



HAL
open science

Diversité des composés terpéniques volatils au sein du genre *Lavandula* : aspects évolutifs et physiologiques

Yann Guitton

► **To cite this version:**

Yann Guitton. Diversité des composés terpéniques volatils au sein du genre *Lavandula* : aspects évolutifs et physiologiques. Biologie végétale. Université Jean Monnet - Saint-Etienne, 2010. Français. NNT : 2010STET4008 . tel-00675866

HAL Id: tel-00675866

<https://theses.hal.science/tel-00675866>

Submitted on 2 Mar 2012

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.

UNIVERSITE DE SAINT-ETIENNE- JEAN-MONNET

Ecole doctorale : Sciences, Ingénierie et Santé

Laboratoire de Biotechnologies Végétales Appliquées aux plantes aromatiques et médicinales

THÈSE DE DOCTORAT

pour l'obtention du grade de

Docteur en Sciences de l'Université de Saint-Etienne - Jean-Monnet

Discipline : Biologie et Physiologie Végétales

Diversité des composés terpéniques volatils au sein du
genre *Lavandula* : aspects évolutifs et physiologiques

Par

Yann Guitton

Thèse dirigée par le Professeur Laurent LEGENDRE

Soutenue le **21 décembre 2010** devant le jury :

M. Francis KARST, Professeur, Université de Strasbourg

Rapporteur

M. Benoit ST-PIERRE, Professeur, Université François Rabelais de Tours

Rapporteur

Mme. Anne BOUSQUET-MELOU, Maître de conférences, Université Aix-Marseille

Examinatrice

M Laurent LEGENDRE, Professeur, UJM, St-Etienne

Directeur de thèse





Adresse du Laboratoire :

Laboratoire de biotechnologies végétales appliquées aux plantes aromatiques et médicinales

23 rue du Docteur Paul Michelon

42023 Saint-Etienne Cedex 2

France

REMERCIEMENTS

Il est de coutume de placer une citation d'un de nos grands penseurs au cœur d'un manuscrit de thèse. Sans vouloir déroger à la règle, je me permets de la détourner et de vous offrir un extrait de ma vie personnelle, qui, à lui seul résume toute l'abnégation dont a fait preuve ma femme Evelyne durant ces années de thèse.

« Ma chérie, tu te souviens de mon maître de stage en Australie ? Et bien il me propose une thèse sur le parfum des lavandes »

« Chouette ça, on part quand en Australie ? »

« Euh bah, en fait il est de retour en France... à St-Etienne »

« Ah..., bon... et bien en route pour le pays des Verts »

Sur ces quelques mots, nous quitions Quimper et je débutais ma thèse au laboratoire de biotechnologies végétales appliquées aux plantes aromatiques et médicinales.

Lors de mes recherches, j'ai bénéficié d'un encadrement scientifique et technique de qualité et j'en remercie l'ensemble des membres du laboratoire. Merci à vous tous, Chacun d'entre vous m'a donné de son temps, m'a transmis une part de ses compétences. Je vous souhaite à tous bonheur, réussite et tout le bien que vous méritez.

Mes recherches n'auraient pas été possibles sans le concours de partenaires extérieurs au laboratoire. Un grand merci à tous, en particulier à Bernard Pasquier (Directeur du Conservatoire National des Plantes à Parfum, Médicinales, Aromatiques et Industrielles de Milly la Forêt) ou encore à Pierre Danelon (chef d'exploitation au lycée horticole de Romans sur Isère) et le financement de la région Rhône-Alpes. Encore merci à toi, Bernard de nous avoir ouvert les portes de ton jardin et de tes serres et de nous avoir laissé couper, trancher, broyer les feuilles et les fleurs de ta collection unique de lavandes botaniques. Merci également de nous avoir fait partager ta passion pour les plantes. Merci à Pierre Danelon et Caroline Baconnier du lycée horticole de Romans-sur-Isère de nous avoir ouvert pendant 4 ans l'accès aux champs de lavandes fines et lavandins. La mise en place au laboratoire BVpam d'un double de la collection du CNPMAI était également nécessaire et n'aurait pu être faite sans l'engagement de Nicolas Boyer.

Pour mes recherches liés à l'analyse de grands jeux de données, la encore le partenariat avec des chercheurs externes au laboratoire a été fructueux, tout particulièrement grâce aux travaux d'Elodie Courtois (doctorante au Laboratoire Evolution & Diversité Biologique de l'université Paul Sabatier de Toulouse) et au soutien financier du Groupe de Recherche Ecologie Chimique (CNRS Institut Ecologie et Environnement SECTION 29).

Je remercie le professeur C. Mauriat de nous avoir ouvert son laboratoire (Néphrologie – Faculté de médecine de Saint-Etienne) et à Farida Abadja sa doctorante pour son aide dans réalisation des qPCR. Mes collègues Nadine Valot et Sylvain Legrand ont été d'une grande aide lors de la mise au point de l'expérimentation en me faisant partager leurs expériences respectives.

Une grande partie de l'échantillonnage a été effectué en 2005 et 2006 alors que je n'étais pas encore au laboratoire. Ces campagnes de récoltes sont le travail d'Alexandra Vial et Sandra Chatelard. Je remercie donc particulièrement les étudiantes de master qui se sont succédées autour du projet lavande : Alexandra et Sandra bien sûr et également un grand merci à vous Laurène, Audrey et Souhila.

Je tiens à remercier Francis Karst et Benoît St-Pierre d'avoir accepté d'être mes rapporteurs et de juger cette thèse.

Ce travail de recherche a été financé par la région Rhône-Alpes dans le cadre du cluster de recherche n°9 « Quamité des Plantes, Agricultures, Acteurs et Territoires», que je remercie très sincèrement.

Merci à Jacques et à Fanc'h d'avoir pris des nouvelles de leur ancien conseiller agricole, qui les a abandonné en Bretagne pour aller courir les champs de lavandes de la Drôme.

Et enfin, un très grand merci à ma fille Diane, qui a gentiment acceptée de réveiller son papa aux aurores afin qu'il finisse, au plus vite, de rédiger sa thèse.

Merci
Guillem

RESUME

La production de lavande concoure au rayonnement de la région Rhône-Alpes. Les applications de l'huile essentielle (HE) de lavande reposent sur la culture de 3 espèces (*L. angustifolia*, *L. latifolia* et *L. stoechas* et d'un hybride *L. x intermedia*) aux chémotypes marqués (principalement constitués de mono- et sesquiterpènes). Le genre *Lavandula* est un modèle idéal pour comprendre la structuration et l'origine de la diversité des composés organiques volatils (COV) en particulier des terpènes. Les lavandes ont l'avantage d'avoir une aire de distribution large avec des régions bioclimatiques différentes, un nombre d'espèces limité (39) ayant des caractéristiques morphologiques et écologiques variées. Pour caractériser la diversité des COV accumulés dans les espèces du genre et envisager leur évolution, nous avons analysé (GC-MS) les COV de 29 espèces (certaines pour la première fois). Comme souvent chez les plantes, la production de COV dans les inflorescences de lavande est soumise à une régulation spatio-temporelle. L'émission différentielle de COV au cours du temps chez *L. angustifolia* a été relevée par les agriculteurs qui ont observé une qualité d'HE différente suivant la maturité des inflorescences au moment de la récolte. Pour modéliser ces variations et les corrélérer avec des étapes du développement de la plante, nous avons analysé, au niveau chimique (GC-FID) et moléculaire (qPCR), les variations temporelles des principaux COV dans les feuilles et les inflorescences (plusieurs années et cultivars). En amont de ces recherches sur les COV du genre *Lavandula*, différents outils de bioinformatique ont été développés. En particulier, le module « MSeasy » qui permet d'automatiser le rapatriement de données de GC-MS. Ceci constitue un pré-requis pour utiliser dans le futur la lavande comme modèle d'étude des COV chez les Lamiacées.

Mots clés : Lamiacée, genus *Lavandula*, *Lavandula angustifolia*, terpènes, terpène synthase, GC-MS, qPCR, R package, métabolomique

ABSTRACT

English title : Diversity of the volatile terpenic compounds within the genus *Lavandula*: evolutionary and physiological aspects

The lavender production is of significant importance for the international visibility of the french Rhône-Alpes region. Uses of lavender essential oil (EO) are based on the growing of 3 species (*L. angustifolia*, *L. latifolia*, *L. stoechas* and an hybride *L. x intermedia*) with marked chemotypes (mainly constituted of mono- and sesquiterpenes). The genus *Lavandula* is an ideal model for understanding the origin of the diversity of volatile organic compounds (VOCs), especially terpenes. Lavenders have the advantage of having a wide range of distribution areas with different bioclimatic regions, a limited number of species (39) with diverse morphological and ecological characteristics. In order to characterize the diversity of the VOCs accumulated in the genus and consider their evolution, we have analyzed (GC-MS) the VOCs accumulated by 29 species (some for the first time). As often, in plants, the production of VOCs in the inflorescences of lavender is subject to spatial and temporal regulation. The differential emission of VOCs over time in *L. angustifolia* is a well known phenomenon for farmers who have observed a different quality of EO depending on the maturity of the inflorescences at harvest. To correlate these variations with stages of plant development, we have analysed the temporal variations of the main VOCs in leaves and inflorescences (several years and cultivars) at the chemical level (GC-FID) and the molecular level (qPCR). Upstream of this research on the genus *Lavandula* different bioinformatic tools have been developed. In particular, the module MSeasy "which can automate GC-MS data retrieval. This is a prerequisite for using lavender in the future as a model study of VOCs in Lamiaceae.

Key Words : Lamiaceae, genus *Lavandula*, *Lavandula angustifolia*, terpenes, terpene synthase, GC-MS, qPCR, R package, metabolomic

Liste des abréviations

LISTE DES ABREVIATIONS

ACoP	Analyse en coordonnée principale
ACP	Analyse en composante principale
ADN	Acide désoxyribonucléique
ADNc	ADN codant
AI	Indice arithmétique
AOC	Appellation d'origine contrôlée
ARN	Acide ribonucléique
ARNm	ARN messager
BVPAM	Laboratoire de Biotechnologies Végétales appliquées aux Plantes Aromatiques et Médicinales
CIHEF	Comité Interprofessionnel des Huiles Essentielles Françaises
CNPMAI	Le Conservatoire National des Plantes Médicinales, Aromatiques et Industrielles
COST	Conseil d'Orientation Scientifique et Technique
COV	Composés organiques volatils
CPPARM	Le Comité des Plantes à Parfum, Aromatiques et Médicinales
CRIEPPAM	Centre Régionalisé Interprofessionnel d'Expérimentation en Plantes à Parfum, Aromatiques et Médicinales
CTV	Composés terpéniques volatils
DMAPP	Diméthylallyl diphosphate
FID	Détecteur à ionisation de flamme
FPP	Farnésyl diphosphate
GC	Chromatographie en phase gazeuse
GGPP	Géranylgéranyl diphosphate
GPP	Géranyl diphosphate
HCA	Hierarchical clustering analysis
HE	Huile essentielle
IPP	Isopentenyl diphosphate
ITEIPMAI	Institut Technique Interprofessionnel des Plantes à Parfum, Aromatiques et Médicinales
ITS	Internal Transcript Spacer
KI	Indice de Kovats
LPP	Linalyl diphosphate
MEV	Voie du 2-C-methyl-D-erythritol 4-phosphate
MS	Spectrométrie de masse
MVA	Voie du mévalonate
ONIPPAM	Office National Interprofessionnel des Plantes à Parfum, Aromatiques et Médicinales
PCR	Polymerase Chain Reaction
PPAM	Plantes à Parfums Aromatiques et Médicinales
RI	Indice de rétention
RT	Temps de rétention
RT	Transcription inverse
TPS	Terpènes synthases

Table des matières

TABLE DES MATIERES

<u>INTRODUCTION GENERALE</u>	<u>15</u>
<u>I LES COMPOSES TERPENIQUES VOLATILS DES VEGETAUX</u>	<u>15</u>
I-1 DEFINITIONS	15
I-2 ROLES ET ACTIVITES BIOLOGIQUES.....	15
I-3 VARIATIONS SPATIALES ET TEMPORELLES DE LA PRODUCTION ET DE L'EMISSION DES COV	17
I-4 LIEUX DE FORMATION ET D'ACCUMULATION	18
I-5 BIOSYNTHESE DES TERPENES.....	20
I-5.1 <i>Origine et diversité des terpènes</i>	20
I-5.2 <i>Les voies de synthèse de l'IPP et leur compartimentation : impact sur la synthèse des terpènes</i> ..	24
I-5.3 <i>Les terpènes synthèses</i>	27
<u>II LE GENRE LAVANDULA</u>	<u>32</u>
II-1 UTILISATION INDUSTRIELLE	32
II-2 TAXONOMIE ET BIODIVERSITE.....	36
II-3 AIRE DE REPARTITION	50
II-4 DIVERSITE DES COV DANS LE GENRE LAVANDULA	56
II-4.1 <i>Variation interspécifique de la composition en COV</i>	56
II-4.2 <i>Variations intraspécifique de la composition en COV</i>	59
II-4.3 <i>Les gènes de lavandes impliqués dans la synthèse des COV</i>	61
<u>III L'ACQUISITION ET LE TRAITEMENT DE DONNEES DE METABOLOMIQUE: UN PROBLEME DE TAILLE</u>	<u>63</u>
III-1 LA METABOLOMIQUE	63
III-2 LES OUTILS STATISTIQUE POUR LE TRAITEMENT DE GRANDS JEUX DE DONNEES : LE DATA MINING	68
III-3 LA CLASSIFICATION DE DONNEES (<i>CLUSTERING</i>)	69
<u>IV LES BUTS DE LA THESE</u>	<u>70</u>
<u>MATERIELS ET METHODES</u>	<u>73</u>
<u>I MATERIEL VEGETAL</u>	<u>73</u>
I-1 ORIGINES	73
I-2 RECOLTES	73
I-2.1 <i>Lavandes botaniques</i>	73
I-2.2 <i>Lavandes fines et lavandins</i>	76

Table des matières

I-3	CONDITIONS DE CULTURE.....	77
II	METHODES	78
II-1	MICROSCOPIE ELECTRONIQUE A BALAYAGE	78
II-2	ANALYSE DES COMPOSES ORGANIQUES VOLATILS	78
II-2.1	<i>Préparation des échantillons</i>	78
II-2.2	<i>Analyse en GC-FID</i>	79
II-2.3	<i>Analyse en GC-MS</i>	79
II-3	BIOLOGIE MOLECULAIRE	80
II-3.1	<i>Extraction de l'ADN génomique</i>	80
II-3.2	<i>Extraction des ARN</i>	80
II-3.3	<i>Synthèse du premier brin d'ADN complémentaire</i>	81
II-3.4	<i>Dosage et contrôles de la qualité des acides nucléiques</i>	82
II-3.5	<i>Amplification par PCR</i>	84
II-3.6	<i>PCR quantitative</i>	85
II-3.7	<i>Séquençage</i>	86
III	ANALYSE DE DONNEES	87
III-1	GENERALITES.....	87
III-2	ANALYSES CHROMATOGRAPHIQUES.....	87
III-2.1	<i>Identification des composés volatils</i>	87
III-2.2	<i>Quantification des composés volatils</i>	89
III-2.3	<i>Alignement des chromatogrammes</i>	89
III-3	ANALYSE DES PCR QUANTITATIVES	89
III-3.1	<i>Choix des amorces</i>	89
III-3.2	<i>Ratio d'expression</i>	90
III-4	ANALYSES PHYLOGENETIQUES	90
III-4.1	<i>Alignement des séquences nucléotidiques</i>	90
III-4.2	<i>Calculs et visualisation des arbres de phylogénie</i>	90
III-4.3	<i>Superposition des données de chimie et de phylogénie moléculaire</i>	91
III-5	ANALYSES STATISTIQUES	92
	RESULTATS	93
I	CREATION D'UN LOGICIEL DE TRAITEMENT DES DONNEES DE GC-MS (MODULE MSEASY POUR R) .96	
I-1	INTRODUCTION.....	96
I-2	RESULTATS : LES OUTILS DE BIOINFORMATIQUE CREES	103
I-2.1	<i>Les outils de compilation de données de GC-FID et GC-MS</i>	103

Table des matières

I-2.2	<i>Le module MSeasy pour R (article en préparation pour soumission)</i>	107
I-3	DISCUSSION	131
II	<u>EVOLUTION DES COMPOSES ORGANIQUES VOLATILS ACCUMULES PAR LES ESPECES DU GENRE</u>	
	<u>LAVANDULA</u>	<u>136</u>
II-1	INTRODUCTION	136
II-2	RESULTATS.....	138
II-3	DISCUSSION	158
III	<u>ETUDE DES VARIATIONS SPATIO-TEMPORELLES DE L'ACCUMULATION DE TERPENES CHEZ L.</u>	
	<u>ANGUSTIFOLIA ET L. XINTERMEDIA</u>	<u>168</u>
III-1	LA PROBLEMATIQUE	168
III-2	RESULTATS.....	169
III-2.1	<i>Articles</i>	169
III-2.2	<i>Résultats complémentaires</i>	193
III-3	DISCUSSION	202
	<u>CONCLUSIONS ET PERSPECTIVES</u>	<u>207</u>
	<u>BIBLIOGRAPHIE</u>	<u>211</u>
	<u>ANNEXES</u>	<u>231</u>

AVANT-PROPOS

Mettons nous dans l'ambiance, il est midi, nous sommes au mois de juin quelque part sur les contreforts des monts du Vercors ou dans la Provence de Jean Giono. L'air est empli d'un bouquet invisible et enivrant, offert à nos sens par les plantes qui nous entourent. Mais, d'où vient ce parfum ? Pourquoi et comment les plantes produisent-elles ces centaines de substances volatiles ?

Les composés organiques volatils (COV) ont, individuellement ou en mélange, de multiples fonctions. Si certains sont émis par les fleurs dans le but d'attirer les pollinisateurs d'autres sont là pour assurer leur défense face aux herbivores. Ces composés sont sources de plaisirs olfactifs et gustatifs car ils confèrent leurs arômes aux fruits et feuilles dont nous agrémentons nos repas.

Ces molécules de parfum sont globalement regroupées en trois catégories : les terpènes (monoterpènes et sesquiterpènes), les phénylpropanoïdes et les dérivés d'acides gras. Certaines donnent aux extraits de plantes qui les contiennent des propriétés médicinales. La phytothérapie exige la présence de tel ou tel composé dans une huile essentielle (HE). Il en est de même pour les industries de la parfumerie ou de la cosmétique qui recherchent dans ces plantes des sources de produits naturels précis.

Le laboratoire de biotechnologies végétales appliquées aux plantes aromatiques et médicinales (BVpam) travaille depuis de nombreuses années sur les mécanismes de synthèse par les plantes de ces composés volatils. Actuellement, les chercheurs décryptent les mécanismes de production et d'émission de parfum par les roses, les menthes, les sauges et les lavandes. Ces dernières sont l'objet de mes recherches, qui visent, pour la première fois, à appréhender la diversité des composés volatils du type terpène formés au sein des espèces d'un même genre : le genre *Lavandula*. J'ai ensuite focalisé mon attention sur les principales espèces commerciales, la lavande fine et le lavandin, pour montrer que les composés terpéniques volatils diffèrent selon le tissu considéré et le stade de développement de celui-ci. J'ai travaillé à différentes échelles en allant du parfum aux gènes en passant par les enzymes responsables de la formation des terpènes (terpènes synthases). Pour pouvoir faire ce travail, j'ai également développé des outils d'analyses bioinformatiques novateurs qui permettront, je pense, à d'autres équipes travaillant sur les COV d'accélérer leurs études. Par la meilleure

Avant-propos

connaissance de la physiologie de la lavande qu'elles apportent mes recherches expliquent quelques unes des observations faites par les professionnels sur les variations qualitatives et quantitatives des huiles essentielles de *L. angustifolia*. Elles pourraient ainsi aider les producteurs à améliorer la qualité des huiles en ciblant une date de récolte optimale ou bien en facilitant la sélection de cultivars. La description des huiles essentielles de lavandes encore inconnues peut intéresser les producteurs qui souhaiteraient développer de nouveaux marchés. Ces travaux devraient, à terme et pour la première fois, permettre de caractériser et de comparer l'évolution des enzymes productrices de mono et sesquiterpènes à l'intérieur d'un même genre, le genre *Lavandula*. Ces travaux n'auraient pas été possibles sans le soutien de nombreux partenaires dont Bernard Pasquier du CNPMAI de Milly-la-Forêt qui nous a fourni les lavandes botaniques et transmet sa passion pour ces plantes, Pierre Danelon et Caroline Baconnier du lycée horticole de Romans sur Isère ainsi que la région Rhône-Alpes pour le financement de ces recherches via le cluster recherche n°9 « plantes et santé ».

INTRODUCTION GENERALE

I Les Composés terpéniques Volatils des végétaux

I-1 Définitions

Les composés organiques volatils (COV) produits par les plantes représentent 1% des métabolites secondaires connus à ce jour et sont principalement des terpènes, des phénylpropanoïdes/ des benzenoïdes et des dérivés d'acides aminés et d'acides gras. Ces COV sont généralement lipophiles et ont une pression de vapeur élevée. Ils peuvent ainsi facilement traverser les membranes cellulaires et être relâchés dans l'atmosphère ou dans le sol (Dudareva, Negre *et al.* 2006). La diversité et la quantité des COV formés varient selon les espèces et sous l'influence de facteurs biotiques et abiotiques (Pichersky et Gershenzon 2002). On retrouve parfois dans des espèces de familles botaniques éloignées, mais vivant dans un même environnement bioclimatique, des molécules volatiles identiques (Owen, Boissard *et al.* 2001) suggérant que ces COV ont des rôles importants pour les plantes. Cependant, il est possible que certains COV puissent ne pas avoir de fonction propre et ne sont que des co-produits de voies métaboliques essentielles ou sont des reliquats de l'évolution qui ont perdu leur fonction chez certaines espèces (Levin, McDade *et al.* 2003). Parmi, ces COV, notre étude se focalisera plus particulièrement sur les composés terpéniques volatils (CTV).

Le nom terpène dérive de térébenthine. En effet, après distillation de l'oléorésine de pin on obtient de l'essence de térébenthine, une huile essentielle (HE) dont les composés majeurs ont été nommés terpènes (Breitmaier 2006 ; Phillips et Croteau 1999). Les terpènes, sont une classe d'hydrocarbures produits par de nombreuses plantes. Ce sont des substances naturelles organiques, volatiles ou non, qui représentent une grande famille de molécules du métabolisme primaire et secondaire.

I-2 Rôles et activités biologiques

Les COV sont un moyen de communication utilisé par les plantes pour interagir avec leur environnement. Ces volatils sont libérés dans l'atmosphère par les feuilles, les fleurs et les fruits et dans le sol par les racines (Dudareva et Negre 2005; Rasmann, Kollner *et al.* 2005; Van Dam, Harvey *et al.* 2003). Ils assurent à la plante une protection contre les herbivores, les

Introduction

pathogènes (Wink 1988; Wink 2003), les trop fortes chaleurs et les stress oxydatifs (Vickers, Gershenzon *et al.* 2009). Ils favorisent la reproduction en attirant les pollinisateurs (Pichersky et Gershenzon 2002) et les animaux qui disperseront les graines (Slaughter 1999) ainsi que la colonisation des milieux en limitant la croissance d'autres plantes par des mécanismes d'allélopathie (Muller 1966). Les COV permettent à la plante de communiquer avec les plantes voisines (Baldwin, Halitschke *et al.* 2006) et également entre ses propres organes (Heil et Silva Bueno 2007). On ajoutera à cela que les émissions de COV par une plante ou un organe au moment d'une attaque par un herbivore n'ont pas nécessairement un effet de défense immédiat, mais préparent la plante à de futures attaques. Ce phénomène est connu sous le nom de priming (Dudareva, Negre *et al.* 2006; Ton, D'Alessandro *et al.* 2007; Unsicker, Kunert *et al.* 2009). La figure 1 reprend ces différents rôles des COV dans les interactions des plantes avec leur environnement.

