un groupement méthoxy dans la structure. En soumettant la formule brute ( $C_{26}H_{26}O_{14}$ ) de de l'ion moléculaire à plusieurs banques de données (Knapsack, HMDB, DNP, Reaxys) et en recoupant avec les informations indiquées ci-dessus, deux structures nous sont apparues pertinentes : l'onogénin 7-O- $\beta$ -D-glucoside 4''-malonate et l'irisolidone 7-O- $\beta$ -D-glucoside-6''-O-malonate. Cependant, nous n'avons pu trouver dans la littérature des informations permettant de préciser, d'infirmer ou de confirmer cette assignation. Pour la première proposition de structure, une fragmentation est décrite dans la littérature et ne correspond pas à celle observée (Gampe et al. 2019). Néanmoins, la source d'ionisation employée dans notre étude (ESI) est différente de celle utilisée dans l'article (une source de type FAB). Pour la deuxième structure aucun spectre MS/MS n'a été trouvé. En revanche, le spectre UV de la partie aglycone a été trouvé et est approchant (Farag et al. 2007).

A partir de ces informations, nous ne pouvons dire si les deux ions cibles M169T511 et M109T511 sont des fragments de cette molécule. Il faudrait fragmenter l'ion moléculaire  $[M+H]^+$  à 563,1387 u pour avoir plus d'information. Des fragments à 221,1905 u et 203,1797 u sont également présents sur le spectre MS (mais absents de celui de la MSe). Le fragment à 203,1797 u découle de celui à 221,1905 u avec une perte d'eau ( $-18$  u). Néanmoins, au vu des formules brutes, l'ion à 169,1229 u ( $C_{10}H_{16}O_2$ ) ne provient pas de l'ion à 221,1905 u ( $C_{15}H_{24}O$ ).

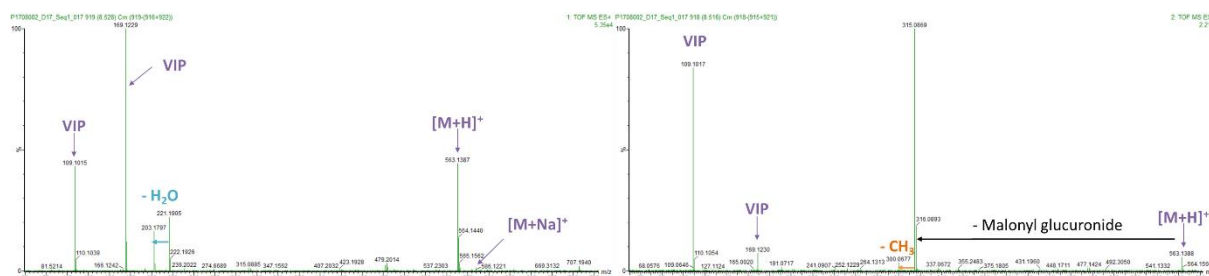


Figure 107 : A gauche, agrandissement du spectre MS annoté du pic présent à 8,52 min correspondant à l'étude des ions discriminants M169T511 et M109T511 ; et à droite, agrandissement du spectre MSe du même composé.

Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

Tableau 40 : Hypothèses d'identification pour les composés cibles contenus dans les extraits d'inflorescences

VIP	Rt (min)	[M+H] <sup>+</sup> m/z expérimental	Formule brute [M]	Δm (ppm) expérimental et théorique	m/z fragments	m/z adduits	Hypothèse identification	Banques <i>Lavandula</i>
<b>M223T548</b>	9,125	241,1790*	C <sub>14</sub> H <sub>24</sub> O <sub>3</sub>	-5,8	223,1700; 205,1596; 177,1639; 147,1179	263,1621 [M+Na] <sup>+</sup>		Multi.
<b>M169T587</b>	9,774	169,1224*	C <sub>10</sub> H <sub>16</sub> O <sub>2</sub>	-3,0	151,1120; 133,1015; 123,1173	191,1056 [M+Na] <sup>+</sup>		Multi.
<b>M273T386</b>	6,432	273,1133	C <sub>16</sub> H <sub>16</sub> O <sub>4</sub>	2,2	x	x		/
<b>M353T386</b>	6,432	389,123	C <sub>20</sub> H <sub>20</sub> O <sub>8</sub>	-1,5	371,1135; 353,1030; 325,1085; 195,0647;177,055; 149,0603; 121,0655	x	Dérivé acide férulique	/
<b>M109T484</b>	8,064	109,0654	C <sub>7</sub> H <sub>8</sub> O	0,9	91,0554	x		/
<b>M593T819</b>	13,669	593,2756*	C <sub>34</sub> H <sub>36</sub> N <sub>4</sub> O <sub>5</sub>	-0,5	533,2543	615,2572 [M+Na] <sup>+</sup>	Phéophorbide A	/
<b>M125T587</b>	9,774	125,0965	C <sub>8</sub> H <sub>12</sub> O	-0,8	107,0858	x		/
<b>M107T584</b>	9,73	107,0862	C <sub>8</sub> H <sub>10</sub>	0,9	91,0553	x		/
<b>M223T633</b>	10,563	223,1701*	C <sub>14</sub> H <sub>22</sub> O <sub>2</sub>	1,3	205,1599; 165,1283; 137,134	245,1522[M+Na] <sup>+</sup> ; 261,1266[M+K] <sup>+</sup> ; 467,323[2M+Na] <sup>+</sup>	Dérivé du caryophyllène	Multi.
<b>M111T587</b>	9,774	111,081	C <sub>7</sub> H <sub>10</sub> O	0,0	93,0704	x		/
<b>M109T510</b>	8,503	109,0651	C <sub>7</sub> H <sub>8</sub> O	-1,8		x		/
<b>M207T702</b>	11,704	207,1751	C <sub>14</sub> H <sub>22</sub> O	0,5		x		/
<b>M277T676</b>	11,265	277,2177	C <sub>18</sub> H <sub>28</sub> O <sub>2</sub>	3,2	x	x		/

Chapitre 4 : Résultats des analyses statistiques, propositions de structure et discussions

<b>M169T511</b>							onogenin 7-O- $\beta$ -D-glucoside 4''-malonate	Multi.
<b>M109T511</b>	8,528	563,1387*	C <sub>26</sub> H <sub>26</sub> O <sub>14</sub>	-2,5	315,0869; 300,0677	585,1129[M+Na] <sup>+</sup>	ou irisolidone 7-O- $\beta$ -D-glucoside-6''-O-malonate	/
<b>M139T539</b>	8,976	139,0759	C <sub>8</sub> H <sub>10</sub> O <sub>2</sub>	0,0		x		Hydroxybenzoic acid isomer 1
<b>M221T442</b>	7,375	221,1908	C <sub>15</sub> H <sub>24</sub> O	1,4		x		Multi.
<b>M151T610</b>	10,169	151,1127*	C <sub>10</sub> H <sub>14</sub> O	2,6	133,1023 ; 123,1176	173,0947 [M+Na] <sup>+</sup>		Multi.
<b>M463T357</b>	5,959	463,0875*	C <sub>21</sub> H <sub>18</sub> O <sub>12</sub>	-0,4	287,0564; 153,0180	925,1655[2M+H] <sup>+</sup> ; 485,0676[M+Na] <sup>+</sup>	Luteolin 7-O-glucuronide	Luteolin 7-O-glucuronide
<b>M262T571</b>	9,528	280,2266*	C <sub>17</sub> H <sub>29</sub> NO <sub>2</sub>	-3,9	262,2164	302,2085[M+Na] <sup>+</sup>	Stemaphylline	/

\* = hypothèse forte de l'ion moléculaire

Multi. = correspondance multiple avec la banque des composés organiques volatils

/ = aucune correspondance avec les deux banques du genre *Lavandula*



### II.B.3. Comparaison des deux listes de biomarqueurs cibles en fonction de l'état symptomatique des plants

En comparant les composés cibles obtenus ci-dessus pour les extraits de feuille et d'inflorescence, nous avons pu constater la présence de cinq ions communs aux deux listes. Ces ions particuliers sont : M593T819, M139T539, M221T442, M223T548 et M207T702.

L'ion M593T819 ressort comme un ion à fort potentiel puisqu'il est discriminant à la fois dans les extraits d'inflorescences et dans les extraits des parties chlorophylliennes. De plus pour les extraits de feuille, il apparaît comme un composé discriminant dans les quatre variétés de l'étude. Il est caractéristique des plants asymptomatiques pour l'ensemble de ces matrices. Une hypothèse de structure a pu être émise pour ce composé, celle du phéophorbide A, un dérivé de chlorophylle.

L'ion M223T548 est un autre candidat intéressant puisqu'il apparaît dans six des huit matrices en fonction de l'état symptomatique. Les deux listes (des 50 composés les plus discriminants) dont il est absent sont celles des extraits d'inflorescences d'abrial et de feuille de grosso. De surcroît, il obtient un très bon score au classement dans les deux catégories avec 66 pour les extraits d'inflorescences pour une note VIP de 3, et 57 pour les extraits de feuilles pour une note également de 3. Il est un marqueur de l'état symptomatique des plants. Malheureusement, aucune correspondance n'a pu être trouvée avec la banque des composés non volatils et nous n'avons pu émettre une hypothèse de structure à partir de son spectre de masse.

Les trois autres ions communs aux deux listes sont également caractéristiques des plants symptomatiques. L'ion M139T539 serait un isomère de l'acide hydroxybenzoïque mais le spectre de masse expérimentale, n'est pas en adéquation avec la fragmentation décrite dans la littérature pour ce composé et aucune autre hypothèse n'a pu être émise.

## III. Discussions et perspectives

Les résultats sur l'étude du caractère sensible et tolérant des variétés a permis d'obtenir dans tous les cas, des groupes nettement distincts dès les analyses descriptives. Les analyses discriminantes sont toutes valides avec des valeurs de  $R^2$  et  $Q^2$  très élevées ( $> 0,9$ ) et aucun sur-apprentissage n'a été détecté.

Les résultats obtenus pour l'étude en fonction du caractère symptomatique des plants sont plus mitigés. Cette performance moindre est potentiellement liée au nombre d'échantillons inférieur. Les valeurs de  $R^2$  et  $Q^2$  sont inférieures et généralement comprises entre 0,9 et 0,7. Elles restent néanmoins toutes supérieures à la limite de validation de 0,5. Plusieurs tests de permutations ont montré un sur-apprentissage, c'est le cas pour les matrices d'inflorescences des variétés grosso et maillette.

En comparant l'ensemble des résultats, plusieurs tendances semblent se dégager :

- Dans l'étude comparant les variétés, c'est à dire concernant le caractère sensible/tolérant des plants, les variétés sensibles (7713 et abrial) apparaissent comme ayant une variabilité chimique plus élevée que les variétés dites tolérantes dont les analyses sont très groupées sur les ACP. Cette observation est particulièrement vraie pour les échantillons d'inflorescences.
- D'ailleurs dans l'étude sur l'influence du caractère sanitaire, nous avons pu constater que les variétés « sensibles » (7713 et abrial) obtiennent une séparation plus marquée que les variétés dites « tolérantes » (maillette et grosso). De façon générale, les résultats des variétés dites « sensibles » sont meilleurs (séparation des groupes biologiques plus marquée, valeurs de  $R^2$  et  $Q^2$  plus élevées, absence de sur-apprentissage).

- Des tendances se dégagent également selon les organes de la plante :
  - les distinctions en fonction de l'état symptomatique sont plus marquées pour les parties chlorophylliennes
  - à contrario, la différence chimique entre les variétés est plus importante pour les échantillons d'inflorescences.

Une grande variabilité de la charge en phytoplasme du Stolbur a été observée (voir chapitre 1). L'hypothèse a été émise par les agronomes du CRIEPPAM que le phytoplasme n'est pas réparti de façon homogène au sein d'un plant infecté. C'est l'une des raisons pour laquelle des réplicats intra-plant ont été effectués et que les prélèvements ont été réalisés à des endroits distincts des plants. Cependant, rien ne nous laisse penser que la variabilité dans la composition chimique soit plus importante pour les groupes symptomatiques. Seules les matrices correspondant aux feuilles de la variété de lavande maillette et aux inflorescences de la variété de lavandin grosso ont des groupes symptomatiques (sur les graphiques des individus des ACP) plus dispersés que les asymptomatiques. Les autres jeux de données ne sont pas concluants.

En ce qui concerne, l'étape de caractérisation la plupart des composés les plus discriminants sont peu voire très peu intenses. De surcroît, plusieurs spectres comportent un nombre d'ions élevés laissant suspecter de nombreuses co-élutions. Tous ces éléments complexifient l'interprétation des spectres de masses. De manière à faciliter le processus d'identification, les pools des différents lots d'extraction et d'analyse devraient être analysés en mode négatif en vue d'aider à la détermination des ions moléculaires en confrontant les spectres de masses obtenus pour les deux modes. De même, il apparaît intéressant de réaliser des analyses MS/MS afin d'obtenir une fragmentation ciblée et s'affranchir des problèmes de co-élution. Une fois des hypothèses d'identification « fortes » établies, l'étape suivante consisterait à injecter les standards pour confirmer les hypothèses formulées précédemment.

Au vu des spectres de masses, ils semblent que de nombreuses fragmentations soient redondantes. Il apparaîtrait judicieux pour cette étude, de mettre en place un réseau moléculaire. Pour ce faire, il est nécessaire de réaliser des analyses en MS/MS sur les marqueurs cibles puis construire un réseau à partir des données générées sur des plateformes en ligne tels que GNPS ou MetGem (voir partie Annotation, identifications chimiques des métabolites et réseaux moléculaires du chapitre 1). L'élaboration d'un réseau moléculaire permettra de faciliter la caractérisation des composés analogues (c'est-à-dire ayant des schémas de fragmentation proches). Si des hypothèses d'identification sérieuses ont pu être générées grâce au réseau, il est pertinent à ce stade d'injecter des standards dans les mêmes conditions analytiques pour infirmer ou confirmer ces nouvelles hypothèses structurales. A priori plusieurs composés présents dans les extraits sont sous formes glycosylées par conséquent la synthèse de standard devra peut-être être envisagée selon la disponibilité des molécules.

Les composés cibles n'ayant pu être identifiés précédemment devront faire l'objet d'un travail de purification et d'élucidation structurale par RMN. Les signaux en masse obtenus pour les ions discriminants sont souvent faibles voire très faibles. Il faut s'attendre à ce que ces analytes soient présents en faible quantité dans les extraits donc à une étape de purification longue et fastidieuse.

Il serait également intéressant de réaliser le même travail d'identification sur les composés les plus discriminants pour chaque matrice obtenue.

Parmi les composés discriminants de l'état symptomatique des plants de lavande et de lavandin dans les organes chlorophylliens, plusieurs phéophorbides (A, B et B éthyle ester) ont été putativement identifiés. Ces dérivés de chlorophylles ont été peu étudiés pour leurs activités biologiques hormis sur

les mammifères (activité antitumorale, immunosuppresseur, toxique, ...) (Vencl et al. 2009). Néanmoins, des propriétés répulsives, photosensibilisantes et insecticides ont été reportées pour certains phéophorbides. En outre, Vencl *et al.* ont mis en évidence un rôle en écologie chimique des phéophorbides A et B. Ils ont observé que la « coquille » protectrice des larves de *C. alternans* permettait de repousser les fourmis *A. lacrymosa* (prédateurs naturels de ces larves) lorsque celles-ci entraient en contact avec la « coquille ». Ils ont ensuite montré que les molécules responsables de cet effet répulsif sont le phéophorbide A en premier lieu mais aussi le phéophorbide B. Au vu du comportement des fourmis, les auteurs ont émis l'hypothèse que le phéophorbide A est irritant et désoriente les insectes (Vencl et al. 2009). Les phéophorbides ont également été rapportés comme photosensibilisant pour les larves de plusieurs insectes, entraînant des nécroses dans les intestins puis la mort de ces dernières après exposition à la lumière (Azizullah et al. 2014). Dans le cas des larves *C. alternans*, les auteurs ont remarqué que les larves se nourrissaient en évitant de s'exposer au soleil direct comme si elles se prémunissaient elles-mêmes de l'action photosensibilisante du phéophorbide A (Vencl et al. 2009). A noter qu'il existe aussi un brevet sur le phéophorbide A méthyle ester dans le but d'être employé comme insecticide pour tuer les larves des nématodes ou les larves qui sont normalement résistantes aux insecticides conventionnels. Il agit en inhibant la voie métabolique des stéroïdes chez les insectes (Yong Seok et al. 2010).

En ce qui concerne les dérivés de chlorophylles de manière plus générale, il a été montré, dans le cas d'une attaque d'un virus, qu'une plante résistante va voir la dégradation des chlorophylles diminuer lors de cette attaque et la synthèse de chlorophylle augmenter. A l'inverse, une plante sensible au virus verra le phénomène inverse se produire (Zanardo et al. 2019). Un autre article mené cette fois-ci sur les dérivés de chlorophylles présents dans les feuilles déverdies de pommiers et d'abricotiers infectés par un phytoplasme (respectivement *Candidatus Phytoplasma mali* et *Candidatus Phytoplasma prunorum*), a montré que la dégradation de la chlorophylle est plus importante et précoce chez les arbres infectés. Cette étude met en évidence que l'infection par le phytoplasme active la voie de l'oxygénase du phéophorbide A/phyllobilin (PaO/PB). En revanche, l'étude n'a pas mis en évidence la présence de phéophorbide, ce sont des phyllobilins qui ont été caractérisés. La raison de l'activation de cette voie enzymatique reste inconnue. Plusieurs hypothèses sont possibles, à savoir la conséquence indirecte du stress causé par l'infection, une stratégie de la part du pathogène pour obtenir les nutriments stockés dans les chloroplastes ou pour empêcher une réponse immunitaire de la part de l'hôte (Mittelberger et al. 2017).

Au vu des renseignements trouvés dans la littérature, il apparaît intéressant de tester le comportement de *Hyalesthes obsoletus* face au trois phéophorbides décrits dans cette thèse et particulièrement le phéophorbide A.



# Conclusion générale

L'objectif principal de cette thèse était de mettre en place une étude métabolomique globale adaptée aux composés organiques non volatiles (CONVs) sur des échantillons de lavande (*Lavandula angustifolia*) et de lavandin (*Lavandula x intermedia*). Ce projet s'inscrivait dans une meilleure compréhension des mécanismes de défenses de la plante induits à la suite de l'infection par le phytoplasme du Stolbur. Le but recherché était double :

- Le premier étant l'identification des biomarqueurs de l'état sanitaire des plants, c'est-à-dire en fonction de leur caractère symptomatique ou asymptomatique.
- Le deuxième étant l'identification des biomarqueurs du caractère sensible ou tolérant des variétés face au dépérissement.

Afin de réaliser cette étude, nous avons en premier lieu mis place des tests préliminaires par HPLC-UV (chapitre 2). La finalité de ces premiers tests était de déterminer le solvant d'extraction le plus adapté à la matrice végétale étudiée, mais aussi d'établir la méthode d'extraction et l'ensemble des étapes de préparation des échantillons à savoir la conservation, le broyage et le séchage. Pour ce faire nous avons dû répondre à plusieurs contraintes. En effet, la méthode se devait d'être compatible avec le prélèvement en champ, ainsi qu'avec la matrice étudiée. Le protocole établi se devait de limiter le nombre d'étapes et d'être le plus simple possible afin d'introduire le moins de biais lors de la préparation des extraits (et ainsi assurer la reproductibilité de l'étape de préparation), tout en ayant une cadence de réalisation élevée afin de gérer un nombre important d'extraits (1429). Enfin la réalisation d'une étude métabolomique non ciblée impose d'obtenir une méthode permettant de couvrir au maximum le métabolome.

Face à ces contraintes, le design expérimental adopté et mis en place pour l'étude comprend :

- Une première étape d'échantillonnage directement en champ avec trempe immédiate des échantillons prélevés dans l'azote liquide afin d'arrêter le métabolisme de la plante.
- Les échantillons ont ensuite été transportés dans de la carboglace et stockés dans un congélateur à  $-80^{\circ}\text{C}$ ,
- Jusqu'à leur séchage par lyophilisation
- Les échantillons sont ensuite stockés à température ambiante et à l'abri de la lumière,
- Jusqu'à leur extraction. La méthode choisie est une extraction solide-liquide assistée par ultrasons. Cette méthode a été choisie pour sa simplicité, son efficacité et sa rapidité. Le solvant d'extraction est un mélange à 50/50 en volume entre de l'éthanol et de l'eau. Les mélanges de solvant permettant d'extraire une plus large gamme d'analyte.
- Après extraction, les extraits sont rapidement analysés.

Après les tests préliminaires, une méthode analytique a été mise en place par UPLC-HRMS (QTOF) (chapitre 2). L'objectif ici était d'obtenir :

- une méthode répétable, afin d'éviter toute introduction de biais expérimentaux néfastes pour la qualité des résultats
- une méthode adaptée à tous les profils chimiques. Rappelons que nous avons des échantillons provenant de 4 variétés différentes (7713, maillette, abrial et grosso), de deux organes différents (parties chlorophylliennes et inflorescences) et de différents états sanitaires (symptomatique et asymptomatique).

## Conclusion générale

- une bonne séparation, afin de limiter le nombre de co-élutions et ainsi faciliter le traitement des données a posteriori
- Minimiser la durée des séquences. La réduction de la durée est en partie en confrontation avec le point précédent. Néanmoins, il était impératif que la durée d'une analyse soit restreinte au maximum car l'allongement des séquences va automatiquement conduire à une dérive analytique.

Malgré nos efforts une perte de signal a été constatée au cours des séquences d'analyse. En revanche, la méthode mise en place a conduit à une faible voire aucune dérive temporelle et massique.

Une fois les analyses effectuées, nous avons conçu un processus de traitement des données (chapitre 3). Le traitement des données que nous avons mis en place, se développe en 5 grandes étapes :

- L'extraction des données spectrales. Cette première étape consiste à transformer les données brutes obtenues après les analyses en matrice en deux dimensions. Elle a été développée et mise en place sur la plateforme en ligne W4M-Galaxy. Elle comprend les étapes de détection des ions, d'alignement des ions, la correction des temps de rétention, et l'attribution des aires aux pics manquants.
- Le nettoyage des données. Dans cette deuxième étape, deux types de nettoyage sont à distinguer et ont été réalisés : le nettoyage sur les ions et celui sur les échantillons. Le premier comprend la suppression des données redondantes telles que les isotopes, les fragments ou les adduits ; ainsi que la suppression des contaminants identifiés grâce à la réalisation d'une extraction à blanc ; mais aussi la suppression des variables non significatives c'est-à-dire n'ayant pas ou peu d'influence sur les compositions chimiques. Le second type de nettoyage sur les échantillons correspond à la suppression des données aberrantes.
- La correction et la préparation aux analyses statistiques. Cette étape a été réalisée sur la plateforme en ligne Metaboanalyst. Les transformations que nous avons appliquées comprennent une normalisation par les pools ayant pour objectifs de corriger la dérive analytique ; ainsi qu'une transformation logarithmique et un « auto-scaling » qui est un centrage et une réduction des données employées afin de donner un poids équivalent aux données.
- Les analyses statistiques et l'obtention des composés discriminants pour chaque jeu de données. Pour chaque jeu de données ont été réalisées une analyse descriptive (ACP) et une analyse discriminante (PLS-DA) toujours grâce à la plateforme en ligne Metaboanalyst. Les PLS-DA, nous ont permis de déterminer les ions les plus discriminants pour chaque matrice ainsi que le caractère discriminant pour chacun des ions.
- L'identification des biomarqueurs. Les composés discriminants révélés par les PLS-DA ont été confrontés à la banque de CONVs du genre *Lavandula*, élaborée grâce à une vieille bibliographie continue, afin de générer des hypothèses de structures.

Cette banque de données comprend près de 230 molécules et a fait l'objet d'une publication dans le journal *Phytochemistry reviews* (Héral et al. 2020).

Pour finir, dans le dernier chapitre de ce manuscrit, a été présenté l'ensemble des résultats obtenus au cours de cette thèse. De manière générale, les résultats montrent une variabilité résiduelle, c'est-à-dire qui n'a pu être imputée à un facteur de l'étude, élevée. Il est probable que ce problème ainsi que la réduction du nombre d'échantillons dans les jeux de données soient la cause des résultats mitigés pour l'étude portant sur le caractère symptomatique et asymptotique des plants. Les jeux de données des extraits d'inflorescences de la variété de lavande maillette et de la variété de lavandin grosso ont montré un sur-apprentissage rendant l'exploitation des données non fiables malgré les indicateurs  $R^2$  et  $Q^2$  qui sont supérieurs à 0,5. Tous les autres jeux de données sont valides et les tests de permutations ne permettent pas de conclure à un sur-apprentissage du modèle. Dans le cas de la

## Conclusion générale

variété 7713, une différenciation des deux groupes symptomatique et asymptomatique a pu être observée pour les deux organes de la plante. Les extraits des parties chlorophylliennes des variétés maillette et abrial ont une différenciation entre les deux groupes, néanmoins elle est peu marquée. Les jeux de données correspondants aux inflorescences de la variété abrial et aux parties chlorophylliennes de la variété de lavandin grosso présentent quant à eux, un léger chevauchement des ellipses sur les graphiques des individus des PLS-DA.

Pour l'étude concernant le caractère sensible ou tolérant des variétés face au dépérissement, les analyses statistiques sont toutes valides et sans sur-apprentissages. Les analyses ont mis en évidence une différence de signature chimique entre les variétés sensibles et les variétés tolérantes pour tous les organes de la plante. A noter que la différence est plus marquée pour les inflorescences que pour les parties chlorophylliennes.

Parmi les ions les plus discriminants de chaque jeu de données, nous avons pu distinguer plusieurs ions communs à différentes matrices que nous avons appelées composés cibles. Dans l'étude du caractère sensible ou tolérant des variétés, 6 ions cibles ont pu être mis en évidence dans les parties chlorophylliennes. En revanche, aucun n'a pu être trouvé pour les extraits des inflorescences. Dans la partie se penchant sur la variabilité chimique en fonction de l'état sanitaire des plants, nous avons pu mettre en évidence 26 ions cibles pour les parties chlorophylliennes et 20 pour les inflorescences. De plus, 5 ions communs aux deux organes ont pu être trouvés, à savoir : M593T819, M139T539, M221T442, M223T548 et M207T702. L'ion M593T819 est un marqueur des plants asymptomatiques. Les quatre autres ions sont caractéristiques de l'état symptomatique des plants et l'ion M139T539 serait un isomère de l'acide hydroxybenzoïque. L'ion M593T819 a, quant à lui, été potentiellement identifié comme le phéophorbide A. Deux autres dérivés de chlorophylles, le phéophorbide B et le phéophorbide B éthyle ester ont été putativement identifiés parmi les composés caractéristiques de l'état asymptomatique dans les parties chlorophylliennes (lieu de ponction alimentaire de *H. obsoletus*) des plants. Or ces composés peuvent s'avérer intéressants car les phéophorbides ont des propriétés répulsives et photosensibilisantes contre les insectes.

Pour conclure, nous avons pu mettre en place une stratégie expérimentale et de traitement des données adaptée au sujet d'étude afin de réaliser une étude métabolomique globale. Cette étude a permis de mettre en évidence des différences de compositions chimiques entre des variétés dites sensibles et des variétés dites tolérantes au phytoplasme du Stolbur, et ce sur les deux organes de la plante étudiés (les parties chlorophylliennes et les inflorescences). Elle a également permis de montrer des différences de signatures chimiques entre des plants symptomatiques et asymptomatiques pour plusieurs des variétés étudiées, notamment pour les variétés dites sensibles au dépérissement (7713 pour la lavande et abrial pour le lavandin). Elle a permis de mettre en avant une famille de composés, les phéophorbides qui peuvent s'avérer une piste particulièrement intéressante pour la lutte contre le dépérissement des lavanderaies en France.

Cette thèse ouvre également la voie à d'autres travaux de recherches, en premier lieu un long travail d'analyse phytochimique pour l'identification des biomarqueurs mis en évidence dans cette étude. Il serait en particulier intéressant de mettre en place un réseau moléculaire. Elle trace également la voie pour réaliser une nouvelle étude similaire sur les échantillons récoltés lors de la campagne 2018, afin de confirmer ou infirmer les résultats obtenus. Une nouvelle étude métabolomique sur des extraits d'acétate d'éthyle peut également être envisagée afin de couvrir une partie du métabolome ayant des propriétés plus lipophiles. Il serait toujours intéressant de réaliser une étude métabolomique sur les racines des plants de lavande et de lavandin sur lesquelles se développent les larves de cicadelle. Néanmoins, il faudrait résoudre le problème d'approvisionnement en matière première. Enfin, il apparaît intéressant de tester l'activité répulsive des phéophorbides sur les larves et les adultes de *H. obsoletus*.





# Valorisation des travaux de recherches

## Publication :

HÉRAL B., STIERLIN É., MICHEL T., FERNANDEZ X. Phytochemicals from the genus *Lavandula*: a review. *Phytochemistry reviews*, 2020. <https://doi.org/10.1007/s11101-020-09719-z>.

## Communication orale :

HÉRAL B., Florence NICOLÈ, FERNANDEZ X., MICHEL T. Étude métabolomique des interactions chimiques dans le dépérissement de la lavande. Journées de Métabolomique en Région Sud-Provence-Alpes-Côte d'azur (MétaSUD), 12-13 juin 2019, Toulon, France.

## Communications posters:

HÉRAL B., Florence NICOLÈ, FERNANDEZ X., MICHEL T. Metabolomic study of the chemical interactions in lavender's yellow disease. Douzièmes journées scientifiques du RFMF (Réseau Francophone de Métabolomique et de Fluxomique), 21 au 23 mai 2019, Clermont-Ferrand, France.

HÉRAL B., FERNANDEZ X., MICHEL T. Study of the chemical interactions in lavender's yellow disease by metabolomic approach. Onzièmes journées scientifiques du RFMF, 23 au 25 mai 2018, Liège, Belgique.

# Annexes

Annexe 1 Composés non volatils caractérisés ou partiellement identifiés dans le genre *Lavandula*

Famille chimique	Nom trivial <sup>(a)</sup> (autre nom)	Formule brute	Organe plante			Espèces	Références
			Partie aérienne		Racine		
			Inflo.	Feuille			
<b>Aldéhydes</b>	3,4-dihydroxybenzaldehyde	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>			+	<i>L. angustifolia</i>	(Yadikar et al. 2017)
	citric acid	C <sub>6</sub> H <sub>8</sub> O <sub>7</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	gluconic/galactonic acid	C <sub>6</sub> H <sub>12</sub> O <sub>7</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	hydroxy-jasmonic acid hexoside isomer 1	C <sub>18</sub> H <sub>28</sub> O <sub>9</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
<b>Acides carboxyliques</b>	hydroxy-jasmonic acid hexoside isomer 2	C <sub>18</sub> H <sub>28</sub> O <sub>9</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	hydroxy-jasmonic acid sulfate	C <sub>12</sub> H <sub>18</sub> O <sub>7</sub> S			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	isocitric acid	C <sub>6</sub> H <sub>8</sub> O <sub>7</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	trihydroxy-octadecadienoic acid	C <sub>18</sub> H <sub>32</sub> O <sub>5</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
<b>Cétones</b>	(iso)lariciresinol hexoside	C <sub>26</sub> H <sub>34</sub> O <sub>11</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
<b>Monoterpénoïdes</b>	nepetanudoside B	C <sub>16</sub> H <sub>22</sub> O <sub>9</sub>			+	<i>L. dentata</i>	(Contreras et al. 2018)
<b>Sequiterpénoïdes</b>	(1R,6R,7R,8S,11R)-1,6-dihydroxy-4,11-dimethyl-germacran-4(5),10(14)-dien-8,12-olide	C <sub>15</sub> H <sub>22</sub> O <sub>4</sub>			+	<i>L. stoechas</i>	(Contreras et al. 2018)
	longipin-2-ene-7β,9α-diol-1-one	C <sub>15</sub> H <sub>22</sub> O <sub>3</sub>			+	<i>L. stoechas</i>	(Ulubelen et al. 1988)

## Annexes

	(Tricyclo[5.4.0.02,8]undec-10-en-9-one, 3,5-dihydroxy-2,6,6,11-tetramethyl-, (1R,2R,3R,5R,7S,8R)-)							
	longipin-2-ene-7 $\beta$ ,9 $\alpha$ -diol-1-one 9-monoacetate							
	(tricyclo[5.4.0.02,8]undec-10-en-9-one, 3-(acetyloxy)-5-hydroxy-2,6,6,11-tetramethyl-, stereoisomer (9CI))	C <sub>17</sub> H <sub>24</sub> O <sub>4</sub>		+	<i>L. stoechas</i>	(Ulubelen et al. 1988)		
<b>Diterpénoïdes</b>	abietatriene	C <sub>20</sub> H <sub>30</sub>			+	<i>L. x intermedia</i> <i>L. pedunculata</i> <i>L. stoechas</i> <i>L. canariensis</i>	(Stierlin et al. 2019) (Breton Funes and Jaudenes Ruiz de Atauri 1986; Topçu et al. 2001; Lis-Balchin 2002)	
	$\alpha$ -amyrin	C <sub>30</sub> H <sub>50</sub> O		+		+	<i>L. pedunculata</i> <i>L. canariensis</i>	(Breton Funes and Jaudenes Ruiz de Atauri 1986; Lis-Balchin 2002)
	$\beta$ -amyrin	C <sub>30</sub> H <sub>50</sub> O			+		<i>L. pedunculata</i> <i>L. canariensis</i>	(Breton Funes and Jaudenes Ruiz de Atauri 1986; Lis-Balchin 2002)
	$\beta$ -amyrin acetate	C <sub>32</sub> H <sub>52</sub> O <sub>2</sub>				+	<i>L. stoechas</i>	(Topçu et al. 2001)
<b>Triterpénoïdes</b>	betulin	C <sub>30</sub> H <sub>50</sub> O <sub>2</sub>			+	+	<i>L. spica</i> <i>L. stoechas</i>	(Papanov et al. 1992; Topçu et al. 2001; Lis-Balchin 2002)
	betulinic acid	C <sub>30</sub> H <sub>48</sub> O <sub>3</sub>			+	+	<i>L. dentata</i> <i>L. spica</i> <i>L. stoechas</i>	(Papanov et al. 1992; Topçu et al. 2001; Lis-Balchin 2002; Contreras et al. 2018)
	3 $\beta$ ,24-dihydroxyolean-12-ene	C <sub>29</sub> H <sub>48</sub> O <sub>2</sub> <sup>(b)</sup>				+	<i>L. stoechas</i>	(Topçu et al. 2001)
	formylursolic acid	C <sub>31</sub> H <sub>48</sub> O <sub>4</sub>			+		<i>L. angustifolia</i> <i>L. spica</i>	(Papanov et al. 1992; Lis-Balchin 2002)
	16 $\beta$ -hydroxylupeol-3-O-myristate	C <sub>46</sub> H <sub>80</sub> O <sub>3</sub>				+	<i>L. stoechas</i>	(Topçu et al. 2001)
	16 $\beta$ -hydroxylupeol-3-O-palmitate	C <sub>44</sub> H <sub>76</sub> O <sub>3</sub>				+	<i>L. stoechas</i>	(Topçu et al. 2001)

## Annexes

18-hydroxy-27-norolean-12,14-dien-30-al-28-oic acid	C <sub>29</sub> H <sub>40</sub> O <sub>5</sub>			+	<i>L. stoechas</i>	(Topçu et al. 2001)
3β-hydroxy-1-oxo-olean-12-ene-30-al-28-oic acid	C <sub>30</sub> H <sub>44</sub> O <sub>5</sub>			+	<i>L. stoechas</i>	(Topçu et al. 2001)
23-hydroxytormentic acid	C <sub>30</sub> H <sub>48</sub> O <sub>6</sub>		+		<i>L. canariensis</i>	(Breton Funes and Jaudenes Ruiz de Atauri 1986; Lis-Balchin 2002)
2α-hydroxyursolic acid	C <sub>30</sub> H <sub>48</sub> O <sub>4</sub>		+		<i>L. canariensis</i>	(Breton Funes and Jaudenes Ruiz de Atauri 1986; Lis-Balchin 2002)
hydroxyursolic acid isomer 1	C <sub>30</sub> H <sub>48</sub> O <sub>4</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
hydroxyursolic acid isomer 2	C <sub>30</sub> H <sub>48</sub> O <sub>4</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
hydroxyursolic acid isomer 3	C <sub>30</sub> H <sub>48</sub> O <sub>4</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
hydroxyursolic acid isomer 4	C <sub>30</sub> H <sub>48</sub> O <sub>4</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
micromeric acid	C <sub>30</sub> H <sub>46</sub> O <sub>3</sub>			+	<i>L. dentata</i> <i>L. pedunculata</i> <i>L. stoechas</i>	(Lis-Balchin 2002; Contreras et al. 2018)
monogynol A <i>cis</i> -coumaryl ester	C <sub>38</sub> H <sub>56</sub> O <sub>4</sub> <sup>(b)</sup>			+	<i>L. stoechas</i>	(Topçu et al. 2001)
monogynol A <i>trans</i> -coumaryl ester	C <sub>38</sub> H <sub>56</sub> O <sub>4</sub> <sup>(b)</sup>			+	<i>L. stoechas</i>	(Topçu et al. 2001)
oleanolic acid	C <sub>30</sub> H <sub>48</sub> O <sub>3</sub>		+	+	<i>L. canariensis</i> <i>L. dentata</i> <i>L. stoechas</i>	(Breton Funes and Jaudenes Ruiz de Atauri 1986; Lis-Balchin 2002; Contreras et al. 2018)

## Annexes

oleanolic acid 3 $\beta$ -acetate	C <sub>32</sub> H <sub>50</sub> O <sub>4</sub>			+	<i>L. stoechas</i>	(Topçu et al. 2001)
11-oxo- $\beta$ -amyryn	C <sub>30</sub> H <sub>48</sub> O <sub>2</sub>			+	<i>L. stoechas</i>	(Topçu et al. 2001)
3-oxo-12-ursene-28-oic acid	C <sub>30</sub> H <sub>46</sub> O <sub>3</sub>		+		<i>L. angustifolia</i> <i>L. spica</i>	(Papanov et al. 1992; Lis-Balchin 2002)
pomolic acid	C <sub>30</sub> H <sub>48</sub> O <sub>4</sub>		+		<i>L. angustifolia</i> <i>L. spica</i>	(Papanov et al. 1992; Lis-Balchin 2002)
$\beta$ -sitosterol	C <sub>29</sub> H <sub>50</sub> O			+	<i>L. stoechas</i>	(Topçu et al. 2001)
$\beta$ -sitosterol acetate	C <sub>31</sub> H <sub>52</sub> O <sub>2</sub>			+	<i>L. stoechas</i>	(Topçu et al. 2001)
sitosteryl-3 $\beta$ -D-glucoside	C <sub>35</sub> H <sub>60</sub> O <sub>6</sub>		+	+	<i>L. canariensis</i>	(Breton Funes and Jaudenes Ruiz de Atauri 1986)
tormentic acid					<i>L. dentata</i>	
(2a,3b,19a-trihydroxyurs-12-en-28-oic acid)	C <sub>30</sub> H <sub>48</sub> O <sub>5</sub>		+		<i>L. stoechas</i>	(Contreras et al. 2018)
trihydroxyursolic acid isomer 1	C <sub>30</sub> H <sub>48</sub> O <sub>6</sub>		+		<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
trihydroxyursolic acid isomer 2	C <sub>30</sub> H <sub>48</sub> O <sub>6</sub>		+		<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
trihydroxyursolic acid isomer 3	C <sub>30</sub> H <sub>48</sub> O <sub>6</sub>		+		<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
trihydroxyursolic acid isomer 4	C <sub>30</sub> H <sub>48</sub> O <sub>6</sub>		+		<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
ursolic acid	C <sub>30</sub> H <sub>48</sub> O <sub>3</sub>		+	+	<i>L. angustifolia</i> <i>L. canariensis</i> <i>L. dentata</i> <i>L. pedunculata</i> <i>L. stoechas</i>	(Breton Funes and Jaudenes Ruiz de Atauri 1986; Lis- Balchin 2002; Spiridon et al. 2011; Contreras et al. 2018)
3- <i>epi</i> ursolic acid	C <sub>30</sub> H <sub>48</sub> O <sub>3</sub>		+		<i>L. angustifolia</i>	(Lis-Balchin 2002)

## Annexes

	ursolic acid lactone	C <sub>30</sub> H <sub>46</sub> O <sub>3</sub>			+	<i>L. angustifolia</i> <i>L. spica</i>	(Papanov et al. 1992; Lis-Balchin 2002)	
	uvaol	C <sub>30</sub> H <sub>50</sub> O <sub>2</sub>				<i>L. pedunculata</i>	(Lis-Balchin 2002)	
<b>Acides hydroxybenzoïques et dérivés</b>	dihydroxybenzoic acid di-pentoside	C <sub>17</sub> H <sub>22</sub> O <sub>12</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)	
	4-hydroxybenzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>			+	<i>L. angustifolia</i>	(Yadikar et al. 2017)	
	hydroxybenzoic acid isomer 1	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)	
	hydroxybenzoic acid isomer 2	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>			+	<i>L. stoechas</i>	(Contreras et al. 2018)	
	4-hydroxybenzoic acid 4-(6-O-sulfo)glucoside	C <sub>13</sub> H <sub>16</sub> O <sub>11</sub> S			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)	
	4-(1-hydroxy-1-methylethyl)benzoic acid	C <sub>10</sub> H <sub>12</sub> O <sub>3</sub>			+	<i>L. angustifolia</i>	(Yadikar et al. 2018)	
	methyl dihydroxybenzoic acid derivative	C <sub>18</sub> H <sub>24</sub> O <sub>13</sub>			+	<i>L. stoechas</i>	(Contreras et al. 2018)	
	protocatechuic acid (3,4-dihydroxybenzoic acid)	C <sub>7</sub> H <sub>6</sub> O <sub>4</sub>			+	<i>L. dentata</i> <i>L. x intermedia</i> <i>L. stoechas</i>	(Torras-Claveria et al. 2007; Yadikar et al. 2017; Contreras et al. 2018)	
	vanillic acid	C <sub>8</sub> H <sub>8</sub> O <sub>4</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)	
	vanilic acid sulfoquinovoside	C <sub>14</sub> H <sub>18</sub> O <sub>12</sub> S			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)	
<b>Acides hydroxycinnamiques et dérivés</b>	caffeic acid					<i>L. angustifolia</i>	(Torras-Claveria et al. 2007; Spiridon et al. 2011; Yadikar et al. 2017)	
	((E)-3-(3,4-dihydroxyphenyl)prop-2-enoic acid)	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>		+	+	+	<i>L. x intermedia</i>	
	caffeic acid derivative 1	C <sub>21</sub> H <sub>30</sub> O <sub>13</sub>				+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	caffeic acid derivative 2	C <sub>27</sub> H <sub>34</sub> O <sub>12</sub>				+	<i>L. dentata</i>	(Contreras et al. 2018)

					<i>L. stoechas</i>	
caffeic acid derivative 3	C <sub>29</sub> H <sub>24</sub> O <sub>13</sub>			+	<i>L. dentata</i>	(Contreras et al. 2018)
caffeic acid-O-glucoside isomer 1	C <sub>15</sub> H <sub>18</sub> O <sub>9</sub> <sup>(b)</sup>			+	<i>L. angustifolia</i> <i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
caffeic acid-O-glucoside isomer 2	C <sub>15</sub> H <sub>18</sub> O <sub>9</sub> <sup>(b)</sup>			+	<i>L. angustifolia</i> <i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
caffeoyl feruloyl tartaric acid	C <sub>23</sub> H <sub>20</sub> O <sub>12</sub>			+	<i>L. stoechas</i>	(Contreras et al. 2018)
caffeoyl glucose	C <sub>15</sub> H <sub>18</sub> O <sub>9</sub> <sup>(b)</sup>	+	+		<i>L. angustifolia</i>	(Areias et al. 2000)
caffeoyl hexoside derivative	C <sub>20</sub> H <sub>28</sub> O <sub>13</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
caffeoyl hexoside isomer 1	C <sub>15</sub> H <sub>18</sub> O <sub>9</sub>			+	<i>L. stoechas</i>	(Contreras et al. 2018)
caffeoyl hexoside isomer 2	C <sub>15</sub> H <sub>18</sub> O <sub>9</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
6-caffeoylsucrose	C <sub>21</sub> H <sub>28</sub> O <sub>14</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
caffeoyl tartaric acid (caftaric acid)	C <sub>13</sub> H <sub>12</sub> O <sub>9</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
chlorogenic acid (3-caffeoylquinic acid)	C <sub>16</sub> H <sub>18</sub> O <sub>9</sub>	+	+	+	<i>L. angustifolia</i>	(Lis-Balchin 2002; Torras-Claveria et al. 2007; Spiridon et al. 2011; Contreras et al. 2018)
					<i>L. dentata</i>	
					<i>L. x intermedia</i>	
					<i>L. lanata</i>	
					<i>L. minutolii</i>	
					<i>L. multifida</i>	
<i>L. pinnata</i>						
<i>L. stoechas</i>						
o-coumaric acid	C <sub>9</sub> H <sub>8</sub> O <sub>3</sub>	+		+	<i>L. angustifolia</i>	(Areias et al. 2000; Yadikar et al. 2017)
coumaric acid-O-glucoside isomer 1	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub> <sup>(b)</sup>			+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)



## Annexes

coumaric acid-O-glucoside isomer 2	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
coumaric acid hexoside	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub> <sup>(b)</sup>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016)
coumaric acid hexoside isomer 1	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
coumaric acid hexoside isomer 2	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
coumaric acid pentoside hexoside	C <sub>20</sub> H <sub>26</sub> O <sub>12</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
dicafeoyl quinic acid	C <sub>25</sub> H <sub>24</sub> O <sub>12</sub> <sup>(b)</sup>	+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)
dicafeoyl tartaric acid (chicoric acid)	C <sub>22</sub> H <sub>18</sub> O <sub>12</sub>		+	<i>L. stoechas</i>	(Contreras et al. 2018)
dihydrocaffeic acid	C <sub>9</sub> H <sub>10</sub> O <sub>4</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
(E,E)-2-(3,4-dihydroxyphenyl) ethenyl ester of 3-(3,4-dihydroxyphenyl)-2- propenoic acid	C <sub>17</sub> H <sub>14</sub> O <sub>6</sub>		Callus	<i>L. angustifolia</i>	(Banthorpe et al. 1985)
3-(3,4-dihydroxyphenyl)-2- hydroxypropanoic acid	C <sub>9</sub> H <sub>10</sub> O <sub>5</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016; Contreras et al. 2018)
3,4-dihydroxyphenyllactic acid hexoside	C <sub>15</sub> H <sub>20</sub> O <sub>10</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
4,4'-dimethoxy-2,2'-di-O-β-D- glucopyranosyl-truxinate	C <sub>32</sub> H <sub>40</sub> O <sub>18</sub>		+	<i>L. angustifolia</i>	(Ablikim et al. 2019)
ferulic acid	C <sub>10</sub> H <sub>10</sub> O <sub>4</sub>	+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)
ferulic acid derivative	C <sub>28</sub> H <sub>36</sub> O <sub>12</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
ferulic acid-4-O-glucoside	C <sub>16</sub> H <sub>20</sub> O <sub>9</sub> <sup>(b)</sup>	+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)

## Annexes

ferulic acid-O-glucoside isomer 1	C <sub>16</sub> H <sub>20</sub> O <sub>9</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
ferulic acid-O-glucoside isomer 2	C <sub>16</sub> H <sub>20</sub> O <sub>9</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
ferulic acid hexoside	C <sub>16</sub> H <sub>20</sub> O <sub>9</sub>		+	<i>L. dentata</i>	(Contreras et al. 2018)
feruloyl quinic acid	C <sub>17</sub> H <sub>20</sub> O <sub>9</sub> <sup>(b)</sup>	+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)
feruloyl tartaric acid (fertaric acid)	C <sub>14</sub> H <sub>14</sub> O <sub>9</sub>		+	<i>L. stoechas</i>	(Contreras et al. 2018)
<i>trans</i> -2-β-D-glucopyranosyloxy-4-methoxycinnamic acid	C <sub>16</sub> H <sub>20</sub> O <sub>9</sub> <sup>(b)</sup>		+	<i>L. angustifolia</i>	(Ablikim et al. 2019)
2-O-glucosilcoumaric acid	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub> <sup>(b)</sup>	+		<i>L. angustifolia</i>	(Areias et al. 2000)
2-O-β-D-glucosylosyl-4-methoxybenzene propanoic acid	C <sub>16</sub> H <sub>22</sub> O <sub>9</sub>		+	<i>L. angustifolia</i>	(Ablikim et al. 2019)
2-(β-glucosyloxy)- <i>cis</i> -cinnamic acid butyl ester	C <sub>19</sub> H <sub>26</sub> O <sub>8</sub> <sup>(b)</sup>		+	<i>L. angustifolia</i>	(Ablikim et al. 2019)
2-(β-D-glucosyloxy)- <i>trans</i> -cinnamic acid butyl ester	C <sub>19</sub> H <sub>26</sub> O <sub>8</sub> <sup>(b)</sup>		+	<i>L. angustifolia</i>	(Ablikim et al. 2019)
4-hydroxycinnamic acid ( <i>p</i> -coumaric acid)	C <sub>9</sub> H <sub>8</sub> O <sub>3</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2017; Ablikim et al. 2019)
hydroxyhydrocinnamic acid glucoside	C <sub>15</sub> H <sub>20</sub> O <sub>8</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016; Contreras et al. 2018)
β-(4-hydroxyphenyl)lactic acid	C <sub>9</sub> H <sub>10</sub> O <sub>4</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2017)
isosalvianolic acid A	C <sub>26</sub> H <sub>22</sub> O <sub>10</sub>		+	<i>L. stoechas</i>	(Contreras et al. 2018)
isosalvianolic acid C	C <sub>26</sub> H <sub>20</sub> O <sub>10</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
lavandudiphenyl A	C <sub>25</sub> H <sub>21</sub> O <sub>8</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
lavandudiphenyl B	C <sub>25</sub> H <sub>21</sub> O <sub>8</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
lavandufluoren	C <sub>26</sub> H <sub>21</sub> O <sub>9</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
lavandufurandiol	C <sub>25</sub> H <sub>19</sub> O <sub>9</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
lavandunat	C <sub>12</sub> H <sub>13</sub> O <sub>6</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2018)

## Annexes

<i>cis</i> -melilotoside	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub>		+	<i>L. angustifolia</i>	(Ablikim et al. 2019)
<i>trans</i> -melilotoside	C <sub>15</sub> H <sub>18</sub> O <sub>8</sub>			+	<i>L. angustifolia</i> (Ablikim et al. 2019)
2-methoxyphenylpropionic acid	C <sub>10</sub> H <sub>12</sub> O <sub>3</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2017)
methyl caffeate (methyl (E)-3-(3,4-dihydroxyphenyl)prop-2-enoate)	C <sub>10</sub> H <sub>10</sub> O <sub>4</sub>		+	<i>L. angustifolia</i> <i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016; Yadikar et al. 2017; Contreras et al. 2018)
<i>cis</i> -4-O-methyl caffeic acid octanol ester	C <sub>18</sub> H <sub>24</sub> O <sub>6</sub> <sup>(b)</sup>			+	<i>L. stoechas</i> (Topçu et al. 2001)
<i>trans</i> -4-O-methyl caffeic acid octanol ester	C <sub>18</sub> H <sub>24</sub> O <sub>6</sub> <sup>(b)</sup>			+	<i>L. stoechas</i> (Topçu et al. 2001)
methyl 3-(3,4-dihydroxyphenyl)propanoate	C <sub>10</sub> H <sub>12</sub> O <sub>4</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
methyl-2-hydroxy-3-(3,4-dihydroxyphenyl)propanoate	C <sub>10</sub> H <sub>12</sub> O <sub>5</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2017)
methyl-2-hydroxy-3-(4-hydroxyphenyl)propanoate	C <sub>10</sub> H <sub>12</sub> O <sub>4</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2017)
methyl-3-(2-hydroxy-4-methoxyphenyl)propanoate	C <sub>11</sub> H <sub>14</sub> O <sub>4</sub>		+	<i>L. angustifolia</i>	(Yadikar et al. 2017)
methyl melitrate	C <sub>28</sub> H <sub>24</sub> O <sub>12</sub>		+	<i>L. stoechas</i>	(Contreras et al. 2018)
methyl rosmarinate	C <sub>19</sub> H <sub>18</sub> O <sub>8</sub>		+	<i>L. dentata</i> <i>L. stoechas</i> <i>L. angustifolia</i>	(Contreras et al. 2018; Ablikim et al. 2019)
nepetoidin A ((Z,E)-[2-(3,5-dihydroxyphenyl)ethenyl] 3-(3,4-dihydroxyphenyl)-2- propanoate)	C <sub>17</sub> H <sub>14</sub> O <sub>6</sub>		+	<i>L. x allardii</i> <i>L. angustifolia</i> <i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. brevidens</i> <i>L. buchii</i> <i>L. canariensis</i>	(Grayer et al. 2003)