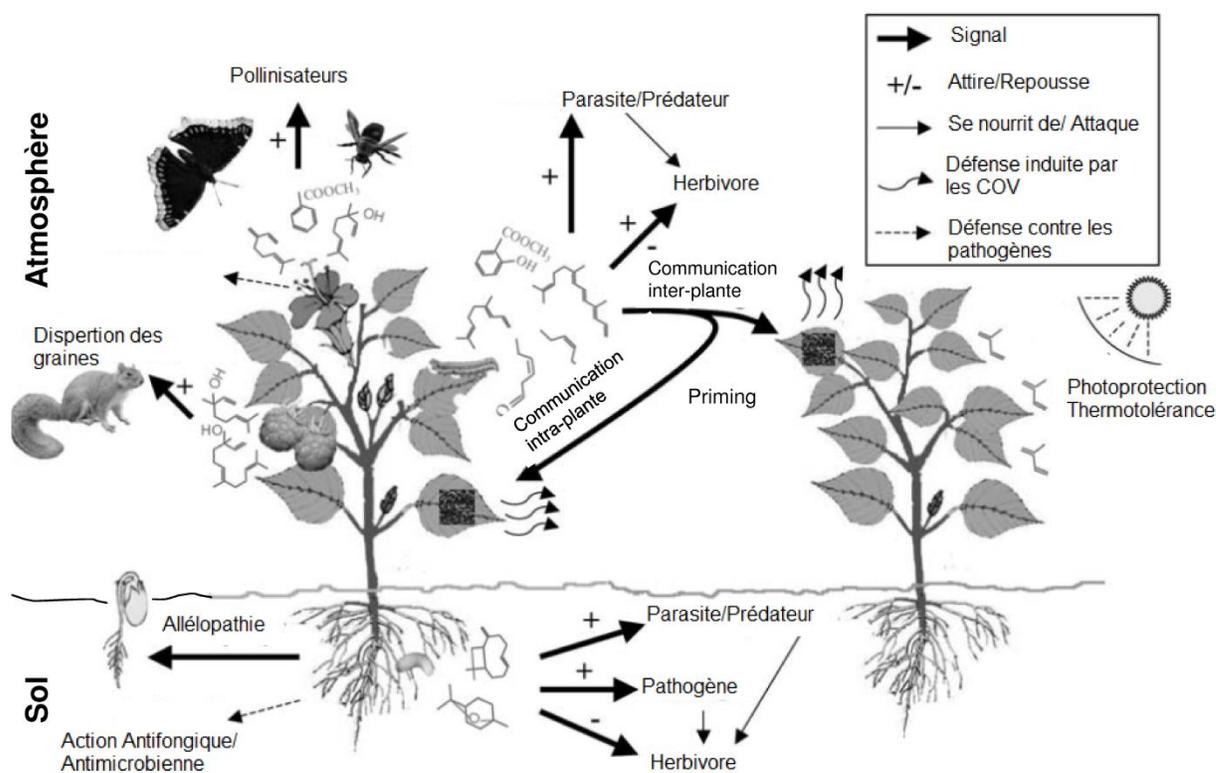


Figure 1 : Schéma des interactions des plantes avec leur environnement via les COV (modifié d'après Dudareva, Negre *et al.* 2006)

Parmi ces COV, les CTV sont particulièrement impliqués dans les interactions des plantes avec l'environnement et jouent un rôle important dans la croissance et le développement des plantes, puisqu'ils conduisent notamment à la synthèse de certaines hormones végétales comme les gibbérellines. Ces CTV ont également un rôle adaptatif important dans les

Introduction

interactions biotiques et abiotiques (Pichersky et Gershenzon 2002). Les terpènes sont entre autre impliqués dans l'attraction des pollinisateurs (Schnee, Kollner *et al.* 2002) et dans la défense directe et indirecte contre les insectes et les microorganismes (Martin, Tholl *et al.* 2002; Tholl 2006). La production et l'émission de CTV par les plantes sont aussi influencées par des facteurs abiotiques comme la lumière (Loreto, Barta *et al.* 2006; Yanhe, Craker *et al.* 1995), la température (Clark et Menary 1980; Duriyaprapan, Britten *et al.* 1986; Fahlén 1997), la fertilisation (Gouinguene et Turlings 2002). Ces molécules jouent également un rôle dans la protection contre le stress photooxydatif et la thermotolérance (Peñuelas et Munné-Bosch 2005).

Certains CTV présents dans les huiles essentielles de plantes sont recherchés car ils contribuent aux propriétés médicinales des HE. Par exemple, une activité antiseptique significative est due à la présence de terpènes comme le camphre (Santoyo, Cavero *et al.* 2005) ou le menthol (Imai, Osawa *et al.* 2001). Chez la menthe, l'huile essentielle agit aussi comme antispasmodique sur les muscles lisses des vertébrés (Lis-Balchin et Hart 1999) Par ailleurs, certaines huiles essentielles ont une activité antifongique et répulsive contre certains insectes comme les moustiques (Padilha de Paula *et al.*, 2003).

Les nombreux rôles et activités biologiques des CTV sont liés à la diversité et la variabilité des terpènes formés par les plantes.

I-3 Variations spatiales et temporelles de la production et de l'émission des COV

Dans diverses espèces de plantes il a été démontré que la composition en COV variait, qualitativement et quantitativement, au niveau de la plante entière en fonction de la saison (Angioni, Barra *et al.* 2006; Atti-Santos, Pansera *et al.* 2004; Johnson, Kazantzis *et al.* 2004) et au niveau de la plante ou de certains de ses organes durant leur développement (Dudareva, Negre *et al.* 2006; Sangwan, Farooqi *et al.* 2001). Ainsi, les fruits émettent des COV essentiellement lorsqu'ils sont à maturité (Aharoni, Giri *et al.* 2004). Les racines produisent par exemple des COV lors d'attaques par des herbivores (Chen, Ro *et al.* 2004; Rasmann et Turlings 2008). Dans les feuilles la composition en COV varie, avec l'âge et la position des feuilles chez *Cymbopogon flexuosus* (Singh, Luthra *et al.* 1989), chez *Artemisia sp.* (Holm, Laakso *et al.* 1997), chez *Salvia officinalis* (Croteau, Felton *et al.* 1981), chez *Mentha piperita* (Gershenzon, McConkey *et al.* 2000; Voirin et Bayet 1996) ou encore chez *Lavandula*

Introduction

latifolia (Muñoz-Bertomeu, Ros *et al.* 2008). Dans les fleurs, on observe des modifications dans l'émission de COV avec, chez la rose, un pic d'émission lors de l'ouverture du bouton floral (Bergougnoux, Caissard *et al.* 2007) ainsi que des variations dans la journée avec un maximum dans les périodes diurnes (Hendel-Rahmanim, Masci *et al.* 2007). Les mêmes variations sont observées chez *Antirrhinum sp.* (Dudareva, Martin *et al.* 2003), la fleur cessant même de produire certains COV après pollinisation (Negre, Kish *et al.* 2003). L'émission des COV par la fleur coïncide ainsi avec la période de présence des pollinisateurs et la maturité sexuelle de la plante.

I-4 Lieux de formation et d'accumulation

Selon les espèces, différentes structures plus ou moins complexes sont responsables de la biosynthèse, du stockage et de l'émission des COV. Chez de nombreuses plantes à fleurs, la synthèse et l'émission des COV sont localisées dans des cellules épidermiques coniques qui couvrent les pétales. Chez les Orchidacées ces cellules spécialisées sont localisées à certains endroits du pétale. On parle alors d'osmophores. La caractérisation des cellules coniques et leur implication dans la sécrétion de COV sont bien caractérisées chez la rose qui constitue donc un excellent modèle pour l'étude des cellules sécrétrices du pétale (Bergougnoux, Caissard *et al.* 2007). De nombreux tissus végétaux possèdent des cavités sécrétrices à la surface de l'épiderme ou à l'intérieur de l'épicarpe comme les fruits de *Citrus* ou les canaux résinifères des conifères (Fahn 1988). Chez les Gymnospermes, des structures sécrétrices plus ou moins spécialisées ont été décrites. On trouve par exemple chez *Thuja plicata* des cellules sécrétrices isolées réparties dans le bois, chez *Abies grandis* des cellules en forme d'ampoules (ou de citernes) séparées des autres qui déversent leurs sécrétions dans une lumière commune et chez *Picea sp* et *Pinus sp* le tronc et l'écorce comportent de véritables canaux résinifères (Trapp et Croteau 2001a).

Les trichomes glandulaires constituent un autre type de structure de sécrétion et/ou d'accumulation des CTV. Ces trichomes ont particulièrement été étudiés chez certaines familles botaniques telles que les Solanacées et les Lamiacées (Fahn 1988).

Chez les Lamiacées aromatiques on trouve deux types de glandes sécrétrices : les glandes peltées et les glandes capitées (Perrin et Colson 1986) (Figure 2 et Figure 3). Ces trichomes sécréteurs sont typiquement composés de quelques cellules épidermiques basales qui par division donnent de une à plusieurs cellule(s) de pied de taille variable surmontées d'une à

Introduction

plusieurs cellules de tête sécrétrices (Werker 1993). On observe par exemple chez *L. angustifolia*, huit cellules de tête pour les glandes peltées (Figure 4) et une ou deux cellules de tête pour les trichomes capités (Couderc-Le-Vaillant, Segur-Fantino *et al.* 1990). L'HE est accumulée sous la cuticule de ces cellules de tête (Figure 5) (Iriti, Colnaghi *et al.* 2006). La séquestration des CTV dans l'espace sous-cuticulaire s'explique en partie par le fait que certains terpènes sont phytotoxiques et qu'il est donc indispensable de les accumuler dans des structures externes spécialisées (Brown, Hegarty *et al.* 1987; Wagner 1991).

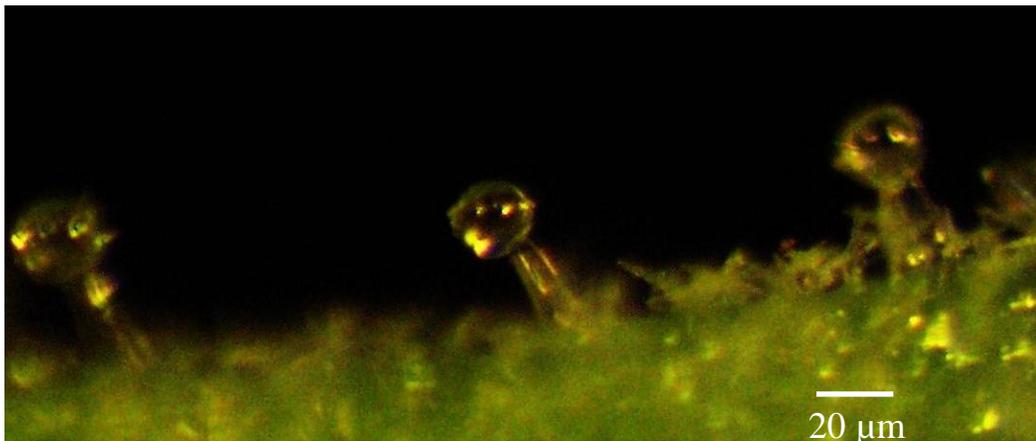


Figure 2 : Glandes capitées sur des feuilles de *L. multifida* (microscopie optique)

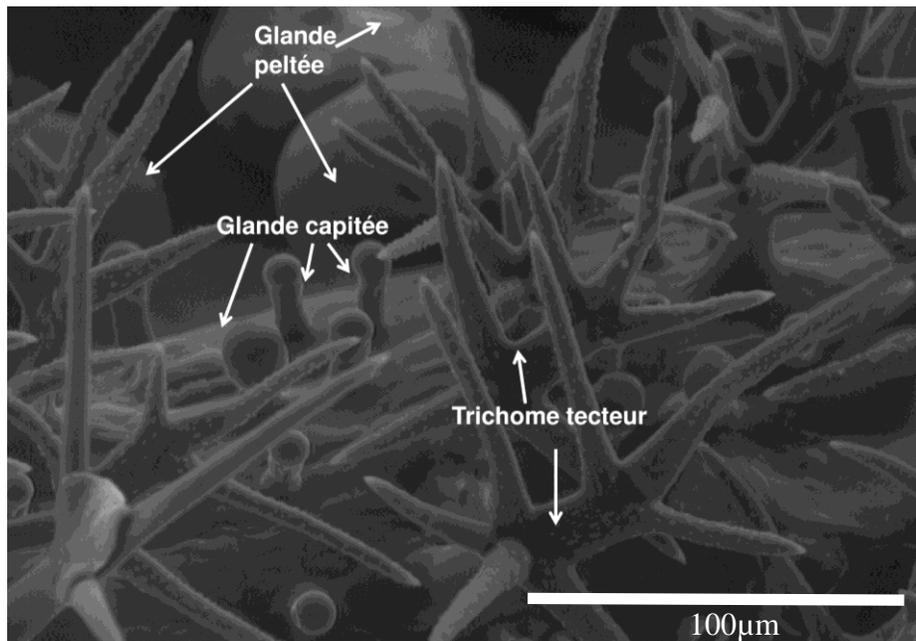


Figure 3 : Trois types de trichomes sur calice de *L. angustifolia* (MEB x500 12kV)

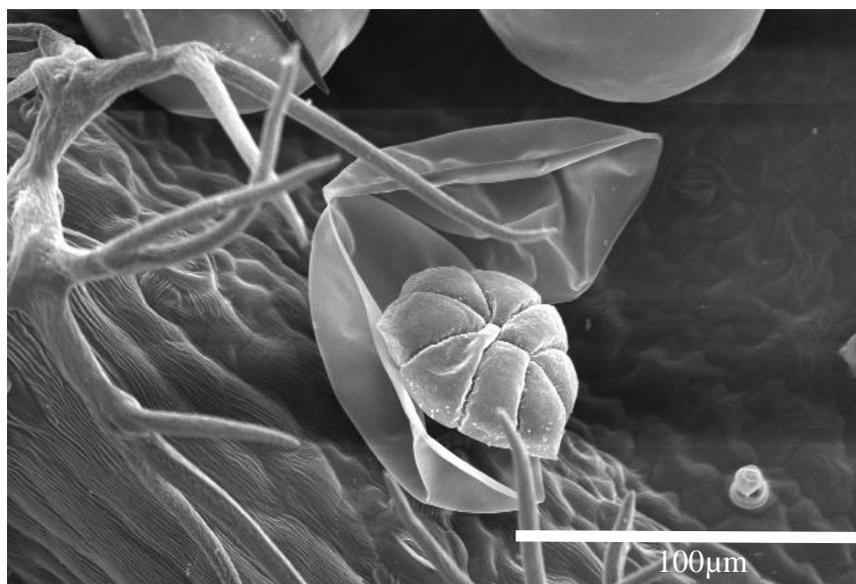


Figure 4 : Détail des 8 cellules sécrétrices d'une glande peltée sur calice de *L. lanata* (MEB x500 8kV)



Figure 5 : Schéma d'un trichome pelté de lavande (d'après Iriti, Colnaghi *et al.* 2006)

Parmi les COV, les terpènes sont des composés très étudiés et l'origine de leur diversité, leur formation et même l'évolution des enzymes responsables de leur formation sont largement décrites dans la littérature. Du fait de la forte production de CTV dans des cellules localisées et externes, donc facilement accessibles, les Lamiacées et notamment la menthe, la sauge ou encore le thym ont servi de modèle pour l'étude de la biosynthèse de ces terpènes et notamment des enzymes responsables de leur squelette carboné : les terpènes synthases (Alonso, Rajaonarivony *et al.* 1992).

I-5 Biosynthèse des terpènes

I-5.1 Origine et diversité des terpènes

A ce jour, entre 30 000 et 40 000 (Kampranis, Ioannidis *et al.* 2007) terpènes ont été identifiés. Les composés de la famille des terpènes ont des structures très variées, allant d'une

Introduction

simple chaîne linéaire d'hydrocarbones jusqu'à des agencements complexes de cycles carbonés (Connolly et Hill 1991). Ces terpènes, aux structures pourtant si variées, ont tous en commun d'être constitués de multiples sous unités d'isoprène (C_5H_8) d'où l'appellation d'isoprénoides donnée également aux terpènes. La formation du squelette carboné de tous ces terpènes est catalysée par des enzymes ; les terpènes synthases (TPS). Ces enzymes synthétisent une grande diversité de squelettes carbonés à partir de l'ionisation de substrats prenyl-diphosphates catalysée par la présence de cations métalliques divalents comme le Mg^{2+} ou le Mn^{2+} (Dewick 2002). Ces substrats sont biosynthétisés à partir de précurseurs communs dont le squelette principal a 5 carbones (C_5) : l'isopentenyl diphosphate (IPP) et son isomère le diméthylallyl diphosphate (DMAPP). L'IPP et le DMAPP sont en équilibre grâce à la simple activité d'une isomérase (Figure 6) (Rohmer et Rohmer 1999).

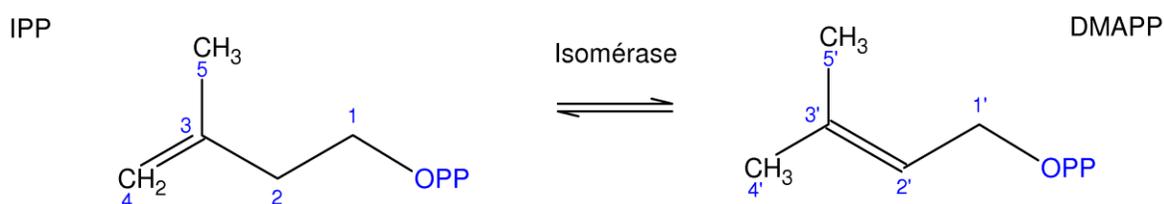


Figure 6 : Schéma de l'IPP et du DMAPP avec la numérotation des carbones

La diversité de ces composés repose sur différents facteurs tels que la longueur de la chaîne carbonnée, le type de condensation des précurseurs et le type de cyclisation auxquels s'ajoutent des modifications secondaires. Ainsi, les terpènes formés par les TPS sont classés par rapport au nombre d'unités en C_5 qui les constitue. On parle de monoterpènes, sesquiterpènes et diterpènes pour les molécules respectivement en C_{10} , C_{15} et C_{20} formées respectivement à partir des substrats prenyl-diphosphates suivants : géranyl diphosphate (GPP), farnésyl diphosphate (FPP) et géranylgéranyl diphosphate (GGPP). Ces précurseurs de terpènes sont synthétisés par des enzymes de la famille des prényltransférases par condensation de l'IPP avec le DMAPP, le GPP ou le FPP respectivement (Figure 7).

Introduction

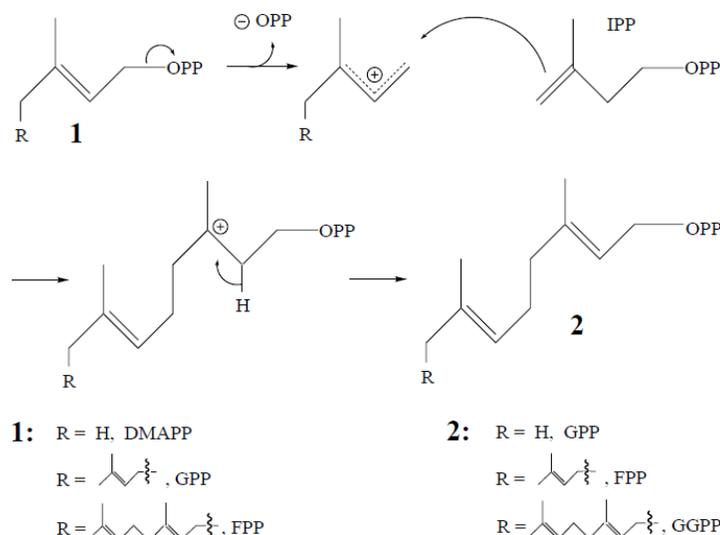


Figure 7 : Mécanisme de réaction des prényltransférases illustrant la liaison faite par la GPP synthase, la FPP synthase et la GGPP synthase d'après (Burke, Wildung *et al.* 1999)

La figure 7 montre que le GPP est formé par une réaction d'ionisation du DMAPP (perte du groupe OPP) suivie d'une réaction de liaison entre le carbone (1') du DMAPP et le carbone (4) de l'IPP, on parle aussi de liaison « tête-queue ». Il en est de même pour le FPP à partir du GPP et pour le GGPP à partir du FPP. Les molécules formées à partir de ces précurseurs issus de liaison (1'-4) sont des terpènes réguliers. Les terpènes irréguliers utilisent les mêmes précurseurs mais sont issus soit de cyclopropanation (c1'-2-3), soit de liaison latérale (1' - 2), soit de cyclobutanation (c1'-2-3-2') (Thulasiram, Erickson *et al.* 2008). Parmi les terpènes irréguliers on trouve le squalène et le phytoène à l'origine des stéroïdes et des caroténoïdes qui sont tous deux issus de condensations (1-1' - « tête-tête » puis cyclisation 2-3 de deux FPP et de deux GGPP, respectivement) (Thulasiram, Erickson *et al.* 2008; Zhao, Inoue *et al.* 2003). Certaines espèces de lavandes produisent un terpène irrégulier, lavandulol qui, chez le chrysanthème, est formé à partir du lavandulyl diphosphate est formé par condensation (1' - 2) de deux DMAPP (Rivera, Swedlund *et al.* 2001).

Introduction

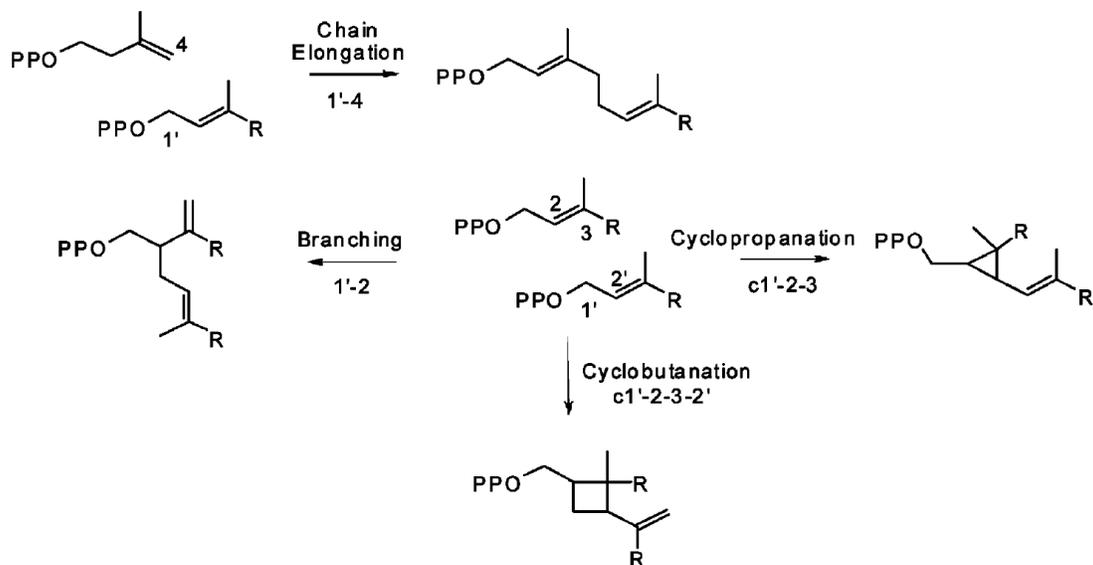


Figure 8 : Principales réactions de liaison à l'origine de la diversité des squelettes carbonés des terpènes (modifié d'après Thulasiram, Erickson *et al.* 2008)

Les TPS catalysent la transformation d'un prenyl diphosphate acyclique (GPP, FPP) en une multitude de formes acycliques, mono ou polycycliques (Figure 9, Degenhardt, Köllner *et al.* 2009). Pour les monoterpènes, les formes les plus communes sont les monoterpènes acycliques comme le linalol, les monoterpènes cycliques avec un cycle à 6 carbones, comme le limonène et les monoterpènes bicycliques comme les pinènes. Par ailleurs, il existe des monoterpènes hétérocycliques appelés iridanes ayant un cycle à 5 carbones (Hallahan 2000). Pour les sesquiterpènes les formes les plus communes sont acyclique, mono-, bi- et tricyclique (Dudareva, Negre *et al.* 2006; Fraga 2007).