			<i>L. christiana</i> <i>L. coronopifolia</i> <i>L. dentata</i> <i>L. dhofarensis</i> <i>L. x heterophylla</i> <i>L. lanata</i> <i>L. latifolia</i> <i>L. mairei</i> <i>L. maroccana</i> <i>L. minutolii</i> <i>L. multifida</i> <i>L. pinnata</i> <i>L. pubescens</i> <i>L. rotundifolia</i> <i>L. stoechas</i> <i>L. tenuisecta</i> <i>L. viridis</i>	
<hr/>				
nepetoidin B ((Z,E)-[2-(3,4-dihydroxyphenyl)ethenyl] 3-(3,4-dihydroxyphenyl)-2-propenoate or (Z,E)-2-(3,4- dihydroxyphenyl)ethenyl ester of 3- (3,4-dihydroxyphenyl)-2-propenoic acid)	$C_{17}H_{14}O_6$	+	<i>L. x allardii</i> <i>L. angustifolia</i> <i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. brevidens</i> <i>L. buchii</i> <i>L. canariensis</i> <i>L. christiana</i> <i>L. coronopifolia</i> <i>L. dentata</i> <i>L. dhofarensis</i> <i>L. x heterophylla</i>	(Banthorpe et al. 1985; Lis-Balchin 2002; Grayer et al. 2003)

					<i>L. lanata</i> <i>L. latifolia</i> <i>L. mairei</i> <i>L. maroccana</i> <i>L. minutolii</i> <i>L. multifida</i> <i>L. pinnata</i> <i>L. pubescens</i> <i>L. rotundifolia</i> <i>L. stoechas</i> <i>L. tenuisecta</i> <i>L. viridis</i>	
rosmarinic acid	C <sub>18</sub> H <sub>16</sub> O <sub>8</sub>	+	+	+	<i>L. angustifolia</i> <i>L. canariensis</i> <i>L. dentata</i> <i>L. x intermedia</i> <i>L. lanata</i> <i>L. pedunculata</i> <i>L. pinnata</i> <i>L. stoechas</i>	(Areias et al. 2000; Lis-Balchin 2002; Torras-Claveria et al. 2007; Spiridon et al. 2011; Algieri et al. 2016; Yadikar et al. 2018; Contreras et al. 2018; Ablikim et al. 2019)
rosmarinic acid hexoside	C <sub>24</sub> H <sub>26</sub> O <sub>13</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
rosmarinic acid methyl ester	C <sub>19</sub> H <sub>18</sub> O <sub>8</sub>			+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
sagerinic acid	C <sub>36</sub> H <sub>32</sub> O <sub>16</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
salvianolic acid A	C <sub>26</sub> H <sub>22</sub> O <sub>10</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)

	salvianolic acid B (2R)-2-[(E)-3-[(2R,3R)-3-[1-carboxy-2-(3,4-dihydroxyphenyl)ethoxy]carbonyl-2-(3,4-dihydroxyphenyl)-7-hydroxy-2,3-dihydro-1-benzofuran-4-yl]prop-2-enoyl]oxy-3-(3,4-dihydroxyphenyl)propanoic acid)	C <sub>36</sub> H <sub>30</sub> O <sub>16</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016; Contreras et al. 2018)
	salvianolic acid C	C <sub>26</sub> H <sub>20</sub> O <sub>10</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	salvianolic acid C isomer	C <sub>26</sub> H <sub>20</sub> O <sub>10</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	3,4,5-trihydroxycinnamic acid-O-glucoside	C <sub>15</sub> H <sub>20</sub> O <sub>10</sub> <sup>(b)</sup>			+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
	yunnaneic acid D	C <sub>27</sub> H <sub>24</sub> O <sub>12</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	yunnaneic acid E isomer 1	C <sub>27</sub> H <sub>24</sub> O <sub>14</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	yunnaneic acid E isomer 2	C <sub>27</sub> H <sub>24</sub> O <sub>14</sub>			+	<i>L. dentata</i>	(Contreras et al. 2018)
	yunnaneic acid E isomer 3	C <sub>27</sub> H <sub>24</sub> O <sub>14</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	yunnaneic acid F	C <sub>29</sub> H <sub>26</sub> O <sub>14</sub>			+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016; Contreras et al. 2018)
<b>Flavanones</b>	eriodictyol-O-hexoside	C <sub>21</sub> H <sub>18</sub> O <sub>10</sub> <sup>(b)</sup>			+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
<b>Flavones</b>	apigenin (5,7-dihydroxy-2-(4-hydroxyphenyl)chromen-4-one)	C <sub>15</sub> H <sub>10</sub> O <sub>5</sub>	+	+	+	<i>L. angustifolia</i> <i>L. dentata</i> <i>L. x intermedia</i> <i>L. lanata</i>	(Upson et al. 2000; Areias et al. 2000; Lis-Balchin 2002; Torrass-Claveria et al. 2007;

Annexes

						<i>L. latifolia</i> <i>L. minutolii</i> <i>L. stoechas</i> <i>L. viridis</i>	Yadikar et al. 2017; Contreras et al. 2018)
apigenin di-C-hexoside	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>				+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016)
apigenin di-C-hexoside isomer 1	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>				+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
apigenin di-C-hexoside isomer 2	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>				+	<i>L. dentata</i>	(Contreras et al. 2018)
apigenin di-C-hexoside isomer 3	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>				+	<i>L. dentata</i>	(Contreras et al. 2018)
apigenin 7-O-glucoside	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>		+	+	+	<i>L. angustifolia</i> <i>L. dentata</i> <i>L. lanata</i> <i>L. latifolia</i> <i>L. stoechas</i> <i>L. viridis</i> <i>S. atriplicifolia</i>	(Upson et al. 2000; Areias et al. 2000; Lis- Balchin 2002; Algieri et al. 2016; Contreras et al. 2018)
apigenin 7-O-glucuronide	C <sub>21</sub> H <sub>18</sub> O <sub>11</sub>				+	<i>L. angustifolia</i> <i>L. dentata</i> <i>L. lanata</i> <i>L. latifolia</i> <i>L. stoechas</i> <i>L. viridis</i> <i>S. atriplicifolia</i>	(Upson et al. 2000; Lis- Balchin 2002; Contreras et al. 2018)
apigenin glucuronide hexoside	C <sub>27</sub> H <sub>28</sub> O <sub>16</sub>				+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
derivative of apigenin 7-O-glucoside or apigenin 7-O-glucuronide					+	<i>L. angustifolia</i> <i>L. dentata</i>	(Upson et al. 2000)

				<i>L. lanata</i> <i>L. latifolia</i> <i>L. mairei</i> <i>L. stoechas</i> <i>L. viridis</i> <i>S. atriplicifolia</i>	
apigenin C-hexoside isomer 1	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016; Contreras et al. 2018)
apigenin C-hexoside isomer 2	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Algieri et al. 2016; Contreras et al. 2018)
apigenin-O-hexoside	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
apigenin rhamnosyl glucoside		+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)
chrysoeriol (luteolin 3'-O-methy ether)	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	+		<i>L. angustifolia</i>	(Banthorpe et al. 1985)
chrysoeriol-O-glucuronide	C <sub>22</sub> H <sub>20</sub> O <sub>12</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
				<i>L. angustifolia</i> <i>L. lanata</i> <i>L. latifolia</i> <i>L. mairei</i> <i>L. stoechas</i> <i>L. viridis</i>	
chrysoeriol 7-O-glycoside			+		(Upson et al. 2000; Lis- Balchin 2002)
chrysoeriol-O-hexoside	C <sub>22</sub> H <sub>22</sub> O <sub>11</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
genkwanin (apigenin 7-methyl ether)	C <sub>16</sub> H <sub>12</sub> O <sub>5</sub>		+	<i>L. angustifolia</i> <i>L. dentata</i> <i>L. lanata</i> <i>L. latifolia</i>	(Upson et al. 2000; Lis- Balchin 2002; Algieri et al. 2016; Contreras et al. 2018)



Annexes

				<i>L. stoechas</i>	
				<i>L. viridis</i>	
hexosyl pentosyl luteolin		+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)
6-hydroxyluteolin 7-O-glycoside			+	<i>L. dentata</i>	(Upson et al. 2000; Lis-Balchin 2002)
hydroxy-trimethoxyflavone	C <sub>18</sub> H <sub>16</sub> O <sub>6</sub> <sup>(b)</sup>	+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)
hypolaetin di-glucuronide	C <sub>27</sub> H <sub>26</sub> O <sub>19</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
				<i>L. antineae</i>	
				<i>L. aristibracteata</i>	
				<i>L. bipinnata</i>	
				<i>L. buchii</i>	
				<i>L. canariensis</i>	
				<i>L. coronopifolia</i>	
				<i>L. dhofarensis</i>	
				<i>L. mairei</i>	
				<i>L. minutolii</i>	
hypolaetin 7-O-glucoside	C <sub>21</sub> H <sub>20</sub> O <sub>12</sub> <sup>(b)</sup>		+	<i>L. maroccana</i> <i>L. minutolii</i> <i>L. multifida</i> <i>L. nimmoi</i> <i>L. pinnata</i> <i>L. pubescens</i> <i>L. rotundifolia</i> <i>L. subnuda</i> <i>L. tenuisecta</i> <i>S. atriplicifolia</i> <i>S. erythrae</i>	(El-Garf et al. 1999; Upson et al. 2000; Lis-Balchin 2002)
hypolaetin 7-O-glucuronide	C <sub>21</sub> H <sub>18</sub> O <sub>13</sub> <sup>(b)</sup>		+	<i>L. dentata</i>	(Lis-Balchin 2002)

hypolaetin 8-O-glucuronide	$C_{21}H_{18}O_{13}$	+	+	<i>L. antineae</i> <i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. buchii</i> <i>L. canariensis</i> <i>L. coronopifolia</i> <i>L. dentata</i> <i>L. dhofarensis</i> <i>L. maroccana</i> <i>L. minutolii</i> <i>L. multifida</i> <i>L. nimmoi</i> <i>L. pinnata</i> <i>L. pubescens</i> <i>L. rotundifolia</i> <i>L. stoechas</i> <i>L. subnuda</i> <i>L. tenuisecta</i>	(El-Garf et al. 1999; Upson et al. 2000; Lis-Balchin 2002; Contreras et al. 2018)
derivative of hypolaetin 8-O-glucuronide			+	<i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. canariensis</i> <i>L. coronopifolia</i> <i>L. dhofarensis</i> <i>L. maroccana</i> <i>L. rotundifolia</i> <i>L. subnuda</i>	(Upson et al. 2000)
hypolaetin 4'-methyl ether 8-O-glucuronide	$C_{21}H_{20}O_{14}^{(b)}$	+	+	<i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. coronopifolia</i>	(El-Garf et al. 1999; Upson et al. 2000; Lis-Balchin 2002)

				<i>L. dhofarensis</i> <i>L. maroccana</i> <i>L. pubescens</i> <i>L. rotundifolia</i> <i>L. subnuda</i>	
				<i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. canariensis</i> <i>L. coronopifolia</i> <i>L. dentata</i> <i>L. maroccana</i> <i>L. minutolii</i> <i>L. multifida</i> <i>L. pubescens</i> <i>L. rotundifolia</i> <i>L. stoechas</i> <i>L. subnuda</i>	(El-Garf et al. 1999; Upson et al. 2000; Lis- Balchin 2002; Algieri et al. 2016; Contreras et al. 2018)
isoscuteallarein 8-O-glucuronide	C <sub>21</sub> H <sub>18</sub> O <sub>12</sub>	+	+		
				<i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. canariensis</i> <i>L. mairei</i> <i>L. maroccana</i> <i>L. minutolii</i> <i>L. multifida</i> <i>L. rotundifolia</i> <i>L. subnuda</i> <i>S. atriplicifolia</i>	(Upson et al. 2000)
isoscuteallarein 7-O-glycoside			+		
i-vitexin		+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)
luteolin	C <sub>15</sub> H <sub>10</sub> O <sub>6</sub>	+	+	<i>L. angustifolia</i>	

## Annexes

(2-(3,4-dihydroxyphenyl)-5,7-dihydroxychromen-4-one)				<i>L. dentata</i>	(Upson et al. 2000; Areias et al. 2000; Lis-Balchin 2002; Contreras et al. 2018)
				<i>L. stoechas</i>	
luteolin 7,4'-di-glucuronide	C <sub>27</sub> H <sub>26</sub> O <sub>18</sub>		+	<i>L. dentata</i>	(Lis-Balchin 2002; Algieri et al. 2016; Contreras et al. 2018)
				<i>L. stoechas</i>	
luteolin 7,4'-di-O-glucuronide	C <sub>27</sub> H <sub>26</sub> O <sub>18</sub>		+	<i>L. angustifolia</i>	
				<i>L. lanata</i>	(Upson et al. 2000)
				<i>L. latifolia</i>	
luteolin 7,3'-dimethyl ether	C <sub>17</sub> H <sub>14</sub> O <sub>6</sub>		+	<i>L. dentata</i>	(Contreras et al. 2018)
				<i>L. stoechas</i>	
				<i>L. angustifolia</i>	
				<i>L. antineae</i>	
				<i>L. bipinnata</i>	
				<i>L. buchii</i>	
				<i>L. canariensis</i>	
				<i>L. coronopifolia</i>	
				<i>L. dentata</i>	
				<i>L. dhofarensis</i>	(El-Garf et al. 1999; Upson et al. 2000; Lis-Balchin 2002; Algieri et al. 2016; Contreras et al. 2018)
luteolin 7-O-glucoside	C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>		+	<i>L. lanata</i>	
			+	<i>L. latifolia</i>	
				<i>L. mairei</i>	
				<i>L. maroccana</i>	
				<i>L. multifida</i>	
				<i>L. nimmoi</i>	
				<i>L. pinnata</i>	
				<i>L. pubescens</i>	
				<i>L. stoechas</i>	
				<i>L. rotundifolia</i>	

				<i>L. tenuisecta</i> <i>L. viridis</i> <i>S. atriplicifolia</i> <i>S. erythraea</i>	
luteolin 7-O-glucoside-4'-O-glucuronide	$C_{27}H_{28}O_{18}^{(b)}$	+		<i>L. angustifolia</i> <i>L. lanata</i> <i>L. latifolia</i>	(Upson et al. 2000; Lis-Balchin 2002)
				<i>L. angustifolia</i> <i>L. bipinnata</i> <i>L. canariensis</i> <i>L. coronopifolia</i> <i>L. dentata</i> <i>L. dhofarensis</i> <i>L. lanata</i> <i>L. latifolia</i> <i>L. mairei</i> <i>L. maroccana</i> <i>L. multifida</i> <i>L. stoechas</i> <i>L. rotundifolia</i> <i>L. viridis</i> <i>S. atriplicifolia</i> <i>S. erythraea</i>	(Upson et al. 2000; Lis-Balchin 2002; Algeri et al. 2016; Contreras et al. 2018)
luteolin 7-O-glucuronide	$C_{21}H_{18}O_{12}$	+	+		
luteolin glucuronide hexoside isomer 1	$C_{27}H_{28}O_{17}$		+	<i>L. dentata</i>	(Contreras et al. 2018)
luteolin glucuronide hexoside isomer 2	$C_{27}H_{28}O_{17}$		+	<i>L. dentata</i>	(Contreras et al. 2018)
luteolin glucuronide hexoside isomer 3	$C_{27}H_{28}O_{17}$		+	<i>L. dentata</i>	(Contreras et al. 2018)
luteolin-O-glucuronide isomer 1	$C_{21}H_{18}O_{12}^{(b)}$		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)

## Annexes

luteolin-O-glucuronide isomer 2	C <sub>21</sub> H <sub>18</sub> O <sub>12</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)	
luteolin-O-glucuronide isomer 3	C <sub>21</sub> H <sub>18</sub> O <sub>12</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)	
luteolin-O-hexoside isomer 1	C <sub>33</sub> H <sub>40</sub> O <sub>20</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)	
luteolin-O-hexoside isomer 2	C <sub>33</sub> H <sub>40</sub> O <sub>20</sub> <sup>(b)</sup>		+	<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)	
luteolin 7-rutinoside	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Upson et al. 2000; Lis-Balchin 2002)	
salvigenin (scutellarein 6,7,4'-trimethyl ether)	C <sub>18</sub> H <sub>16</sub> O <sub>6</sub>		+	<i>L. aristibracteata</i> <i>L. bipinnata</i> <i>L. minutolii</i> <i>L. stoechas</i> <i>L. viridis</i>	(Upson et al. 2000; Lis-Balchin 2002)	
scutellarein 7-O-glycoside			+	<i>L. dentata</i>	(Upson et al. 2000; Lis-Balchin 2002)	
vicenin-2 (5,7-dihydroxy-2-(4-hydroxyphenyl)- 6,8-bis[(2S,3R,4R,5S,6R)-3,4,5- trihydroxy-6-(hydroxymethyl)oxan-2- yl]chromen-4-one)	C <sub>27</sub> H <sub>30</sub> O <sub>15</sub>		+	<i>L. dentata</i>	(Upson et al. 2000; Lis-Balchin 2002)	
vitexin (apigenin 8-C-glucoside)	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>		+	<i>L. dentata</i> <i>L. stoechas</i> <i>L. viridis</i>	(Upson et al. 2000; Lis-Balchin 2002)	
xanthomicrol (5,4'-dihydroxy-6,7,8- trimetoxyflavone)	C <sub>18</sub> H <sub>16</sub> O <sub>7</sub>		+	+	<i>L. coronopifolia</i> <i>L. mairei</i> <i>L. minutolii</i> <i>L. pubescens</i>	(El-Garf et al. 1999; Upson et al. 2000; Lis-Balchin 2002)

## Annexes

Flavonols	kaempferol-3-coumaryl glucoside	$C_{30}H_{26}O_{13}^{(b)}$	+	+			<i>L. angustifolia</i>	(Areias et al. 2000)
	kaempferol malonyl glucoside	$C_{24}H_{22}O_{14}^{(b)}$	+	+			<i>L. angustifolia</i>	(Upson et al. 2000; Spiridon et al. 2011)
	quercetin-3-O-glucoside (isoquercitrin)	$C_{21}H_{20}O_{12}$			+		<i>L. x intermedia</i>	(Torras-Claveria et al. 2007)
Anthocyanidines	cyanidin 3'-(6"-p-coumarylglucoside)-5-(6'''-malonylglucoside)	$C_{39}H_{38}O_{21}^{(b)}$	+				<i>L. stoechas</i>	(Lis-Balchin 2002)
	delphinidin	$C_{15}H_{11}O_7$	+				<i>L. angustifolia</i>	(Banthorpe et al. 1985)
	delphinidin 3-(6"-p-coumarylglucoside)-5-(4''',6'''-dimalonylglucoside)	$C_{42}H_{39}O_{25}^{(b)}$	+				<i>L. dentata</i>	(Lis-Balchin 2002)
	delphinidin 3-(6"-p-coumarylglucoside)-5-glucoside	$C_{36}H_{37}O_{19}^{(b)}$	+				<i>L. stoechas</i>	(Lis-Balchin 2002)
	delphinidin 3-(6"-p-coumarylglucoside)-5-(6'''-malonylglucoside)	$C_{39}H_{38}O_{22}^{(b)}$	+				<i>L. stoechas</i>	(Lis-Balchin 2002)
	malvidin	$C_{17}H_{15}O_7$	+				<i>L. angustifolia</i>	(Banthorpe et al. 1985)
	malvidin 3-(6"-p-coumarylglucoside)-5-(dimalonylglucoside)	$C_{44}H_{43}O_{25}^{(b)}$	+				<i>L. stoechas</i>	(Lis-Balchin 2002)
	malvidin 3-(6"-p-coumarylglucoside)-5-glucoside	$C_{38}H_{41}O_{19}^{(b)}$	+				<i>L. dentata</i>	(Lis-Balchin 2002)
	malvidin 3-(6"-p-coumarylglucoside)-5-(6'''-malonylglucoside)	$C_{41}H_{42}O_{22}^{(b)}$	+				<i>L. dentata</i> <i>L. stoechas</i>	(Lis-Balchin 2002)
Coumarines	coumarin	$C_9H_6O_2$	+	+	+	+	<i>L. angustifolia</i> <i>L. dentata</i> <i>L. latifolia</i>	(Banthorpe et al. 1985; Areias et al. 2000; Lis-Balchin 2002; Yadikar et al. 2017)
	coumarin glycosided						<i>L. angustifolia</i>	(Lis-Balchin 2002)
	6,7-dihydroxycoumarin (esculetin)	$C_9H_6O_4$			+		<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
	herniarin	$C_{10}H_8O_3$	+	+	+	+	<i>L. angustifolia</i>	

Annexes

	(7-methoxycoumarin)			<i>L. dentata</i>	(Banthorpe et al. 1985; Areias et al. 2000; Lis-Balchin 2002; Yadikar et al. 2017)
				<i>L. latifolia</i>	
	herniarin glycosiled		+	<i>L. angustifolia</i>	(Lis-Balchin 2002)
	lavandupyrone A	C <sub>26</sub> H <sub>19</sub> O <sub>10</sub>	+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
	lavandupyrone B	C <sub>17</sub> H <sub>13</sub> O <sub>7</sub>	+	<i>L. angustifolia</i>	(Yadikar et al. 2018)
	umbelliferone (7-hydroxycoumarin)	C <sub>9</sub> H <sub>6</sub> O <sub>3</sub>	+	<i>L. dentata</i>	(Lis-Balchin 2002)
<b>Dérivés benzofuranes</b>	ebenfuran IV	C <sub>22</sub> H <sub>22</sub> O <sub>6</sub> <sup>(b)</sup>	+	<i>L. angustifolia</i>	(Tang et al. 2017)
	4-(5-(2-hydroxyethyl)-7-methoxy-3-methylbenzofuran-2-yl) phenol	C <sub>18</sub> H <sub>18</sub> O <sub>4</sub>	+	<i>L. angustifolia</i>	(Tang et al. 2017)
	3-hydroxy-1-(6-methoxy-1,3-dihydroisobenzofuran-5-yl) propan-1-one	C <sub>12</sub> H <sub>14</sub> O <sub>4</sub> <sup>(b)</sup>	+	<i>L. angustifolia</i>	(Zhou et al. 2017)
	5-hydroxymethyl-6-prenyl-isobenzofuran-1(3H)-one	C <sub>14</sub> H <sub>16</sub> O <sub>3</sub> <sup>(b)</sup>	+	<i>L. angustifolia</i>	(Zhou et al. 2017)
	iteafuranal B ((E)-2-(3-hydroxy-4-methoxyphenyl)-3-formyl-4-hydroxy-5-(prop-1-en-1-yl)benzofuran)	C <sub>19</sub> H <sub>16</sub> O <sub>5</sub> <sup>(b)</sup>	+	<i>L. angustifolia</i>	(Tang et al. 2017)
	lavandulactone A (6-hydroxymethyl-5-prenyl-isobenzofuran-1(3H)-one)	C <sub>14</sub> H <sub>16</sub> O <sub>3</sub>	+	<i>L. angustifolia</i>	(Zhou et al. 2017)
	lavandulactone B (6-methoxy-5-(3-methyl-2-oxobut-3-enyl)-isobenzofuran-1(3H)-one)	C <sub>14</sub> H <sub>14</sub> O <sub>4</sub>	+	<i>L. angustifolia</i>	(Zhou et al. 2017)
	lavandulactone C	C <sub>14</sub> H <sub>14</sub> O <sub>4</sub>	+	<i>L. angustifolia</i>	(Zhou et al. 2017)
	2-(2'-methoxy-4'-hydroxy)-aryl-3-methyl-6-hydroxy-benzofuran	C <sub>16</sub> H <sub>14</sub> O <sub>4</sub> <sup>(b)</sup>	+	<i>L. angustifolia</i>	(Tang et al. 2017)



## Annexes

	2-(6-methoxy-2-(4-methoxyphenyl)-3-methylbenzofuran-5-yl) ethanol	C <sub>19</sub> H <sub>20</sub> O <sub>4</sub>		+	<i>L. angustifolia</i>	(Tang et al. 2017)
	2-(7-methoxy-2-(4-methoxyphenyl)-3-methylbenzofuran-5-yl) ethanol	C <sub>19</sub> H <sub>20</sub> O <sub>4</sub>		+	<i>L. angustifolia</i>	(Tang et al. 2017)
	5-methoxy-6-(2-oxopropyl)isobenzofuran-1(3H)-one	C <sub>12</sub> H <sub>14</sub> O <sub>4</sub> <sup>(b)</sup>		+	<i>L. angustifolia</i>	(Zhou et al. 2017)
<b>Composés azotés</b>	2-N-phenylaminonaphthalene	C <sub>16</sub> H <sub>13</sub> N		+	<i>L. vera</i>	(Papanov et al. 1985; Lis-Balchin 2002)
	disaccharide	C <sub>12</sub> H <sub>22</sub> O <sub>11</sub>		+	<i>L. dentata</i> <i>L. stoechas</i>	(Contreras et al. 2018)
<b>Autres</b>	galloyl-HHDP-glucose					
	(galloyl-hexahydroxydiphenoyl-glucose)	C <sub>27</sub> H <sub>22</sub> O <sub>18</sub> <sup>(b)</sup>	+	+	<i>L. angustifolia</i>	(Spiridon et al. 2011)

<sup>(a)</sup> les noms des molécules sont en anglais

<sup>(b)</sup> Information non fournie dans les publications

Inflo.: inflorescence

Indet.: indéterminé

## Annexe 2 Tableau des solvants et réactifs employés durant cette thèse

Etude	Type	Nom	Abréviation	Caractéristique	Fournisseur
Tests préliminaires	Solvant analytique	Eau		Grade analytique HPLC-GC	Sigma-Aldrich
		Acétonitrile	AcN	Grade analytique HPLC-GC	Sigma-Aldrich
		Isopropanol	iPrOH	Grade analytique HPLC-GC	Sigma-Aldrich
	Solvant extraction	Acétate d'éthyle	AcOEt	Grade analytique HPLC-GC	Sigma-Aldrich
		Acétone	Ac	Grade analytique HPLC-GC	Sigma-Aldrich
		Dichlorométhane	DCM	Grade analytique HPLC-GC	Sigma-Aldrich
		Eau		Grade analytique HPLC-GC	Sigma-Aldrich
		Ethanol	EtOH	Grade analytique HPLC-GC	Sigma-Aldrich
		Ether diéthylique	DE	Grade analytique HPLC-GC	Sigma-Aldrich
		Hexane	H	Grade analytique HPLC-GC	Sigma-Aldrich
		Méthanol	MeOH	Grade analytique HPLC-GC	Sigma-Aldrich
	Réactifs	Acide formique	FA	Grade analytique HPLC-GC	Sigma-Aldrich
Etude métabolomique	Solvant analytique	Acétonitrile	AcN	Grade ULC/MS	Biosolve
		Eau purifiée		<i>Système Milli-Q Barnstead Easypure II ThermoScientific</i>	
	Solvant extraction	Eau		Grade analytique (LiChrosolv®) pour LC-MS	Merck
		Ethanol absolu	EtOH		Merck
	Réactifs	Acide formique (acide méthanoïque)	FA	Grade ULC/MS	Biosolve
		Leucine encéphaline (lockspray)	LE		Waters
	Standard	Caféine	Caf	Grade analytique (pureté ≥ 99,0%)	Sigma-Aldrich

## Annexe 3 Matériel végétal usité dans cette étude et sa provenance

Etapes	Matériel végétal	Provenance
Tests prétraitements des échantillons	Lot A plants n°2 et 3	Jardinerie (06)
Détermination du solvant d'extraction	Lot A plant n°1	Jardinerie (06)
Etude métabolomique	Echantillonnage en champs 2017	Plateau de Sault (84)

Annexe 4 Tableau des colonnes utilisées et leurs spécifications techniques

Instrument	Colonne	Porosité (Å)	Taille des particules (µm)	Dimension (mm) lxd	Mode de séparation	Phase	Support	Applications	Constructeur
HPLC	Luna C18	100	5	150 x 4,6	Phase inverse	C 18 avec end-capping non polaire (TMS)	Silice entièrement poreuse	Séparation des composés hydrophobiques	Phenomenex
	Kinetex C18	100	2,6	100 x 4,6		C 18 avec end-capping non polaire (TMS)	Silice de type core-shell	Polaire et apolaire dans solution aqueuse	Phenomenex
	Kinetex PFP					PFP avec end-capping non polaire (TMS)		Molécule polaire, hydrophobique, aromatique	Phenomenex
	Synergi C12 Max-RP	80	4	250 x 4,6		C 12 avec end-capping non polaire (TMS)	Silice entièrement poreuse	Favorise la rétention des composés hydrophobes et la sélectivité stérique	Phenomenex
	Synergi C18 fusion-RP					C 18 avec end-capping polaire à faible densité de ligand intégré (TMS + polaire)		Séparation de mélange de composés polaires et apolaires. Elle se distingue par une grande capacité à lier et partager des liaisons hydrogènes.	Phenomenex
UPLC	Kinetex PFP	100	1,7	100 x 2,1	Phase Inverse	PFP endcappée TMS	Silice de type core-shell	Molécule polaire, hydrophobique, aromatique	Phenomenex
	Kinetex PFP F5	100	1,7	100 x 2,1					Phenomenex

## Annexes

	Acquity UPLC HSS T3 (C18)	100	1,8	100 x 2,1		C18 trifonctionnalisé (T3) servant d'endcapping	Silice à résistance élevée	Adapté aux molécules apolaires, mais conçu pour les molécules polaires jusqu'à très polaires	Waters
--	------------------------------------	-----	-----	-----------	--	--	----------------------------------	--	--------

BEH = Ethylene Bridged Hybrid

PFP = pentafluorophenyl

TMS = triméthyl silane

HSS (= High Strength Silica)

## Annexe 5 Paramètres analytiques employés lors de la conception et de la mise en place de l'étude métabolomique

Nom méthode		Généraliste								Polaire 1					Polaire 2				
Instrument		HPLC-UV-DEDL								HPLC-UV-DEDL					HPLC-UV-DEDL				
Paramètres	Débit (ml/min)	1								1					1				
	Volume injection (µl)	10								10					10				
	Stop time (min)	60								60					60				
	Post time (min)	10								10					10				
Gradient	Temps (min)	0	5	40	50	55	68	70	0	5	40	50	60	0	5	40	50	60	
	Eau 0,1 FA (%)	95	95	55					95	95	55			95	95	65			
	AcN 0,1 FA (%)	5	5	45	100	100			5	5	45	100	100	5	5	35	100	100	
	iPrOH (%)						100	100											
Utilisation		Méthode généraliste								Fraction polaire					Fraction polaire				
Colonne		Luna C18								Kinetex PFP					Kinetex PFP				

Annexes

Nom méthode		Lipophile 1				Lipophile 2				Lipophile 3				Lipophile 4					
Instrument		HPLC-UV-DEDL				HPLC-UV-DEDL				HPLC-UV-DEDL				HPLC-UV-DEDL					
Paramètres	Débit (ml/min)	1				1				1				1					
	Volume injection (µl)	10				10				10				10					
	Stop time (min)	60				60				65				110					
	Post time (min)	10				10				10				10					
Gradient	Temps (min)	0	30	50	60	0	30	55	60	0	30	60	65	0	30	55	70	95	110
	Eau 0,1 FA (%)	70				70				70				70					
	AcN 0,1 FA (%)	30	100			30	100			30	100			30	100				
	Acétate d'éthyle (%)																	100	100
	iPrOH (%)			100	100			100	100			100	100			100	100		
Utilisation		Fraction peu polaire				Fraction peu polaire				Fraction peu polaire				Fraction peu polaire					
Colonne		Luna C18				Luna C18 Acclaim C30				Luna C18				Acclaim C30					

## Annexes

Nom		Inflo 1					Inflo 2					
Instrument		HPLC-UV-DEDL					HPLC-UV-DEDL					
Paramètres	Débit (ml/min)	1					1					
	Volume injection (µl)	10					10					
	Stop time (min)	60					60					
	Post time (min)	10					10					
Gradient	Temps (min)	0	5	40	55	60	0	5	20	40	55	60
	Eau 0,1 FA (%)	95	95	75			95	95	82	75		
	AcN 0,1 FA (%)	5	5	25	100	100	5	5	18	25	100	100
Utilisation		Fraction polaire					Fraction polaire					
Colonne		Kinetex PFP					Kinetex PFP					



Annexes

Nom méthode		UPLC_KinetexPFP_Grad01							UPLC_HSST3_Grad02							UPLC_HSST3_Grad03						
Instrument		UPLC-HRMS							UPLC-HRMS							UPLC-HRMS						
Paramètres	Débit (ml/min)	0,4							0,4							0,4						
	Volume injection (µl)	1							1							1						
	Stop time (min)	/							/							/						
	Post time (min)	(3)							(2)							(2)						
Gradient	Temps (min)	0	1,5	6	12	16	16,10	19	0	1	2,5	13	17	17,10	19	0	1	7	13	17	17,10	19
	Eau à 0,1 FA (%)	95	95	65			95	95	90	90	70			95	95	95	95	70			95	95
	AcN 0,1 FA (%)	5	5	35	100	100	5	5	10	10	30	100	100	5	5	5	5	30	100	100	5	5
Utilisation		Fraction polaire							Fraction polaire							Fraction polaire						
Colonne		Kinetex PFP							C18 HSS T3							C18 HSS T3						

## Annexes

Nom méthode		UPLC_ HSST3_Grad04								UPLC_ HSST3_Grad05								UPLC_ HSST3_Grad_final							
Instrument		UPLC-HRMS								UPLC-HRMS								UPLC-HRMS							
Paramètres	Débit (ml/min)	0,45								0,60								0,40							
	Volume injection (µl)	1								1								1							
	Stop time (min)	/								/								/							
	Post time (min)	(2)								(3)								(1,85)							
Gradient	Temps (min)	0	1	7	13	17	17,10	19	0	0,5	5	10	14	14,10	17	0	1	7	13	16,65	16,75	18,5			
	Eau à 0,1 FA (%)	95	95	70			95	95	95	95	70			95	95	95	95	70			95	95			
	AcN 0,1 FA (%)	5	5	30	100	100	5	5	5	5	30	100	100	5	5	5	5	30	100	100	5	5			
Utilisation		Fraction polaire								Fraction polaire								Fraction polaire							
Colonne		C18 HSS T3								C18 HSS T3								C18 HSS T3							

FA = acide formique (HCOOH)

AcN = acétonitrile (C<sub>2</sub>H<sub>3</sub>N)

iPrOH = isopropanol (CH<sub>3</sub>CHOHCH<sub>3</sub>)

## Annexe 6 : Matériel végétal, protocole d'extraction et d'analyse usité lors de la détermination du solvant d'extraction

### Matériel végétal :

Le matériel végétal employé est un plant adulte de *Grossa* asymptotique et supposé sain, acheté en jardinerie (plant n°1 du lot A) le 6 janvier 2017. Le plant ne présentait pas d'inflorescence. Le plant entier a été mis à sécher sur du papier à température ambiante et non exposé à la lumière directe pendant 10 jours. Les racines ont été préalablement débarrassées de leur terre et lavées à l'eau. Le plant a été partagé en trois parties : les feuilles et les tiges chlorophylliennes (F), les parties boisées (PB) et les racines (R). Les feuilles et les racines ont été broyées avec un mixeur domestique (type broyeur à couteau) Kenwood. Les parties boisées plus dures, ont été broyées à l'aide d'un broyeur à marteau (Polymix de Kinematica).

### Extraction :

L'extraction des échantillons a été réalisée à partir de 20 ml de solvant et de 200 mg de matière végétale. Les flacons sont passés brièvement au vortex (LSE vortex mixer de Corning) puis mis au bain à ultra-sons pendant 30 min. Ensuite, chaque extrait est passé sur papier filtre et le filtrat est évaporé à l'aide d'un évaporateur rotatif. La température du bain est fixée à 40 °C, hormis pour l'extrait hydroéthanolique où la température du bain a été montée à 50 °C. La température du bain est maintenue à une température inférieure à 50 °C afin d'éviter toute dégradation des composés. Une fois l'évaporation terminée, les résidus sont laissés sous hotte une nuit.

### Préparation des extraits pour analyse :

Les résidus sont dissous dans leur solvant dilution de façon à obtenir une concentration à 10 mg/ml et passés aux ultra-sons pendant 3 min pour ce faire. Dans le cas où la masse de résidu était faible, les résidus ont été dissous dans 2 ml de leur solvant d'extraction, transvasés puis évaporés sous flux d'azote (cas des extraits hexanique, diéthyle éther et acétate d'éthyle des parties boisées et des racines). Les résidus sont dissous avec le solvant de dilution correspondant pour obtenir la concentration souhaitée de 10 mg/ml ou à défaut diluer dans 100 µL (cas des extraits racines).

Tableau 41 : Correspondance entre les extraits (solvant d'extraction) et les solvants pour solubiliser les extraits avant analyse.

Solvant d'extraction	Solvant de dilution	Concentration à 10 mg/ml			Filtration		
		Feuilles	Parties boisées	Racines	Feuilles	Parties boisées	Racines
Hexane	THF / Méthanol (1/1)	oui	oui <sup>(a)</sup>	non <sup>(a)</sup>	non	non	non
Dichlorométhane		oui			oui		
Diéthyle éther		oui	oui <sup>(a)</sup>	non <sup>(a)</sup>	oui	non	non
Acétate d'éthyle		oui	oui <sup>(a)</sup>	non <sup>(a)</sup>	oui	non	non
Méthanol	Méthanol	oui	oui	oui	oui	oui	oui
Acétone		non <sup>(b)</sup>			oui		
Eau / Ethanol (1/1)	Eau / Méthanol (1/1)	oui	oui	oui	oui	oui	oui

THF = Tétrahydrofurane

## Annexes

(a) Extraits ayant dû être redilués et évaporés lors de la préparation

(b) Extrait avec une concentration de 7,35 mg/ml (erreur lors de la dilution)

Si nécessaire, les extraits ont été filtrés avant injection sur des filtres PVDF avec une porosité de 0,2  $\mu\text{m}$  et de 25 mm de diamètre. Après analyses, les extraits sont conservés au réfrigérateur.

## Annexe 7 Matériel végétal, protocole d'extraction et d'analyse pour les tests préliminaires sur le prétraitement des échantillons

### Matériel végétal :

Les tests préliminaires ont été réalisés avec des plants de grosso sans inflorescence (période hivernale) asymptomatique et supposé sain. Ces expériences ont été menées sur les parties aériennes comprenant uniquement les feuilles et les tiges chlorophylliennes du plant. Les conditions 1 à 5 ont été menées sur le plant n°2. La condition n°6 effectuée a posteriori a été réalisée à partir d'un nouveau plant (n°3).

### Matériels :

- Bain à ultra-sons : Fisher Scientific FB15049
- Centrifugeuse Sigma Fisher BioBlock Scientific 2-6.
- HPLC-UV-DEDL : Agilent technologies 1200 Series
- Colonnes : Kinetex PFP (Tableau des colonnes utilisées et leurs spécifications techniques Annexe 4)

### Extraction :

Les extractions ont été réalisées avec un mélange eau/éthanol (1/1). Six répliqués d'extractions ont été réalisés et randomisés pour chaque condition.

- 1) Prendre des tubes Eppendorf de 2 ml, noter les codes échantillon sur le côté et le couvercle.
- 2) Pour les conditions 1, 5 et 6 congeler les tubes (avec de l'azote liquide ou placer au préalable les tubes dans le congélateur). Afin de conserver les échantillons toujours congeler jusqu'à l'extraction.
- 3) Dans les tubes, peser 50 mg de matrice végétale (pour les conditions 1, 2, 3 et 4). Pour les conditions 5 et 6 introduire 86,3 mg de matrice végétale dans les tubes.
- 4) Ajouter 1,0 ml du solvant d'extraction et fermer les tubes Eppendorf.
- 5) Passer les tubes quelques secondes au vortex.
- 6) Mettre les tubes dans un bain à ultra-sons pendant 30 min sans chauffage.
- 7) Mettre les tubes Eppendorfs de façon symétrique dans la centrifugeuse et lancer la centrifugation à 4000 rpm pendant 10 min.
- 8) Récupérer les surnageants et jeter les résidus.

### Analyses par HPLC :

La durée du prétraitement est différente pour chaque condition. Cependant, on veillera au maximum à réaliser le broyage, l'extraction et à lancer les analyses de tous les répliqués d'une même condition sur une période maximale de 24 h.

- 1) Filtrer les échantillons sur filtre PVDF de porosité 0,2  $\mu\text{m}$  et de diamètre  $\varnothing = 13 \text{ mm}$ .
- 2) Mettre les extraits dans des vials.
- 3) Injecter en HPLC.

Tous les échantillons réalisés avec le mélange eau-éthanol seront analysés avec la colonne Kinetex PFP (Annexe 4) et la méthode Polaire 2 (Annexe 5). Tous les échantillons sont analysés une fois. Un blanc

## Annexes

avec le solvant d'extraction est effectué à chaque début de séquence. Après analyse, les vials sont conservés au congélateur.

## Annexe 8 Protocole broyeur à billes

**Matériels :**

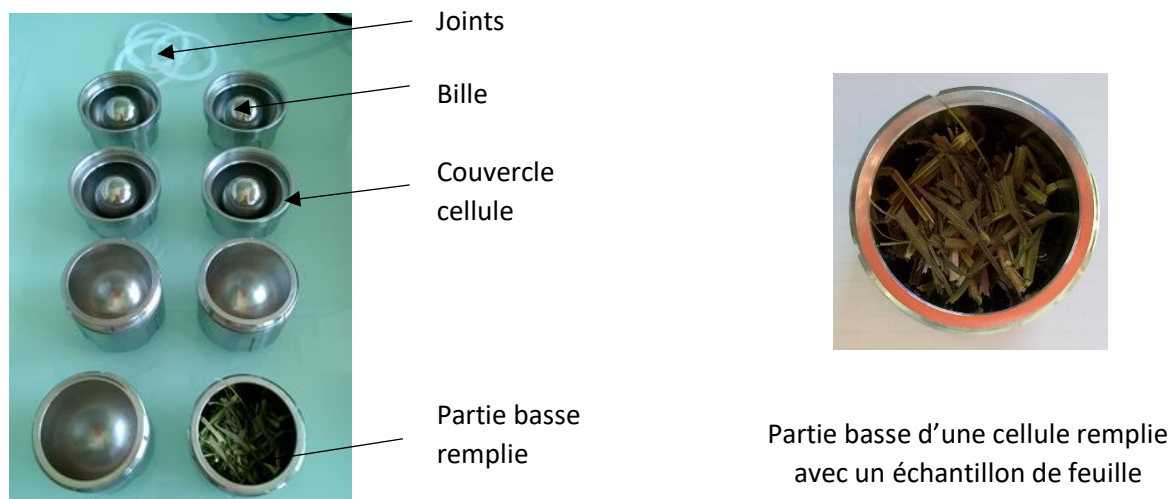
Broyeur à boules MM 301 de chez Retsch, utilisation de 4 cellules en acier de 50 ml, avec une grosse bille en acier (diamètres de 25 mm) et un joint en plastique par cellule.

**Broyage :**

Une fois les cellules propres et le joint correctement positionné (sous le pas de vis de la partie haute de la cellule). Mettre des gants propres. Remplir la partie basse de la cellule de broyage à moitié. Pour cela couper l'échantillon à la main (ou à l'aide de ciseaux secs et nettoyés à l'éthanol) en petits bouts d'environ 1 cm de longueur (surtout pour les tiges). Attention, le broyat s'agglomère au fond de la cellule et/ou est moins fin si la cellule est trop remplie. Entre chaque échantillon laver les gants en se « lavant les mains » avec de l'éthanol technique puis s'essuyer les mains avec du sopalin propre, vérifier que les gants soient bien secs. Les échantillons sont préparés par 4 et constituent une série. Les gants sont changés entre chaque série. Dans le cas où un échantillon est trop important pour être contenu dans une seule cellule, réaliser le broyage en plusieurs fois (adopter l'ordre de broyage de la série) mais conserver la même cellule (afin de limiter le nettoyage). Une légère perte de broyat lors du broyage est à noter lorsqu'une même cellule est réutilisée.

Remarque : selon la nature de l'échantillon, il est possible de remplir plus ou moins la cellule

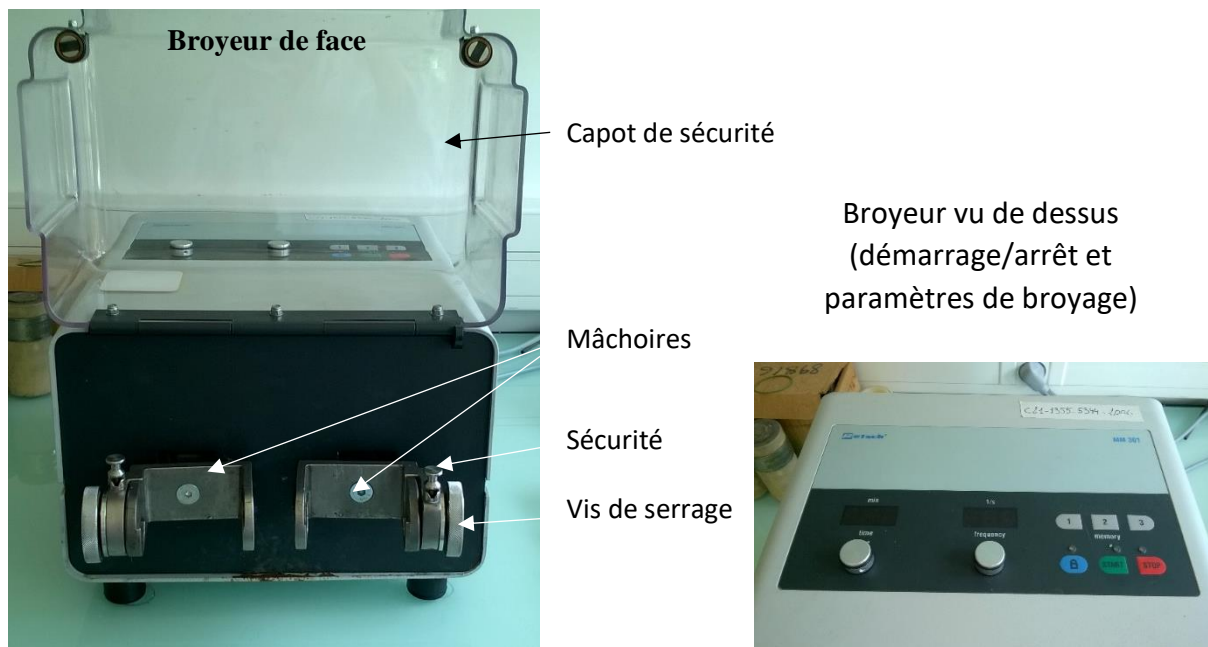
Mettre une grosse bille et fermer la cellule. Marquer les cellules avec le nom de l'échantillon pour éviter toute confusion.



Placer et fixer les cellules dans les mâchoires. Abaisser le capot de sécurité. Régler les paramètres du broyeur et démarrer le broyage.

Le temps et la fréquence de broyage utilisés pour les échantillons sont les suivants :

Type échantillon	Temps (min)	Fréquence (1/s)
Feuille-tige (F)	0.40	30.0 (max)
Inflorescence	0.30	30.0 (max)



Une fois le broyage terminé, lever le capot et enlever les cellules.

#### Récupération des échantillons :

Les échantillons sont réduits en poudre fine volatile. Il est recommandé de ne pas se mettre sous hotte pour les récupérer pour éviter leur dispersion et les pertes de matière. Il est conseillé d'utiliser un quart de feuille de papier blanche A4 pliée en son centre pour récupérer les échantillons. Se positionner au-dessus de la feuille lors de l'ouverture des cellules. A l'aide d'une spatule gratter la cellule et la bille et récupérer un maximum de quantité. Transférer l'échantillon à l'aide de la feuille de papier. Jeter la feuille après avoir récupéré l'échantillon.



#### Nettoyage :

Les cellules, les joints, les billes, et la spatule sont lavés à grandes eaux, frottés à la main ou avec une brosse pour le pas de vis. En l'absence d'air comprimé, le séchage du matériel a été fait grâce à de l'éthanol technique et à l'utilisation d'essuie-tout.

Attention, si le séchage n'est pas optimum, l'échantillon aura tendance à former des agglomérats (à coller).



## Annexe 9 Protocole d'extraction de l'étude métabolomique

Pour chaque lot d'extraction et d'analyse de l'étude métabolomique, l'ordre d'extraction est randomisé et les extractions sont réalisées successivement afin d'assurer une reproductibilité maximale.

### Préparation solvant d'extraction :

Le solvant d'extraction utilisé est un mélange eau/éthanol (1/1) en volume. Le standard interne est directement dissous dans le solvant lors de la préparation. La préparation a été renouvelée pour chaque lot d'extraction et d'analyse.

Le standard interne utilisé pour l'étude est la caféine.

La caféine employée est une caféine anhydre de grade analytique (pureté  $\geq 99,0\%$ ) produite par Sigma-Aldrich. Les solvants d'extraction sont de l'éthanol absolu et de l'eau de grade analytique (LiChrosolv®) pour LC-MS de chez Merck.

Préparation d'un total de 500 ml de solvant contenant la caféine :

- 1) Dans un flacon d'un litre, mettre 250 ml d'éthanol (plus grande solubilité de la caféine dans l'éthanol que dans l'eau).
- 2) Peser 2,5 mg de standard interne
- 3) La caféine est pesée dans un petit morceau de papier d'aluminium plié en son centre, afin de faciliter le transfert, et déposé dans une coupelle de pesée). Il est à noter qu'il est intéressant d'introduire l'un des solvants puis la caféine puis de rincer les bords du flacon avec le deuxième pour récupérer toute la caféine qui se serait éventuellement déposée sur les bords. La préparation a été réalisée dans un flacon de 1L.
- 4) Verser 250 ml d'eau (en rinçant les parois du flacon pour récupérer la caféine qui se serait éventuellement déposée sur les bords)
- 5) Fermer le flacon et agiter manuellement
- 6) Vérifier la qualité de la préparation (ie absence de contaminants) en effectuant une analyse « blanc solvant ».

Faire attention au nettoyage impeccable de la verrerie.

### Matériels extraction :

- Vortex : Wise Mix VM-10 (Serflam)
- Bain à ultra-sons : Wise Clean WUC-A06H
- Centrifugeuse : WiseSpin CF10 (Serflam)

### Protocole extraction :

- 1) Préparer le solvant d'extraction au préalable (voir ci-dessus)
- 2) Prendre des tubes Eppendorf de 2 ml, et noter les codes échantillon sur le côté et le couvercle. Manipuler les Eppendorf avec des gants nitriles pour éviter la contamination.