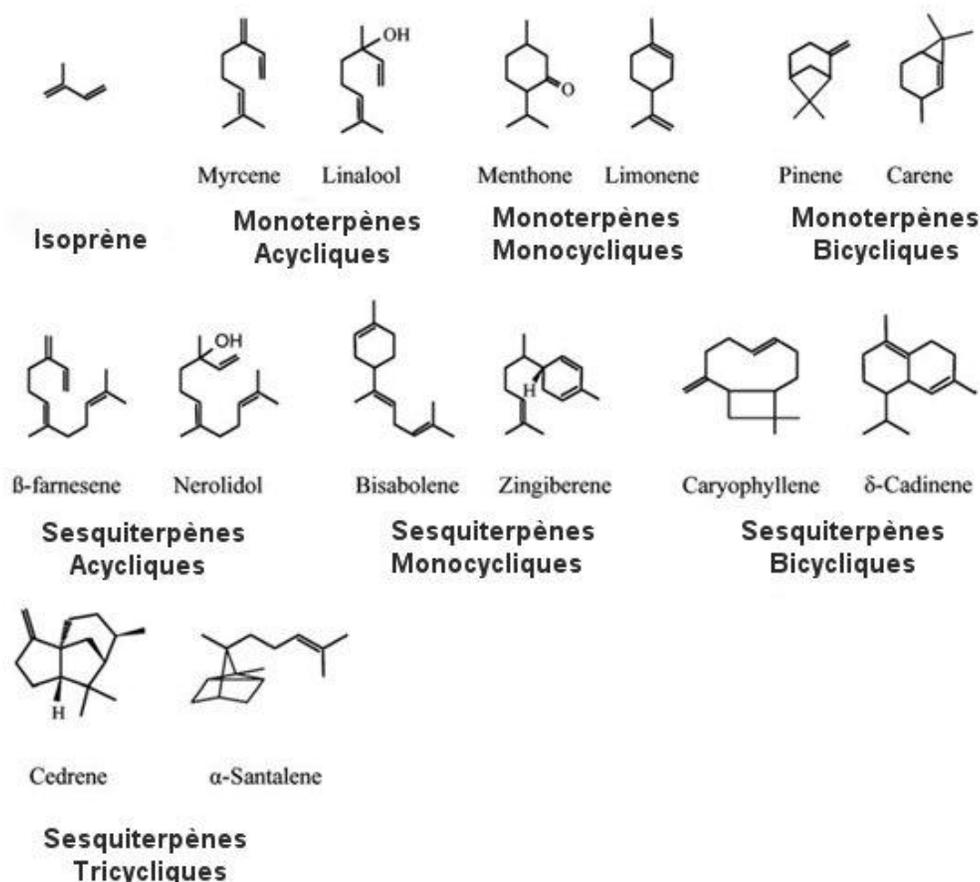


Figure 9 : Schémas des principales formes de squelettes carbonés des mono et sesquiterpènes (modifié d'après Dudareva, Negre *et al.* 2006)

En conclusion, la grande diversité des terpènes s'explique par les nombreuses possibilités de réarrangements des chaînes carbonées qui les constituent. Elle s'explique également par les modifications secondaires (acétylation, méthylation, oxydation, hydroxylation, réduction...) que ces molécules peuvent subir après leur formation sous l'action d'autres familles d'enzymes. Ainsi, chez le fraisier, l'*alpha*-pinène est modifié par une cytochrome P₄₅₀ hydroxylase (CytP₄₅₀) en myrténol puis ce dernier est transformé en acétate de myrtényl par une alcool acétyl transférase (AAT) (Aharoni, Giri *et al.* 2004). Ces modifications aboutissent généralement à une augmentation de la volatilité et une variation des propriétés olfactives des composés.

I-5.2 Les voies de synthèse de l'IPP et leur compartimentation : impact sur la synthèse des terpènes

Pendant plusieurs dizaines d'années, il a été admis que l'IPP et le DMAPP étaient synthétisés dans le cytosol à partir du mévalonate via la voie du mévalonate (MVA).

Introduction

Cependant, différentes études de surexpression de l'enzyme HMGR impliquée dans la voie du mévalonate ne conduisaient pas à une augmentation de la synthèse des monoterpènes. La découverte, dans les bactéries, d'une autre voie de synthèse de l'IPP indépendante du mévalonate a permis de comprendre *a posteriori* ces résultats (Rohmer, Knani *et al.* 1993). La présence de cette voie chez les plantes fut confirmée quelques années après par Arigoni et Lichtenthaler (Arigoni, Sagner *et al.* 1997 ; Lichtenthaler 1999). Il existe donc deux voies différentes dans la nature pour synthétiser l'IPP et le DMAPP. La voie cytosolique du mévalonate (MVA) à partir de l'acetyl-CoA, et la voie plastidiale du 2-C-methyl-D-erythritol 4-phosphate (MEP) à partir du 1-deoxyxylulose 5-phosphate (Rohmer 2007). Il a été démontré que ces deux voies métaboliques ne sont pas totalement disjointes et que des flux d'IPP entre le plaste et le cytosol existent (Hemmerlin, Hoeffler *et al.* 2003) et sont notamment régulés par la lumière (Rodriguez-Concepcion, Fores *et al.* 2004). La figure 10 montre également que quelle que soit l'origine de l'IPP (et du DMAPP), il existe chez les plantes une compartimentation cellulaire de la biosynthèse des terpènes. Les mono, di et tétraterpènes sont synthétisés dans les plastes alors que les sesqui et triterpènes sont d'origine cytosolique.

Introduction

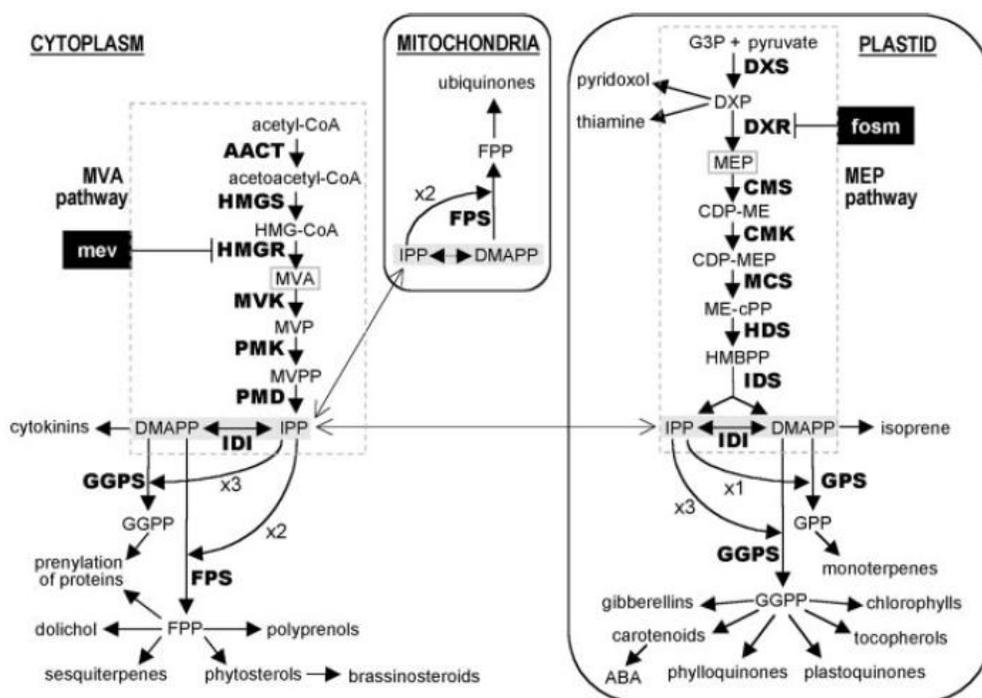


Figure 10 : Voies de synthèses des isoprénoïdes dans les plantes (modifié d'après Rodriguez-Concepcion, Fores *et al.* 2004)

ABA, acide abscisique; CDP-ME, 4-diphosphocytidyl-méthylerythritol; CDP-MEP, CDP-ME 2-phosphate; DMAPP, diméthylallyl diphosphate; DXP, deoxyxylulose 5-phosphate; FPP, farnésyl diphosphate; G3P, glyceraldéhyde 3-phosphate; GGPP, geranylgeranyl diphosphate; GPP, geranyl diphosphate; HMBPP, hydroxyméthylbutenyl 4-diphosphate; HMG-CoA, hydroxyméthylglutaryl coenzyme A; IPP, isopentényl diphosphate; ME-cPP, méthylerythritol 2,4-cyclodiphosphate; MEP, méthylerythritol 4-phosphate; MVA, acide mévalonique; MVP, 5-phosphomévalonate; MVPP, 5-diphosphomévalonate. Les enzymes sont indiquées en gras: AACT, acétoacétyl CoA thiolase (EC 2.3.1.9); CMK, CDP-ME kinase (EC 2.7.1.148); CMS, CDP-ME synthase (EC 2.7.7.60); DXR, DXP reductoisomérase (EC 1.1.1.267); DXS, DXP synthase (EC 4.1.3.37); IDS, IPP/DMAPP synthase; HDS, HMBPP synthase (EC 1.17.4.3); IDI, IPP isomérase (EC 5.3.3.2); HMGR, HMG-CoA réductase (EC 1.1.1.88); HMGS, HMG-CoA synthase (EC 4.1.3.5); MCS, ME-cPP synthase (EC 4.6.1.12); MVK, MVA kinase (EC 2.7.1.36); PMD, MVPP décarboxylase (EC 4.1.1.33); PMK, MVP kinase (EC 2.7.4.2);. Les étapes inhibées par la mevinoline (MEV) et la fosmidomycine (FSM) sont indiquées.

I-5.3 Les terpènes synthases

Classification

De nombreuses mono et sesquiterpènes synthases de plantes ont été isolées et caractérisées dans les organes reproducteurs et végétatifs d'Angiospermes et de Gymnospermes (revu dans Degenhardt, Köllner *et al.* 2009). Un des points communs entre ces enzymes est qu'elles sont, en général, capables de synthétiser plusieurs terpènes différents (Steele, Crock *et al.* 1998). Par exemple, la limonène synthase de menthe a pour produit majoritaire le limonène mais génère aussi des co-produits : le myrcène, l' α - et le β -pinène (McGarvey et Croteau 1995). On distingue deux classes de terpènes synthases définies selon leurs structures et leurs mécanismes réactionnels : la classe 1 et la classe 2. A partir des précurseurs (GPP, FPP, GGPP) les terpènes synthases de classe 1 vont, grâce à des cations métalliques (Mg^{2+} , Mn^{2+}) enlever le groupement pyrophosphate (rupture de la liaison C-OOP) et produire un carbocation qui, par des réactions secondaires de réarrangement, va pouvoir générer une grande diversité de terpènes. Le mécanisme réactionnel de classe 2 diffère dans le mécanisme de formation du carbocation. Le carbocation est obtenu par protonation d'une double liaison (carbone-carbone) ou d'un groupement époxyde (oxygène ponté sur une liaison carbone-carbone pour former un hétérocycle à 3) (Christianson 2008). Par analyse phylogénétique des séquences d'acides aminés, 7 sous-familles de TPS ont été définies (TPS-a à TPS-g) (Bohlmann, Martin *et al.* 2000; Bohlmann, Meyer-Gauen *et al.* 1998; Dudareva, Martin *et al.* 2003). Les TPS d'une même sous-famille ont des séquences d'acides aminés proches. Cependant, on ne peut pas déduire la fonction d'une TPS de la comparaison de sa séquence avec celles de TPS dont la fonction a été caractérisée (Aubourg, Aubourg *et al.* 2002; Bohlmann, Meyer-Gauen *et al.* 1998). En effet, deux TPS d'une même espèce ont souvent des séquences plus proches entre elles que deux séquences de même fonction mais d'espèces différentes. On distingue malgré tout un certain nombre de structures communes aux TPS.

Structures

En règle générale les monoterpènes synthases ont entre 600 et 650 acides aminés alors que les sesquiterpènes synthases sont plus courtes avec entre 550 et 580 acides aminés. Cette différence de taille s'explique par la présence, chez les monoterpènes synthases, d'un peptide d'adressage au plaste sur la partie N-terminale de la séquence. Ce peptide assure le transfert de

Introduction

la protéine du cytosol vers l'intérieur des plastes. Il est composé d'un grand nombre de résidus sérine (S) et thréonine (T), mais aucune séquence conservée d'acides aminés n'a été trouvée entre différentes séquences d'adressage (Bohlmann, Steele *et al.* 1997). Quatre TPS ont fait l'objet d'analyse cristallographique afin d'établir des corrélations entre la structure et la fonction de ces enzymes. Ces TPS sont : une sesquiterpène synthase (5-Epi-aristolochène) de *Nicotiana tabacum* (Starks, Back *et al.* 1997), et trois monoterpènes synthases, la bornyl diphosphate de *Salvia officinalis* (Whittington, Wise *et al.* 2002), la 1,8-cinéole synthase de *Salvia fruticosa* (Kampranis, Ioannidis *et al.* 2007) et la limonène synthase de *Mentha spicata* (Hyatt, Youn *et al.* 2007). Il a ainsi été démontré que, si ces TPS ont des fonctions différentes elles n'en partagent pas moins des structures communes. Les TPS des sous-familles TPS-a, b et d sont composées de deux domaines structuraux : un domaine C-terminal contenant le site catalytique et un domaine N-terminal semblable en structure à certaines glycosyl hydrolases (Trapp et Croteau 2001b) qui aurait un rôle dans le repliement correct de la protéine et dans la forme du site catalytique de la partie C-terminale (Kollner, Schnee *et al.* 2004). Le site catalytique de la partie C-terminal est une poche hydrophobe formée de 6 hélices-alpha et organisée de manière à laisser la partie diphosphate du substrat à l'entrée du site catalytique ou elle sera en contact avec le cation métallique (Mg^{2+}, Mn^{2+} pour les enzymes d'Angiospermes et K^+, Mn^{2+}, Fe^{2+} pour certaines enzymes de Gymnospermes) catalysant la rupture de la liaison C-OOP (Boeckelmann 2008).

Chez ces enzymes, on trouve différents motifs conservés :

- Il existe un domaine riche en aspartate (D) ; le motif DD_{xx}D localisé à l'entrée du site catalytique. Ce dernier est caractéristique des TPS. Ce motif, à cette place, est essentiel à la formation des terpènes car, comme l'ont montré des expériences de mutagénèse dirigée, il permet la fixation du cation métallique nécessaire à la rupture de la liaison C-OOP et l'orientation correcte du substrat dans le site catalytique (Cane, Xue *et al.* 1996; Lesburg, Zhai *et al.* 1997; Starks, Back *et al.* 1997). Il existe un contre exemple à cette séquence conservée, chez une sesquiterpène synthase ayant un motif ND_{xx}D en lieu et place du DD_{xx}D (Prosser, Altug *et al.* 2004). Cependant, cette modification n'affecte pas, chez cette enzyme, sa capacité catalytique.
- Un second site de fixation du cation métallique est parfois présent (de manière moins conservée) à l'opposé du motif DD_{xx}D à l'entrée du site actif. Il s'agit parfois d'un

Introduction

second motif DD_{xx}D (Steele, Crock *et al.* 1998) ou bien d'un motif NSE/DTE (*i.e.* (N,D)D(L,I,V)_x(S,T)_{xxx}E) (Christianson 2006; Degenhardt, Köllner *et al.* 2009).

- Par ailleurs, 35 acides aminés en amont du motif DD_{xx}D on trouve un motif conservé R_xR qui empêche les attaques nucléophiles sur l'intermédiaire carbocationique issu de l'ionisation du substrat prenyl-diphosphate (Starks, Back *et al.* 1997).
- A environ 60 acides aminés de l'extrémité N-terminale de la séquence des TPS, on trouve un motif conservé RR (ou RR_{x8}WD). Les acides aminés situés avant ce motif RR n'ont pas d'impact sur la fonction de l'enzyme. Par contre l'absence du motif RR empêche la fixation du GPP mais pas celle du linalyl diphosphate (LPP). Ceci suggère que ce motif RR participe à l'isomérisation du GPP en un composé cyclisable probablement via l'*alpha*-terpinyl cation (Williams, McGarvey *et al.* 1998). Ce motif pourra donc être absent chez les synthases formant des monoterpènes acycliques à partir du linalyl cation issu du LPP ou directement à partir du géranyl cation (Degenhardt, Köllner *et al.* 2009).
- Certaines TPS possèdent au niveau du site catalytique un motif LQLYEASFLL (Landmann, Fink *et al.* 2007; McGeady et Croteau 1995; Wise, Savage *et al.* 1998). Ce dernier n'est, pour le moment, pas associé à une fonction particulière

La composition en acides aminés du site actif ayant un rôle prépondérant sur la fonction (Greenhagen et Chappell 2001; Greenhagen, O'Maille *et al.* 2006; Köllner, Schnee *et al.* 2004), il a, dernièrement, été fait une étude phylogénétique des TPS basée uniquement sur les acides aminés des sites actifs des mono et sesquiterpènes synthases caractérisées. Cette étude a démontré qu'à défaut de donner la fonction des TPS, ou de discriminer les TPS à un produit des TPS à multiples produits, on pouvait tout de même faire le lien entre la séquence du site actif et la possibilité de former des terpènes cycliques ou acyclique. La séquence du site actif renseignerait aussi sur le carbocation préférentiellement formé (Degenhardt, Köllner *et al.* 2009). Les auteurs concluent qu'une meilleure connaissance des TPS au sein d'autres genres ainsi que sur les forces cinétiques et thermodynamiques en jeu, aiderait à résoudre les relations structure-fonction des TPS.

Evolution

Les ressemblances entre TPS de Gymnospermes et d'Angiospermes laissent penser à une origine évolutive commune (Back et Chappell 1995; Colby, Alonso *et al.* 1993). Les premières preuves vinrent avec les travaux de Bohlmann *et al* en 1998 et la classification des TPS en 6 sous-familles. Ensuite, les études de phylogénie moléculaire montrèrent que l'histoire évolutive de la grande famille multigénique des TPS s'est faite à partir d'une duplication d'un gène du métabolisme primaire qui existait chez une plante avant la séparation des Angiospermes et des Gymnospermes (- 300 Ma). Ainsi, une copie d'un gène ancestral d'une TPS, probablement une diterpène synthase impliquée dans le métabolisme primaire, a muté et progressivement évolué pour donner les TPS impliquées dans le métabolisme secondaire. L'autre copie du gène ancestral est restée conservée et est probablement à l'origine des TPS responsables de la formation des gibbérellines (Trapp et Croteau 2001b). Ce type de mutations et néofonctionnalisation de gènes après des évènements de duplications est fréquent, mais, seules celles procurant un avantage évolutif sont conservées. C'est ainsi que certains évènements évolutifs vont avoir un impact sur la production de plusieurs molécules en même temps. Ainsi, par exemple, une simple modification d'acides aminés dans le site catalytique d'une TPS formant plusieurs produits peut modifier les produits formés à partir du même substrat (Katoh, Hyatt *et al.* 2004) ou bien en permettant à la TPS d'accepter un nouveau substrat (e.g du FPP au lieu du GPP Kampranis, Ioannidis *et al.* 2007, ou bien du (Z,Z)-FPP au lieu du (E,E)-FPP; Sallaud, Rontein *et al.* 2009). Certaines TPS peuvent naturellement accepter divers substrats (Landmann, Fink *et al.* 2007; Schnee, Kollner *et al.* 2002). Ainsi, la perte du peptide d'adressage chloroplastique d'une monoterpène synthase peut conduire à la formation d'une sesquiterpène synthase en plaçant cette nouvelle synthase en contact avec le FPP présent dans le cytoplasme par opposition au GPP présent dans le chloroplaste (Landmann, Fink *et al.* 2007). Les études sur l'évolution des TPS montrent que les TPS de Gymnospermes et d'Angiospermes ont évolué indépendamment et que l'acquisition de fonctions communes entre des TPS de plantes de ces deux groupes reflète une convergence évolutive (Degenhardt, Köllner *et al.* 2009; Pichersky et Gang 2000). En parallèle, d'autres familles d'enzymes en évoluant ont eut un impact sur la formation de terpènes par les plantes (Pichersky et Gang 2000). L'évolution de certaines FPP synthases semble avoir conduit à la formation de TPS capables de former des terpènes irréguliers (Hemmerlin, Rivera *et al.* 2003; Rivera, Swedlund *et al.* 2001). De même, des modifications dans les enzymes responsables

Introduction

des transformations secondaires des terpènes, comme par exemple l'inactivation d'une CytP₄₅₀ hydroxylase, peut conduire à la perte de certains terpènes (Aharoni, Giri *et al.* 2004; Schwab 2003).

Certes, l'apparition de nouvelles molécules peut générer un avantage évolutif et peut conduire au maintien dans le génome des nouveaux gènes impliqués dans leur production. Mais, comme le rôle des terpènes est parfois lié aux quantités de molécules produites, l'avantage évolutif peut également venir de modifications dans les mécanismes de régulation de l'expression des TPS (Theis et Lerdau 2003).

Expression et régulation des terpènes synthases

L'expression des TPS est souvent modifiée pendant le développement de la plante et en réponse à des stress biotiques et abiotiques. Il a été démontré que la régulation de l'expression des TPS étaient principalement transcriptionnelle (Tholl 2006). Chez *Mentha piperita* la transcription de monoterpènes synthases et de cytP₄₅₀ aussi impliquées dans la voie de biosynthèse des terpènes s'effectue de manière coordonnée durant le développement de la feuille (McConkey, Gershenzon *et al.* 2000). Ceci est également vérifié chez *Antirrhinum majus* pour les gènes impliqués dans la synthèse du linalol et du nérolidol dans les pétales (Nagegowda, Gutensohn *et al.* 2008) et chez *Citrus sinensis* pour les gènes impliqués dans la synthèse du valencène lors de la maturation des fruits (Sharon-Asa, Shalit *et al.* 2003). Ceci suggère que la régulation de la production de terpènes par les plantes est largement effectuée au niveau transcriptionnel. Paradoxalement, un seul facteur de transcription de la famille WRKY, agissant sur une sesquiterpène synthase à été mis en évidence chez le coton (Xu, Wang *et al.* 2004). Certaines TPS sont régulées au niveau post-transcriptionnel ou post-traductionnel chez *Arabidopsis thaliana* (Tholl, Chen *et al.* 2005) et chez *Antirrhinum majus* (Nagegowda, Gutensohn *et al.* 2008). La production de terpènes semble donc être régulée à différents niveaux selon les espèces et les synthases concernées (Besser, Harper *et al.* 2009).

II Le genre *Lavandula*

II-1 Utilisation industrielle

La production d'essences aromatiques par les plantes a un grand intérêt économique. En France, les Plantes à Parfum Aromatiques et Médicinales (PPAM) représentaient en 2006 un marché de 85 Millions d'euros (ONIPPAM 2006). Parmi celles-ci, la lavande (*L. angustifolia*) et le lavandin (*L. × intermedia*) sont connus et récoltés depuis des siècles pour divers usages allant de la parfumerie à la médecine. Les vertus et applications des huiles essentielles (HE) de lavandes sont décrites dans des ouvrages depuis l'antiquité et le moyen âge.

L'huile essentielle de lavande est utilisée dans l'industrie de la lessive et de la savonnerie, ainsi qu'en parfumerie. La lavande est également employée en herboristerie, en aromathérapie et est considérée comme une plante médicinale pour l'action de son huile. En effet, celle-ci est utilisée pour soigner des plaies et brûlures superficielles et présente des effets sédatifs, antibactériens, antifongiques, antidépresseurs (Cavanagh et Wilkinson 2002) et anti-inflammatoire (Sosa, Altinier *et al.* 2005). Les propriétés médicinales et le parfum des huiles essentielles de lavandes sont principalement attribués aux composés organiques volatils de la famille des terpènes. Ce sont les monoterpènes et les sesquiterpènes qui donnent à la lavande son parfum caractéristique (Flores, Blanch *et al.* 2005) et à ces huiles leurs propriétés. La proportion des principaux terpènes dans l'HE est un critère d'évaluation de la qualité de l'HE (AFNOR 1996) et joue un rôle dans le choix de telle ou telle variété par l'herboriste ou l'industriel (Grayer, Kite *et al.* 1996). Par exemple, le lavandin (*L. × intermedia* Emeric ex Loisel), hybride naturel entre *L. latifolia* et *L. angustifolia*, à un meilleur rendement de production d'HE vis-à-vis de son parent *L. angustifolia* ce qui en fait une espèce de choix pour l'industrie. Mais, son HE contient une grande quantité de camphre, un monoterpène, ce qui joue en sa défaveur auprès des parfumeurs.

En France, ce n'est que depuis les années 1920, à la demande des parfumeurs de Grasse, qu'a été entreprise la culture des lavandes. Avant cette date, la production d'HE était liée à l'activité de cueillette de lavandes sauvages faite par les bergers et quelques paysans dans les *baïassières* (nom local des lavanderaies naturelles) des collines de Provence. Jusqu'aux années 1970, la France a gardé une place prépondérante parmi les pays producteurs d'HE de lavandes et lavandins. Cependant, avec l'arrivée de la Bulgarie et de la Russie sur le

Introduction

marché des HE, suivie par celle de la Chine, au début des années 1980, la production française voit la liste de ses concurrents directs augmenter considérablement (Upson et Andrews 2004). Pour faire face à une concurrence internationale de plus en plus forte les producteurs français se sont organisés au sein d'une filière PPAM et ont mis en place un label de qualité : l'appellation d'origine contrôlée (AOC) « Huile essentielle de lavande de Haute-Provence » qui a été reconnue par un décret du 14 décembre 1981. Cette appellation permet d'identifier une production de très haute qualité concernant uniquement l'huile essentielle de lavande de population. Les plantations doivent être situées dans un territoire déterminé à une altitude minimale de 800 m (par dérogation à 600 m pour quelques communes plus septentrionales de la Drôme). L'aire géographique ne concerne que 284 communes des Alpes-de-Haute-Provence, des Hautes-Alpes, de la Drôme et du Vaucluse. De plus, afin d'obtenir la mention AOC, l'huile essentielle est soumise à un examen olfactif puis à des analyses physico-chimiques qui déterminent si sa composition s'inscrit dans les normes de qualité retenues par le décret.