## Annexes

- 3) Dans les Eppendorf, à l'aide d'une petite spatule (nettoyée à l'éthanol technique entre chaque échantillon et séchée à l'air comprimé), peser 50 mg de matrice végétale (matière sèche broyée).
- 4) A l'aide d'une propipette, ajouter 1,0 ml de solvant correspondant comme indiquer dans le tableau ci-dessous et fermer les Eppendorf. Pour chaque condition respecter l'ordre de randomisation.
- 5) Passer les échantillons quelques secondes au vortex. (Environ 30 s)
- 6) Mettre les Eppendorf dans un bain à ultra-sons pendant 30 min sans chauffage.

Remarques : malgré l'absence de chauffage, les ultra-sons provoquent une augmentation de la température du bain à ultra-sons. La température est contrôlée grâce à un thermomètre digital et l'eau est renouvelée régulièrement afin de maintenir le bain à une température comprise entre 30 et 40 °C. La moitié du bac est siphonnée et remplacée par de l'eau froide (eau du robinet) dès que la température avoisine les 40°C, soit en moyenne un renouvellement toutes les 1 h 30.

Deux plaques d'extraction contenant 12 tubes chacune ont été employées afin d'extraire simultanément plusieurs échantillons<sup>1</sup>. Les plaques sont introduites à 15 min d'intervalle pour assurer un roulement.

- 7) Mettre les Eppendorf dans la centrifugeuse.
- 8) Lancer la centrifugation à 10 000 rpm pendant 5 min.
- 9) Récupérer les surnageants et jeter les résidus.
- 10) Mettre les extraits au moins une nuit dans le frigo (12 h) ou dans le congélateur à -18 °C (cette étape permet de faire précipiter les acides gras).
- 11) Centrifuger les extraits à 13 500 rpm pendant 5 min.
- 12) Récupérer le surnageant et le mettre dans un vial.

Les extraits sont conservés dans le frigo (à 4,6 °C) jusqu'à analyse puis congelés (à - 18 °C) après.

Afin de traiter dans un laps de temps relativement restreint toutes les extractions d'un même lot, les pesées, l'étiquetage des vials, et le marquage des tubes Eppendorf ont été réalisés au préalable. Toutes les manipulations sont réalisées avec des gants en nitrile pour éviter toute contamination. Les pesées ont été réalisées par lot (A, B, C puis D) dans l'ordre des échantillons.



Figure 108 : De gauche à droite, photographie du bain à ultra-sons lors de la réalisation de deux jeux d'extraction en simultanément, photographie d'une plaque d'extraction vue de dessous à droite, pool des extraits de feuille de lavande, et pool des extraits d'inflorescences de lavande.

### **Constitution des pools :**

Un pool a été réalisé pour chaque lot d'extraction et d'analyse.

A l'aide d'une micropipette, prélever 50 µl pour chaque extrait et les rassembler dans un flacon propre (de 20 ml dans notre cas). Le prélèvement doit idéalement se faire au milieu et à mi-hauteur du vial. Changer d'embout entre chaque extrait pour ne pas « contaminer » le suivant. Une fois le pool constitué le flacon est rebouché et parafilmé puis agité manuellement. Les flacons de pool sont maintenus à l'abri de la lumière (flacon ambré ou papier d'aluminium) et dans le réfrigérateur le temps des analyses puis ont été congelés à -18 °C.

Pour les analyses, le pool a été divisé en 8 vials, c'est-à-dire un vial par rack analysé. Bien que la température du rack soit maintenue constante, nous avons choisi de multiplier les vials de pool afin d'éviter tout phénomène de concentration. Pour la constitution des vials, agiter le flacon du pool manuellement, prélever 1,5 ml au centre du flacon pour chaque vial.

## Annexe 10 : Paramètres UPLC-HRMS (QTOF)

**Paramètres UHPLC-HRMS :**

Les analyses ont été effectuées sur un système Acquity UPLC® couplé à un spectromètre de masse à temps de vol XEVO-G2-QTOF de la marque Waters Corporation (Manchester, United Kingdom). Le système chromatographique a été équipé de la colonne ACQUITY UPLC HSS T3 de chez Waters (100 Å ; 1,8µm ; 2,1 x 100 mm) et de la pré colonne correspondante (VanGuard HSS T3 1,8 µm). Le volume d'injection a été fixé 1 µL. Le porte-échantillon a été maintenu à une température de 15 °C et la colonne a été réglée à 40 °C. Le débit d'élution est de 0,4 ml/min et la durée totale de l'analyse incluant le retour aux conditions initiales est de 18,5 min. L'eau (solvant A) et l'acétonitrile (solvant B) ont tous les deux été acidifiés par ajout de 0,1 % d'acide formique. L'élution débute par un plateau à 5 % de B durant une minute, suivi d'un gradient d'élution de 6 min pour atteindre 30 % de B. Le gradient d'élution se poursuit pour atteindre 100 % (B) à 13 min, un palier est observé jusqu'à 16,65 min puis un retour aux conditions initiales est effectué et maintenu pendant presque 2 min afin de nettoyer la colonne avant la prochaine injection.

Les paramètres du spectromètre de masse sont : tension du capillaire 1,85 kV ; tension du cône d'échantillon 25 V, tension du cône d'extraction 4 V, température de la source 135°C, et température de désolvatation 550°C. Le gaz de désolvatation est de l'azote et a un débit de 1000 L/h. Le gaz de collision est de l'argon et son débit est de 40L/h. Le mode d'acquisition employé est la MS<sup>E</sup>, et l'ionisation a été effectuée en mode positif. La gamme de masse employée est de 50 à 1200 Da. L'énergie de collision pour la fonction n°1 (MS1) est de 6 eV et pour la fonction 2 (MS2) une rampe de 10 à 40 V a été appliquée. La molécule employée pour le lockspray est la leucine encéphaline. La leucine encéphaline est un petit peptide permettant d'assurer la calibration du spectromètre de masse en continu. L'ion moléculaire qui produit en mode positif à un rapport  $m/z$  de 556,2771 Da.

Les données ont été acquises en mode centroïde.

**Gradient d'élution :**

Temps (min)	0	1	7	13	16,65	16,75	18,5
A (%)	95	95	70	0	0	95	95
B (%)	5	5	30	100	100	5	5

Solvant A = eau + 0,1 % d'acide formique

Solvant B = acétonitrile + 0,1 % d'acide formique

**Solvants d'analyses :**

Les solvants d'élution employés sont de l'eau et de l'acétonitrile acidifié à 0,1% d'acide formique. L'eau purifiée est obtenue par un système Milli-Q Barnstead Easypure II. L'acétonitrile et l'acide formique sont de grade ULC/MS de chez Biosolve.

## Annexe 11 : Protocole suppression des variables non significatives

La procédure est avant tout réalisée sous MetaboAnalyst et le logiciel excel, mais peut également faire appel au logiciel R. Le principe général consiste à réaliser un T-test avec une limite de p-value = 0,05. Tous les ions ayant une valeur supérieure sont supprimés.

### Procédure :

- 1) Rendre la matrice utilisable sous MetaboAnalyst (à faire sous excel), c'est-à-dire :
  - a) Mettre les échantillons en ligne ie transposer la matrice
  - b) Ajouter une colonne en deuxième position. Réaliser une jointure (voir procédure jointure ci-dessous) pour inclure les métadonnées à savoir les variétés ou l'état symptomatique des plants selon le type de matrice à réaliser.
  - c) Changer le format en .csv (séparateur point-virgule)
- 2) Normaliser la matrice (sous MetaboAnalyst)
  - a) Importer les données
  - b) Filtrer les données : aucun filtre à appliquer
  - c) Normalisation des données :
    - Sample normalization = pool group,
    - Data transformation = log
    - Data scaling = auto-scaling
- 3) Déterminer si le jeu de données suit une distribution normale
  - a) Dans MetaboAnalyst, après la normalisation regarder les graphiques de densité pour les échantillons et les ions. Si les courbes sont de forme gaussienne, les données suivent une distribution normale et un test paramétrique peut être utilisé. Dans le contraire, c'est le test non-paramétrique qui sera employé.
  - b) S'il est difficile de conclure, réaliser un test de normalité sous R (voir détails ci-dessous).
- 4) Faire un T-test (sous MetaboAnalyst)
  - a) Exporter la matrice normalisée précédemment. Sous Excel, supprimer les pools.
  - b) Ré-importer la matrice dans MetaboAnalyst.
  - c) Faire un T-test (paramétrique ou non-paramétrique selon les résultats du test sur la distribution des données) sur la matrice normalisée et sans les pools. (Par conséquent, n'appliquer aucune transformation ou normalisation ou filtration des données).
- 5) Suppression des variables non significatives :
  - a) Si le nombre de points à exclure est faible : supprimer dans la matrice de données, tous les ions ayant une p-value > 0,05 (en bleu sur le graphique)
  - a bis) Si le nombre de points à supprimer est important, exporter le fichier excel contenant la liste des ions significatifs. Faire une jointure de tableau (cf. ci-dessous).
- 6) Vérifier sur le graphique que tous les points décrits comme non-significatifs le soient bien. Pour ce faire :

## Annexes

- a) Vérifier que le nombre d'ions sur le graphique et le nombre d'ions supprimé de l'excel soient les mêmes.
- b) Sur le graphique, cliquer sur chaque point décrit comme non-significatif pour faire apparaître ses boxplots ou boîtes à moustache et confirmer que celui-ci est bien non-significatif. Si l'ion est non-significatif, les boxplots pour les différents groupes biologiques (ex : les 2 variétés) doivent être très rapprochés.

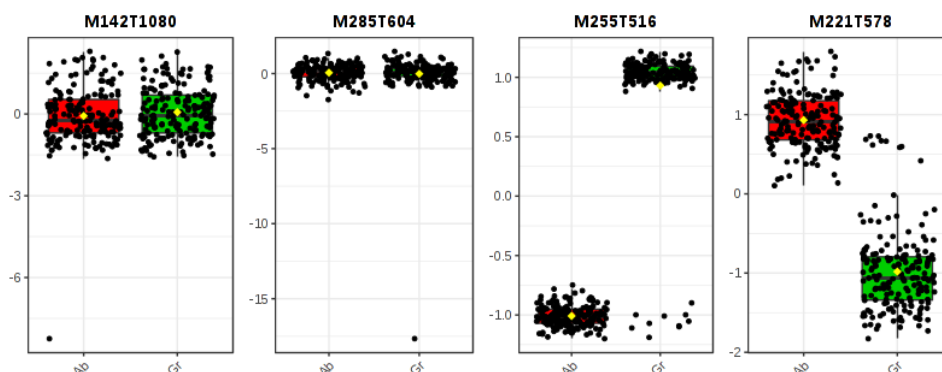


Figure 109 : A gauche, *boxplot* de deux ions non significatifs ; à droite, *boxplot* de deux ions significatifs

### Test de normalité : sous R

- 1) Extraire la matrice normalisée sous MetaboAnalyst
- 2) Préparer la matrice pour le test :
  - a) Supprimer les pools
  - b) Transposer la matrice :
    - En ligne, le nom des échantillons
    - En colonnes, les ions
  - c) Rajouter les colonnes suivantes :
    - 1 colonne = le nombre de pics
    - 1 colonne = quantité totale (somme des aires)
    - 1 colonne par catégorie = les facteurs de l'étude (*i.e.* statut ou variété)

Remarque : pour pouvoir faire les calculs, remplacer les points par des virgules dans la matrice de données.

- d) Copier et coller uniquement les valeurs des colonnes sample et les 3 colonnes rajoutées dans un nouveau tableur
- e) Enregistrer au format .csv (séparateur point-virgule) ce nouveau tableur

	A	B	C	D	E	F
1		statut	var	Area total		
2	P1708002_SE	infecte	7713	-4,20331575		
3	P1708002_SE	infecte	7713	-56,7128091		
4	P1708002_SE	sain	7713	-9,19111803		
5	P1708002_SE	infecte	7713	-25,1232731		
6	P1708002_SE	sain	7713	-25,6458948		
7	P1708002_SE	sain	7713	-5,26944123		
8	P1708002_SE	sain	7713	-8,30568971		
9	P1708002_SE	infecte	7713	-75,5458997		
10	P1708002_SE	sain	7713	-115,574519		
11	P1708002_SE	infecte	7713	-41,8651791		
12	P1708002_SE	sain	7713	-11,1623667		
13	P1708002_SE	infecte	7713	-46,9069621		
14	P1708002_SE	sain	7713	-127,22363		
15	P1708002_SE	sain	7713	35,9070236		
16	P1708002_SE	infecte	7713	-39,1190644		
17	P1708002_SE	sain	7713	4,99066492		
18	P1708002_SE	sain	7713	-3,3432464		
19	P1708002_SE	infecte	7713	-51,6226976		
20	P1708002_SE	infecte	7713	-54,0917384		
21	P1708002_SE	infecte	7713	-54,355428		
22	P1708002_SE	sain	7713	10,7861015		
23	P1708002_SE	infecte	7713	-38,9032403		
24	P1708002_SE	infecte	7713	-11,6820044		

Figure 110 : Tableur type préparé pour réaliser un test de normalité sous R

3) Dans R, appeler la matrice

- a) Utiliser la fonction : `tapply(X, Y, shapiro.test)`, pour chaque facteur d'étude et variables regardés :

```
tapply(t$Peak.total.number, t$statut, shapiro.test)
```

```
tapply(t$Peak.total.number, t$var, shapiro.test)
```

```
tapply(t$Area.total, t$statut, shapiro.test)
```

```
tapply(t$Area.total, t$var, shapiro.test)
```

4) Si la p-value > 0,05 la distribution est normale.

### Faire une jointure :

En sortie du T-test de MetaboAnalyst, une liste des ions significatifs est obtenue. Une jointure entre cette liste et la matrice précédemment corrigée est alors réalisée sous excel, afin d'obtenir une matrice ne contenant que les ions significatifs. Le principe de la jointure est d'appeler pour chaque ion significatif la ligne de la matrice de données lui correspondant.

### Procédure :

- 1) Dans un nouveau classeur, copier dans une feuille la matrice de données, dans une autre la liste des ions significatifs.
- 2) Créer une troisième feuille pour réaliser la jointure. Dans cette feuille : copier la liste des noms des échantillons (*i.e.* la première ligne de la matrice de données).
- 3) Sélectionner l'intégralité des cases de la matrice de données sauf la 1ère ligne. Nommer la matrice (en faisant clic droit puis définir un nom). Par exemple, la nommer MatriceA17. Attention à l'orthographe qui doit être strictement la même (majuscule comprise) dans la fonction.
- 4) Dans l'onglet jointure, copier dans la première colonne en-dessous de « name » (*i.e.* en A2) la liste des ions significatifs.

## Annexes

- 5) Dans un nouvel onglet ou l'onglet contenant les résultats du T-test par exemple, créer une liste d'incrémentation en ligne *i.e.* allant de 1 en 1 jusqu'au nombre total d'échantillons + 1. Il va nous permettre d'implémenter le nombre de colonnes. Par conséquent, cette liste ne sera utile qu'à partir de 2 (la 1ère colonne étant celle des ions dans la matrice finale).

Astuce : copier la première ligne de la matrice (nom des échantillons) et implémenter la ligne en dessous)

- 6) Toujours dans l'onglet jointure, utiliser la fonction RECHERCHEV (mettre l'ion que l'on cherche et bloquer la colonne ; le nom de la matrice de données ici MatriceA17 ; la liste d'implémentation précédemment créée et qui correspond au numéro de la colonne et bloquer la ligne ; FAUX car on cherche une correspondance exacte entre les deux listes d'ions).  
Ce qui nous donne pour B2 =RECHERCHEV(\$A2;MatriceA17;Implémentation!B\$2;FAUX)

- 7) Etendre la fonction en colonne puis en ligne (double clic).

- 8) Vérifier sur quelques intensités la correspondance.

- 9) Copier-coller (les valeurs) de la matrice dans un nouveau fichier et l'enregistrer.

A noter que ceci, est une description d'une procédure de jointure complète. Dans le cas de la suppression des ions significatifs, il est plus facile de réaliser la jointure inverse, c'est-à-dire de joindre les métadonnées dans la matrice plutôt que d'appeler toutes les intensités.



## Annexe 12 : Script de base pour la recherche de correspondance entre les VIP et la base de données

Ce script a été codé avec le langage Python. Les données qui peuvent ou doivent être modifiées dans le script ont été mises en violet. Alpha est la différence de masse autorisée pour qu'il y ait correspondance.

### Script :

```
alpha = 0.05
banque_de_donnee = 'BanqueCOnVs.asc'
VIP = '77Inflo.asc'
nom_fichier_sortie = 'Résultats_COnVs_77Inflo.csv'

# Composés discriminants
liste_exp = []
with open(VIP, 'r') as filin:
    for line in filin:
        line0 = line.replace('\n', '')
        line1 = line0.replace(',', ".")
        line2 = line1.split(';')
        line2[1] = float(line2[1]) - 1.007825
        liste_exp.append(line2)

# Banque de données
liste_theo = []
with open(banque_de_donnee, 'r') as filin:
    for line in filin:
        line0 = line.replace('\n', '')
        line1 = line0.replace(',', ".")
        line2 = line1.split(';')
        line2[0] = float(line2[0])
        line3 = [line2[1], line2[0]]
        liste_theo.append(line3)

search = []
for exp in liste_exp:
    for theo in liste_theo:
        masse_min = exp[1] - alpha
        masse_max = exp[1] + alpha
        if (theo[1] < masse_min) or (theo[1] > masse_max):
            continue
        else:
            search.append(f"{theo[0]};{theo[1]};{exp[0]};{exp[1]}")

with open(nom_fichier_sortie, 'w') as filout:
    for line in search:
        filout.write(line + "\n")
```

### Annexe 13 : Tableau des 50 composés les plus discriminants pour chaque matrice de données

**Caractère sensible/tolérant des variétés :**

Feuille de lavande					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M165T585	7713	2,1649	165,0915		
M155T407	7713	2,1644	155,1071	Protocatechuic acid	0,0207
M155T341	7713	2,1615	155,1071	Protocatechuic acid	0,0207
M327T705	Ma	2,1604	327,2301		
M133T571	7713	2,1586	133,1017		
M167T469	7713	2,1574	167,1071		
M151T705	Ma	2,1571	151,1122		
M287T705	Ma	2,1568	287,2377	Isoscutellarein	0,0092
				Luteolin	0,0092
				Kaempferol	0,0092
				Scutellarein	0,0102
M299T685	Ma	2,1484	299,1990		
M141T428	7713	2,1479	141,0915		
M167T274	7713	2,1470	167,0707		
M459T276	Ma	2,1267	459,1837	Coumaric acid pentoside hexoside	0,0327
M109T705	Ma	2,1267	109,0653		
M169T446	7713	2,1025	169,1228		
M127T302	7713	2,0998	127,0759		
M153T407	7713	2,0830	153,0914		
M195T715	7713	2,0513	195,1749	Methyl caffeate	0,0189
				Ferulic acid	0,0189
M151T189	7713	2,0505	151,1122		
M355T324	Ma	2,0460	355,1731		
M373T226	Ma	2,0429	373,1835		
M287T768	Ma	2,0271	287,2376	Scutellarein	0,0102
				Isoscutellarein	0,0092
				Luteolin	0,0092
				Kaempferol	0,0092
M151T255	7713	2,0250	151,1122		
M151T768	Ma	2,0235	151,1122		
M185T333_2	7713	2,0207	185,1178		
M141T347	7713	2,0161	141,0915		
M99T302	7713	1,9954	99,0810		
M357T779	7713	1,9934	357,2404		

## Annexes

M207T319	7713	1,9785	207,1385		
M355T889	Ma	1,9721	355,2996	Chlorogenic acid	0,0183
M397T443	7713	1,9701	397,1472		
M441T366	Ma	1,9695	441,1732		
M139T567	7713	1,9624	139,1122	3.4-Dihydroxybenzaldehyde	0,0156
				4-Hydroxybenzoic acid	0,0156
M153T380	7713	1,9598	153,1278		
M109T664	Ma	1,9548	109,0653		
M151T664	Ma	1,9299	151,1122		
M207T333	7713	1,9079	207,0998		
M207T360	Ma	1,9040	207,1019		
M109T341	7713	1,9008	109,1017		
M285T604	Ma	1,8987	285,0766	Apigenin 7-methyl ether	0,0017
M129T342	7713	1,8821	129,0915		
M389T809	Ma	1,8806	389,2663		
M165T373	Ma	1,8622	165,0914		
M369T768	Ma	1,8596	369,2406		
M153T519	Ma	1,8589	153,1279		
M109T768	Ma	1,8511	109,0653		
M303T765	Ma	1,8363	303,2329	Hypolaetin	0,0130
M121T549	Ma	1,8346	121,1016		
M183T517	7713	1,8225	183,1021	Dihydrocaffeic acid	0,0355
M287T813	Ma	1,8117	287,2376	Scutellarein	0,0102
				Isoscutellarein	0,0092
				Luteolin	0,0092
				Kaempferol	0,0092
M361T283	7713	1,8084	361,0924	3.4-Dihydroxyphenyllactic acid hexoside	0,0232
				Trihydroxycinnamic acid-O-glucoside	0,0211
				Rosmarinic acid (trans-rosmarinic acid)	0,0015

Inflorescence de lavande					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M109T664	Ma	2,1225	109,0654		
M167T275	7713	2,1168	167,0708		
M155T407	7713	2,1163	155,1072	Protocatechuic acid	0,0207
M299T710	7713	2,1148	299,2006		
M109T705	Ma	2,1142	109,0654		
M287T705	Ma	2,1090	287,2378	Scutellarein	0,0100
				Isoscutellarein	0,0090

## Annexes

				Luteolin	0,0090
				Kaempferol	0,0090
M327T705	Ma	2,1076	327,2303		
M155T341	7713	2,1070	155,1072	Protocatechuic acid	0,0207
M137T341	7713	2,1058	137,0966		
M167T469	7713	2,1005	167,1072		
M357T776	7713	2,0989	357,2406		
M127T303	7713	2,0943	127,0759		
M109T767	Ma	2,0914	109,0653		
M183T517	7713	2,0875	183,1021	Dihydrocaffeic acid	0,0355
M133T571	7713	2,0864	133,1017		
M165T585	7713	2,0745	165,0915		
M169T446	7713	2,0739	169,1229		
M287T767	Ma	2,0709	287,2378	Scutellarein	0,0101
				Isoscutellarein	0,0091
				Luteolin	0,0091
				Kaempferol	0,0091
M185T333_1	7713	2,0495	185,0814		
M141T348	7713	2,0350	141,0915		
M139T567	7713	2,0288	139,1122	3,4-Dihydroxy-benzaldehyde	0,0156
				4-Hydroxybenzoic acid	0,0156
M133T255	7713	2,0026	133,1017		
M141T429	7713	2,0006	141,0916		
M151T705	Ma	1,9878	151,1123		
M429T818	Ma	1,9868	429,2975		
M149T315	Ma	1,9780	149,0967		
M153T325	Ma	1,9610	153,1279		
M231T705	Ma	1,9494	231,1751		
M194T557	7713	1,9453	194,1546		
M99T303	7713	1,9154	99,0810		
M194T587	7713	1,9137	194,1547		
M151T664	Ma	1,9079	151,1123		
M305T664	Ma	1,9007	305,2486		
M133T189	7713	1,8745	133,1018		
M153T407	7713	1,8733	153,0916		
M109T739	Ma	1,8685	109,0654		
M369T767	Ma	1,8628	369,2406		
M453T720	7713	1,8573	453,3364		
M147T299	Ma	1,8500	147,0447		
M123T299	Ma	1,8444	123,0446		
M95T341_1	7713	1,8433	95,0497		
M185T333_2	7713	1,8391	185,1179		
M151T767	Ma	1,8327	151,1123		

## Annexes

M165T374	Ma	1,8292	165,0915		
M463T380	7713	1,8245	463,1233	Luteolin 7-O-glucuronide	0,0349
				Isoscutellarein 8-O-glucuronide	0,0353
				Luteolin-O-glucuronide isomer 1, 2 or 3	0,0357
M373T226	Ma	1,8192	373,1837		
M477T497	7713	1,7912	477,1392	Caffeoyl hexoside derivative	0,0226
				Chrysoeriol-O-glucuronide	0,0359
M399T711	7713	1,7903	399,2146		
M95T341_2	7713	1,7891	95,0861		
M129T342	7713	1,7869	129,0916		

Feuille de lavandin					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M121T727	Ab	2,5590	121,1017		
M195T634	Ab	2,5337	195,1385		
M219T517	Gr	2,5267	219,1751		
M237T517	Gr	2,5047	237,1857		
M191T517	Gr	2,5001	191,1800		
M237T525	Gr	2,4967	237,1856		
M221T645	Ab	2,4691	221,1908		
M221T483	Gr	2,4496	221,1906		
M219T696	Gr	2,4459	219,1749		
M219T525	Gr	2,4335	219,1749		
M221T527	Gr	2,4071	221,0815		
M219T766	Ab	2,3177	219,1750		
M361T283	Gr	2,3155	361,0922	3.4-Dihydroxyphenyllactic acid hexoside	0,0234
				Trihydroxycinnamic acid-O-glucoside	0,0213
				Rosmarinic acid (trans-rosmarinic acid)	0,0017
M181T609	Ab	2,3047	181,1229	4-(1-hydroxy-1-methylethyl)benzoic acid	0,0364
				Caffeic acid	0,0450
M255T517	Gr	2,2881	255,1962		
M237T688	Ab	2,2623	237,1857		
M155T341	Ab	2,2390	155,1072	Protocatechuic acid	0,0207
M203T727	Ab	2,2296	203,1801		
M147T727	Ab	2,1987	147,1173	Coumarin	0,0355
M93T727	Ab	2,1979	93,0705		
M235T524	Gr	2,1442	235,1699		
M235T532	Gr	2,1395	235,1700		