En 2009, les surfaces cultivées en France étaient de près de 15 000 ha pour le lavandin et 4000 ha pour la lavande. Les principales régions productrices sont les Alpes de Haute Provence, les Hautes Alpes, la Drôme et le Vaucluse. Elles concentrent 94% de la production (Figure 11). Les autres départements producteurs sont l'Ardèche, le Gard, le Lot et les Bouches-du-Rhône.

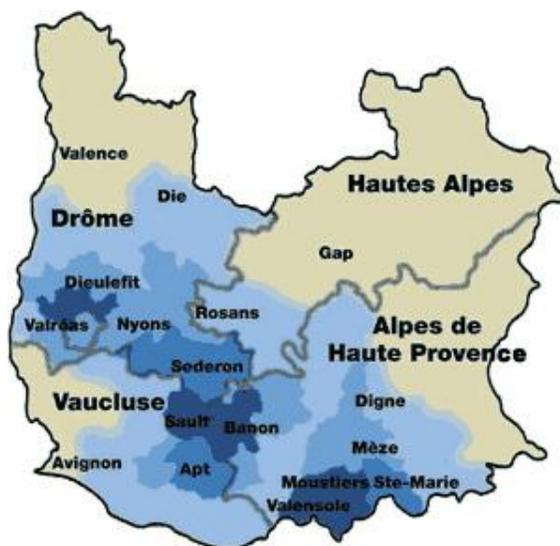


Figure 11 : Principales zones de production d'HE de lavande et lavandin en France (source CPPARM)
Le dégradé de bleu indique l'importance de la production.

Les surfaces de lavandin sont majoritairement plantées avec le cultivar 'Grosso' viennent ensuite 'Abrial' 'Super' et 'Sumian' (L'Agriculture drômoise 2007). Sur les 1036

Introduction

tonnes vendues en 2009, 838 tonnes (80%) sont issues de 'Grosso'. Les lavandes fines sont réparties en deux groupes : les lavandes clonales et les lavandes de population. Au total 40 tonnes d'HE de lavandes fines ont été vendues en 2009, dont 30 tonnes pour les lavandes clonales (Krausz et Chaisse 2009). Jusqu'à la fin des années 1990, seules deux variétés de lavandes étaient cultivées par les producteurs français (Maillette et Matheronne). Plus récemment, de nouvelles variétés comme 77-13, Diva ou encore C15-50 ont été sélectionnées par les centres techniques et les chambres d'agriculture de la Drôme et du Vaucluse (Chaisse et Fontaine 2009).

Ainsi, la production française d'huile essentielle de lavandes et lavandins représente respectivement 50% et 80% de la production mondiale et concourt largement au rayonnement international des régions Rhône-Alpes et Provence-Alpes-Côte-d'Azur. Mais, au-delà des aspects de volumes d'HE échangés sur les marchés mondiaux, la culture de la lavande participe fortement à l'image du sud de la France et attire de nombreux touristes. Au total, le chiffre d'affaire annuel induit directement ou indirectement, tous secteurs confondus (agriculture, industrie, services) dépasse plusieurs milliards d'euros. Le tourisme est lui aussi à compter parmi les retombées indirectes de la culture des lavandes. Une enquête de l'Office National Interprofessionnel des Plantes à Parfum, Aromatiques et Médicinales (ONIPPAM) montre que 75% des sondés choisissent la lavande à égalité avec le soleil comme identifiant de la Provence. En effet, que seraient les paysages provençaux sans les champs de lavande ? L'importance économique des PPAM, et, entre autre des lavandes, explique que de nombreux acteurs se soient engagés pour pérenniser et développer la filière PPAM en France.

En France, la filière de production d'HE de PPAM, dont font partie la lavande et le lavandin, est animée par différents organismes. De la culture de la plante à la commercialisation des produits, chacun d'eux met ses compétences au service des producteurs et des industriels de la filière afin d'en assurer le développement. Le Conservatoire National des Plantes Médicinales, Aromatiques et Industrielles (CNPMAI) participe à la conservation des espèces, à des activités pédagogiques et vend des plants et des semences de PPAM. La sélection de nouvelles variétés, les expérimentations pour valider la qualité des HE produites sur différents sites et l'appui technique pour les professionnels des PPAM sont assurés par l'Institut Technique Interprofessionnel des Plantes à Parfum, Aromatiques et Médicinales (ITEIPMAI). C'est le Conseil d'Orientation Scientifique et Technique (COST) de cet institut qui anime et coordonne actuellement les diverses activités de recherche et de développement

Introduction

de la filière. Les innovations techniques tant sur l'itinéraire cultural que sur la récolte et la distillation sont en grande partie issues des travaux du Centre Régionalisé Interprofessionnel d'Expérimentation en Plantes à Parfum, Aromatiques et Médicinales (CRIEPPAM). La veille réglementaire et économique est faite par ONIPPAM. Depuis le décret no 2009-340, du 27 mars 2009 l'ONIPPAM fait partie de FranceAgriMer. FranceAgriMer est l'établissement national des produits de l'agriculture et de la mer, créé le 1er avril 2009. Il a en charge la gestion des filières des grandes cultures, de l'élevage, de la pêche, de l'aquaculture, des vins, des fruits et légumes, de l'horticulture, des plantes à parfum, aromatiques et médicinales. L'ONIPPAM est accompagné dans cette tâche par les chambres d'agriculture et le Comité Interprofessionnel des Huiles Essentielles Françaises (CIHEF). Le comité est chargé de développer l'utilisation des huiles essentielles françaises sur le marché mondial, d'étudier et de promouvoir toutes mesures d'ordre scientifique et technique visant à améliorer la compétitivité et la qualité de la production française et de favoriser la promotion et la mise en marché des huiles essentielles dont l'Interprofession à la charge. Le Comité des Plantes à Parfum, Aromatiques et Médicinales (CPPARM, anciennement CEPPARM) est un comité économique qui mène des actions en faveur des groupements de producteurs pour une meilleure connaissance des marchés, la coordination d'actions commerciales, la mise en place de programmes techniques ou d'études économiques. La filière dispose également d'un appui financier régional Rhône-Alpes à travers le Pôle d'Expérimentation et de Progrès des Plantes à Parfums Aromatiques et Médicinales (PEP PPAM) et national à travers le pôle de compétitivité Plantes Arômes Senteurs Saveurs (pôle PASS) et le pôle Végépolis pour promouvoir la culture des PPAM. Grâce aux travaux de ces organismes, la filière a pu se mettre en conformité avec la réglementation REACH (enRegistrement, Evaluation et Autorisation des substances Chimiques, règlement CE 1907/2006), mettre en place des actions en faveur du développement durable via les actions du CIHEF, les essais de désherbage mécanique (sans intervention de produits phytosanitaires) et le travail sur les économies d'énergies fossiles (récupération de chaleur lors de la distillation...) initiés par le CRIEPPAM. Un des défis majeurs reste la lutte contre le dépérissement ; cette maladie à phytoplasmes affecte les lavandes et les lavandins. Dans cette optique, le CRIEPPAM a sélectionné des variétés tolérantes (Diva, Clara).

La filière PPAM encourage donc les recherches qui permettent de mieux connaître et maîtriser la production des COV par les plantes aromatiques et médicinales.

II-2 Taxonomie et biodiversité

Les espèces du genre *Lavandula* appartiennent à la famille des *Lamiaceae*. Cette famille qui regroupe 7200 espèces réparties dans 236 genres se caractérise généralement par des plantes à la tige quadrangulaire, aux feuilles opposées et décussées et au calice persistant entourant, à maturité, un tétrakène. La famille des *Lamiaceae* est subdivisée en 7 sous-familles. Le genre *Lavandula* appartient à la sous-famille des *Nepetoideae* du fait du caractère hexaperturé des grains de pollen. La sous-famille des *Nepetoideae* est divisée en 3 tribus dont celle des *Ocimeae* dans laquelle sont placées les lavandes. Le genre *Lavandula* a été rattaché à cette tribu sur la base d'une étude phylogénétique conduite à partir de l'étude du polymorphisme de séquences d'ADN chloroplastique (Paton, Springate *et al.* 2004). Bien que le rattachement des lavandes à la tribu des *Ocimeae* soit encore l'objet de recherches, cette filiation appuierait l'hypothèse selon laquelle l'origine de diversification des lavandes serait localisée en Asie comme les autres genres de cette tribu. L'histoire taxonomique du genre *Lavandula* a été particulièrement bien décrite dans « The genus *Lavandula* » (Upson et Andrews 2004). Brièvement, dès la renaissance déjà 5 espèces de lavande étaient reconnues. Le nombre d'espèces décrites n'a cessé d'augmenter au fil des explorations botaniques dans le Maghreb et l'Asie (Tableau 1). Ces études ont conduit à reconnaître actuellement 39 espèces différentes réparties en trois sous-genres *Fabricia*, *Sabaudia* et *Lavandula* et huit sections (Tableau 2). Cette structuration taxonomique est appuyée aussi par une analyse phylogénétique du polymorphisme de séquences ITS (*Internal Transcribed Spacer* –séquence interne traduite) (Figure 12).

Introduction

Tableau 1 : Historique des principales évolutions de la classification du genre *Lavandula* (modifié d'après Gingins de la Sarraz 1826; Upson 1997)

Date	Auteurs	Observations majeures
Pline	76 avant JC	séparation de <i>L.stoechas</i> et de <i>L.angustifolia</i> (probablement)
John Gerard	1597	séparation en deux groupes <i>Lavandula</i> et <i>Stoechas</i> avec chez <i>Lavandula</i> : <i>lavandula flore caeruleo</i> (<i>L.latifolia</i>) <i>lavandula flore albo</i> (<i>L.xintermedia</i> forme blanche) <i>lavandula minor siue spica</i> (<i>L.angustifolia</i>) avec chez <i>Stoechas</i> : <i>Stoechas sive spica hortulana</i> (<i>L.stoechas</i> ssp <i>stoechas</i>) <i>Stoechas multifida</i> (<i>L.multifida</i>) <i>Stoechas folio serato</i> (<i>L.dentata</i>) <i>Stoechas summis cauliculis nudis</i> (<i>L.pedunculata</i>)
Joseph Tournefort	1700	<i>L.multifida</i> placée dans le groupe <i>Lavandula</i>
Carl Linné	1753	inclusion des <i>Stoechas</i> dans le genre <i>Lavandula</i> ajout de <i>L.viridis</i>
Michel Adanson	1763-1764	classification en trois genres en retirant <i>L.multifida</i> des <i>Stoechas</i> pour le placer dans les <i>Fabricia</i> 1 <i>Fabricia</i> 2 <i>Lavandula</i> 3 <i>Stoechas</i>
Philip Miller	1768	ajout de <i>L.canariensis</i> (mais dans les sous genre <i>Lavandula</i>) car il ne tient pas compte des travaux de Adanson
Johannes D. Lundmark	1780	ajout de <i>L.pinnata</i> et description de variétés pour <i>L.spica</i> (<i>L.latifolia</i>), <i>L.stoechas</i> et <i>L.multifida</i>
Baron Frédéric Charles Jean de Gingins de la Sarraz	1826	ajout d'espèces et d'hybrides, classification en 3 sections: <i>Stoechas</i> <i>L.stoechas</i> L. <i>L.pedunculata</i> Cav. <i>L.viridis</i> L'Hér. <i>L.dentata</i> L. <i>Spica (Lavandula)</i> <i>L.heterophylla</i> Poir. (<i>L.dentata</i> x <i>L.latifolia</i>) <i>L.pyrenaica</i> DC. (<i>L.angustifolia</i> ssp <i>pyrenaica</i>) <i>L.vera</i> DC. emend Gingins (<i>L.angustifolia</i> Miller) <i>L.spica</i> DC. emend Gingins (<i>L.latifolia</i>) <i>Pterostoechas (Fabricia de Adanson)</i> <i>L.pinnata</i> L.f. <i>L.coronopifolia</i> Poir. <i>L.abrotanoides</i> Lam. (<i>L.canariensis</i>) <i>L.multifida</i> L.
George Bentham	1832-1836	le genre <i>Lavandula</i> est placé dans la tribus des <i>Ocymoiideae</i> ajout d'une nouvelle section : <i>Chaetostachys</i> ajout d'espèces: <i>L.burmanni</i> <i>L.rotundifolia</i>
D. A. Chaytor	1937	ajout d'une nouvelle section : <i>Subnudae</i> ajout d'une nouvelle espèce: <i>L.somaliensis</i> compilation des travaux précédents avec 28 espèces classées en 5 sections
Arnoldo Rozeira	1949	révision et modification de la section <i>Stoechas</i> par la création de deux sous section <i>Dentata</i> et <i>Stoechas</i>
Devetak et Cenci	1963	classification des sections dans un ordre reflétant une possible phylogénie avec deux grandes branches 1 <i>Spica (Lavandula)</i> <i>Stoechas</i> 2 <i>Pterostoechas</i> <i>Subnudae</i> <i>Chaetostachys</i>
El-Gazzar et Watson	1968	même conclusions que Devetak et Cenci sur la base de caractères morphologiques
Antony G. Miller	1985	révision des espèces d'Arabie et du nord-est de l'Afrique ajout d'espèces: <i>L.aritibracteata</i> <i>L.citriodora</i> <i>L.dhofarensis</i> <i>L.galgalloensis</i> <i>L.hasikensis</i> suggère qu' <i>L.atriplicifolia</i> est liées au genre <i>Sabaudia</i>
Maria C. León-Arencibia	1987	étude des espèces de îles canaries
Maria Suárez-Cervera	1985-1989	création d'une section <i>Dentata</i> utilisation de techniques modernes (Microscopie électronique à balayage, Caryotypes...)
Tim Upson	1997	classification en trois sous-genres et huit section utilisation de techniques de biologie moléculaire

Introduction

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

Genre	sous genre	section		
<i>Lavandula</i>	<i>Lavandula</i>	<i>Lavandula</i>	<i>L.angustifolia</i> Mill.	
			<i>L.lanata</i> Medik.	
			<i>L.latifolia</i> Boiss.	
		<i>Dentatae</i>	<i>L.dentata</i> L.	
			<i>Stoechas</i>	<i>L.pedunculata</i> (Mill) Cav.
				<i>L.stoechas</i> L.
		<i>L.viridis</i> L'Hér.		
		<i>Fabricia</i>	<i>Pterostoechas</i>	<i>L.multifida</i> L.
				<i>L.canariensis</i> Mill.
	<i>L.minutolii</i> Bolle.			
	<i>L.bramwellii</i> Upson & S.Andrews			
	<i>L.pinnata</i> L.f.			
	<i>L.buchii</i> Webb & Berthel.			
	<i>L.rotundifolia</i> Benth.			
	<i>L.maroccana</i> Murb.			
	<i>L.tenuisecta</i> Coss. Ex Ball			
	<i>L.rejadalii</i> Upson & Jury			
	<i>L.mairei</i> Humbert			
	<i>L.coronopifolia</i> Poir.			
	<i>L.saharica</i> Upson & Jury			
	<i>L.antinea</i> Maire			
	<i>L.pubescens</i> Decne.			
	<i>L.citriodora</i> A.G. Mill.			
	<i>Subnudae</i>		<i>L.subnuda</i> Benth.	
			<i>L.macra</i> Baker	
			<i>L.dhofarensis</i> A.G. Mill.	
			<i>L.samhanensis</i> Upson & S.Andrews	
			<i>L.setifera</i> T. Anderson	
			<i>L.qishnensis</i> Upson & S.Andrews	
			<i>L.nimmoi</i> Benth.	
			<i>L.galgalloensis</i> A.G. Mill.	
			<i>L.aristibracteata</i> A.G. Mill.	
			<i>L.somaliensis</i> Chaytor	
	<i>Chaetostachys</i>		<i>L.bipinnata</i> (Roth) Kuntz	
			<i>L.gibsonii</i> J. Graham	
	<i>Hasikenses</i>		<i>L.haskinensis</i> A.G. Mill.	
			<i>L.sublepidota</i> Rech.f.	
	<i>Sabaudia</i>		<i>Sabaudia</i>	<i>L.atriplicifolia</i> Benth.
				<i>L.erythraeae</i> (Chiov.) Cufod

Introduction

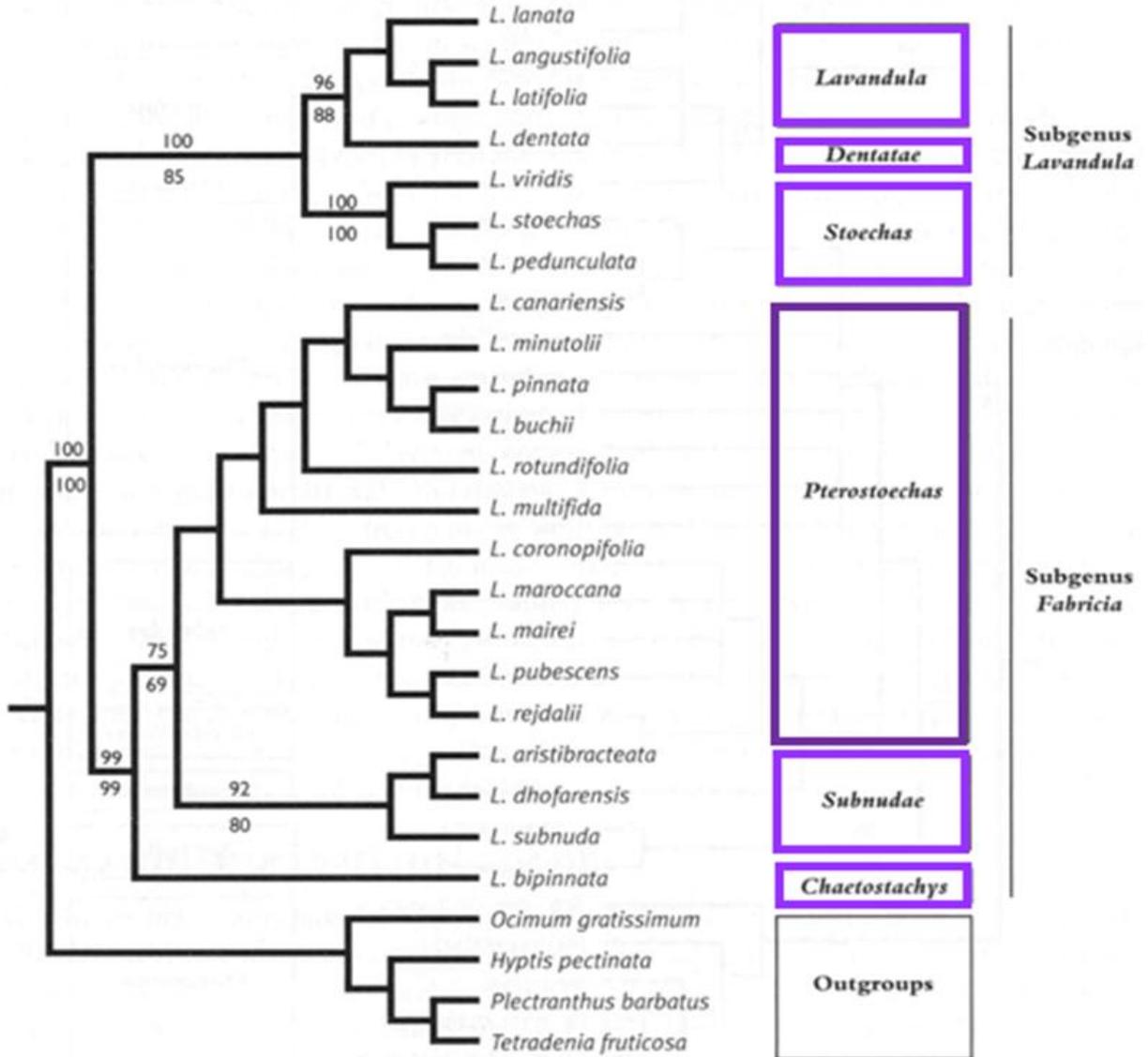


Figure 12 : Arbre de classification des espèces du genre *Lavandula* basé sur les séquences nucléaire ITS (modifié d'après Upson et Andrews 2004)

Introduction

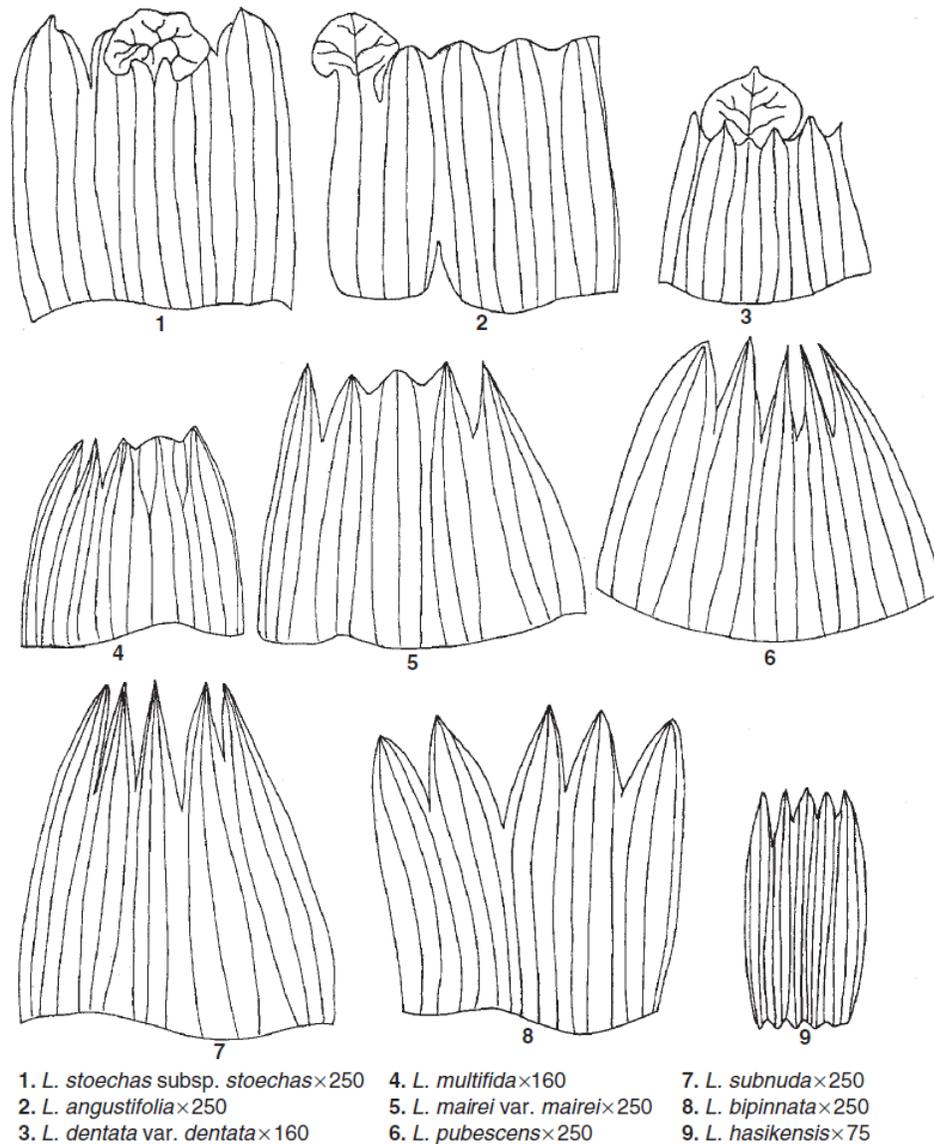


Figure 13: Variations de la morphologie du calice dans certaines espèces du genre *Lavandula* (modifié d'après Tim Upson dans Lis-Balchin 2002)

Botaniquement, le genre *Lavandula* dans la Flore de Coste (Coste et Flahault 1998) se différencie des autres genres de la Famille des *Lamiaceae* par des étamines et un style inclus totalement dans le tube de la corolle. Dans *Flora Europea* le caractère morphologique qui distingue le genre *Lavandula* est l'appendice obovale qui recouvre la pointe du calice (Tutin, Heywood *et al.* 1972). Ce caractère est discriminant si on ne considère que les lavandes européennes mais il n'est pas pertinent, car absent, chez de nombreuses espèces africaines ou asiatiques (Figure 13, Lis-Balchin 2002).

La description morphologique des lavandes a été particulièrement détaillée par Upson et Andrews (2004). Parmi les caractères qui permettent le mieux de distinguer les différentes

Introduction

espèces nous retiendrons la morphologie des feuilles, la forme de l'inflorescence, des bractées, du calice, et de la corolle même si d'autres caractères moins facilement visibles entrent aussi en compte dans la détermination des espèces comme le nombre de chromosomes, la forme des grains de pollen (Suarez-Cervera et Seoane-Camba 1986) ou encore celle des graines.

La morphologie des feuilles dans le genre *Lavandula* est très variable (Figure 14). Alors que les feuilles du sous-genre *Lavandula* et *Sabaudia* sont simples et allongées, celles du sous-genre *Fabricia* sont très diversement lobées et dentées.

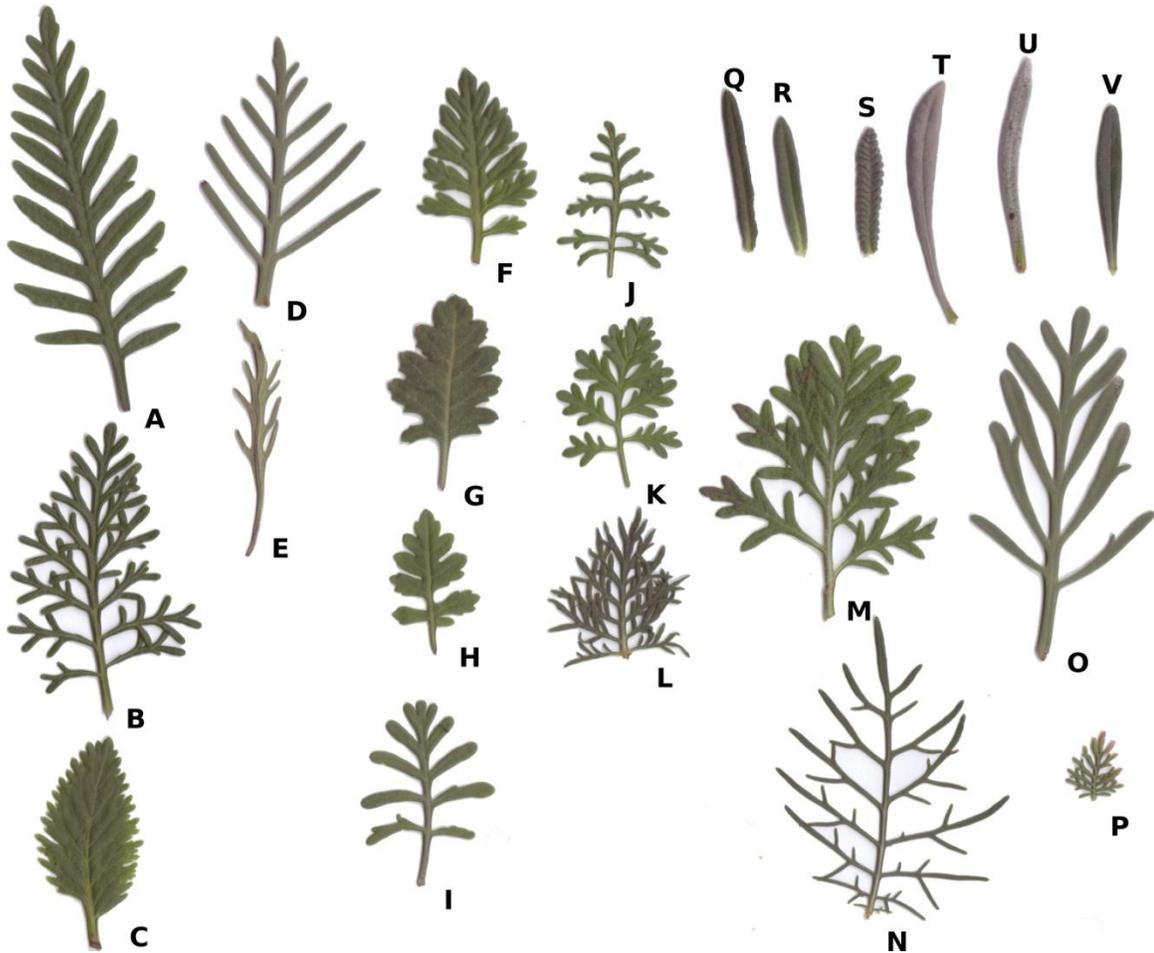


Figure 14 : Variation de la morphologie des feuilles des espèces du genre *Lavandula*.

A. *L. minutolii* B. *L. canariensis* subsp *canariae* C. *L. rotundifolia* D. *L. buchii* var *buchii* E. *L. buchii* var *tolpidifolia* F. *L. aristibracteata* G. *L. samhanensis* H. *L. dhofarensis* I *L. citriodora* J. *L. maroccana* K. *L. marei* L. *L. coronopifolia* M. *L. rejdalii* N. *L. bipinnata* O. *L. pinnata* P. *L. antinea* Q. *L. viridis* R. *L. pedunculata* S. *L. dentata* T. *L. latifolia* U. *L. lanata* V. *L. angustifolia*

Introduction

A l'opposé des feuilles, la structure de l'inflorescence constitue un caractère commun à l'ensemble des lavandes. Les fleurs de lavandes sont organisées en une inflorescence composée mixte ressemblant à un épi de cymes (mais dont les fleurs ont un pédicelle), appelé encore thyrsespiciforme. L'inflorescence principale ressemble donc à un épi plus ou moins lâche. L'inflorescence secondaire est une cyme. Celle-ci est bipare et scorpioïde dans le sous-genre *Lavandula* ou uniflore dans les sous-genres *Sabaudia* et *Fabricia* (Figure 15) (Lis-Balchin 2002; Upson 1997). Dans la majorité des espèces du genre, les bractées et les cymes sont opposées décussées, ce qui donne à l'épi composé une forme quadrangulaire. Cette forme de thyrsespiciforme est particulièrement bien visible dans la section *Pterostoechas* du sous-genre *Fabricia*. A l'opposé, dans le sous-genre *Lavandula* les cymes à plusieurs fleurs donnent un aspect cylindrique à ce thyrses (Figure 16). Dans les autres sections du sous-genre *Fabricia* ainsi que dans le sous-genre *Sabaudia* les bractées et les cymes sont organisées en spirale le long de l'axe floral (Figure 16) (Upson et Andrews 2004).

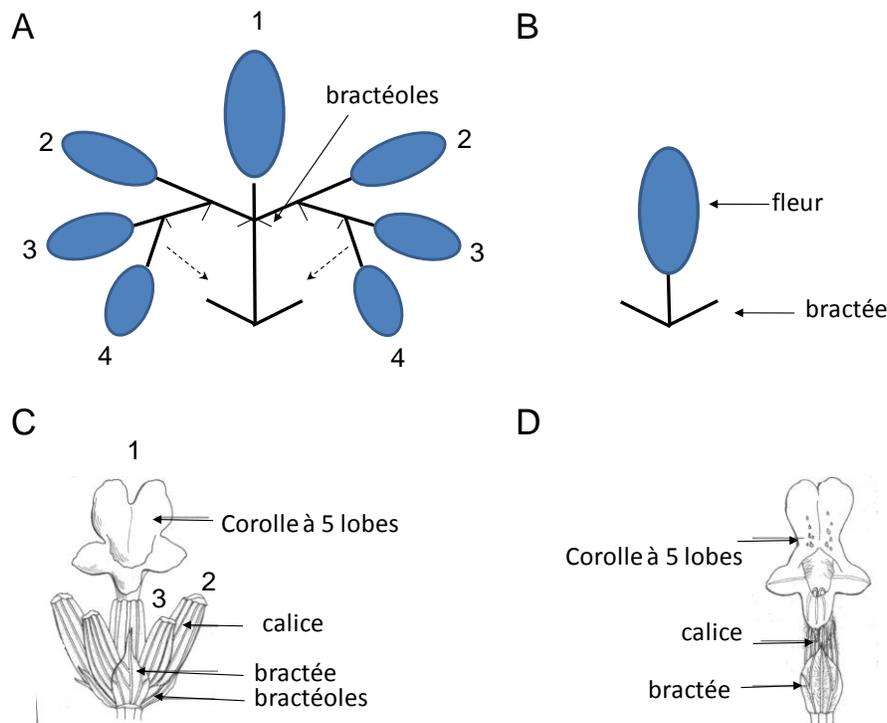


Figure 15 : Schémas des inflorescences du genre *Lavandula* (d'après Upson 1997).

A : une cyme à plusieurs fleurs (bipare et scorpioïde) avec l'ordre d'apparition des fleurs dans la cyme représenté par les numéros 1 à 4. B : une cyme à fleur unique. C : dessin d'une cyme à plusieurs fleurs. D : dessin d'une cyme à fleur unique. Schémas et dessins d'après (Upson et Andrews 2004)

Introduction

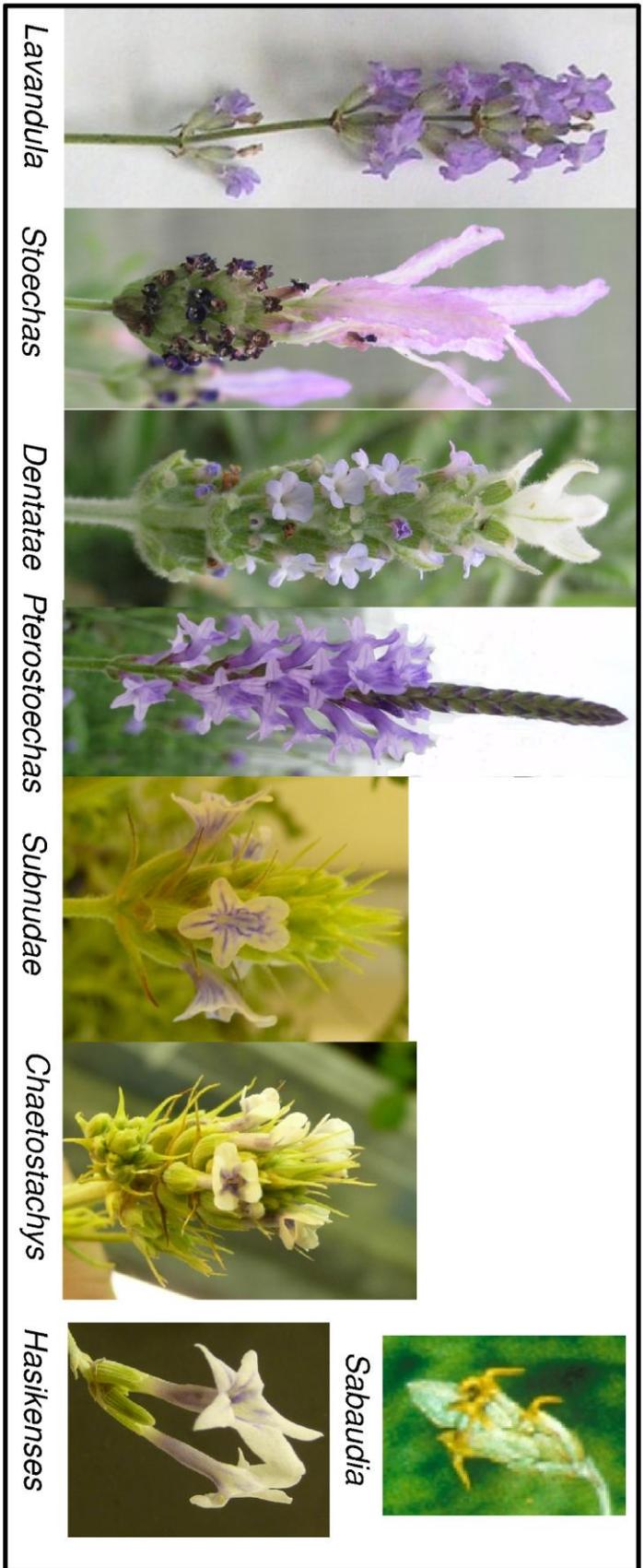


Figure 16 : Photographies de thyrses représentatives de 8 sections du genre *Lavandula*
La photo de la section *Sabaudia* est de Tim Upson (Upson 2009)

Introduction

Les bractées constituent aussi un caractère de détermination facile à observer et dont la forme et la taille est très variable chez les lavandes. Les bractées sont situées à la base de chaque cyme, et aussi parfois au sommet de l'inflorescence, comme dans les sections *Dentatae* et *Stoechas* (Figure 16). Ces dernières sont alors colorées et ont probablement un rôle attractif pour les pollinisateurs (Herrera 1997). Les bractées à la base de chaque cyme peuvent être ovales, triangulaires ou trapézoïdales, la pointe étant plus ou moins effilée. Dans les sections *Dentatae* et *Stoechas* les bractées sont trapézoïdales avec une pointe courte alors qu'elles sont triangulaires dans la section *Subnudae* avec une pointe très allongée. La véneration des bractées permet de reconnaître les espèces du sous-genre *Fabricia* qui ont des nervures parallèles alors que les espèces des deux autres sous-genres ont des nervures réticulées (Figure 17).

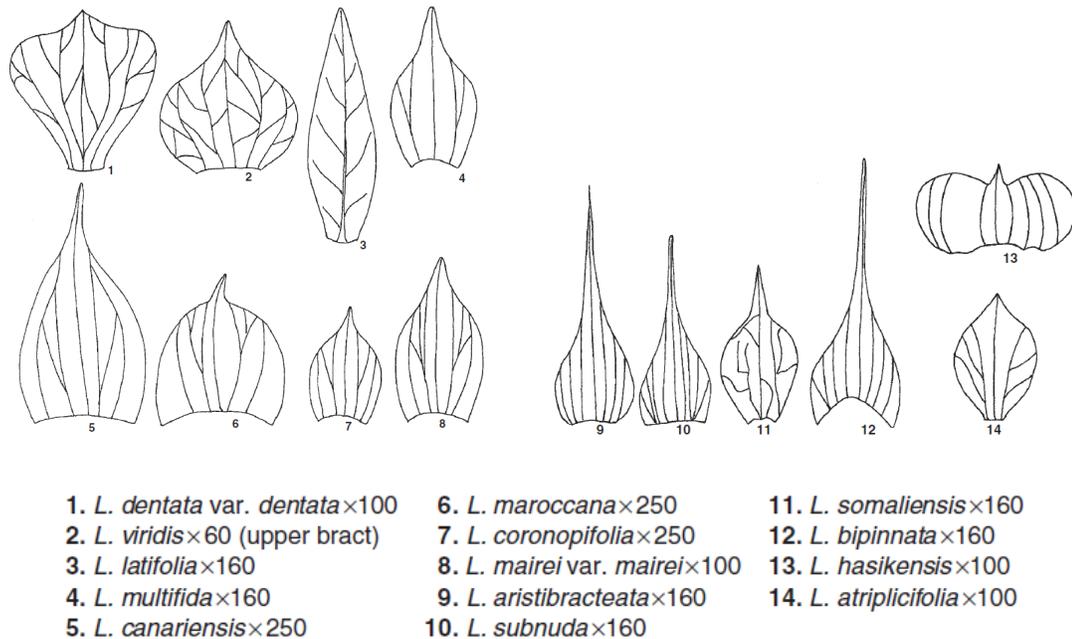


Figure 17 : Variations de la morphologie des bractées dans les espèces du genre *Lavandula* (modifié d'après Tim Upson dans Lis-Balchin 2002)

Introduction

Le calice des lavandes est tubulaire en raison de la fusion des sépales et s'ouvre à l'extrémité en deux lèvres constituées respectivement de 3 lobes dorsaux et deux lobes ventraux. Le sous-genre *Sabaudia* fait exception à cette organisation puisque dans ce groupe le calice est régulier sans zygomorphie bilabée apparente. Dans le sous-genre *Lavandula* le lobe médian de la lèvre supérieure forme une extension qui recouvre l'ouverture du calice avant l'épanouissement de la fleur ainsi qu'après la tombée de la corolle. Chez *L. angustifolia* le calice est le site majeur de production de terpènes et de nombreuses glandes (trichomes sécréteurs) peuvent être observées dans chacun des sillons délimités par deux nervures (Figure 18) (Couderc-Le-Vaillant, Segur-Fantino *et al.* 1990). Enfin, la présence de poils tecteurs (trichomes tecteurs) plus ou moins denses et longs constitue un caractère systématique interspécifique. Chez *L. angustifolia* ces trichomes tecteurs portent plusieurs branches (Figure 18). Une étude approfondie des trichomes montre qu'il existe chez *L. angustifolia* 5 types de trichomes tecteurs et 4 types de trichomes sécréteurs (dont 6 sont spécifiques de certains organes) et que cette diversité de trichomes pourrait être utilisée pour aider à l'identification des espèces, surtout au niveau infra spécifique et au repérage d'hybrides (Figure 19, Couderc-Le-Vaillant, Segur-Fantino *et al.* 1990)

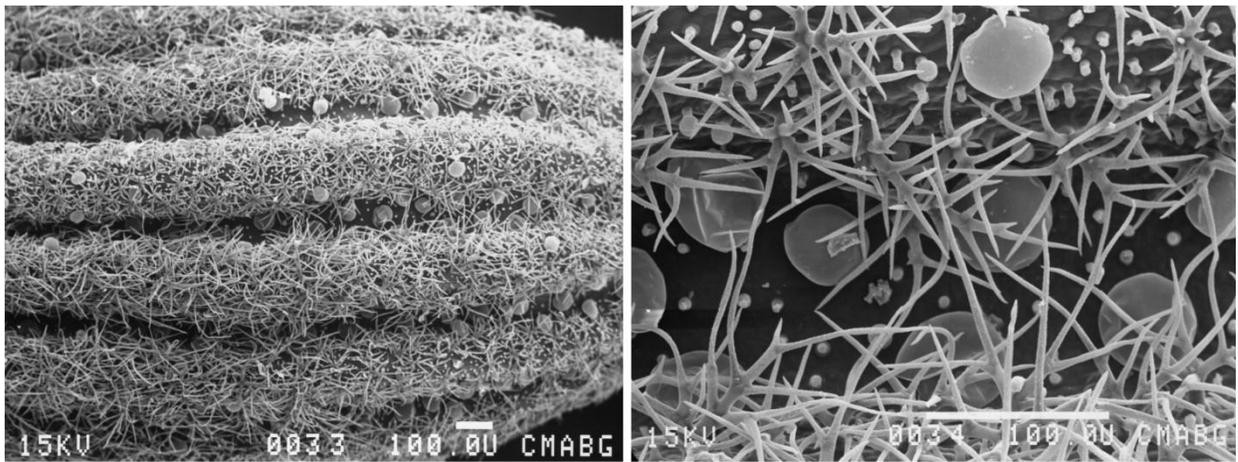
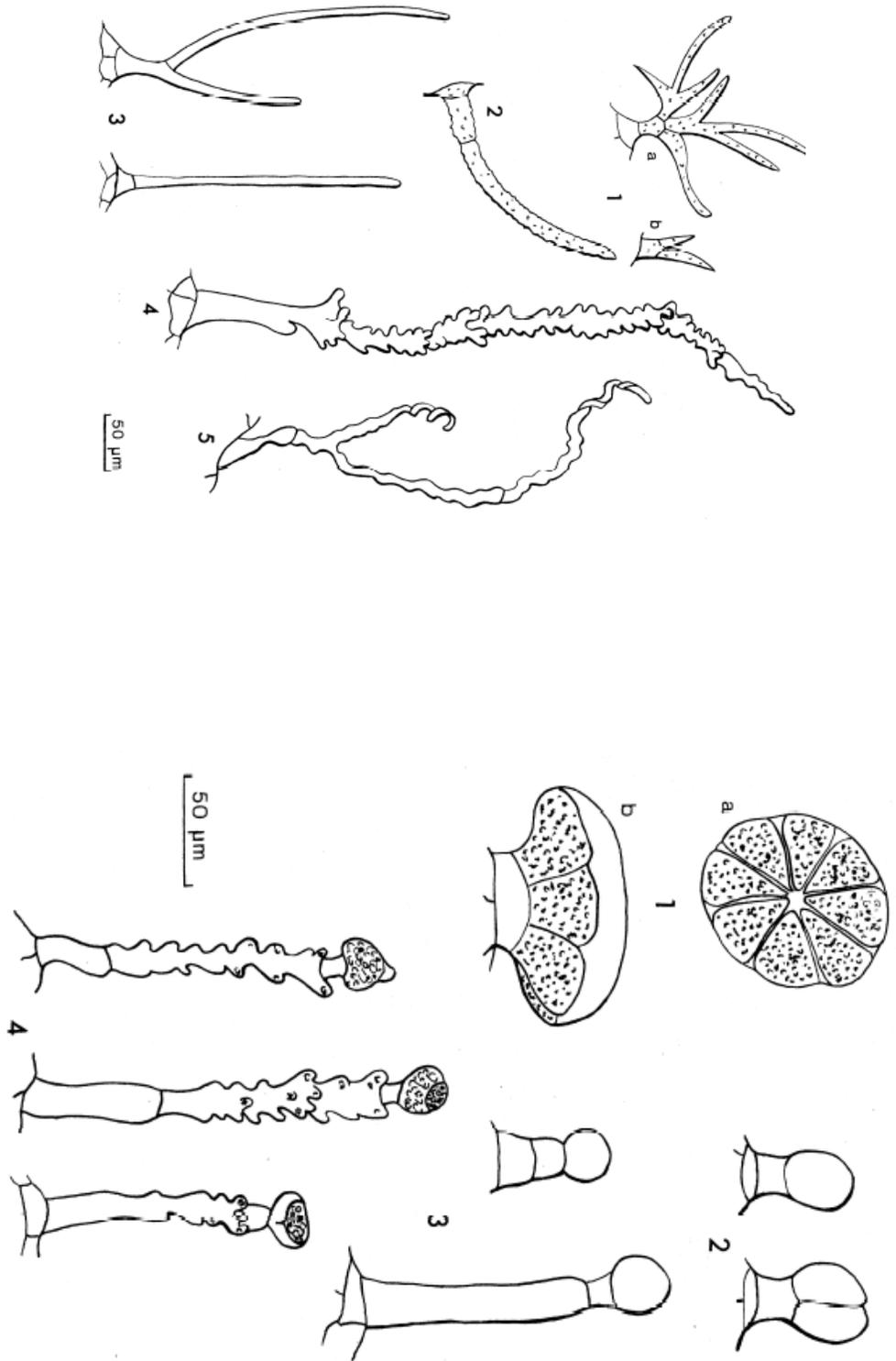


Figure 18 : Calice et trichomes de *L. angustifolia* (MEB x 100 15 kV)

A gauche : sillons sur un calice de *L. angustifolia* laissant voir les glandes sécrétrices (trichomes sécréteurs) et les poils (trichomes tecteurs). A droite : vue détaillée d'un sillon laissant voir les trichomes tecteurs et deux types de glandes sécrétrices (peltées et capitées). Les barres blanches représentent 100 μ m.