## Annexes

M153T609	Ab	2,1115	153,1279		
M191T525	Gr	2,1081	191,1800		
M369T768	Gr	2,1034	369,2405		
M203T750	Ab	2,0820	203,1801		
M133T571	Ab	2,0090	133,1017		
M193T623	Ab	2,0005	193,1230		
M251T638	Gr	2,0005	251,1651		
M149T283	Gr	1,9659	149,0602		
M269T638	Gr	1,9612	269,1756		
M447T399	Ab	1,9348	447,0922	Apigenin 7-O-glucuronide	0,0025
M255T525	Gr	1,9304	255,1962		
M249T689	Ab	1,8793	249,1857		
M263T702	Ab	1,8730	263,2013		
M167T274	Ab	1,8572	167,0707		
M163T283	Gr	1,8446	163,0394		
M123T235	Gr	1,7976	123,0447		
M287T806	Gr	1,7884	287,2378	Scutellarein	0,0101
				Isoscutellarein	0,0091
				Luteolin	0,0091
				Kaempferol	0,0091
M147T235	Gr	1,7862	147,0446		
M203T647	Ab	1,7849	203,1801		
M253T642	Gr	1,7802	253,1806		
M121T235	Gr	1,7746	121,0653		
M433T397	Ab	1,7451	433,1129	Apigenin 7-O-glucoside	0,0009
				Apigenin-O-hexoside	0,0006
				Apigenin C-hexoside isomer 1 or 2	0,0006
				I-vitexin	0,0005
M163T329	Gr	1,7447	163,0395		
M219T748	Ab	1,7150	219,1750		
M201T635	Ab	1,6841	201,1644		
M113T228	Ab	1,6575	113,0603		
M155T407	Ab	1,6376	155,1071	Protocatechuic acid	0,0207
M197T383	Gr	1,6296	197,1179	Methyl 3-(3,4-dihydroxyphenyl) propanoate	0,0365

Inflorescence de lavandin					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	Δm
M264T748	Gr	1,7903	264,2330		
M191T516	Gr	1,7841	191,1800		
M139T606	Gr	1,7826	139,1123	3,4-Dihydroxy-benzaldehyde	0,0155

## Annexes

				4-Hydroxybenzoic acid	0,0155
M255T516	Gr	1,7782	255,1963		
M361T283	Gr	1,7634	361,0924	3.4-Dihydroxyphenyllactic acid hexoside	0,0232
				Trihydroxycinnamic acid-O-glucoside	0,0211
				Rosmarinic acid (trans-rosmarinic acid)	0,0015
M309T646	Gr	1,7603	309,1681		
M237T516	Gr	1,7542	237,1858		
M219T525	Gr	1,7503	219,1751		
M249T688	Ab	1,7426	249,1857		
M95T748	Gr	1,7345	95,0861		
M203T482	Gr	1,7323	203,1801		
M253T524	Gr	1,7275	253,1806		
M261T646	Gr	1,7252	261,1833		
M221T578	Ab	1,7225	221,1908		
M221T482	Gr	1,7216	221,1907		
M147T646	Gr	1,7196	147,1173	Coumarin	0,0355
M203T750_1	Ab	1,7178	203,1802		
M195T634	Ab	1,7164	195,1386		
M235T532	Gr	1,7099	235,1701		
M150T748	Gr	1,7090	150,1363		
M105T646	Gr	1,7031	105,0704		
M201T516	Gr	1,6992	201,1644		
M203T614	Gr	1,6953	203,1802		
M153T519	Ab	1,6945	153,1280		
M147T750	Ab	1,6861	147,1173	Coumarin	0,0355
M153T505	Ab	1,6855	153,1279		
M149T554	Ab	1,6772	149,0965		
M263T702	Ab	1,6756	263,2015		
M185T319	Ab	1,6752	185,1178		
M137T617	Gr	1,6742	137,1330		
M121T748	Gr	1,6727	121,1017		
M209T516	Gr	1,6685	209,1906		
M191T619	Ab	1,6670	191,1436		
M113T228	Ab	1,6665	113,0603		
M181T283	Gr	1,6659	181,0501	4-(1-hydroxy-1-methylethyl) benzoic acid	0,0363
M163T606	Gr	1,6642	163,1486	Umbelliferone	0,0008
M237T525	Gr	1,6594	237,1858		
M179T628	Gr	1,6594	179,1436		
M149T748	Gr	1,6592	149,1330	Cinnamic acid	0,0358

## Annexes

M159T516	Gr	1,6528	159,1173		
M255T458	Gr	1,6526	255,1963		
M331T614	Gr	1,6498	331,1184		
M163T283	Gr	1,6492	163,0395		
M265T436	Ab	1,6430	265,1806		
M109T748	Gr	1,6429	109,1017		
M655T304	Gr	1,6405	655,1149	Hypolaetin di-glucuronide	0,0011
M149T753	Gr	1,6358	149,1330	Cinnamic acid	0,0358
M163T492	Ab	1,6344	163,0758		
M221T750	Ab	1,6336	221,1909		
M447T399	Ab	1,6333	447,0924	Apigenin 7-O-glucuronide	0,0025

### Etat sanitaire des plants (caractère symptomatique/asymptomatique) :

Feuille de lavande 7713					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M85T31	Symp	2,2278	84,9602		
M126T31	Symp	2,1496	125,9868		
M475T31	Symp	1,9844	474,8335		
M221T442	Symp	1,9582	221,1907		
M163T361	Symp	1,9402	163,0394		
M223T548	Symp	1,9137	223,1699		
M593T819	Asy	1,9077	593,2765		
M135T636	Symp	1,9047	135,0809		
M121T635	Symp	1,8916	121,0653		
M219T748	Symp	1,8872	219,1749		
M179T306	Asy	1,8859	179,0708	6.7-Dihydroxycoumarin	0,0360
M101T67	Asy	1,8721	101,0238		
M200T36	Symp	1,8667	200,0447		
M137T306	Asy	1,8576	137,0602		
M139T53	Asy	1,8565	139,0030	Hydroxybenzoic acid isomer 1	0,0375
				Hydroxybenzoic acid isomer 2	0,0371
M197T306	Asy	1,8547	197,0814	Gluconic /galactonic acid	0,0145
				Methyl 3-(3.4-dihydroxyphenyl) propanoate	0,0001
M223T609	Symp	1,8501	223,1699		
M524T36	Symp	1,8439	524,1506		
M219T31	Symp	1,8416	218,9110		
M589T31	Symp	1,8400	588,8145		
M345T31	Symp	1,8318	344,8760		
M167T272	Symp	1,8318	167,1071		
M607T779	Asy	1,7814	607,2558		



## Annexes

M263T548	Symp	1,7566	263,1626		
M179T486	Asy	1,7565	179,1436		
M256T31	Symp	1,7538	255,9447		
M151T31	Asy	1,7496	151,0357		
M129T53	Asy	1,7470	129,0187		
M237T636	Symp	1,7413	237,1856		
M139T49	Asy	1,7375	139,0031	Hydroxybenzoic acid isomer 1	0,0374
				Hydroxybenzoic acid isomer 2	0,0370
M129T49	Asy	1,7313	129,0187		
M291T608	Symp	1,7284	291,1574		
M235T475	Asy	1,6723	235,1335		
M207T702	Symp	1,6719	207,1750		
M223T634	Symp	1,6612	223,1698		
M279T777	Asy	1,6381	279,2327		
M549T409	Symp	1,6230	549,1243		
M215T53	Asy	1,6051	215,0167		
M130T31	Symp	1,6025	129,9796		
M293T38	Asy	1,6012	293,0642		
M153T147	Symp	1,5988	153,0551		
M193T253	Asy	1,5845	193,0500	Citric acid / isocitric acid	0,0142
M399T486	Asy	1,5552	399,1991		
M397T443	Symp	1,5519	397,1472		
M292T37	Asy	1,5420	292,1040		
M127T38	Symp	1,5419	127,0395		
M111T49	Asy	1,5323	111,0081		
M537T679	Asy	1,5309	537,3037		
M207T692	Symp	1,5219	207,1749		
M635T836	Asy	1,5170	635,2870		

Inflorescence de lavande 7713					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M139T54	Asy	2,1437	139,0031	Hydroxybenzoic acid isomer 1	0,0374
				Hydroxybenzoic acid isomer 2	0,0370
M181T574	Asy	2,1393	181,1230	4-(1-hydroxy-1-methylethyl) benzoic acid	0,0365
				Caffeic acid	0,0449
M169T587	Asy	2,0853	169,1229		
M401T389	Symp	2,0512	401,1599		
M107T584	Asy	2,0188	107,0861		
M109T484	Asy	2,0120	109,0653		
M533T224	Asy	2,0025	533,1881		
M353T386	Symp	1,9997	353,1027		

## Annexes

M273T386	Symp	1,9568	273,1131		
M277T676	Asy	1,9549	277,2171		
M625T803	Asy	1,9256	625,2665		
M279T777	Asy	1,9196	279,2328		
M237T636	Symp	1,9142	237,1858		
M607T779	Asy	1,8763	607,2559		
M519T443	Symp	1,8577	519,1137		
M389T386	Symp	1,8452	389,1235		
M129T54	Asy	1,8437	129,0188		
M205T389	Symp	1,8360	205,0866		
M593T819	Asy	1,8229	593,2765		
M455T797	Symp	1,8122	455,3520	3-oxo-12-ursene-28-oic acid	0,0005
				Micromeric acid	0,0020
M713T224	Asy	1,7994	713,2309	4,4'-dimethoxy-2,2'-di-O- $\beta$ -D-glucopyranosyl-truxinate	0,0016
M135T636	Symp	1,7983	135,0810		
M263T548	Symp	1,7912	263,1627		
M223T548	Symp	1,7865	223,1700		
M383T148	Symp	1,7755	383,0952		
M291T608	Symp	1,7699	291,1577		
M139T49	Asy	1,7690	139,0031	Hydroxybenzoic acid isomer 1	0,0374
				Hydroxybenzoic acid isomer 2	0,0370
M125T587	Asy	1,7639	125,0966		
M207T360	Symp	1,7593	207,1020		
M107T409	Asy	1,7454	107,0861		
M129T49	Asy	1,7273	129,0188		
M277T672	Asy	1,7140	277,2171		
M221T442	Symp	1,7009	221,1908		
M111T587	Asy	1,6970	111,0809		
M325T386	Symp	1,6903	325,1080		
M163T361	Symp	1,6793	163,0395		
M139T539	Symp	1,6780	139,0758	Hydroxybenzoic acid isomer 1	0,0353
				Hydroxybenzoic acid isomer 2	0,0357
M101T67	Asy	1,6677	101,0238		
M153T409	Asy	1,6657	153,1280		
M271T397	Symp	1,6595	271,0611	Apigenin	0,0001
M194T587	Asy	1,6432	194,1547		
M353T362	Asy	1,6343	353,1578		
M107T562	Asy	1,6307	107,0861		
M245T634	Symp	1,6285	245,1521		
M207T702	Symp	1,6268	207,1751		
M219T636	Symp	1,6109	219,1752		
M223T634	Symp	1,6085	223,1700		

## Annexes

M354T386	Symp	1,5932	354,1061		
M151T610	Asy	1,5901	151,1124		
M463T357	Symp	1,5871	463,0875	Luteolin 7-O-glucuronide	0,0009
				Isoscutellarein 8-O-glucuronide	0,0005
				Luteolin-O-glucuronide isomer 1 or 2 or 3	0,0001

Feuille de lavande maillette					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M181T609	Symp	2,3754	181,1228	4-(1-hydroxy-1-methylethyl)benzoic acid	0,0364
				Caffeic acid	0,0450
M233T604	Asy	2,2982	233,1155		
M195T635	Symp	2,2853	195,1385		
M179T486	Asy	2,2722	179,1436		
M399T486	Asy	2,2214	399,1991		
M209T278	Asy	2,1286	209,1542		
M237T688	Symp	2,0511	237,1856		
M137T400	Symp	1,9795	137,0966		
M163T361	Symp	1,9425	163,0394		
M201T31_2	Asy	1,9381	200,9735		
M85T31	Symp	1,8989	84,9602		
M253T433	Asy	1,8980	253,1417		
M153T147	Symp	1,8674	153,0551		
M335T632	Asy	1,8433	335,1838		
M287T489	Symp	1,8391	287,0558	Luteolin	0,0005
M621T871	Asy	1,8359	621,3077		
M126T31	Symp	1,8322	125,9868		
M209T375	Symp	1,7944	209,1542		
M593T819	Asy	1,7803	593,2765		
M139T539	Symp	1,7747	139,0758	Hydroxybenzoic acid isomer 1	0,0353
				Hydroxybenzoic acid isomer 2	0,0357
M309T36	Asy	1,7720	309,1301		
M193T253	Asy	1,7720	193,0500	Citric acid/isocitric acid	0,0142
M183T31	Asy	1,7678	182,9628		
M153T433	Asy	1,7578	153,1278		
M112T31	Asy	1,7573	112,0070		
M235T475	Asy	1,7523	235,1335		
M128T31	Asy	1,7509	128,0197		
M223T548	Symp	1,7452	223,1699		
M635T836	Asy	1,7342	635,2870		
M111T31	Asy	1,7320	111,0105		

## Annexes

M467T284	Symp	1,7092	467,1524		
M133T400	Symp	1,6833	133,1017		
M299T685	Symp	1,6795	299,1990		
M181T147	Symp	1,6690	181,0500	4-(1-hydroxy-1-methylethyl)benzoic acid	0,0364
M207T280	Asy	1,6497	207,1386		
M373T226	Symp	1,6460	373,1835		
M263T702	Symp	1,6428	263,2013		
M425T688	Symp	1,6388	425,3414		
M483T253	Symp	1,6283	483,1473		
M433T396	Symp	1,6211	433,1130	Apigenin 7-O-glucoside	0,0007
				Apigenin C-hexoside isomer 1 (or isomer 2)	0,0014 (0,0015)
				Apigenin-O-hexoside	0,0005
M248T36	Asy	1,6155	248,1136		
M607T779	Asy	1,6124	607,2558		
M193T399	Asy	1,6122	193,1208		
M333T628	Asy	1,6120	333,2042		
M234T36	Asy	1,5928	234,0980		
M397T443	Symp	1,5834	397,1472		
M277T676	Asy	1,5817	277,2170		
M151T31	Asy	1,5553	151,0357		
M191T446	Symp	1,5515	191,1437		
M191T400	Symp	1,5493	191,1436		

Inflorescence de lavande maillette					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M273T386	Symp	2,2459	273,1131		
M353T386	Symp	2,2346	353,1027		
M169T587	Asy	2,2115	169,1229		
M151T587	Asy	2,1617	151,1123		
M152T587	Asy	2,1217	152,1175		
M125T587	Asy	2,0852	125,0966		
M111T587	Asy	2,0550	111,0809		
M287T489	Symp	2,0407	287,0560	Luteolin	0,0005
M209T450	Asy	1,9759	209,1158		
M93T584	Asy	1,9463	93,0704		
M136T584	Asy	1,9311	136,1208		
M123T587	Asy	1,9283	123,1173		
M109T484	Asy	1,9105	109,0653		
M151T610	Asy	1,8937	151,1124		
M221T773	Asy	1,8709	221,1907		

## Annexes

M463T357	Symp	1,8429	463,0875	Luteolin 7-O-glucuronide	0,0009
				Isoscutellarein 8-O-glucuronide	0,0005
				Luteolin-O-glucuronide isomer 1, 2 or 3	0,0001
M137T718	Asy	1,8422	137,1331		
M287T705	Asy	1,8306	287,2378	Scutellarein	0,0100
				Isoscutellarein	0,0090
				Luteolin	0,0090
				Kaempferol	0,0090
M109T664	Asy	1,8212	109,0654		
M208T656	Symp	1,8151	208,1340		
M235T597	Asy	1,8064	235,1312		
M217T718	Asy	1,7904	217,1959		
M109T705	Asy	1,7798	109,0654		
M81T625	Asy	1,7785	81,0705		
M323T379	Symp	1,7732	323,0920		
M175T751	Asy	1,7709	175,1487		
M153T519	Asy	1,7635	153,1280		
M91T584	Asy	1,7568	91,0548		
M95T625	Asy	1,7517	95,0861		
M107T584	Asy	1,7509	107,0861		
M137T625	Asy	1,7466	137,1331		
M251T584	Asy	1,7366	251,1262		
M153T505	Asy	1,7358	153,1280		
M197T383	Asy	1,7188	197,1180	Methyl 3-(3.4-dihydroxyphenyl) propanoate	0,0365
M223T548	Symp	1,7167	223,1700		
M151T664	Asy	1,7136	151,1123		
M163T262	Symp	1,6997	163,0395		
M204T751	Asy	1,6923	204,1836		
M135T584	Asy	1,6852	135,1174		
M109T610	Asy	1,6596	109,0654		
M195T635	Symp	1,6548	195,1386		
M262T571	Asy	1,6415	262,2174		
M203T751	Asy	1,6298	203,1803		
M153T562	Asy	1,6246	153,1280		
M151T705	Asy	1,6114	151,1123		
M447T400	Symp	1,6079	447,0924	Apigenin 7-O-glucuronide	0,0023
M147T751	Asy	1,5914	147,1174	Coumarin	0,0354
M455T658	Symp	1,5888	455,3518	Micromeric acid	0,0022
				3-oxo-12-ursene-28-oic acid	0,0007
M93T505	Asy	1,5803	93,0705		
M81T718	Asy	1,5792	81,0705		

Feuille de lavandin abrial					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M593T819	Asy	2,2115	593,2765		
M169T512	Symp	2,1574	169,1229		
M167T549	Symp	2,1287	167,1071		
M193T253	Asy	2,1271	193,0501	Citric acid/isocitric acid	0,0142
M181T573	Asy	2,0834	181,1229	Caffeic acid	0,0449
				4-(1-hydroxy-1-methylethyl) benzoic acid	0,0365
M179T486	Asy	2,0359	179,1436		
M309T37	Asy	1,9941	309,1303		
M607T779	Asy	1,9471	607,2558		
M195T634	Symp	1,9406	195,1385		
M609T768	Asy	1,9166	609,2714		
M293T372	Asy	1,9089	293,0818		
M171T372	Asy	1,9032	171,0446		
M591T816	Asy	1,8842	591,2605		
M399T486	Asy	1,8670	399,1990		
M223T608	Symp	1,8595	223,1700		
M620T870	Asy	1,8541	620,2997		
M621T870	Asy	1,8452	621,3076		
M197T486	Asy	1,8225	197,1542		
M643T870	Asy	1,8015	643,2896		
M310T37	Asy	1,7723	310,1150		
M234T37	Asy	1,7539	234,0982		
M335T93	Asy	1,7499	335,0745		
M223T548	Symp	1,7182	223,1700		
M133T549	Symp	1,7091	133,1017		
M635T836	Asy	1,7057	635,2869		
M266T36	Asy	1,6736	266,1246		
M259T793	Asy	1,6421	259,2064		
M147T443_4	Asy	1,6329	147,2625		
M130T36	Asy	1,6261	130,0506		
M221T645	Symp	1,6163	221,1908		
M625T803	Asy	1,6123	625,2664		
M139T539	Symp	1,6018	139,0759	Hydroxybenzoic acid isomer 1	0,0353
				Hydroxybenzoic acid isomer 2	0,0357
M147T443_1	Asy	1,5895	147,0448		
M449T357	Symp	1,5887	449,1081	Methyl dihydroxybenzoic acid derivative	0,0228
				Luteolin 7-O-glucoside	0,0006
M165T544	Asy	1,5878	165,1279	o-Coumaric acid	0,0400

## Annexes

				4-Hydroxycinnamic acid	0,0400
M221T442	Symp	1,5865	221,1907		
M433T397	Symp	1,5750	433,1129	Apigenin C-hexoside isomer 2	0,0016
				Apigenin C-hexoside isomer 1	0,0015
				Apigenin 7-O-glucoside	0,0009
				Apigenin-O-hexoside	0,0006
				I-vitexin	0,0005
M177T503_3	Asy	1,5747	177,2303		
M153T609	Symp	1,5705	153,1279		
M501T270	Symp	1,5567	501,1604		
M209T278	Asy	1,5525	209,1542		
M453T615	Symp	1,5514	453,3363		
M463T356	Symp	1,5495	463,0874	Luteolin 7-O-glucuronide	0,0010
				Isoscutellarein 8-O-glucuronide	0,0006
				Luteolin-O-glucuronide isomer 1, 2 or 3	0,0002
M147T376	Asy	1,5442	147,0445		
M221T578	Symp	1,5265	221,1907		
M177T270	Symp	1,5179	177,0552		
M248T36	Asy	1,5157	248,1139		
M82T30_2	Asy	1,5147	82,0154		
M153T270	Symp	1,5146	153,0551		
M291T586	Asy	1,5022	291,1961		

Inflorescence de lavandin abrial					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M151T483	Asy	2,9174	151,1123		
M109T483	Asy	2,8491	109,0653		
M197T382	Asy	2,8319	197,1179	Methyl 3-(3.4-dihydroxyphenyl) propanoate	0,0365
M181T573	Asy	2,8025	181,1229	Caffeic acid	0,0449
				4-(1-hydroxy-1-methylethyl)benzoic acid	0,0365
M195T596	Asy	2,3258	195,1386		
M223T633	Asy	2,1341	223,1701		
M221T578	Symp	2,1236	221,1908		
M221T772	Asy	2,0479	221,1907		
M235T596	Asy	2,0371	235,1312		
M107T415	Asy	2,0273	107,0861		
M433T396	Symp	2,0231	433,1131	Apigenin 7-O-glucoside	0,0007
				Apigenin-O-hexoside	0,0004
				Apigenin C-hexoside isomer 1 or 2	0,0004

## Annexes

				I-vitexin	0,0003
M133T483	Asy	2,0178	133,1017		
M593T819	Asy	1,9799	593,2765		
M109T510	Asy	1,9657	109,0653		
M169T511	Symp	1,9325	169,1229		
M179T776	Asy	1,9295	179,1436		
M153T596	Asy	1,9046	153,1280		
M153T415	Asy	1,8855	153,1280		
M111T586	Asy	1,8854	111,0810		
M109T511	Symp	1,8828	109,1017		
M204T37	Symp	1,8642	204,0876		
M147T442_1	Asy	1,8634	147,0447		
M205T610	Symp	1,8342	205,1594		
M191T498	Asy	1,8308	191,1435		
M335T93	Asy	1,8272	335,0745		
M221T690	Asy	1,8049	221,1907		
M125T586	Asy	1,8038	125,0966		
M146T386	Symp	1,8008	146,0606		
M130T36	Asy	1,7288	130,0506		
M123T483	Asy	1,7125	123,1173		
M133T586	Asy	1,7025	133,1017		
M262T571	Asy	1,6986	262,2175		
M607T778	Asy	1,6963	607,2558		
M771T360	Symp	1,6957	771,1777		
M449T395	Symp	1,6771	449,1078	Luteolin 7-O-glucoside	0,0008
				Methyl dihydroxybenzoic acid derivative	0,0230
M99T37	Symp	1,6743	98,9849		
M251T463	Asy	1,6580	251,1650		
M277T676	Asy	1,6444	277,2170		
M196T588	Asy	1,6390	196,1702		
M116T39	Asy	1,6369	116,0714		
M331T347	Symp	1,6319	331,1548		
M279T776	Asy	1,6246	279,2327		
M353T558	Symp	1,6098	353,1578		
M293T371	Asy	1,5982	293,0818		
M91T31	Symp	1,5873	90,5266		
M441T368	Asy	1,5873	441,1731		
M463T356	Symp	1,5767	463,0877	Luteolin 7-O-glucuronide	0,0007
				Isoscutellarein 8-O-glucuronide	0,0003
				Luteolin-O-glucuronide isomer 1, 2 or 3	0,0001
M147T36	Asy	1,5765	147,0771		
M477T410	Symp	1,5655	477,1030	Chrysoeriol-O-glucuronide	0,0003



M271T396	Symp	1,5593	271,0610	Apigenin	0,0002
----------	------	--------	----------	----------	--------

Feuille de lavandin grosso					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M167T356	Symp	2,5316	167,1072		
M181T573	Asy	2,3700	181,1229	Caffeic acid	0,0449
				4-(1-hydroxy-1-methylethyl)benzoic acid	0,0365
M453T615	Asy	2,2148	453,3363		
M237T517	Symp	2,1983	237,1857		
M167T549	Symp	2,1379	167,1071		
M151T442	Symp	2,1378	151,1122		
M471T711	Asy	2,0741	471,3468		
M209T273	Symp	2,0596	209,1178		
M237T635	Symp	2,0391	237,1858		
M227T272	Symp	2,0019	227,1285		
M153T368	Symp	1,9923	153,1279		
M201T635	Symp	1,9858	201,1644		
M207T702	Symp	1,9753	207,1751		
M153T609	Symp	1,9456	153,1279		
M139T330	Symp	1,9416	139,1123	3,4-Dihydroxybenzaldehyde	0,0156
				4-Hydroxybenzoic acid	0,0156
M183T338	Symp	1,9233	183,1021	Dihydrocaffeic acid	0,0355
M137T604	Symp	1,9228	137,1330		
M211T285	Asy	1,9211	211,1699		
M285T604	Symp	1,8856	285,0765	Apigenin 7-methyl ether	0,0017
M153T177	Symp	1,8670	153,1278		
M203T750	Symp	1,8420	203,1801		
M593T819	Asy	1,7856	593,2765		
M113T90	Symp	1,7729	113,0239		
M151T278	Symp	1,7711	151,1122		
M169T512	Symp	1,7678	169,1229		
M109T538	Symp	1,7455	109,0653		
M195T245	Asy	1,7434	195,0884	Methyl caffeate	0,0215
M219T517	Symp	1,7365	219,1751		
M437T789	Asy	1,7259	437,3414		
M133T549	Symp	1,7241	133,1017		
M203T693	Symp	1,7219	203,1437		
M219T748	Symp	1,7162	219,1750		
M249T440	Symp	1,6979	249,1493		
M139T356	Symp	1,6895	139,1123	3,4-Dihydroxybenzaldehyde	0,0155
				4-Hydroxybenzoic acid	0,0155