Figure 19 : Les trichomes de *L. angustifolia* (modifié d'après Condere-Le-Vaillant, Segur-Fantino *et al.*, 1990)

A : Cinq types de trichomes tecteurs 1a sur la corolle (face externe) 1b sur le style, 2 sur la zone d'insertion de l'anthere 3 sur le calice 4 sur la corolle face interne 5 sur la zone de déhiscence de l'anthere B : 1 glande peltée sur calice (a dessus b coupe transversale) 2 capitulé court sur feuille 3 capitulé long sur corolle face interne 4 sécréteurs bosselés sur la corolle à différents stades de la sécrétion



Introduction

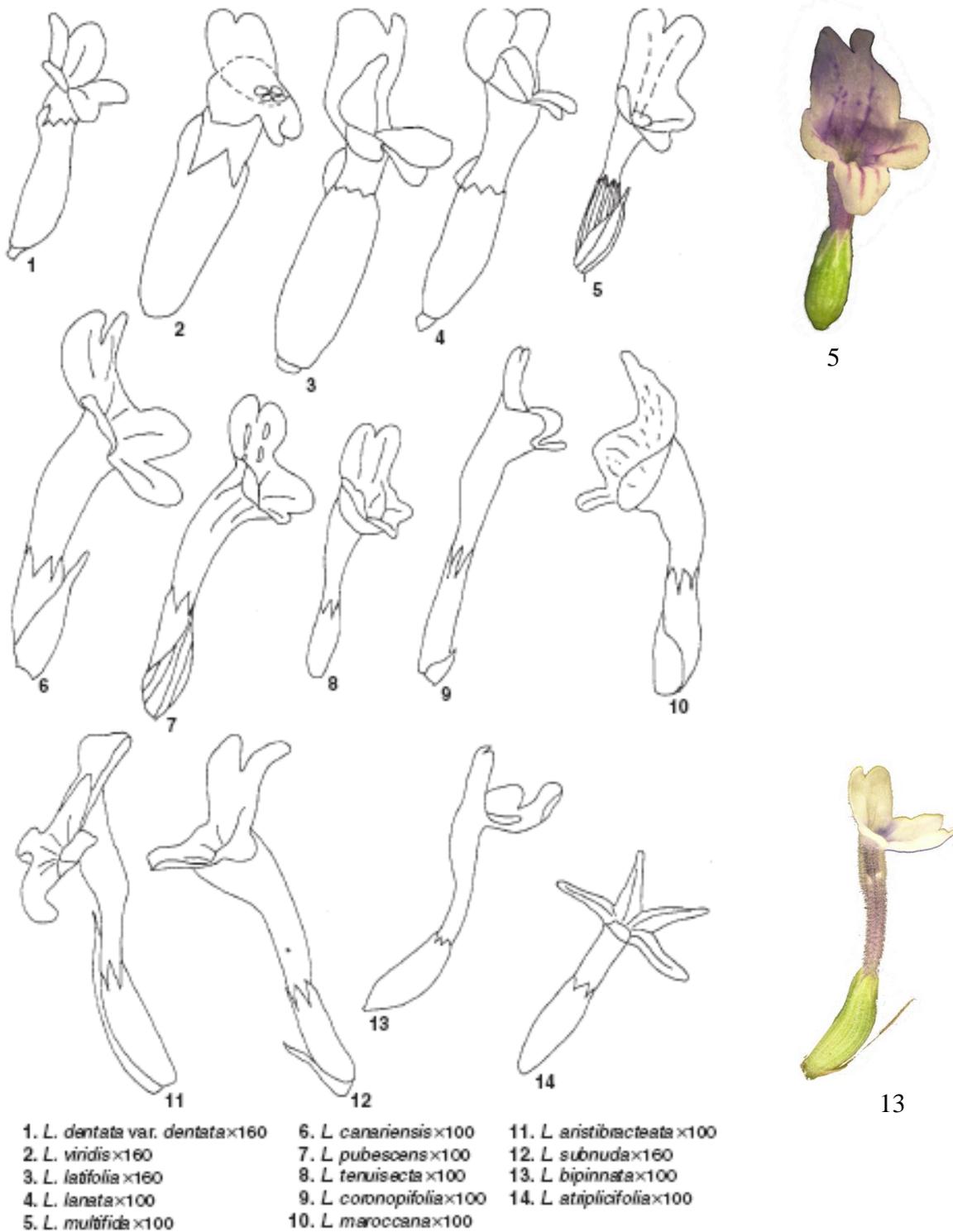


Figure 20 : Variations de la morphologie des fleurs des espèces du genre *Lavandula* (D'après Upson dans Lis-Balchin 2002)

Introduction



Figure 21 : Détails de la corolle de *L. stoechas* et *L. pedunculata*

La corolle des lavandes est constituée de 5 pétales soudés formant à la base un tube et à l'extrémité une structure bilabiée. Le sous-genre *Sabaudia* fait exception à cette règle. Les fleurs sont alors constituées de 5 pétales soudés en étoile. Dans les autres sous-genres, la lèvre supérieure est formée de 2 pétales soudés et la lèvre inférieure de 3 pétales formant des lobes égaux. L'épiderme de la corolle est constitué de cellules coniques et de glandes sécrétrices, celles-ci étant présentes uniquement sur la face inférieure. La partie du tube de la corolle qui est à l'extérieur du calice est plus ou moins longue et permet de distinguer la section *Stoechas* ou le tube est presque totalement inséré dans le calice, les sections *Lavandula* et *Dentatae* où la longueur du tube visible est égale à celle de la corolle et enfin toutes les sections de sous-genre *Fabricia* ou la partie visible du tube est supérieure à deux fois la longueur du calice (Figure 20). La forme de la corolle permet également de distinguer *L. stoechas* de *L. pedunculata*. *L. stoechas* possède des pétales de taille égale alors que ceux de *L. pedunculata* sont de tailles différentes (Figure 21). La plupart des lavandes ont une corolle de couleur qui va du bleu au violet avec parfois des fleurs roses (Figure 20, Figure 21). Certaines lavandes aux fleurs blanches pourraient être la conséquence de mutations. Enfin les fleurs des espèces du sous-genre *Sabaudia* sont jaunes-marrons (Figure 16).

Ainsi, il apparaît à la description de ces quelques caractères que le polymorphisme morphologique est important dans le genre *Lavandula*. Afin de permettre une meilleure reconnaissance des sous-genres une clé simple de détermination est proposée au travers des figures des pages 41, 43, 47 et 48. Ces variations anatomiques interspécifiques sont probablement le reflet d'une aire de répartition très importante avec des adaptations différentes à des milieux et des pollinisateurs variés.

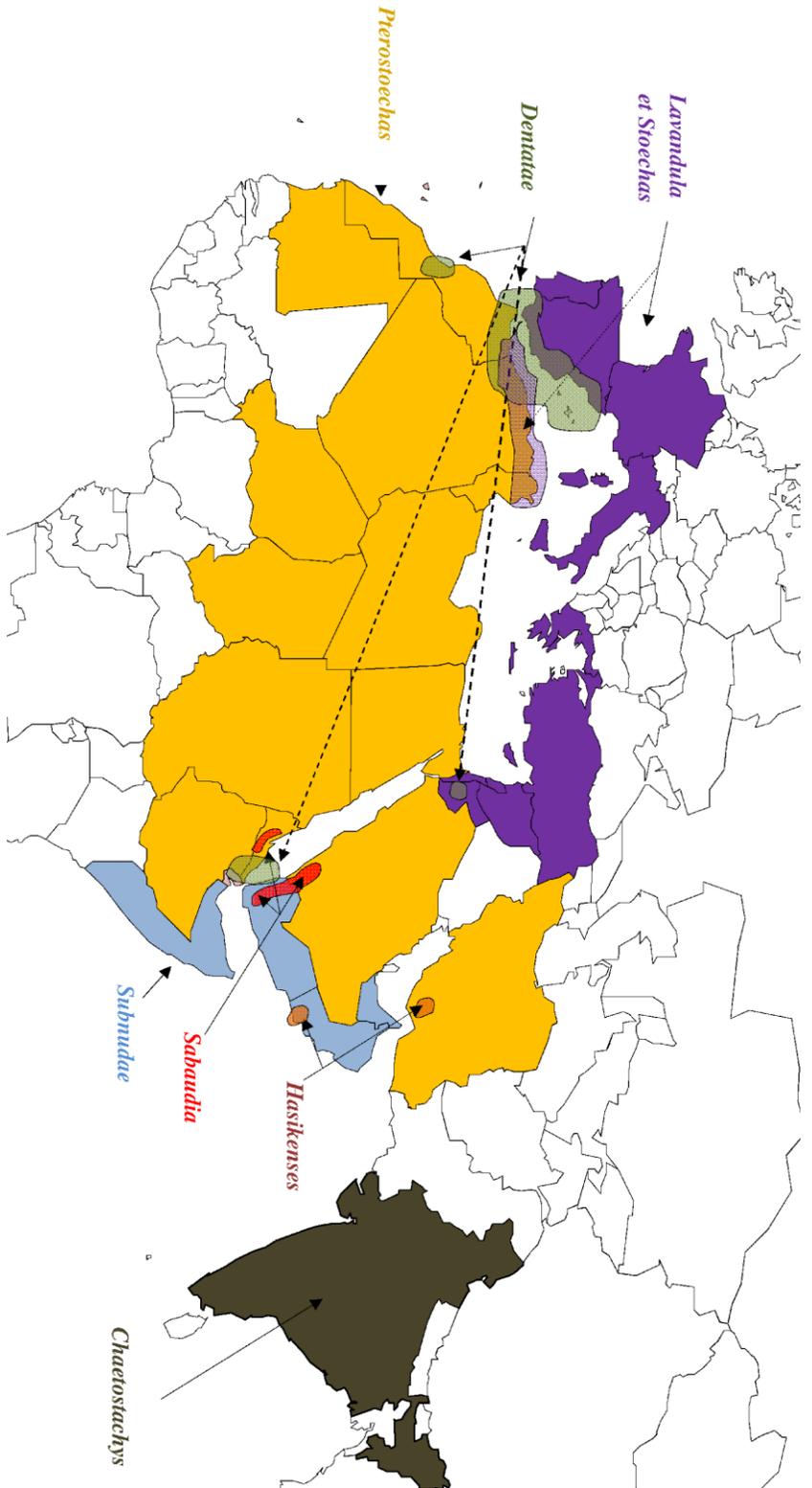


Figure 22: Schematisation de l'aire de répartition des sections du genre *Lavandula*
 Les pays sont de la couleur de la section majoritairement présente. Pour certaines sections dont la répartition est plus restreinte, une zone de couleur indique les principales zones de présence. **Jaune** : Section *Pterostochas*, **Violet** : Sections *Lavandula* et *Stoechas*, **Vert foncé** : Section *Chaetostachys*, **Bleu** : Section *Subnudae*.
Vert : Section *Dentatae*, **Rouge** : Section *Sabaudia* et **Orange** : Section *Hasikenses*.

II-3 Aire de répartition

Les espèces du genre *Lavandula* sont pour l'essentiel présentes dans les régions méditerranéennes avec, au nord de la Méditerranée, des espèces du Portugal jusqu'au Proche Orient et, au sud, des représentants du genre à travers les pays d'Afrique du nord jusqu'aux pays du Moyen Orient (des abords de la mer rouge à l'ouest de l'Iran, Figure 22). On trouve également des espèces de lavandes à l'ouest des côtes africaines sur les îles de la Macaronésie (*i.e.* îles du Cap Vert, des Canaries, de Madère, des Açores et Selvagem, Figure 23) exclusion faite des deux dernières sur lesquelles il ne semble pas y avoir de lavandes, ainsi qu'au nord-est de l'Afrique tropicale (Somalie, Ethiopie, Erythrée et Soudan). Enfin, il existe deux espèces de lavandes en Inde qui créent une disjonction du genre à l'est (Tableau 3, Figure 22).

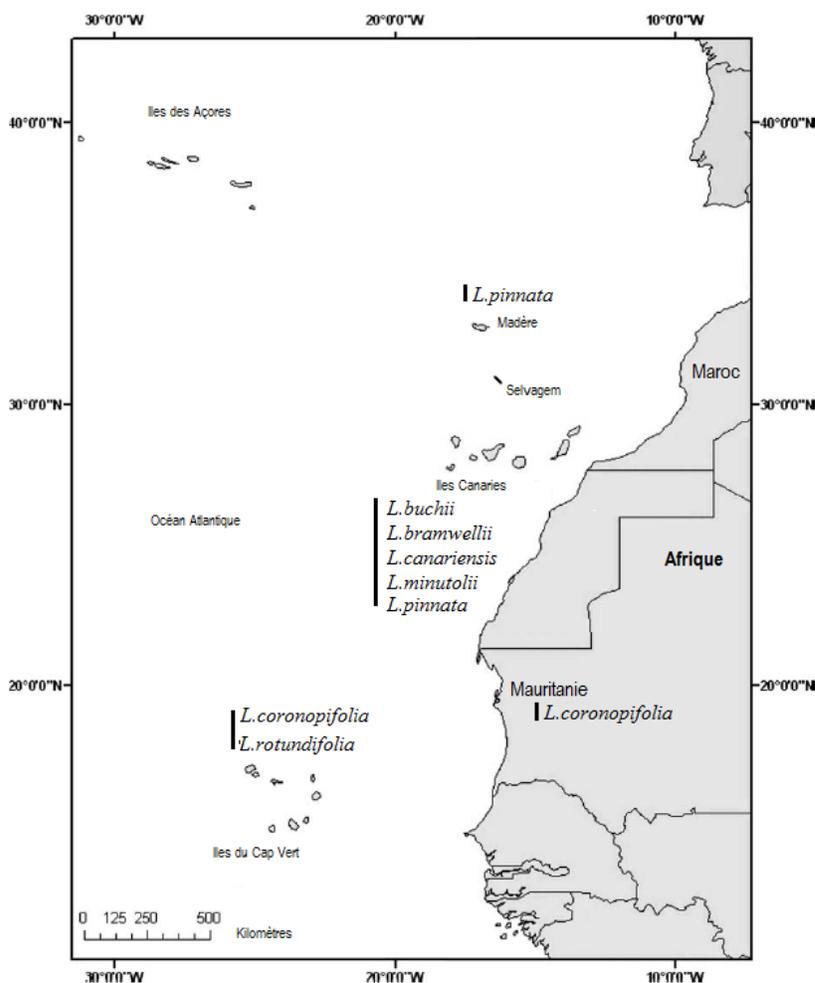


Figure 23 : Carte des îles de la Macaronésie et de la répartition des espèces de lavande (modifié d'après Allan, Francisco-Ortega *et al.* 2004)

Introduction

Tableau 3 : Liste des origines géographiques des espèces du genre *Lavandula*.

* espèces non étudiées lors de ces recherches.

		Origines géographiques
1	<i>L. angustifolia ssp angustifolia</i>	Espagne, France, Italie
	<i>L. angustifolia ssp pyrenaica</i>	Espagne, France
2	<i>L. antinea</i>	Algérie, Niger, Tchad, Soudan
3	<i>L. aristibracteata</i>	Somalie
4	<i>L. atriplicifolia</i>	Yemen, Arabie Saoudite
5	<i>L. bipinnata</i>	Inde
6	<i>L. bramwellii*</i>	Iles Canaries
7	<i>L. buchii</i>	Iles Canaries
8	<i>L. canariensis</i>	Iles Canaries
9	<i>L. citriodora</i>	Arabie Saoudite, Yemen
10	<i>L. coronopifolia</i>	Maroc, Mauritanie, Algérie, Niger, Tchad, Ethiopie, Yemen, Soudan, Egypte, Arabie Saoudite, Iran, Cap Vert
11	<i>L. dentata</i>	Maroc, Espagne, Algérie, Ethiopie, Yemen, Arabie Saoudite, Erythrée, Jordanie
12	<i>L. dhofararensis</i>	Oman
13	<i>L. erythraeae*</i>	Erythrée
14	<i>L. galgalloensis*</i>	Somalie
15	<i>L. gibsonii*</i>	Inde
16	<i>L. haskinensis</i>	Oman
17	<i>L. lanata</i>	Espagne
18	<i>L. latifolia</i>	Espagne, France, Italie
19	<i>L. macra</i>	Yemen, Oman, Somalie
20	<i>L. mairei</i>	Maroc
21	<i>L. maroccana</i>	Maroc
22	<i>L. minutolii</i>	Iles Canaries
23	<i>L. multifida</i>	Portugal, Espagne, Maroc, Algérie, Tunisie, Libye, Egypte, Soudan, Italie
24	<i>L. nimmoi</i>	Ile Socotra (Yemen)
25	<i>L. pedunculata</i>	Maroc, Espagne, Turquie, Portugal
26	<i>L. pinnata</i>	Iles Canaries, Madère
27	<i>L. pubescens</i>	Egypte, Jordanie, Israël, Erythrée, Yemen, Arabie Saoudite
28	<i>L. qishnensis*</i>	Yemen
29	<i>L. rejadalii</i>	Maroc
30	<i>L. rotundifolia</i>	Cap Vert
31	<i>L. saharica*</i>	Algérie
32	<i>L. samhanensis</i>	Oman
33	<i>L. setifera*</i>	Yemen, Somalie
34	<i>L. somaliensis*</i>	Somalie
35	<i>L. stoechas ssp luisieri</i>	Espagne, Portugal
	<i>L. stoechas ssp stoechas</i>	Maroc, Espagne, Algérie, Tunisie, Liban, Israël, Grèce, Sardaigne, Turquie, France, Portugal
36	<i>L. sublepidota*</i>	Iran
37	<i>L. subnuda</i>	Emirats Arabes unis, Oman
38	<i>L. tenuisecta</i>	Maroc
39	<i>L. viridis</i>	Portugal, Espagne

Cette aire de répartition ainsi que la position basale du genre *Lavandula* dans la tribu des *Ocimeae* laisse penser que les lavandes dérivent d'un ancêtre asiatique (Paton, Springate *et al.* 2004). On retrouve d'ailleurs des espèces de 5 des 8 sections du genre *Lavandula* dans la zone sud-ouest de l'Asie (Figure 22) parmi celles-là, les sections *Subnuda* et *Sabaudia* sont à la base de l'arbre de phylogénie moléculaire obtenu avec des séquences d'ADN nucléaire (*i.e.* ITS, Figure 12). Ces mêmes données moléculaires placent également la section *Chaetostachys* à la base du genre. Tout ceci permet de soutenir l'hypothèse d'un centre d'endémisme situé au sud de la mer Téthys au niveau des zones actuelles du nord-est de la Corne de l'Afrique, du sud de l'Arabie et de l'île de Socotra (Yémen) (Upson et Andrews 2004). Ces zones étaient

Introduction

liées au continent indien (Figure 24) jusqu'à la fin du jurassique (-160 Ma) (Birse, Bott *et al.* 1997). Ces observations pourraient, à première vue, laisser penser à une origine très ancienne du genre, trop ancienne, car les Angiospermes n'existaient pas encore.

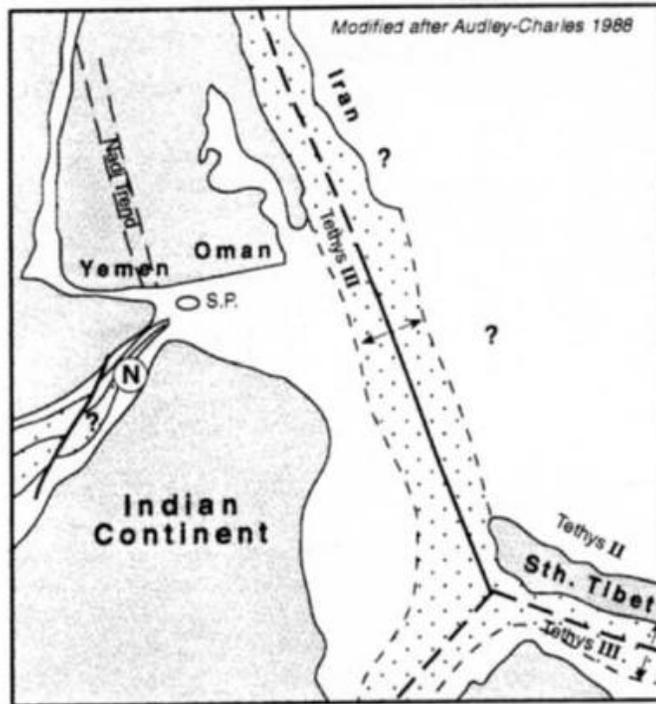


Figure 24 : Reconstruction de la zone d'origine probable du genre *Lavandula* à la fin du Jurassique (modifié d'après Birse, Bott *et al.* 1997)

Ⓝ Bassin Nord de la Somalie et S.P : plateforme de Socotra

De surcroît, les premiers fossiles de pollens hexacolpés, probablement un *Ocimum*, ont été trouvés au Rajasthan et sont datés du début de l'Eocène (-56 Ma) (Kar 1996 dans Harley, Atkins *et al.* 2004). D'après la chronologie de Paton *et al.* 2004 le genre *Lavandula* se serait diversifié après le genre *Ocimum*, ceci interdit une origine aussi ancienne que le Jurassique et alimente plutôt l'hypothèse d'une diversification plus récente, probablement lors du Miocène (- 23 Ma) (Harley, Atkins *et al.* 2004). Le climat était alors chaud et humide sur l'ensemble de l'aire de répartition actuelle du genre, avant une aridification de certaines zones dont le Sahara et l'Arabie (Costeur et Legendre 2008). La paléogéographie montre clairement les liens possibles entre l'ensemble des zones de l'aire de répartition actuelle du genre (Meulenkaamp et Sissingh 2003). Davis et Hedge (1971) remarquent que le genre *Lavandula* est un exemple de lien floristique entre le sud-ouest de l'Asie et le nord-ouest de l'Afrique au même titre que les genres *Helianthemum*, *Cuminum* et *Salvia*. Actuellement, le plus grand nombre d'espèces de lavandes (44%) est dans la zone ouest de l'aire de répartition et représente certainement une

Introduction

diversification plus récente. Le Maroc et les îles de la Macaronésie ne comptent pas moins de 11 des 16 espèces de *Pterostoechas* (Figure 25), faisant de cette zone un centre de diversification majeur (Upson et Jury 2002). Par ailleurs, la phylogénie moléculaire du genre *Lavandula* (Figure 12) montre une faible résolution des valeurs de bootstrap due à un faible nombre de caractères informatifs entre les espèces de cette section ce qui suggère une diversification récente (après la dernière ère glaciaire). Cette hypothèse est basée sur les conclusions de travaux analogues sur les *Reseda* qui ont une aire de répartition similaire (Martín-Bravo, Meimberg *et al.* 2007).

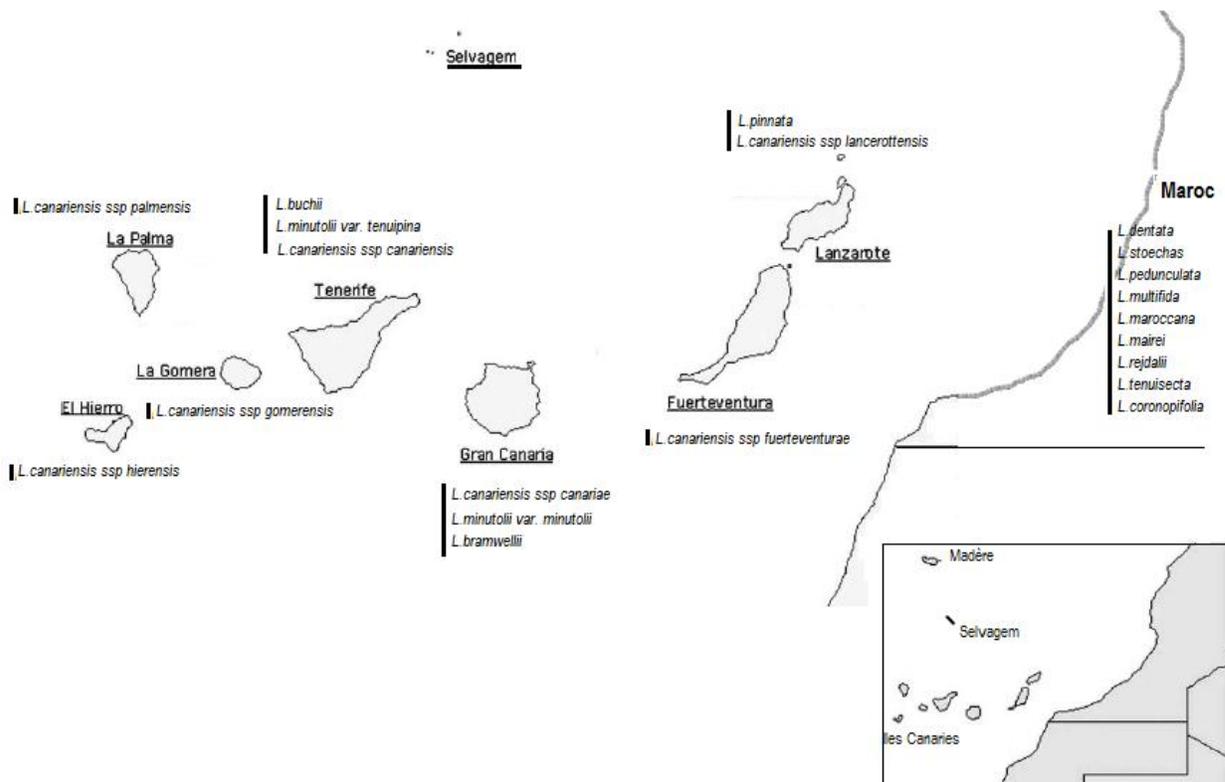


Figure 25: Carte détaillée de la répartition des espèces de lavandes sur les îles Canaries et au Maroc (modifié d'après Allan, Francisco-Ortega *et al.* 2004)

La section *Pterostoechas* fait clairement le lien entre l'est et l'ouest de l'aire de répartition (Figure 22). Cette section est présente dans la majeure partie des pays de la zone phytogéographique Saharo-Sindian (Figure 26), pour des cartes détaillées voir (Emberger, Gausson *et al.* 1963). Le lien entre phytogéographie et les sections est également visible pour les sections *Lavandula*, *Stoechas* qui sont inféodées à la zone Méditerranéenne et pour les espèces de *Sabaudia*, *Subnudae* et *Chaetostachys* qui sont, elles, présentes dans les régions de la zone Sudao-Deccanien (Figure 26).