## Annexes

M287T806	Symp	1,6721	287,2378	Scutellarein	0,0101
				Isoscutellarein	0,0091
				Luteolin	0,0091
				Kaempferol	0,0091
M455T706	Asy	1,6431	455,3519	Micromeric acid	0,0022
				3-oxo-12-ursene-28-oic acid	0,0007
M107T278	Symp	1,6329	107,0861		
M139T539	Symp	1,6324	139,0759	Hydroxybenzoic acid isomer 1	0,0353
				Hydroxybenzoic acid isomer 2	0,0357
M217T615	Symp	1,6275	217,1594		
M331T348	Asy	1,6078	331,1547		
M607T779	Asy	1,5768	607,2558		
M121T356	Symp	1,5635	121,1017		
M135T719	Asy	1,5491	135,1174		
M135T312	Symp	1,5390	135,1173		
M209T278	Asy	1,5362	209,1542		
M237T525	Symp	1,5319	237,1856		
M221T442	Symp	1,5240	221,1907		
M157T145	Symp	1,4979	157,0500		
M167T336	Symp	1,4936	167,1072		
M221T578	Symp	1,4844	221,1907		

Inflorescence de lavandin grosso					
Features	Caractère discriminant	VIP score	m/z	Correspondance banques de données	
				Proposition structure	$\Delta m$
M223T608	Symp	2,6843	223,1701		
M207T702	Symp	2,6438	207,1751		
M191T516	Symp	2,6363	191,1800		
M237T635	Symp	2,6336	237,1859		
M219T747	Symp	2,6117	219,1752		
M219T635	Symp	2,5946	219,1752		
M223T548	Symp	2,5852	223,1701		
M249T439	Symp	2,5830	249,1494		
M203T693	Symp	2,5803	203,1437		
M135T635	Symp	2,5717	135,0810		
M139T356	Symp	2,5587	139,1123	3.4-Dihydroxy-benzaldehyde	0,0156
				4-Hydroxybenzoic acid	0,0156
M183T337	Symp	2,5331	183,1022	Dihydrocaffeic acid	0,0355
M191T747	Symp	2,5312	191,1436		
M121T635	Symp	2,5311	121,0653		
M237T516	Symp	2,5179	237,1858		
M165T436	Symp	2,3985	165,0915		
M209T593	Symp	2,3771	209,1543		

## Annexes

M193T538	Symp	2,3435	193,1229		
M167T548	Symp	2,3433	167,1071		
M221T482	Symp	2,3256	221,1907		
M139T539	Symp	2,3083	139,0759	Hydroxybenzoic acid isomer 1	0,0353
				Hydroxybenzoic acid isomer 2	0,0357
M145T635	Symp	2,2988	145,1017		
M139T329	Symp	2,2895	139,1123	3.4-Dihydroxybenzaldehyde	0,0156
				4-Hydroxybenzoic acid	0,0156
M301T717	Symp	2,2441	301,2170		
M219T696	Symp	2,2349	219,1751		
M199T565	Symp	2,2279	199,1487	3-(3.4-Dihydroxyphenyl)-2-hydroxypropanoic acid	0,0332
M125T314	Symp	2,2069	125,0602		
M109T510	Symp	2,2000	109,0653		
M221T442	Symp	2,1540	221,1908		
M163T635	Symp	2,1432	163,1122	Umbelliferone	0,0356
M535T400	Symp	2,1378	535,1085	Kaempferol malonyl glucoside	0,0003
M223T633	Symp	2,1231	223,1701		
M109T511	Symp	2,1072	109,1017		
M169T511	Symp	2,0964	169,1229		
M137T610	Asy	2,0052	137,1330		
M205T753	Symp	1,9717	205,1958		
M149T753	Symp	1,9281	149,1330	Cinnamic acid	0,0358
M121T753	Symp	1,9176	121,1017		
M203T699	Asy	1,9024	203,1802		
M95T624	Asy	1,8880	95,0861		
M353T385	Symp	1,8792	353,1025		
M109T537	Symp	1,8754	109,0653		
M181T651	Asy	1,8628	181,0500	4-(1-hydroxy-1-methylethyl)benzoic acid	0,0364
M137T624	Asy	1,8380	137,1331		
M203T646	Symp	1,8364	203,1802		
M273T385	Symp	1,8274	273,1130		
M191T624	Asy	1,8239	191,1800		
M81T624	Asy	1,8228	81,0705		
M147T646	Symp	1,8224	147,1173	Coumarin	0,0355
M217T624	Asy	1,8108	217,1958		

Ma = maillette, Ab = abrial, Gr = grosso, Symp = symptomatique, Asy = asymptotique

$\Delta m$  = différence de masse entre la masse mesurée et la masse de la correspondance trouvée dans la banque de composés non volatils (CONVs)

Multi = correspondance multiple

## Références bibliographiques

- Ablikim G, Bobakulov K, Li J, et al (2019) Two new glucoside derivatives of truxinic and cinnamic acids from *Lavandula angustifolia* mill. *Natural Product Research* 1–9. <https://doi.org/10.1080/14786419.2019.1684283>
- Algieri F, Rodriguez-Nogales A, Vezza T, et al (2016) Anti-inflammatory activity of hydroalcoholic extracts of *Lavandula dentata* L. and *Lavandula stoechas* L. *Journal of Ethnopharmacology* 190:142–158. <https://doi.org/10.1016/j.jep.2016.05.063>
- Allen D, McWhinney B (2019) Quadrupole time-of-flight mass spectrometry: a paradigm shift in toxicology screening applications. *CBR* 40:135–146. <https://doi.org/10.33176/AACB-19-00023>
- Alonso A, Marsal S, Julià A (2015) Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol* 3:. <https://doi.org/10.3389/fbioe.2015.00023>
- Álvarez-Sánchez B, Priego-Capote F, Castro MDL de (2010) Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC* 29:120–127. <https://doi.org/10.1016/j.trac.2009.12.004>
- ANSES (2012) Avis de l'agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail relatif à la réalisation d'une analyse de risque phytosanitaire sur les phytoplasmes des arbres fruitiers. ANSES, Maisons-Alfort
- Areias FM, Valentão P, Andrade PB, et al (2000) HPLC-DAD analysis of phenolic compounds from lavender and its application to quality control. *Journal of Liquid Chromatography & Related Technologies* 23:2563–2572. <https://doi.org/10.1081/JLC-100100510>
- Aszyk J, Bylinski H, Namiesnik J, Kot-Wasik A (2018) Main strategies, analytical trends and challenges in LC-MS and ambient mass spectrometry-based metabolomics. *TrAC* 108:278–295. <https://doi.org/10.1016/j.trac.2018.09.010>
- Attoma A (2013) Développement de méthodes bidimensionnelles en ligne LCxLC-MS pour l'analyse de composés chargés. Thèse de doctorat en chimie, Claude Bernard Lyon 1
- Azizullah A, Rehman ZU, Ali I, et al (2014) Chlorophyll derivatives can be an efficient weapon in the fight against dengue. *Parasitol Res* 113:4321–4326. <https://doi.org/10.1007/s00436-014-4175-3>
- Bagnères A-G, Hossaert-Mckey M (eds) (2016) *Chemical Ecology*. ISTE Ltd and John Wiley & Sons, Inc
- Banerjee P, Erehman J, Gohlke B-O, et al (2015) Super Natural II - A database of natural products. *Nucleic Acids Research* 43:D935–D939. <https://doi.org/10.1093/nar/gku886>
- Banerjee S, Mazumdar S (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International Journal of Analytical Chemistry* 41. <https://doi.org/10.1155/2012/282574>
- Banthorpe DV, Bilyard HJ, Watson DG (1985) Pigment formation by callus of *Lavandula angustifolia*. *Phytochemistry* 24:2677–2680. [https://doi.org/10.1016/S0031-9422\(00\)80692-5](https://doi.org/10.1016/S0031-9422(00)80692-5)
- Barbery M (2017) Relevés météorologiques à la station de Carpentras (84) du 05/07/2017. In: meteo60. [https://www.meteo60.fr/stations-relevés/station-jour?station\\_id=-07586&date=05/07/2017](https://www.meteo60.fr/stations-relevés/station-jour?station_id=-07586&date=05/07/2017)
- Bécard S (2012) Développements méthodologiques en spectrométrie de masse et en mobilité ionique pour l'étude d'assemblage supramoléculaires en biologie. Thèse de doctorat en chimie, Strasbourg
- Beck JJ, Alborn HT, Block AK, et al (2018) Interactions Among Plants, Insects, and Microbes: Elucidation of Inter-Organismal Chemical Communications in Agricultural Ecology. *J Agric Food Chem* 66:6663–6674. <https://doi.org/10.1021/acs.jafc.8b01763>
- Bernatoniene J, Cizauskaite U, Ivanauskas L, et al (2016) Novel approaches to optimize extraction processes of ursolic, oleanolic and rosmarinic acids from *Rosmarinus officinalis* leaves. *Industrial Crops and Products* 84:72–79. <https://doi.org/10.1016/j.indcrop.2016.01.031>

## Références bibliographiques

- Bertsch A, Gröpl C, Reinert K, Kohlbacher O (2011) OpenMS and TOPP: open source software for LC-MS data analysis. In: Hamacher M, Eisenacher M, Stephan C (eds) Data Mining in Proteomics. Humana Press, Totowa, NJ, pp 353–367
- Bijttebier S, Van der Auwera A, Foubert K, et al (2016) Bridging the gap between comprehensive extraction protocols in plant metabolomics studies and method validation. *Analytica Chimica Acta* 935:136–150. <https://doi.org/10.1016/j.aca.2016.06.047>
- Bogdanovic A, Tadic V, Arsic I, et al (2016) Supercritical and high pressure subcritical fluid extraction from Lemon balm (*Melissa officinalis* L., Lamiaceae). *The Journal of Supercritical Fluids* 107:234–242. <https://doi.org/10.1016/j.supflu.2015.09.008>
- Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185–193. <https://doi.org/10.1093/bioinformatics/19.2.185>
- Bonnefille B (2017) Evaluation de l'approche métabolomique pour l'étude de la métabolisation et des effets du diclofénac chez la moule méditerranéenne. Thèse de doctorat en science de l'eau, Montpellier
- Borrás Linares I, Arráez-Román D, Herrero M, et al (2011) Comparison of different extraction procedures for the comprehensive characterization of bioactive phenolic compounds in *Rosmarinus officinalis* by reversed-phase high-performance liquid chromatography with diode array detection coupled to electrospray time-of-flight mass spectrometry. *Journal of Chromatography A* 1218:7682–7690. <https://doi.org/10.1016/j.chroma.2011.07.021>
- Breton Funes JL, Jaudenes Ruiz de Atauri I (1986) Triterpenes from *Lavandula canariensis*. *Journal of Natural Products* 49:937–949
- Broadhurst D, Goodacre R, Reinke SN, et al (2018) Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* 14:17. <https://doi.org/10.1007/s11306-018-1367-3>
- Brotas V, Plante-Cuny M-R (1996) Identification et quantification des pigments chlorophylliens et caroténoïdes des sédiments marins : un protocole d'analyse par HPLC. *Oceanologica Acta* 19:623–634
- Brown M, Wedge DC, Goodacre R, et al (2011) Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. *Bioinformatics* 27:1108–1112. <https://doi.org/10.1093/bioinformatics/btr079>
- Buckingham J, Munasinghe VRN (2015) Dictionary of flavonoids, Taylor&Francis Group. CRC Press
- Cabana R, Silva LR, Valentão P, et al (2013) Effect of different extraction methodologies on the recovery of bioactive metabolites from *Satureja parvifolia* (Phil.) Epling (Lamiaceae). *Industrial Crops and Products* 48:49–56. <https://doi.org/10.1016/j.indcrop.2013.04.003>
- Caspi R, Billington R, Keseler IM, et al (2020) The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Research* 48:D445–D453. <https://doi.org/10.1093/nar/gkz862>
- Chaillan C (2013) Production et marchés des huiles essentielles de lavandes et lavandins. Les synthèses de FranceAgriMer - Plantes à parfum, aromatiques et médicinales
- Chatelard E, Proffit M, Nicolè F, et al (2017) Etude des perceptions olfactives chez la cicadelle *Hyalesthes obsoletus*, hôte intermédiaire de la maladie du dépérissement des lavandes. Université de Saint-Etienne
- Chee C-F, Lee HB, Ong HC, Ho AS-H (2005) Photocytotoxic pheophorbide-related compounds from *Aglaonema simplex*. *C&B* 2:1648–1655. <https://doi.org/10.1002/cbdv.200590134>
- Chen J, Liu Z, Fan S, et al (2014) Combined application of NMR- and GC-MS-based metabolomics yields a superior urinary biomarker panel for bipolar disorder. *Sci Rep* 4:5855. <https://doi.org/10.1038/srep05855>
- Chen K, Ríos JJ, Roca M, Pérez-Gálvez A (2015) Development of an accurate and high-throughput methodology for structural comprehension of chlorophylls derivatives. (II) Dephytylated

## Références bibliographiques

- derivatives. *Journal of Chromatography A* 1412:90–99.  
<https://doi.org/10.1016/j.chroma.2015.08.007>
- Chernushevich IV, Loboda AV, Thomson BA (2001) An introduction to quadrupole–time-of-flight mass spectrometry. *J Mass Spectrom* 36:849–865. <https://doi.org/10.1002/jms.207>
- Chetwynd AJ, David A (2018) A review of nanoscale LC-ESI for metabolomics and its potential to enhance the metabolome coverage. *Talanta* 182:380–390.  
<https://doi.org/10.1016/j.talanta.2018.01.084>
- Chirita R-I (2009) Développement de nouvelles méthodes séparatives compatibles avec une détection par spectrométrie de masse et par électrochimie pour l'analyse des traces de catécholamines et molécules apparentées. Thèse de doctorat en chimie, d'Orléans
- Chong J, Wishart DS, Xia J (2019) Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Current Protocols in Bioinformatics* 68:.  
<https://doi.org/10.1002/cpbi.86>
- Cohen J (1992) Statistical Power Analysis. *Curr Dir Psychol Sci* 1:98–101.  
<https://doi.org/10.1111/1467-8721.ep10768783>
- Conord C, Despinasse Y, Moja S, et al (2012) Menaces sur l'or bleu - étude de la diversité génétique des lavandes fines sauvages et cultivées. Limoges
- Contreras M del M, Algieri F, Rodriguez-Nogales A, et al (2018) Phytochemical profiling of anti-inflammatory *Lavandula* extracts via RP-HPLC-DAD-QTOF-MS and -MS/MS: assessment of their qualitative and quantitative differences. *ELECTROPHORESIS* 39:1284–1293.  
<https://doi.org/10.1002/elps.201700393>
- Cottret L, Frainay C, Chazalviel M, et al (2018) MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Research* 46:W495–W502.  
<https://doi.org/10.1093/nar/gky301>
- Courant F, Antignac J-P, Dervilly-Pinel G, Le Bizec B (2014) Basics of mass spectrometry based metabolomics. *Proteomics* 14:2369–2388. <https://doi.org/10.1002/pmic.201400255>
- Courant F, Giraudeau P (2018) La métabolomique: enjeux technologiques et scientifiques - Démarche analytique et technologiques: du plan d'expérience au traitement des données
- Creek DJ, Jankevics A, Burgess KEV, et al (2012) IDEOM: an Excel interface for analysis of LC–MS-based metabolomics data. *Bioinformatics* 28:1048–1049.  
<https://doi.org/10.1093/bioinformatics/bts069>
- CRIEPPAM (2013) Mesure de l'appétence et de l'attractivité des lavande et lavandins vis-à-vis de *Hyalesthes obsoletus*. CRIEPPAM
- CRIEPPAM (2012) Mesure de l'appétence et de l'attractivité des lavande et lavandins vis-à-vis de *Hyalesthes obsoletus*. CRIEPPAM
- CRIEPPAM (2019) Compte-rendu technique CRIEPPAM 2019 - Réduire l'impact des maladies en lavanderaies. CRIEPPAM
- Croville G (2017) Séquençage et PCR à haut débit : application à la détection et la caractérisation d'agents pathogènes respiratoires aviaires et au contrôle de pureté microbiologique des vaccins. Thèse de doctorat en science de la santé - Biologie cellulaire et moléculaire, Paris Sciences et Lettres
- Crozier A, Clifford MN, Ashiara H (eds) (2006) Plant secondary metabolites: occurrence, structure and role in the human diet
- Cui Q, Lewis IA, Hegeman AD, et al (2008) Metabolite identification via the Madison Metabolomics Consortium Database. *Nature Biotechnology* 26:162–164. <https://doi.org/10.1038/nbt0208-162>
- Cuyckens F, Claeys M (2004) Mass spectrometry in the structural analysis of flavonoids. *J Mass Spectrom* 39:1–15. <https://doi.org/DOI:10.1002/jms.585>
- Davidson RL, Weber RJM, Liu H, et al (2016) Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. *GigaSci* 5:10. <https://doi.org/10.1186/s13742-016-0115-8>

## Références bibliographiques

- Davis GDJ, Vasanthi AHR (2011) Seaweed metabolite database (SWMD): a database of natural compounds from marine algae. *Bioinformatics* 5:361–364
- de Pascual-T J, Caballero E, Caballero C, Machin G (1983) Constituents of the essential oil of *Lavandula latifolia*. *Phytochemistry* 22:1033–1034. [https://doi.org/10.1016/0031-9422\(83\)85052-3](https://doi.org/10.1016/0031-9422(83)85052-3)
- De Rijke E, Out P, Niessen WMA, et al (2006) Analytical separation and detection methods for flavonoids. *J Chromatogr A* 1112:31–63. <https://doi.org/doi:10.1016/j.chroma.2006.01.019>
- Degtyarenko K, de Matos P, Ennis M, et al (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research* 36:D344–D350. <https://doi.org/10.1093/nar/gkm791>
- DeHaven CD, Evans AM, Dai H, Lawton KA (2010) Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J Cheminform* 2:12. <https://doi.org/10.1186/1758-2946-2-9>
- Délégation nationale de FranceAgriMer (Volx) (2020) Marché des plantes à parfum, aromatiques et médicinales - Panorama 2018. FranceAgriMer
- Demarque DP, Dusi RG, de Sousa FDM, et al (2020) Mass spectrometry-based metabolomics approach in the isolation of bioactive natural products. *Sci Rep* 10:1051. <https://doi.org/10.1038/s41598-020-58046-y>
- Despinasse Y (2015) Diversité chimique et caractérisation de l'impact du stress hydrique chez les lavandes. Thèse de doctorat en science - Biologie végétale, Saint Etienne - Jean Monnet
- Dietmair S, Timmins NE, Gray PP, et al (2010) Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. *Analytical Biochemistry* 404:155–164. <https://doi.org/10.1016/j.ab.2010.04.031>
- Draper J, Enot DP, Parker D, et al (2009) Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour “rules.” *BMC Bioinformatics* 10:227. <https://doi.org/10.1186/1471-2105-10-227>
- Dudzik D, Barbas-Bernardos C, García A, Barbas C (2018) Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *Journal of Pharmaceutical and Biomedical Analysis* 147:149–173. <https://doi.org/10.1016/j.jpba.2017.07.044>
- El-Aneed A, Cohen A, Banoub J (2009) Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Applied Spectroscopy Reviews* 44:210–230. <https://doi.org/10.1080/05704920902717872>
- El-Garf I, Grayer RJ, Kite GC, Veitch NC (1999) Hypolaetin 8-O-glucuronide and related flavonoids from *Lavandula coronopifolia* and *L. pubescens*. *Biochemical Systematics and Ecology* 27:4
- Engel J, Gerretzen J, Szymańska E, et al (2013) Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry* 50:96–106. <https://doi.org/10.1016/j.trac.2013.04.015>
- Ernst M, Silva DB, Silva RR, et al (2014) Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing. *Nat Prod Rep* 31:784. <https://doi.org/10.1039/c3np70086k>
- Ettre LS (1993) Nomenclature for chromatography (IUPAC recommendation 1993). *Pure & Appl Chem* 65:819–872
- Fahy E, Sud M, Cotter D, Subramaniam S (2007) LIPID MAPS online tools for lipid research. *Nucleic Acids Research* 35:W606–W612. <https://doi.org/10.1093/nar/gkm324>
- Fan TW-M (2012) Considerations of sample preparation for metabolomics investigation. In: Fan TW-M, Lane AN, Higashi RM (eds) *The Handbook of Metabolomics*. Humana Press, Totowa, NJ, pp 7–27
- Farag MA, Huhman DV, Lei Z, Sumner LW (2007) Metabolic profiling and systematic identification of flavonoids and isoflavonoids in roots and cell suspension cultures of *Medicago truncatula* using HPLC–UV–ESI–MS and GC–MS. *Phytochemistry* 68:342–354. <https://doi.org/10.1016/j.phytochem.2006.10.023>
- Favre L (2017) Caractérisation par analyse métabolomique de biomarqueurs bactériens au sein de biofilms marins. Thèse de doctorat en biochimie, Toulon

## Références bibliographiques

- Fekete S, Olah E, Fekete J (2012) Fast liquid chromatography: the domination of core-shell and very fine particles. *Journal of Chromatography A* 1228:57–71.  
<https://doi.org/10.1016/j.chroma.2011.09.050>
- Fernández-Albert F, Llorach R, Andrés-Lacueva C, Perera A (2014) An R package to analyse LC/MS metabolomic data: MAIT (Metabolite Automatic Identification Toolkit). *Bioinformatics* 30:1937–1939. <https://doi.org/10.1093/bioinformatics/btu136>
- Fotakis C, Tsigirimi D, Tsiaka T, et al (2016) Metabolic and antioxidant profiles of herbal infusions and decoctions. *Food Chemistry* 211:963–971.  
<https://doi.org/10.1016/j.foodchem.2016.05.124>
- Fox Ramos AE, Evanno L, Poupon E, et al (2019) Natural products targeting strategies involving molecular networking: different manners, one goal. *Nat Prod Rep* 36:960–980.  
<https://doi.org/10.1039/C9NP00006B>
- Francis F (2003) Interactions tritrophiques : étude du modèle Brassicaceae - Pucerons - Coccinelle prédatrice. Thèse de doctorat en sciences agronomiques et ingénierie biologique, des sciences agronomiques de Gembloux
- Furey A, Moriarty M, Bane V, et al (2013) Ion suppression; a critical review on causes, evaluation, prevention and applications. *Talanta* 115:104–122.  
<https://doi.org/10.1016/j.talanta.2013.03.048>
- Gampe N, Darcsi A, Nagyné Nedves A, et al (2019) Phytochemical analysis of *Ononis arvensis* L. by liquid chromatography coupled with mass spectrometry. *J Mass Spectrom* 54:121–133.  
<https://doi.org/10.1002/jms.4308>
- Gaskell SJ (1997) Electrospray: principles and practice. *JOURNAL OF MASS SPECTROMETRY* 32:677–688
- Gasparich GE (2010) Spiroplasmas and phytoplasmas: Microbes associated with plant hosts. *Biologicals* 38:193–203. <https://doi.org/10.1016/j.biologicals.2009.11.007>
- Gaude E, Chignola F, Spiliotopoulos D, et al (2013) *muma*, an R package for metabolomics univariate and multivariate statistical analysis. *CMB* 1:180–189.  
<https://doi.org/10.2174/2213235X11301020005>
- Ghasemi Pirbalouti A, Oraie M, Pouriamehr M, Babadi ES (2013) Effects of drying methods on qualitative and quantitative of the essential oil of Bakhtiari savory (*Satureja bachtiarica* Bunge.). *Industrial Crops and Products* 46:324–327.  
<https://doi.org/10.1016/j.indcrop.2013.02.014>
- Giacomini F, Le Corguille G, Monsoor M, et al (2015) Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* 31:1493–1495.  
<https://doi.org/10.1093/bioinformatics/btu813>
- Giray FH (2018) An Analysis of world lavender oil markets and lessons for Turkey. *Journal of Essential Oil Bearing Plants* 21:1612–1623. <https://doi.org/10.1080/0972060X.2019.1574612>
- Gnonlonfin GJB, Sanni A, Brimer L (2012) Review Scopoletin – A coumarin phytoalexin with medicinal properties. *Critical Reviews in Plant Sciences* 31:47–56.  
<https://doi.org/10.1080/07352689.2011.616039>
- Goleniowski M, Bonfill M, Cusido R, Palazón J (2013) Phenolic acids. In: Ramawat KG, Mérillon J-M (eds) *Natural products*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1951–1973
- Gonella E, Negri I, Marzorati M, et al (2011) Bacterial endosymbiont localization in *Hyalesthes obsoletus*, the insect vector of Bois noir in *Vitis vinifera*. *Appl Environ Microbiol* 77:1423–1435. <https://doi.org/10.1128/AEM.02121-10>
- Gorrochategui E, Jaumot J, Lacorte S, Tauler R (2016) Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: overview and workflow. *TrAC* 82:425–442.  
<https://doi.org/10.1016/j.trac.2016.07.004>
- Gougeon RD, Lucio M, Frommberger M, et al (2009) The chemodiversity of wines can reveal a metabologeography expression of cooperage oak wood. *PNAS* 106:9174–9179.  
<https://doi.org/10.1073/pnas.0901100106>



## Références bibliographiques

- Grand P (2009) *Lavandula Angustifolia* Mill. et son huile essentielle. Thèse d'exercice de docteur en pharmacie, de Nantes - Faculté de pharmacie
- Grayer RJ, Eckert MR, Veitch NC, et al (2003) The chemotaxonomic significance of two bioactive caffeic acid esters, nepetoidins A and B, in the Lamiaceae. *Phytochemistry* 64:519–528. [https://doi.org/10.1016/S0031-9422\(03\)00192-4](https://doi.org/10.1016/S0031-9422(03)00192-4)
- Gromski PS, Muhamadali H, Ellis DI, et al (2015) A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta* 879:10–23. <https://doi.org/10.1016/j.aca.2015.02.012>
- Guillarme D (2014) Évolutions majeures en chromatographie liquide. Techniques de l'ingénieur
- Guillén MD, Cabo N, Burillo J (1996) Characterisation of the essential oils of some cultivated aromatic plants of industrial interest. *Journal of the Science of Food and Agriculture* 70:359–363. [https://doi.org/10.1002/\(SICI\)1097-0010\(199603\)70:3<359::AID-JSFA512>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0010(199603)70:3<359::AID-JSFA512>3.0.CO;2-0)
- Guitton Y (2010) Diversité des composés terpéniques volatils au sein du genre *Lavandula* : aspects évolutifs et physiologiques. Thèse de doctorat en science - Biologie et Physiologie Végétale, de Saint-Etienne - Jean-Monnet
- Guitton Y, Tremblay-Franco M, Le Corguillé G, et al (2017) Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics. *The International Journal of Biochemistry & Cell Biology* 93:89–101. <https://doi.org/10.1016/j.biocel.2017.07.002>
- Hastings J, de Matos P, Dekker A, et al (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Research* 41:D456–D463. <https://doi.org/10.1093/nar/gks1146>
- Haug K, Salek RM, Conesa P, et al (2013) MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Research* 41:D781–D786. <https://doi.org/10.1093/nar/gks1004>
- Héral B, Stierlin É, Fernandez X, Michel T (2020) Phytochemicals from the genus *Lavandula*: a review. *Phytochem Rev*. <https://doi.org/10.1007/s11101-020-09719-z>
- Hervé MR, Nicolè F, Lê Cao K-A (2018) Multivariate analysis of multiple datasets: a practical guide for chemical ecology. *J Chem Ecol* 44:215–234. <https://doi.org/10.1007/s10886-018-0932-6>
- Heuberger AL, Robison FM, Lyons SMA, et al (2014) Evaluating plant immunity using mass spectrometry-based metabolomics workflows. *Front Plant Sci* 5:. <https://doi.org/10.3389/fpls.2014.00291>
- Horai H, Arita M, Kanaya S, et al (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 45:703–714. <https://doi.org/10.1002/jms.1777>
- Hörtensteiner S, Kräutler B (2011) Chlorophyll breakdown in higher plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1807:977–988. <https://doi.org/10.1016/j.bbabi.2010.12.007>
- Jabot C (2017) Développement de stratégies analytiques basées sur la LC-MS/MS pour la recherche de traces de pesticides et métabolites dans des matrices apicoles. Thèse de doctorat de chimie analytique, Claude Bernard Lyon 1
- Jovanović AA, Đorđević VB, Zdunić GM, et al (2017) Optimization of the extraction process of polyphenols from *Thymus serpyllum* L. herb using maceration, heat- and ultrasound-assisted techniques. *Separation and Purification Technology* 179:369–380. <https://doi.org/10.1016/j.seppur.2017.01.055>
- Juin C, Bonnet A, Nicolau E, et al (2015) UPLC–MS<sup>E</sup> profiling of phytoplankton metabolites: application to the identification of pigments and structural analysis of metabolites in *Porphyridium purpureum*. *Marine Drugs* 13:2541–2558. <https://doi.org/10.3390/md13042541>
- Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. 28:27–30. <https://doi.org/10.1093/nar/28.1.27>
- Katajamaa M, Miettinen J, Orešič M (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22:643–636. <https://doi.org/doi:10.1093/bioinformatics/btk039>

## Références bibliographiques

- Kessler N, Neuweger H, Bonte A, et al (2013) MeltDB 2.0-advances of the metabolomics software system. *Bioinformatics* 29:2452–2459. <https://doi.org/10.1093/bioinformatics/btt414>
- Khachik F, Beecher GR, Whittaker NF (1986) Separation, identification, and quantification of the major carotenoid and chlorophyll constituents in extracts of several green vegetables by liquid chromatography. *J Agric Food Chem* 34:603–616. <https://doi.org/10.1021/jf00070a006>
- Khanam H, Shamsuzzaman (2015) Bioactive benzofuran derivatives: A review. *European Journal of Medicinal Chemistry* 97:483–504. <https://doi.org/10.1016/j.ejmech.2014.11.039>
- Kim HK, Verpoorte R (2010) Sample preparation for plant metabolomics: sample preparation for plant metabolomics. *Phytochem Anal* 21:4–13. <https://doi.org/10.1002/pca.1188>
- Kokubun T, D’Costa L (2013) Direct and unbiased information recovery from liquid chromatography–mass spectrometry raw data for phenotype-differentiating metabolites based on screening window coefficient of ion currents. *Anal Chem* 85:8684–8691. <https://doi.org/10.1021/ac401545b>
- Kopka J, Schauer N, Krueger S, et al (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21:1635–1638. <https://doi.org/10.1093/bioinformatics/bti236>
- Kosovac A, Johannesen J, Krstić O, et al (2018) Widespread plant specialization in the polyphagous planthopper *Hyaletthes obsoletus* (Cixiidae), a major vector of stolbur phytoplasma: Evidence of cryptic speciation. *PLoS ONE* 13:30. <https://doi.org/10.1371/journal.pone.0196969>
- Kotz S, Johnson NL (eds) (1992) *Breakthroughs in statistics*, Springer. Springer-Verlag, New York
- Kovacevic V, Simpson MJ (2020) Fundamentals of environmental metabolomics. In: *Environmental Metabolomics*. Elsevier, pp 1–33
- Kuhl C, Tautenhahn R, Böttcher C, et al (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 84:283–289. <https://doi.org/10.1021/ac202450g>
- Kunz G, Roschatt C, Schweigkofler W (2010) Biodiversity of planthoppers (Auchenorrhyncha) in vineyards infected by the Bois noir phytoplasma. *Gredleriana* 10:89–108
- Laatsch H (2014) *AntiBase 2014: the natural compound identifier. Version upgrade*. Wiley-Vch
- Lavilla I, Bendicho C (2017) Chapter 11 - Fundamentals of Ultrasound-Assisted Extraction. In: Dominguez González H, González Muñoz MJ (eds) *Water Extraction of Bioactive Compounds: From Plants to Drug Development*. Elsevier, p 26
- Lê Cao K-A, Le Gall C (2011) Integration and variable selection of “omics” data sets with PLS: a survey. *Journal de la société française de statistique* 152:77–96
- Le Mapihan K (2004) *Caractérisation et classification des phases stationnaires utilisées pour l’analyse CPL de produits pharmaceutiques*. Thèse de doctorat en chimie, Paris VI
- Li B, Tang J, Yang Q, et al (2016) Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Sci Rep* 6:38881. <https://doi.org/10.1038/srep38881>
- Lin L, Lin H, Zhang M, et al (2015) Types, principle, and characteristics of tandem high-resolution mass spectrometry and its applications. *RSC Advances* 130:27. <https://doi.org/10.1039/C5RA22856E>
- Lis-Balchin M (ed) (2002) *Lavender: the genus *Lavandula**. Taylor & Francis
- Liu X, Locasale JW (2017) Metabolomics: a primer. *Trends in Biochemical Sciences* 42:274–284. <https://doi.org/10.1016/j.tibs.2017.01.004>
- Liu X, Zhou L, Shi X, Xu G (2019) New advances in analytical methods for mass spectrometry-based large-scale metabolomics study. *TrAC* 121:115665. <https://doi.org/10.1016/j.trac.2019.115665>
- Lommen A, Kools HJ (2012) MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. *Metabolomics* 8:719–726. <https://doi.org/10.1007/s11306-011-0369-1>
- Louris JN, Brodbelt-Lustig S, Graham R, et al (1990) Ion isolation and sequential stages of mass spectrometry in a quadupole ion trap mass spectrometer. *International journal of mass spectrometry and ion processes* 96:117–137

## Références bibliographiques

- Mak TD, Laiakis EC, Goudarzi M, Fornace AJ (2014) MetaboLyzer: A novel statistical workflow for analyzing postprocessed LC–MS metabolomics data. *Anal Chem* 86:506–513. <https://doi.org/10.1021/ac402477z>
- Marcone C, Bellardi MG, Bertaccini A (2016) Phytoplasma diseases of medicinal and aromatic plants. *Journal of Plant Pathology* 98:379–404
- Markom M, Hasan M, Daud W, et al (2007) Extraction of hydrolysable tannins from *Phyllanthus niruri* Linn.: Effects of solvents and extraction methods. *Separation and Purification Technology* 52:487–496. <https://doi.org/10.1016/j.seppur.2006.06.003>
- Martin J-C, Maillot M, Mazerolles G, et al (2015) Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. *Metabolomics* 11:807–821. <https://doi.org/10.1007/s11306-014-0740-0>
- Martin J-F (2017) LC-MS Pre-processing
- Matsuda F Spectrum PR040220 for Rosmarinic acid. In: MoNA - MassBank of North America. <https://mona.fiehnlab.ucdavis.edu/spectra/display/PR040220>. Accessed 6 Jan 2021
- McMaster MC (2005) LC/MS: a practical user's guide, John Wiley&Sons
- Mehl F, Marti G, Merle P, et al (2014) Integrating metabolomic data from multiple analytical platforms for a comprehensive characterisation of lemon essential oils: Lemon oil characterisation by multiblock metabolomic analysis. *Flavour Fragr J* 30:131–138. <https://doi.org/10.1002/ffj.3230>
- Melamud E, Vastag L, Rabinowitz JD (2010) Metabolomic analysis and visualization engine for LC–MS data. *Anal Chem* 82:9818–9826. <https://doi.org/10.1021/ac1021166>
- Mittelberger C, Yalcinkaya H, Pichler C, et al (2017) Pathogen-induced leaf chlorosis: products of chlorophyll breakdown found in degreened leaves of phytoplasma-infected apple (*Malus × domestica* Borkh.) and apricot (*Prunus armeniaca* L.) trees relate to the pheophorbide *a* oxygenase/phyllobilin pathway. *J Agric Food Chem* 65:2651–2660. <https://doi.org/10.1021/acs.jafc.6b05501>
- Mohn T, Plitzko I, Hamburger M (2009) A comprehensive metabolite profiling of *Isatis tinctoria* leaf extracts. *Phytochemistry* 70:924–934. <https://doi.org/10.1016/j.phytochem.2009.04.019>
- Moja S, Guitton Y, Nicolè F, et al (2016) Genome size and plastid trnK-matK markers give new insights into the evolutionary history of the genus *Lavandula* L. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 150:1216–1224. <https://doi.org/10.1080/11263504.2015.1014006>
- Mulinacci N, Innocenti M, Bellumori M, et al (2011) Storage method, drying processes and extraction procedures strongly affect the phenolic fraction of rosemary leaves: An HPLC/DAD/MS study. *Talanta* 85:167–176. <https://doi.org/10.1016/j.talanta.2011.03.050>
- Müller C, Riederer M (2005) Plant surface properties in chemical ecology. *J Chem Ecol* 31:2621–2651. <https://doi.org/10.1007/s10886-005-7617-7>
- Mushtaq MY, Choi YH, Verpoorte R, Wilson EG (2014) Extraction for metabolomics: access to the metabolome. *Phytochem Anal* 25:291–306. <https://doi.org/10.1002/pca.2505>
- Nishida R (2014) Chemical ecology of insect–plant interactions: ecological significance of plant secondary metabolites. *Bioscience, Biotechnology, and Biochemistry* 78:1–13. <https://doi.org/10.1080/09168451.2014.877836>
- Nothias-Scaglia L-F, Esposito M, Costa J, et al (2015) Les réseaux moléculaires, une approche bio-informatique globale pour interpréter les données de spectrométrie de masse tandem. *Spectra Analyse* 65–70
- Nyamundanda G, Brennan L, Gormley I (2010) Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* 11:571. <https://doi.org/10.1186/1471-2105-11-571>
- Oleszek W, Stochmal A (2010) High performance liquid chromatography of triterpenes (including saponins). In: Waksmundzka-Hajnos M, Sherma J (eds) *High Performance Liquid Chromatography in Phytochemical Analysis*. CRC Press, pp 633–651

## Références bibliographiques

- Olivon F, Elie N, Grelier G, et al (2018) MetGem software for the generation of molecular networks based on the t-SNE algorithm. *Anal Chem* 90:13900–13908. <https://doi.org/10.1021/acs.analchem.8b03099>
- Pagliari L, Martini M, Loschi A, Musetti R (2016) Looking inside phytoplasma-infected sieve elements: A combined microscopy approach using *Arabidopsis thaliana* as a model plant. *Micron* 89:87–97. <https://doi.org/10.1016/j.micron.2016.07.007>
- Papanov GY, Bozov P, Malakov PY (1992) Triterpenoids from *Lavandula spica*. *Phytochemistry* 31:1424–1426
- Papanov GY, Gacs-Baitz E, Malakov PY (1985) 2-N-phenylaminonaphthalene from *Lavandula vera*. *Phytochemistry* 24:3045–3046
- Peters K, Worrlich A, Weinhold A, et al (2018) Current Challenges in Plant Eco-Metabolomics. *IJMS* 19:. <https://doi.org/10.3390/ijms19051385>
- Plumb RS, Johnson KA, Rainville P, et al (2006) UPLC/MS<sup>E</sup>; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun Mass Spectrom* 20:1989–1994. <https://doi.org/10.1002/rcm.2550>
- Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395. <https://doi.org/10.1186/1471-2105-11-395>
- Ramirez-Gaona M, Marcu A, Pon A, et al (2017) YMDB 2.0: a significantly expanded version of the yeast metabolome database. *Nucleic Acids Res* 45:D440–D445. <https://doi.org/10.1093/nar/gkw1058>
- Raro M, Ibáñez M, Gil R, et al (2015) Untargeted metabolomics in doping control : detection of new markers of testosterone misuse by ultrahigh performance liquid chromatography coupled to high-resolution mass spectrometry. *Anal Chem* 87:8373–8380. <https://doi.org/10.1021/acs.analchem.5b02254>
- Reisdorph N, Reisdorph R, Quinn K, Doenges K (2019) Metabolomics mass spectrometry data processing: applications in food analysis. In: *Reference Module in Food Science*. Elsevier, p B9780081005965230000
- Riekeberg E, Powers R (2017) New frontiers in metabolomics: from measurement to insight. *F1000Res* 6:1148. <https://doi.org/10.12688/f1000research.11495.1>
- Riolo P, Minuz RL, Anfora G, et al (2012) Perception of host plant volatiles in *Hyalesthes obsoletus*: behavior, morphology, and electrophysiology. *J Chem Ecol* 38:1017–1030. <https://doi.org/10.1007/s10886-012-0154-2>
- Royer A-L, Escourrou A, Guitton Y, et al (2017) L'approche métabolomique par spectrométrie de masse : évidences, incertitudes et points critiques. *Spectra Analyse* 47–54
- Saint-Lary L (2015) Évaluation de l'approche métabolomique pour l'authentification des extraits naturels utilisés dans le secteur arômes et parfums. Thèse de doctorat en chimie, Nice Sophia-Antipolis
- Sajed T, Marcu A, Ramirez M, et al (2016) ECMDDB 2.0: A richer resource for understanding the biochemistry of *E. coli*. *Nucleic Acids Res* 44:D495–D501. <https://doi.org/10.1093/nar/gkv1060>
- Sakurai N, Ara T, Kanaya S, et al (2013) An application of a relational database system for high-throughput prediction of elemental compositions from accurate mass values. *Bioinformatics* 29:290–291. <https://doi.org/10.1093/bioinformatics/bts660>
- Salem MA, Perez de Souza L, Serag A, et al (2020) Metabolomics in the context of plant natural products research: from sample preparation to metabolite analysis. *Metabolites* 10:37. <https://doi.org/10.3390/metabo10010037>
- Sawada Y, Nakabayashi R, Yamada Y, et al (2012) RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry* 82:38–45. <https://doi.org/10.1016/j.phytochem.2012.07.007>

## Références bibliographiques

- Scheltema RA, Jankevics A, Jansen RC, et al (2011) PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal Chem* 83:2786–2793. <https://doi.org/10.1021/ac2000994>
- Schlöpfer P, Zhang P, Wang C, et al (2017) Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol* 173:2041–2059. <https://doi.org/10.1104/pp.16.01942>
- Schymanski EL, Jeon J, Gulde R, et al (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol* 48:2097–2098. <https://doi.org/10.1021/es5002105>
- Sémétey O, Gaudin J, Danet J-L, et al (2018) Lavender decline in France is associated with chronic infection by lavender-specific strains of “*Candidatus* Phytoplasma solani.” *Appl Environ Microbiol* 84:. <https://doi.org/10.1128/AEM.01507-18>
- Sforza R, Bourgoïn T, Wilson SW, Boudon-Padieu E (1999) Field observations, laboratory rearing and descriptions of immatures of the planthopper *Hyalesthes obsoletus* (Hemiptera: Cixiidae). *European Journal of Entomology* 96:409–418
- Shinbo Y, Nakamura Y, Altaf-UI-Amin M, et al (2006) KNApSACK: A comprehensive species-metabolite relationship database. In: *Biotechnology in Agriculture and Forestry*, K. Saito, R. Dixon, L. Willmitzer. Springer Berlin Heidelberg, pp 165–181
- Silva CL, Câmara JS (2013) Profiling of volatiles in the leaves of Lamiaceae species based on headspace solid phase microextraction and mass spectrometry. *Food Research International* 51:378–387. <https://doi.org/10.1016/j.foodres.2012.12.040>
- Smilde AK, van der Werf MJ, Bijlsma S, et al (2005) Fusion of mass spectrometry-based metabolomics data. *Anal Chem* 77:6729–6736. <https://doi.org/10.1021/ac051080y>
- Smith CA, O’Maille G, Want EJ, et al (2005) METLIN: a metabolite mass spectral database. *Therapeutic Drug Monitoring* 27:747–751. <https://doi.org/10.1097/01.ftd.0000179845.53213.39>
- Smith CA, Want EJ, O’Maille G, et al (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78:779–787. <https://doi.org/10.1021/ac051437y>
- Snyder LR (1974) Classification of the solvent properties of common liquids. *Journal of Chromatography A* 92:223–230. [https://doi.org/10.1016/S0021-9673\(00\)85732-5](https://doi.org/10.1016/S0021-9673(00)85732-5)
- Spicer R, Salek RM, Moreno P, et al (2017) Navigating freely-available software tools for metabolomics analysis. *Metabolomics* 13:106. <https://doi.org/10.1007/s11306-017-1242-7>
- Spiridon I, Colceru S, Anghel N, et al (2011) Antioxidant capacity and total phenolic contents of oregano (*Origanum vulgare*), lavender (*Lavandula angustifolia*) and lemon balm (*Melissa officinalis*) from Romania. *Natural Product Research* 25:1657–1661. <https://doi.org/10.1080/14786419.2010.521502>
- Stalikas CD (2007) Extraction, separation, and detection methods for phenolic acids and flavonoids. *J Sep Sci* 30:3268–3295. <https://doi.org/10.1002/jssc.200700261>
- Stierlin É (2020) Développements analytiques et approche métabolomique pour l’étude des composés volatils d’origine végétale - Application à l’étude du dépérissement de la lavande. Thèse de doctorat en chimie, Côte d’Azur
- Stierlin É, Nicolè F, Fernandez X, Michel T (2019) Development of a headspace solid-phase microextraction gas chromatography-mass spectrometry method to study volatile organic compounds (VOCs) emitted by lavender roots. *Chem Biodiversity* 16:14. <https://doi.org/10.1002/cbdv.201900280>
- Sudria C, Piñol MT, Palazon J, et al (1999) Influence of plant growth regulators on the growth and essential oil content of cultured *Lavandula dentata* plantlets. *Plant Cell, Tissue and Organ Culture* 58:177–184. <https://doi.org/10.1023/A:1006377003962>
- Szymańska E, Saccenti E, Smilde AK, Westerhuis JA (2012) Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 8:S3–S16. <https://doi.org/10.1007/s11306-011-0330-3>

## Références bibliographiques

- Tang S, Shi J, Liu C, et al (2017) Three new arylbenzofurans from *Lavandula angustifolia* and their bioactivities. *Phytochemistry Letters* 19:60–63. <https://doi.org/10.1016/j.phytol.2016.12.005>
- Tautenhahn R, Böttcher C, Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 9:. <https://doi.org/10.1186/1471-2105-9-504>
- Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* 84:5035–5039. <https://doi.org/10.1021/ac300698c>
- Tenenhaus M (1998) La régression PLS - Théorie et pratique. Technip.
- Thévenot EA, Roux A, Xu Y, et al (2015) Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J Proteome Res* 14:3322–3335. <https://doi.org/10.1021/acs.jproteome.5b00354>
- Toma M, Vinatoru M, Paniwnyk L, Mason TJ (2001) Investigation of the effects of ultrasound on vegetal tissues during solvent extraction. *Ultrasonics Sonochemistry* 8:137–142. [https://doi.org/10.1016/S1350-4177\(00\)00033-X](https://doi.org/10.1016/S1350-4177(00)00033-X)
- Topçu G, Ayril MN, Aydın A, et al (2001) Triterpenoids of the roots of *Lavandula stoechas* ssp. *stoechas*. *Pharmazie* 56:892–895
- Torras-Claveria L, Jauregui O, Bastida J, et al (2007) Antioxidant activity and phenolic composition of lavandin (*Lavandula x intermedia* Emeric ex Loiseleur) waste. *J Agric Food Chem* 55:8436–8443. <https://doi.org/10.1021/jf070236n>
- Triba MN, Moyec LL, Amathieu R, et al (2015) PLS/OPLS models in metabolomics: Impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Molecular BioSystems* 11:17. <https://doi.org/10.1039/c4mb00414k>
- Tsugawa H, Cajka T, Kind T, et al (2015) MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* 12:523–526. <https://doi.org/10.1038/nmeth.3393>
- Ulubelen A, Gören N, Olcay Y (1988) Longipinene derivatives from *Lavandula stoechas* subsp. *stoechas*. *Phytochemistry* 27:3966–3967
- Upton T, Andrews S (2004) The genus *Lavandula*. Royal Botanic Gardens
- Upton TM, Grayer RJ, Greenham JR, et al (2000) Leaf favonoids as systematic characters in the genera *Lavandula* and *Sabaudia*. *Biochemical Systematics and Ecology* 28:17
- Van Den Berg RA, Hoefsloot HC, Westerhuis JA, et al (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7:. <https://doi.org/10.1186/1471-2164-7-142>
- van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH (2009) Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *J Proteome Res* 8:5132–5141. <https://doi.org/10.1021/pr900499r>
- Vencl FV, Gómez NE, Ploss K, Boland W (2009) The chlorophyll catabolite, pheophorbide *a*, confers predation resistance in a larval tortoise beetle shield defense. *J Chem Ecol* 35:281–288. <https://doi.org/10.1007/s10886-008-9577-1>
- Vinatoru M (2001) An overview of the ultrasonically assisted extraction of bioactive principles from herbs. *Ultrasonics Sonochemistry* 8:303–313. [https://doi.org/10.1016/S1350-4177\(01\)00071-2](https://doi.org/10.1016/S1350-4177(01)00071-2)
- Wang M, Carver JJ, Phelan VV, et al (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>
- Waters Corporation (2010) Instrumental specification Waters Xevo G2 QTof
- Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin* 1:80–83. <https://doi.org/10.2307/3001968>
- Wishart DS, Tzur D, Knox C, et al (2007) HMDB: the Human Metabolome Database. *Nucleic Acids Research* 35:D521–D526. <https://doi.org/10.1093/nar/gkl923>

## Références bibliographiques

- Wolfender J-L, Marti G, Thomas A, Bertrand S (2015) Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of Chromatography A* 1382:136–164. <https://doi.org/10.1016/j.chroma.2014.10.091>
- Wolfender J-L, Rudaz S, Choi YH, Kim HK (2013) Plant metabolomics: from holistic data to relevant biomarkers. *Current Medicinal Chemistry* 20:1056–1090
- Worley B, Powers R (2013) Multivariate Analysis in Metabolomics. *CMB* 1:92–107. <https://doi.org/10.2174/2213235X11301010092>
- Woronuk G, Demissie Z, Rheault M, Mahmoud S (2011) Biosynthesis and Therapeutic Properties of *Lavandula* Essential Oil Constituents. *Planta Med* 77:7–15. <https://doi.org/10.1055/s-0030-1250136>
- Xia J, Wishart DS (2011) Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protoc* 6:743–760. <https://doi.org/10.1038/nprot.2011.319>
- Yadikar N, Bobakulov K, Li G, Aisa HA (2018) Seven new phenolic compounds from *Lavandula angustifolia*. *Phytochemistry Letters* 23:149–154. <https://doi.org/10.1016/j.phytol.2017.12.005>
- Yadikar N, Bobakulov KhM, Eshbakova KA, Aisa HA (2017) Phenolic compounds from *Lavandula angustifolia*. *Chem Nat Compd* 53:562–564. <https://doi.org/10.1007/s10600-017-2050-z>
- Yong Seok C, Young Kook K, Mun Chual R, et al (2010) An acyl-CoA: choloesterol acyltransferase inhibitor and a therapeutic agent containing pheophorbide A methyl ester, porphyrin-type compound or extracts of *Diospyros kaki* as an effective ingredient for the treatment of larvicide
- Yoshida K, Mori M, Kondo T (2009) Blue flower color development by anthocyanins: from chemical structure to cell physiology. *Nat Prod Rep* 26:884. <https://doi.org/10.1039/b800165k>
- Yvin C (2010) Synthèse des recherches conduites depuis 1994 sur les dépérissements des lavandes et lavandins. iteipmai
- Yvin C, Cansière R, Thierry D, et al (2018) Etude des facteurs sensoriels intervenant dans le choix de plantes et dans le comportement alimentaire des adultes de *Hyalecthes obsoletus* (Echostol). *Innovations Agronomiques* 63:445–456
- Zanardo LG, de Souza GB, Alves MS (2019) Transcriptomics of plant–virus interactions: a review. *Theor Exp Plant Physiol* 31:103–125. <https://doi.org/10.1007/s40626-019-00143-z>
- Zeng Y, Li S, Wang X, et al (2015) Validated LC-MS/MS method for the determination of scopoletin in rat plasma and its application to pharmacokinetic studies. *Molecules* 20:18988–19001. <https://doi.org/10.3390/molecules201018988>
- Zhou M, Xing H-H, Ma H-Y, et al (2017) Three new isobenzofurans from *Lavandula angustifolia* and their bioactivities. *Phytochemistry Letters* 19:156–159. <https://doi.org/10.1016/j.phytol.2016.12.034>