Introduction

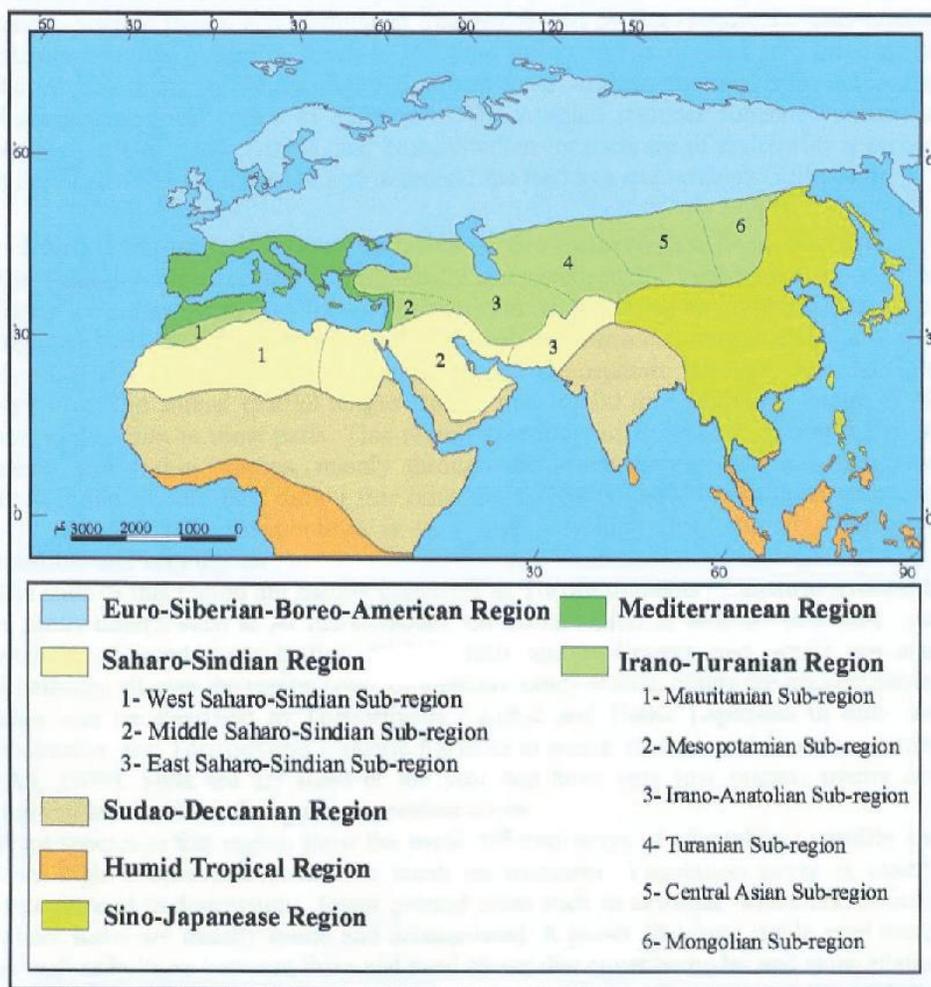


Figure 26 : Grandes régions phytogéographiques (modifié d'après Al-Nafie 2008)

Les deux seules exceptions à cette corrélation entre phytogéographie et aire de répartition sont *L. dentata* et *L. multifida* (Figure 22). Les deux doivent leur répartition aux changements climatiques survenus au cours de l'histoire. En effet, la disposition actuelle de *L. dentata* démontre qu'elle devait être présente sur toute la largeur de l'Afrique du nord et jusqu'au sud de la péninsule arabique. Suite au réchauffement et à la désertification du Sahara au cours de l'Holocène (Davis et Hedge 1971), cette espèce a dû trouver refuge dans les zones de montagne avec, à l'Est : les chaînes de montagne d'Ethiopie et de l'ouest de la péninsule arabique ou bien à l'Ouest : le Haut Atlas. Les zones de montagne étant alors des refuges au climat plus stable, ceci expliquerait que ces régions soient devenues des zones d'endémisme (Verlaque, Médail *et al.* 1997) de *L. dentata*. Pour *L. multifida*, l'explication de l'existence d'une disjonction en Egypte est également à rapprocher de la dessiccation du Sahara qui a suivi la dernière ère glaciaire (10 000 ans avant notre ère). Au nord de la mer méditerranée,

Introduction

ces changements climatiques ont également eut un impact sur les espèces du sous-genre *Lavandula* avec par exemple *L. angustifolia* qui est désormais présente au nord-est de l'Espagne, dans les montagnes des Pyrénées, dans la région des Causses, les Alpes et le sud de l'Italie, mais qui devait être présente sur toute cette zone lors des périodes plus clémentes de la dernière ère glaciaire avant de voir ses zones d'extension réduites par la remontée des températures. Là encore, les zones de montagne ont servi de refuge. La plupart des exemples présentés sont en lien avec les changements climatiques et/ou la tectonique des plaques, mais ceci n'explique pas la présence de lavandes sur les îles volcaniques de l'océan atlantique qui n'ont jamais été en relation avec le continent. Les lavandes ont donc colonisé ces îles via une dispersion à longue distance probablement par des pinsons ou d'autres espèces d'oiseaux granivores (Upson et Andrews 2004). Ceci est tout à fait envisageable, l'île de Fuerteventura n'étant éloignée du continent africain que de 100 km (Nicolas, Fernandez-Palacios *et al.* 1989). Pour les îles du Cap Vert et de Madère les distances sont plus importantes, 540 et 640 km respectivement (Nicolas, Fernandez-Palacios *et al.* 1989). De nombreuses espèces de plantes ont cependant colonisé ces îles, le transport des graines ayant pu être facilité par les Alizés ou les forts courants marins (Juan, Emerson *et al.* 2000). Pour les lavandes, les résultats de biologie moléculaire de Tim Upson (Upson 1997) suggèrent que chaque groupe d'île a été colonisé en une fois, avec un évènement pour la colonisation des îles Canaries suivi par celle de Madère. La colonisation de Madère étant probablement faite via Fuerteventura au regard de la présence de *L. pinnata* sur ces deux îles. Cependant, chez une autre *Lamiaceae* (*Micromeria*) la colonisation de Madère est supposée avoir été faite via l'île de Tenerife plus proche (Meimberg, Abele *et al.* 2006). La colonisation des îles du Cap Vert est, elle, probablement issue d'un autre évènement. La présence de *L. coronopifolia* sur le continent et sur les îles du Cap Vert, les correspondances morphologiques (*i.e.* pédoncule ramifié, forme du lobe du calice) et les hybridations fréquentes entre les deux espèces suggèrent que ces espèces sont liées même si elles vivent dans des zones climatiques différentes (Upson 1997). L'histoire évolutive des espèces végétales de Macaronésie est loin d'être simple et le champ des possibles est vaste (Allan, Francisco-Ortega *et al.* 2004; Freitas et Brehm 2001; Juan, Emerson *et al.* 2000; Meimberg, Abele *et al.* 2006). Les données actuelles de phylogénie ne permettent pas de conclure véritablement sur les routes suivies par les lavandes pour coloniser ces îles ni sur les détails de la radiation, mais elles offrent déjà de bonnes indications. Cet

Introduction

aperçu de la biogéographie du genre *Lavandula* demanderait à être complété par des études au niveau populationnel couvrant l'ensemble de l'aire de répartition.

En conclusion, les espèces du genre *Lavandula* sont différenciables sans ambiguïté et leurs relations phylogénétiques sont bien établies. Elles sont réparties sur des biotopes variés et produisent des COV et certaines dans de si grandes quantités qu'elles ont été mises en culture par l'homme. Cette phylogénie robuste si elle est associée à des données sur les COV produits par un nombre suffisamment grand d'espèces du genre (Levin, McDade *et al.* 2003), fait du genre *Lavandula* un candidat idéal pour la réalisation d'études sur l'origine évolutive de la diversité de ces COV au sein d'un genre. L'étude de la diversité des CTV du genre *Lavandula* constituerait le premier travail sur la diversité des composés terpéniques à l'intérieur d'un genre. La superposition des CTV produits par des espèces proches sur un arbre phylogénétique robuste associée aux connaissances sur les voies de synthèses de ces CTV et sur les gènes des terpènes synthases associées à leur production permettraient, dans le futur, d'appréhender les relations évolutives entre ces différents CTV ainsi que les relations évolutives entre les TPS responsables de leur production.

II-4 Diversité des COV dans le genre *Lavandula*

II-4.1 Variation interspécifique de la composition en COV

Parmi les 39 espèces du genre *Lavandula*, 12 espèces seulement ont fait l'objet d'analyses de volatils (Tableau 4). Les sept espèces du sous-genre *Lavandula* ont été particulièrement étudiées et ce, certainement du fait de l'utilisation commerciale de leurs huiles. A l'opposé, les espèces du sous-genre *Fabricia* et *Sabaudia* ont été jusqu'alors peu, ou pas étudiées. Ainsi, seules les compositions des HE de *L. multifida*, *L. canariensis*, *L. pinnata*, *L. maroccana* et *L. bipinnata* ont été rapportées dans la littérature (Tableau 4). Les données ethnobotaniques nous apprennent par contre que certaines de ces espèces dont la composition en COV est inconnues sont utilisés depuis des siècles comme remèdes (El-Hilaly, Hmammouchi *et al.* 2003; Upson et Andrews 2004). Il ressort de ces études que la majorité des COV formés par les différentes espèces du genre *Lavandula* sont des terpènes. Ainsi plus de 200 monoterpènes et sesquiterpènes différents ont pu être caractérisés dans l'ensemble des espèces analysées. Par ailleurs, certaines espèces telles que *L. stoechas* synthétisent plus de 100 terpènes différents alors que d'autres espèces telle que *L. canariensis* expriment une diversité terpénique beaucoup plus faible (Tableau 4).

Introduction

Ainsi, une solide étude de la diversité des CTV dans le genre *Lavandula* nécessite une analyse des CTV produits par un grand nombre d'espèces de ce genre.

Introduction

Tableau 4 : Les huiles essentielles des espèces du genre *Lavandula* dans la littérature

Sous-genre	Section	Espèce	Références bibliographiques	
Lavandula	Lavandula	<i>L. angustifolia</i> Mill.	(An, Haig et al. 2001; Fakhari, Salehi et al. 2005; Garcia-Vallejo 1992; Kim et Lee 2002; Shellie, Marriott et al. 2000; Shellie, Mondello et al. 2002; Tsuro, Inoue et al. 2001)	
		<i>L. lanata</i> Medik.	(Garcia-Vallejo 1992; Wiesenfeld 1997; Yassaa et Williams 2005)	
		<i>L. latifolia</i> Boiss.	(Boelens 1986; De Pascual-Teresa et Machin 1983; Gamez, Zarzuelo et al. 1988; Munoz-Bertomeu, Arrillaga et al. 2007; Salido, Altarejos et al. 2004)	
	Dentatae	<i>L. dentata</i> L.	(Bousmaha, Boti et al. 2006; Dob, Dahmane et al. 2005; Gamez, Jimenez et al. 1990; Gamez, Zarzuelo et al. 1988; Sudriá, Piñol et al. 1999; Wiesenfeld 1997)	
	Stoechas	<i>L. pedunculata</i> (Mill) Cav.	(Bellakhdar 1997; Garcia-Vallejo 1992)	
		<i>L. stoechas</i> L.	(Angioni, Barra et al. 2006; Baldovini, Muselli et al. 1998; Dob, Dahmane et al. 2006; Gamez, Jimenez et al. 1987; Garcia-Vallejo 1992; Garcia-Vallejo, Garcia-Vallejo et al. 1994; Kokkalou 1988; Lavoine-Hanneguelle et Casabianca 2004; Ristorcelli, Tomi et al. 1998; Skoula, Abidi et al. 1996; Topcu, Ayrál et al. 2001; Ulubelen, Goren et al. 1988; Valentini, Arnold et al. 1993; Wiesenfeld 1997; Zrira et Benjilali 2003)	
		<i>L. viridis</i> L'Hér.	(Garcia-Vallejo 1992; Nogueira et Romano 2002; Wiesenfeld 1997)	
	Fabricia	Pterostoechas	<i>L. multifida</i> L.	(Bellakhdar 1997; Gamez, Jimenez et al. 1987; Garcia-Vallejo 1992; Politi, De Tommasi et al. 2002; Sosa, Altinier et al. 2005; Wiesenfeld 1997)
			<i>L. canariensis</i> Mill.	(Pala-Paul, Brophy et al. 2004)
			<i>L. minutolii</i> Bolle.	Non déterminée
<i>L. bramwellii</i> Upson & S.Andrews			Non déterminée	
<i>L. pinnata</i> L.f.			(Figueiredo, Barroso et al. 1995; Wiesenfeld 1997)	
<i>L. buchii</i> Webb & Berthel.			Non déterminée	
<i>L. rotundifolia</i> Benth.			Non déterminée	
<i>L. maroccana</i> Murb.			(Bellakhdar 1997)	
<i>L. tenuisecta</i> Coss. Ex Ball			Non déterminée	
<i>L. rejadalii</i> Upson & Jury			Non déterminée	
<i>L. mairei</i> Humbert			Non déterminée	
<i>L. coronopifolia</i> Poir.			Non déterminée	
<i>L. saharica</i> Upson & Jury			Non déterminée	
<i>L. antinea</i> Maire			Non déterminée	
<i>L. pubescens</i> Decne.			Non déterminée	
<i>L. citriodora</i> A.G. Mill.	Non déterminée			

Introduction

Suite du Tableau 4

Sous-genre	Section	Espèce	Bibliographie
Fabricia	Chaetostachys	<i>L. bipinnata</i> (Roth) Kuntz	(Hanamanthagouda, Kakkalameli et al.)
		<i>L. gibsonii</i> J. Graham	Non déterminée
	Hasikenses	<i>L. haskinensis</i> A.G. Mill.	Non déterminée
		<i>L. sublepidota</i> Rech.f.	Non déterminée
	Subnudae	<i>L. subnuda</i> Benth.	Non déterminée
		<i>L. macra</i> Baker	Non déterminée
		<i>L. dhofarensis</i> A.G. Mill.	Non déterminée
		<i>L. samhanensis</i> Upson & S.Andrews	Non déterminée
		<i>L. setifera</i> T. Anderson	Non déterminée
		<i>L. qishnensis</i> Upson & S.Andrews	Non déterminée
		<i>L. nimmoi</i> Benth.	Non déterminée
		<i>L. galgalloensis</i> A.G. Mill.	Non déterminée
		<i>L. aristibracteata</i> A.G. Mill.	Non déterminée
<i>L. somaliensis</i> Chaytor	Non déterminée		
Sabaudia	Sabaudia	<i>L. atriplicifolia</i> Benth.	Non déterminée
		<i>L. erythraeae</i> (Chiov.) Cufod	Non déterminée

II-4.2 Variations intraspecificque de la composition en COV

Les études de la diversité des terpènes formés par les espèces du genre *Lavandula* montrent qu'il existe au sein d'une même espèce des populations exprimant des chémotypes différents. Ainsi, des variations intraspécifiques de la composition terpénique ont été rapportées chez les espèces de lavandes d'Espagne (Tableau 5) (Garcia-Vallejo 1992), sur *L. stoechas* ssp. *stoechas* de Crète (Skoula, Abidi *et al.* 1996), sur *L. dentata* d'Algérie (Bousmaha, Boti *et al.* 2006) ou bien encore sur *L. multifida* du Maroc (Bellakhdar, Berrada *et al.* 1985). Ces variations correspondent souvent à la synthèse de composés différents ou à des proportions différentes de certains composés. Ces variations sont généralement soit d'ordre

Introduction

individuelle (variations spatiales et temporelles en page 17) soit d'ordre populationnel. Par exemple, chez *L. stoechas* ssp. *stoechas* de Crète on trouve des populations avec comme premier constituant de l'HE le fenchone et en second le camphre (chénotype fenchone/camphre) et d'autres populations avec en premier le 1,8-cinéole et en second le fenchone (chénotype 1,8-cinéole/fenchone) (Skoula, Abidi *et al.* 1996). Ce dernier chénotype est très particulier car on ne le retrouve dans aucune des autres études faites sur *L. stoechas* ssp. *stoechas* que ce soit en Espagne (Garcia-Vallejo 1992), en Grèce (Kokkalou 1988), à Chypre ou bien même au Maroc (Valentini, Arnold *et al.* 1993). Les chénotypes des espèces du genre *Lavandula* présentes en Espagne ont été étudiés finement dans les travaux de thèse de Maria Garcia-Vallejo (Garcia-Vallejo 1992). La chimie de ces plantes permet de discriminer les espèces et même les sous-espèces, l'exemple le plus marquant étant celui de *L. stoechas* ssp. *luisieri* qui a comme constituants majeurs de son HE trois monoterpènes irréguliers uniques dans le règne végétal : le cis et le *trans-alpha*-nécrodol et le *trans-alpha*-nécrodol acétate (Baldovini, Lavoine-Hanneguelle *et al.* 2005; Garcia-Vallejo, Garcia-Vallejo *et al.* 1994; Lavoine-Hanneguelle et Casabianca 2004; Sanz, Soria *et al.* 2004). Ces composés n'ayant jusqu'alors été détectés que chez certains insectes (Millar, Midland *et al.* 2005). Sur la base des chénotypes rencontrés Maria Garcia-Vallejo parvient même à faire une classification chénotaxonomique des espèces et de certaines sous-espèces du sous-genre *Lavandula* (Tableau 5). Des chénotaxonomies basées sur les terpènes existent chez d'autres Lamiacées : *Thymus* (Granger et Passet 1973), *Ocimum* (Grayer, Kite *et al.* 1996).

Tableau 5 : Résumé de la classification chénotaxonomique des espèces espagnoles de *Lavandula* (modifié d'après Garcia-Vallejo 1992)

Sous genre	Section	Espèce	Sous espèce	Variété	Composants majeurs
<i>Lavandula</i>	<i>Lavandula</i>	<i>L. angustifolia</i>	<i>angustifolia</i>		linalyl acetate>linalool
		<i>L. angustifolia</i>	<i>pyrenaica</i>		linalool>borneol>camphre
		<i>L. lanata</i>			camphre>lavandulol>linalool
		<i>L. latifolia</i>			1.8 cineol>camphre>linalool
	<i>Stoechas</i>	<i>L. pedunculata</i>	<i>pedunculata</i>		fenchone>1.8 cineol>camphre
		<i>L. pedunculata</i>	<i>sampaioana</i>		camphre>fenchone
		<i>L. pedunculata</i>		<i>lusitanica</i>	fenchone>camphre>1.8 cineol
		<i>L. stoechas</i>			fenchone>camphre>1.8 cineol
		<i>L. stoechas</i>	<i>luisieri</i>		trans-a-necrodol, trans-a-necrodyl acetate>1.8 cineol
		<i>L. viridis</i>			1.8 cineol>camphre>alpha-pinene
	<i>Dentatae</i>	<i>L. dentata</i>			1.8 cineol>b-pinene>trans-pinocarveol
<i>Fabricia</i>	<i>Pterostoechas</i>	<i>L. multifida</i>			carvacrol>b-bisabolol

Introduction

Chez *L. latifolia*, des chémotypes différents ont été observés entre des populations vivant dans des zones bioclimatiques différentes (Munoz-Bertomeu, Arrillaga *et al.* 2007). Des conditions environnementales particulières pourraient donc induire une pression de sélection différentielle sur la production de certains terpènes dans des populations de zones bioclimatiques différentes. Une étude sur des populations d'*Origanum microphyllum* (Gotsiou, Naxakis *et al.* 2002) montre que des variations chemotypiques observées dans la nature sont conservées lorsque les individus issus des différentes populations sont transplantés et cultivés sur un même site. Il semble donc que les variations chemotypiques sont soutenues par une base génétique. Chez le basilic, Iijima et ses collaborateurs ont démontré que de petites variations dans la séquence de gènes responsables de la synthèse des principaux terpènes peuvent être responsables du changement de chémotypes (Iijima, Davidovich-Rikanati *et al.* 2004; Iijima, Gang *et al.* 2004). Ainsi une mutation d'un seul nucléotide dans la linalol synthase d'un des cultivars est responsable de la formation d'une protéine non-fonctionnelle et donc de l'absence de linalol dans ce cultivar. Cependant, les mécanismes moléculaires responsables de la variabilité chemotypique chez les espèces aromatiques ne saurait être limités à des mutations ponctuelles sur quelques gènes. Chez les lavandes, aucune étude ne met en parallèle des variations génétiques avec des variations chémotypiques. Ainsi, bien que des études sur la variabilité génétique de populations de *L. stoechas*, de *L. multifida* (étude de 7 iso-enzymes Chograni, Messaoud *et al.* 2008), de *L. angustifolia* (étude de caractères morphologiques Segur-Fantino, Bahuaud *et al.* 1988) ou de *L. latifolia* (par AFLP Herrera et Bazaga 2007) existent, elles n'ont pas été couplées avec des données sur les HE.

II-4.3 Les gènes de lavandes impliqués dans la synthèse des COV

Seulement 4 TPS ont été étudiées chez les lavandes, 3 pour *L. angustifolia*, une limonène synthase (LaLIMs), une linalol synthase (LaLINs) et une bergamotène synthase (LaBERs) (accessions genbank ABB73044, ABB73045 et ABB73046 respectivement Landmann, Fink *et al.* 2007) et une linalol synthase (LIS) chez *L. latifolia* (accession genbank ABD77417.1 non publiée). Ce sont des TPS de la sous-famille TPS-b. LaLIMs et LaBERs sont capables de synthétiser plusieurs terpènes alors que LaLINs ne fait que du linalol. LaLINs et LaLIMs n'acceptent que le GPP comme substrat alors que LaBERs accepte le GPP et le FPP. On retrouve entre LaLIMs et LaBERs un lien évolutif possible par la simple perte de l'adressage cellulaire de la partie N-terminale (Landmann, Fink *et al.* 2007). Ces trois

Introduction

synthases n'expliquent qu'une petite partie de la diversité des terpènes formés par *L. angustifolia* et, par extension, dans les espèces du genre.

Globalement peu de données moléculaires existent pour les lavandes. Fin 2009, 17 séquences protéiques et 27 séquences nucléotidiques étaient disponibles sur GenBank. On retrouve les 4 TPS et leurs séquences nucléotidiques, 2 alcool acyltransférases potentielles et des fragments de gènes (rps16, rbcL, ITS, TrnL-TrnF) issus de travaux de phylogénie sur les *Lamiaceae* (Kaufmann et Wink 1994; Paton, Springate *et al.* 2004) (Trusty, Olmstead *et al.* 2004). Malgré ce manque de données moléculaires, un groupe espagnol travaille depuis de nombreuses années sur la production de terpènes par *L. Latifolia* en utilisant des gènes connus chez *Arabidopsis* ou *Mentha* (Munoz-Bertomeu, Arrillaga *et al.* 2006; Munoz-Bertomeu, Arrillaga *et al.* 2007; Muñoz-Bertomeu, Ros *et al.* 2008; Muñoz-Bertomeu, Sales *et al.* 2007). Ces travaux ont démontré que la surexpression de l'enzyme DXS de la voie du MEP ou de l'enzyme HMGR de la voie du MVA conduisait à l'augmentation de la production de terpènes volatils par *L. latifolia*. Par la surexpression de la limonène synthase de menthe ce groupe a démontré que la production de cette TPS chez *L. latifolia* était, comme chez d'autres plantes comme la menthe, régulée au cours du développement de la feuille (régulations transcriptionnelles et post-transcriptionnelles) (Muñoz-Bertomeu, Ros *et al.* 2008). Ces études permettent de considérer les lavandes comme un modèle pour l'étude de la régulation de la production de terpènes par les plantes. Fin 2009, un groupe canadien a publié une banque de 14 213 séquences obtenues sur des feuilles et des fleurs de *L. angustifolia* (Lane, Boecklemann *et al.*). Cette banque contient, entre autre, des gènes orthologues des gènes impliqués dans les voies du MEP et du MVA, ainsi que 25 séquences de gènes de terpènes synthases potentielles (Lane, Boecklemann *et al.*). Cependant, cette banque n'est pas accessible via GenBank. La banque de séquence qu'ils ont obtenue sera, dès qu'elle sera publiée sur les bases de données, un outil précieux pour la caractérisation de terpènes synthases chez la lavande. Une autre approche pour étudier les TPS est de caractériser la diversité terpénique dans le genre *Lavandula* afin de trouver des espèces qui expriment des terpènes particuliers. Une étude de la diversité terpénique nécessite donc l'analyse chimique d'un grand nombre d'échantillons d'espèces différentes. Pour cette étude nous avons développé des outils de métabolomique et de *data mining* qui méritent quelques définitions.

III L'acquisition et le traitement de données de métabolomique: un problème de taille

III-1 La métabolomique

La métabolomique est une méthode d'exploration très générale des organismes vivants qui vise à décrire de façon exhaustive les différents métabolites primaires et secondaires (i.e. métabolome Oliver, Winson *et al.* 1998) présents aux différentes échelles du vivant (de la cellule à l'organisme). Utilisée dans différents contextes d'études physiologiques, la métabolomique permet de comprendre l'impact des perturbations étudiées sur un système biologique (Sumner, Mendes *et al.* 2003). Elle fait partie, avec la génomique (i.e. la recherche de gènes), la transcriptomique (i.e. l'étude de l'expression des gènes), la protéomique (i.e. l'étude de l'expression des protéines au niveau traductionnel et post-traductionnel) du champ des « omiques » qui vise à comprendre un système biologique en faisant le lien entre le génotype et le phénotype (Fiehn 2002; Goodacre, Vaidyanathan *et al.* 2004) (Figure 27 (d'après Dettmer, Aronov *et al.* 2007)).

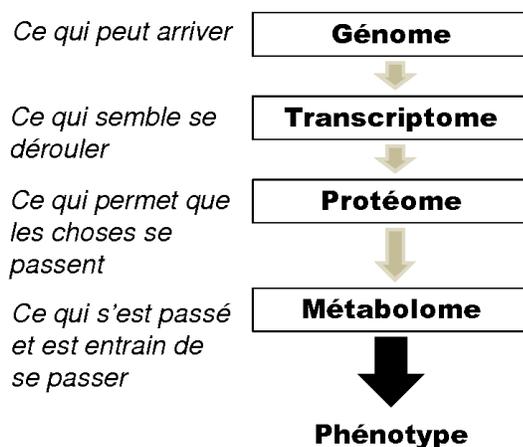


Figure 27 : La cascade des "omiques" (modifiée d'après Dettmer, Aronov *et al.* 2007)

Certains voient dans la métabolomique la technique permettant d'avoir la vue la plus « fonctionnelle » possible parmi les « omiques », la transcriptomique et la protéomique étant parfois incapables de résoudre la fonction d'un gène (Bino, Hall *et al.* 2004; Messerli, Partovi Nia *et al.* 2007; Sumner, Mendes *et al.* 2003). Par exemple, une hausse des ARN messagers ne corrèle pas toujours avec une hausse de la quantité de protéine; une protéine peut de plus être synthétisée mais être non-active. Ainsi, des variations dans le transcriptome ou le protéome ne correspondent pas toujours aux altérations du profil métabolique. On peut ajouter à cela que de

Introduction

nombreuses fonctions de gènes ou de protéines sont des fonctions supposées, respectivement basées sur des similarités de séquences nucléotidiques ou peptidiques (Bino, Hall *et al.* 2004; Sumner, Mendes *et al.* 2003). Il ne faut, cependant, pas oublier qu'un même métabolite peut avoir différentes origines et agir à différents niveaux sur de nombreux processus biologiques, ce qui complexifie l'identification des fonctions des gènes par la métabolomique (Schwab 2003). La métabolomique se veut être une technique sans biais permettant d'identifier et de quantifier l'ensemble des métabolites d'un organisme (Van Dam et Poppy 2008). Mais ceci reste, à ce jour, un objectif non-atteint. En effet, même si le développement rapide des techniques de chimie analytique et la hausse des performances de stockage et de calculs des ordinateurs permettent une hausse constante des capacités d'analyse de métabolites, il n'en reste pas moins que l'analyse complète et simultanée d'un métabolome n'a pas encore pu être achevée (Sumner, Mendes *et al.* 2003). Les progrès dans le domaine de la métabolomique sont également dépendants des avancées en biostatistique et en bioinformatique qui permettront d'analyser rapidement des matrices de données multidimensionnelles de plus en plus complexes (Smilde, Westerhuis *et al.*).

Deux approches sont généralement associées avec la métabolomique : l'empreinte métabolique (*metabolic fingerprinting*) et le profil métabolique (*metabolic profiling*) (Dettmer, Aronov *et al.* 2007; Fiehn 2002).

L'approche par empreintes métaboliques ne cherche pas à identifier chaque molécule précisément mais vise à comparer rapidement des tendances dans des profils métaboliques en réponse à une perturbation. Ceci est généralement obtenu en codant simplement l'absence ou la présence d'une molécule (Fiehn 2002). Cette approche est rapide et permet de discriminer des groupes d'échantillons et/ou d'identifier des régions discriminantes sur les chromatogrammes. Par exemple, des pommes de terres génétiquement modifiées et naturelles ont été comparées par cette technique et il a été mis en évidence des changements inattendus dans les métabolites suite à la transgénèse (Catchpole, Beckmann *et al.* 2005). Ce type d'identification permet ensuite de cibler les recherches sur la quantification de quelques métabolites précis (biomarqueurs).

L'approche par profil métabolique, elle, demande d'identifier et de quantifier l'ensemble des molécules individuelles, même les molécules de structures chimiques inconnues (De Souza, Saunders *et al.* 2006; De Vos, Moco *et al.* 2007). Ceci est généralement fait au sein d'une même famille de molécule ou d'une même voie métabolique (Fiehn 2002). Kaplan *et al*

Introduction

identifient des molécules intervenant dans la résistance au stress lié à la température chez *Arabidopsis* (Kaplan, Kopka *et al.* 2004).

La majorité des outils utilisés en métabolomique sont issus de la chromatographie et de la spectrométrie de masse : chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS), chromatographie liquide et spectrométrie de masse (LC-MS) et autres ainsi que de la résonance magnétique nucléaire (NMR) (De Souza, Saunders *et al.* 2006; Sumner, Mendes *et al.* 2003). Une fois recueillies, les données sont traitées statistiquement, comparées à des bases de données de voies de synthèses ou d'autres types, utilisées pour repérer de nouveaux marqueurs biologiques et dans l'aide au diagnostic. Figure 28 d'après (Dettmer, Aronov *et al.* 2007).

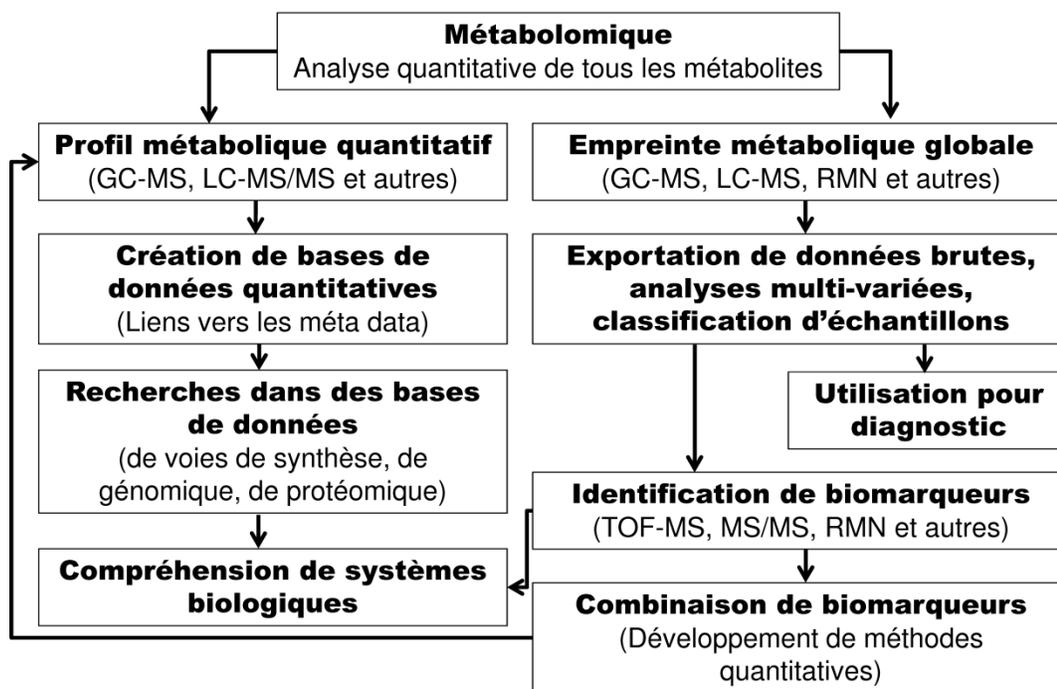


Figure 28: Stratégies d'investigation en métabolomique (modifié d'après Dettmer, Aronov *et al.* 2007).

Peu importe la technique utilisée, le signal recueilli est fait de signaux informatifs (liés à la présence de molécules parfois en mélange car plus ou moins bien séparées) et parasité par un bruit de fond. Chaque pic dans l'intensité du signal est associé à une molécule (un métabolite) et la hauteur du pic est corrélée avec la concentration de la molécule (De Souza, Saunders *et al.* 2006). Un des problèmes communs est qu'une fois analysés les échantillons doivent être comparés. Il faut alors comparer les pics d'un échantillon avec ceux des autres pour obtenir une matrice résumant les informations expérimentales. Cette matrice est en général un tableau dont les lignes sont les échantillons/les individus testés et les colonnes sont les molécules

Introduction

déteectées. La génération de cette matrice fait appel à différentes étapes de traitement de données qu'on regroupe sous le terme de prétraitement (*preprocessing* en anglais). En général, les étapes du prétraitement sont : l'élimination du bruit de fond, la détection, l'intégration des pics et enfin l'alignement des pics entre les différentes analyses pour générer la matrice finale. Si les étapes de nettoyage du bruit de fond, de détection et intégration des pics sont largement automatisées, l'étape d'alignement des profils est, quant à elle, parfois assez fastidieuse car elle est impactée par la technique utilisée, le matériel biologique, les conditions expérimentales, la durée de l'expérimentation... Le choix de la (des) technique(s) d'analyse repose donc sur un compromis entre la rapidité, la sensibilité et la sélectivité nécessaire à l'étude. La NMR est une technique rapide et sélective mais est reconnue pour être peu sensible (Bino, Hall *et al.* 2004). Les techniques de GC-MS ou de LC-MS sont sensibles et sélectives mais sont parfois assez lentes. Elles ont cependant l'avantage d'être d'un coût abordable et d'être de plus en plus rapides avec notamment l'apparition des GC-MS ultra rapides (Sumner, Mendes *et al.* 2003). La GC-MS est d'ailleurs la toute première technique ayant été utilisée pour faire une étude de métabolomique dans les années 1970 (Devaux, Horning dans Sumner, Mendes *et al.* 2003).

La rapidité et le faible coût des techniques permettent de mettre en place des plans expérimentaux complexes et d'assurer des traitements statistiques robustes de l'ensemble des données acquises et ce même si ces plans expérimentaux nécessitent de nombreuses réplifications des expériences sur un échantillonnage de grande taille (De Souza, Saunders *et al.* 2006; Sumner, Mendes *et al.* 2003). Les étapes d'analyse de données (visualisation et interprétation) sont regroupées sous le terme de post-traitement (*postprocessing* en anglais), ce sont généralement des traitements statistiques des données qui font, par exemple, appel à des analyses multivariées comme les analyses en composantes principales (ACP), analyse en coordonnées principales (ACoP), les algorithmes de classification (HCA, K-means)(Van Dam et Poppy 2008) ou encore des combinaisons des ces analyses (Sumner, Mendes *et al.* 2003). Ceci permet également de tenir compte de la variabilité biologique entre échantillons et de contrôler les variations techniques liées à l'appareillage de mesure (Fiehn 2002). Cependant, ces possibilités sont encore trop rarement utilisées bien qu'elles soient vivement recommandées (Sumner, Mendes *et al.* 2003).

La grande variété de techniques, d'appareillages utilisés (GC-MS, LC-MS, NMR...), de domaines de recherche et de types de données recueillis, nécessite une standardisation des

Introduction

méthodes et des formats d'échange de données (Castle, Fiehn *et al.* 2006; Fukusaki et Kobayashi 2005; Lindon, Nicholson *et al.* 2005). Ce besoin de standardisation intervient à tous les niveaux d'expérimentation de métabolomique : de l'acquisition des données brutes à leur traitement statistique final (Bino, Hall *et al.* 2004). Ainsi, l'utilisation des langages standardisés de données (NetCDF, mzDATA, mzXML) (Strömbäck, Hall *et al.* 2007), facilitera les échanges des données entre laboratoires et permettra de comparer plus aisément les expérimentations. Ceci procurera à la métabolomique un plus grand potentiel de découverte (Brazma, Krestyaninova *et al.* 2006). De nombreuses avancées ont été faites dans le prétraitement des données (des techniques d'analyses et l'acquisition qualitative et quantitative) ce qui a permis d'augmenter considérablement les potentialités de la métabolomique. A l'opposé, le post-traitement de ces données demande à être amélioré car ces données s'acquièrent beaucoup plus vite qu'elles ne se traitent créant une saturation des systèmes de traitement informatique (Kearsley, Wallace *et al.* 2005). Actuellement, la compilation des données brutes, issues d'appareils de différentes marques, demande beaucoup de temps et est encore trop souvent faite « manuellement » en utilisant des logiciels du type tableur (ex :Excel, Open office calc...) pour rapatrier les données aux différents formats avant de les regrouper sous un format unique (Duran, Yang *et al.* 2003). La standardisation des formats de données et la création d'outils de bioinformatique et de statistiques dédiés au pré et post-traitement des données brutes de métabolomique sont donc nécessaires et participeront à renforcer l'efficacité des recherches en permettant de traiter de grands jeux de données et en facilitant les échanges entre laboratoires (Tohge et Fernie 2009). Des nombreux groupes de recherche appellent à la mise en place de groupes de travail pour la création des procédures et des outils d'analyse et de traitement des jeux de données de plus en plus grands et complexes issus de la métabolomique (Kumar et Mann 2009; Smilde, Westerhuis *et al.*; Van Dam et Poppy 2008). Des initiatives telles que celle du MIAMET (informations minimum à fournir pour publier des recherches en métabolomique de Bino, Hall *et al.* 2004) ou bien l'initiative pour des standards en métabolomique (MSI de Fiehn, Wohlgemuth *et al.* 2008) participent à l'évolution du domaine de la métabolomique. Tout ceci permettra, à terme, de comprendre des systèmes biologiques complexes. Parmi les avancées significatives dans ces domaines, on citera pour le prétraitement d'analyses de GC/LC-MS: Targetsearch, XCMS et son extension pour la caractérisation de structures chimiques (XCMS²) développés avec le logiciel de statistique gratuit R (Benton, Wong *et al.* 2008; Cuadros-Inostroza, Caldana *et al.* 2009; Smith, Want *et al.* 2006). Metalign, Tagfinder (via Metalign), MathDAMP, MSFACTs sont

Introduction

des outils indépendants et libres (Baran, Kochi *et al.* 2006; Duran, Yang *et al.* 2003; Lommen 2009; Luedemann, Strassburg *et al.* 2008). D'autres auteurs ont développé des algorithmes avec des logiciels commerciaux comme Dixon *et al.* avec MathLab (Dixon, Brereton *et al.* 2006). D'autres logiciels, plus généralistes, permettent à la fois le pré et le post-traitement. On citera MZmine (Katajamaa, Miettinen *et al.* 2006) et d'autres algorithmes sans nom (Steinfath, Groth *et al.* 2008 ; Tikunov, Lommen *et al.* 2005). Certains sont disponibles via des plateformes de calcul sur internet (Revue dans Tohge et Fernie 2009). Plus particulièrement MetaboAnalyst (Xia, Psychogios *et al.* 2009) et MeltDB (Neuweger, Albaum *et al.* 2008) qui permettent le traitement de données de métabolomique issues de différentes expérimentations. Ces plateformes acceptent des données de NMR, de MS ou simplement des listes de molécules avec leurs concentrations et permettent un grand nombre d'analyses statistiques (ACP, classification...). En parallèle, les fabricants d'instruments d'analyses proposent leurs propres logiciels (Agilent – Genespring MS et Mass profiler pro, Water – MarkerLynx, AB Sciex – Markerview). De nombreux outils de « post-traitement », parfois créés pour d'autres domaines, existent et sont adaptés au traitement des informations issues d'expérimentations de métabolomique, la « fouille » de données et plus particulièrement la classification de données, que je développe ci-dessous, en font partie.

III-2 Les outils statistique pour le traitement de grands jeux de données : le *data mining*

La « fouille » de données (*data mining* en anglais) est l'ensemble des méthodes et techniques destinées à l'exploration et à l'analyse de grandes bases de données informatiques, de façon automatique, ou semi-automatique, en vue de détecter dans ces données des règles, des associations, des tendances inconnues ou cachées, des structures particulières restituant l'essentiel de l'information utile tout en réduisant la quantité de données. Les outils de *data mining* permettent de limiter la subjectivité humaine dans le processus de décision lié à l'analyse de jeux de données. Les outils utilisés sont, soit descriptifs, soit prédictifs. Les techniques descriptives ou exploratoires visent à mettre en évidence des informations présentes mais cachées par le volume de données, c'est par exemple le cas des classifications automatiques de données (*Clustering* en anglais). Les techniques prédictives permettent d'extrapoler de nouvelles informations à partir des informations présentes (Tufféry 2005). Grâce aux potentialités de l'informatique moderne ces techniques, qui ne sont pas innovantes expriment tout leur potentiel en extrayant de plus en plus rapidement les relations cachées

entre données aussi nombreuses soient-elles. Ceci fait du *data mining* et, parmi les techniques de *data mining*, de la classification de données un domaine intéressant pour le traitement de données de composition chimique (Bino, Hall *et al.* 2004; Sumner, Mendes *et al.* 2003) et en particulier dans des travaux utilisant la GC-MS (De Souza, Saunders *et al.* 2006).

III-3 La classification de données (*Clustering*)

En statistiques, lorsque l'on souhaite regrouper des objets (individus ou variables) en un nombre limité de groupes homogènes (classes/ segments/ *clusters*), on utilise des techniques de classification de données (clustering, segmentation ou encore classification automatique selon les auteurs Tufféry 2005). Les techniques de *clustering* sont les plus répandues des techniques de *data mining* descriptif. On les utilise lorsqu'on dispose d'un grand volume de données duquel on souhaite distinguer des sous-ensembles homogènes, susceptibles de traitements et d'analyses différenciés. Ces ensembles peuvent être disjoints, chevauchants ou bien organisés de façon hiérarchique. Lorsqu'on souhaite regrouper de façon non-supervisée les objets similaires et séparer les objets ayant des caractéristiques différentes on utilisera les algorithmes de classification non-supervisée. Ces algorithmes sont divisés en deux catégories : la classification par partitionnement et la classification hiérarchique (agglomérative ou divisive). Ces algorithmes permettent de retrouver des relations de parenté et/ou de proximité entre les échantillons. Chaque algorithme cherche à obtenir des *clusters* les plus homogènes possibles tout en maximisant les différences entre les *clusters* (homogénéité intra-groupe et séparation inter-groupes). Cette homogénéité est mesurée par des critères tels que l'inertie (*compactness* en anglais). Cette inertie est mesurée par différents indices mathématiques (Wild et Blankley 2000). Les algorithmes de classification non-supervisée utilisent des calculs de distances entre les données et d'autres calculs de lien pour regrouper les données. Il faut donc pouvoir choisir quel calcul de distance et quel calcul de lien adopter. Les calculs de distance courants sont basés sur des corrélations (par exemple corrélation de Pearson) ou les distances de Minkowski (plus particulièrement la distance euclidienne). Pour les calculs de lien on citera : moyen, simple, complet et ward (Zhao et Karypis 2003). Ces choix sont loin d'être aisés et dépendent des données traitées. Un autre choix crucial est celui du critère d'arrêt du *clustering*. En effet, les algorithmes de classification hiérarchiques divisifs commencent par regrouper toutes les données en un seul *cluster* qu'ils divisent ensuite récursivement jusqu'à créer autant de *clusters* que de données. Les algorithmes de classification agglomératifs font le chemin inverse. Il faut donc pouvoir définir un seuil d'arrêt permettant d'obtenir le *clustering* de la

Introduction

meilleure qualité possible. La qualité du *clustering*, et donc le seuil d'arrêt de cette clusterisation, peut être évaluée par différents indices tels que l'indice silhouette (Rousseeuw 1987).

Les ensembles homogènes obtenus à la fin du *clustering* partagent donc des caractéristiques communes et leur association à un sens. Cette définition est volontairement floue, car la notion de groupes homogènes dont l'association à un sens est profondément dépendante du type d'objets analysés/du type des variables et de l'application finale de cette classification de données (Zhao et Karypis 2003). En biologie, la classification de données est utilisée dans de nombreux domaines comme la médecine (*i.e.* traitement de données cliniques), la génomique, la protéomique (Zhao et Karypis 2005) et en métabolomique (Sumner, Mendes *et al.* 2003) sans oublier la phylogénie (Leebens-Mack, Vision *et al.* 2006; Paton, Springate *et al.* 2004).

IV Les buts de la thèse

L'objectif de mes recherches est de caractériser la diversité des composés terpéniques volatils au sein des espèces du genre *Lavandula* d'un point de vu évolutif et physiologique. Ces travaux associés à une phylogénie moléculaire robuste du genre devant, à terme, permettre de montrer la part que la génétique a dans l'évolution de la production de CTV par les lavandes. En effet, de nombreuses plantes produisent des CTV et, il a été démontré l'importance que ceux-ci avaient pour la survie des végétaux dans leur contexte écologique. D'autre part, la diversité et la quantité de CTV produits sont influencées par de nombreux facteurs biotiques et abiotiques. Cependant, la génétique a un rôle majeur sur le type de CTV qui sont produit et sur le contrôle de son accumulation spatio-temporelle. L'étude de l'évolution de la production de ces composés terpéniques volatils au sein d'un même genre, bien qu'elle puisse apporter de nouvelles connaissances sur les mécanismes évolutifs, n'est que très peu étudiée (Levin, McDade *et al.* 2003). Lorsque de telles études ont été entreprises elles se sont heurtées à la difficulté de disposer dans un même endroit, et donc dans les mêmes conditions environnementales, d'un nombre suffisamment grand d'espèces d'un genre pour être représentatif de la diversité du genre (Levin, McDade *et al.* 2003). Azuma *et al* travaillèrent sur 25 des 128 espèces du genre *Magnolia* pour caractériser la divergence évolutive du parfum des fleurs au sein du genre (Azuma, Thien *et al.* 1999). Mais les données de chimie des plantes étaient d'origines différentes (*i.e.* travaux de Azuma de 1997 et de Thien

Introduction

de 1975). Une autre difficulté dans ce type d'étude est le besoin de développer des outils d'analyse et de traitement de données capables de travailler sur des données issues d'échantillons de différents niveaux taxonomiques et de différents organes (échantillonnage stratifié). Il n'en reste pas moins que les terpènes et les TPS sont très bien connus (voies de synthèse, régulation, évolution), ce qui en fait des cibles privilégiées pour une telle étude.

Mes recherches sur la diversité des CTV du genre *Lavandula* se sont articulées autour des 3 questions suivantes :

- Comment optimiser les outils de prétraitement des données de GC-MS pour analyser un grand échantillonnage d'extraits de CTV à plusieurs niveaux taxonomiques et dans différents organes de lavandes ?
- Quel rapport existe-t-il entre les CTV accumulés par les différents taxons du genre *Lavandula* et les relations phylogénétiques entre ces taxons ?
- La production de CTV est-elle modifiée lors du développement de l'inflorescence de la lavande ?

Matériel et Méthodes

MATERIELS ET METHODES

Pour ce chapitre, j'ai souhaité être aussi précis que possible pour permettre à un étudiant novice de refaire les expériences et d'analyser les résultats. Cependant, pour ne pas alourdir ce chapitre, les codes fournisseurs des principaux réactifs et matériels sont listés dans l'Annexe II : Références produits.

I Matériel végétal

I-1 Origines

Pour mes recherches, j'ai travaillé sur des lavandes originaires de deux sites :

- Les lavandes botaniques étudiées (Tableau 6) sont celles de la collection entretenue par Bernard Pasquier au conservatoire national des plantes médicinales aromatiques et industrielles de Milly-la-forêt (CNPMAI). Leur identification a été faite par Bernard Pasquier sur la base des critères définis par Tim Upson (Upson et Andrews 2004).

- Les lavandes fines (*L. angustifolia*) et les lavandins (*L. × intermedia*) proviennent de la collection du lycée horticole de Romans-sur-Isère. Trois cultivars par espèce ont été suivis ; 'Matheronne', 'C15-50' et '77-13' pour *L. angustifolia* et les cultivars 'Super', 'Certitude' et 'Grosso' pour *L. × intermedia*.

I-2 Récoltes

I-2.1 Lavandes botaniques

Les récoltes de lavandes botaniques ont eut lieu en septembre 2006 au CNPMAI et en juillet 2008 sur des plantes âgées de 1 à 2 ans.

En 2006, des feuilles de 30 taxons (26 espèces ainsi que 2 hybrides et 2 sous-espèces) ont été récoltées et immédiatement placées dans un tube contenant du silica gel. Une fois au laboratoire, les échantillons ont été placés dans l'azote liquide et conservés à -80°C jusqu'au jour de l'extraction de leur ADN pour l'étude de phylogénie moléculaire (Tableau 6).

Matériel et Méthodes

En 2006 et 2008, des feuilles et, si disponibles, des inflorescences ont été prélevées sur divers taxons et stockées dans un papier humide pour leur transport au laboratoire. Dès leur arrivée au laboratoire, les CTV de ces échantillons ont été extraits et analysés par chromatographie en phase gazeuse et spectrométrie de masse (GC-MS) (Tableau 6).

Cet échantillonnage stratifié (Figure 29) nous permet de couvrir l'ensemble de l'aire de répartition ainsi que deux sous-genres et 7 des 8 sections du genre. Lors des analyses de CTV, plusieurs échantillons de fleurs et inflorescences ont été prélevés sur des plantes différentes afin d'estimer une variabilité éventuelle entre plantes appartenant au même taxon.

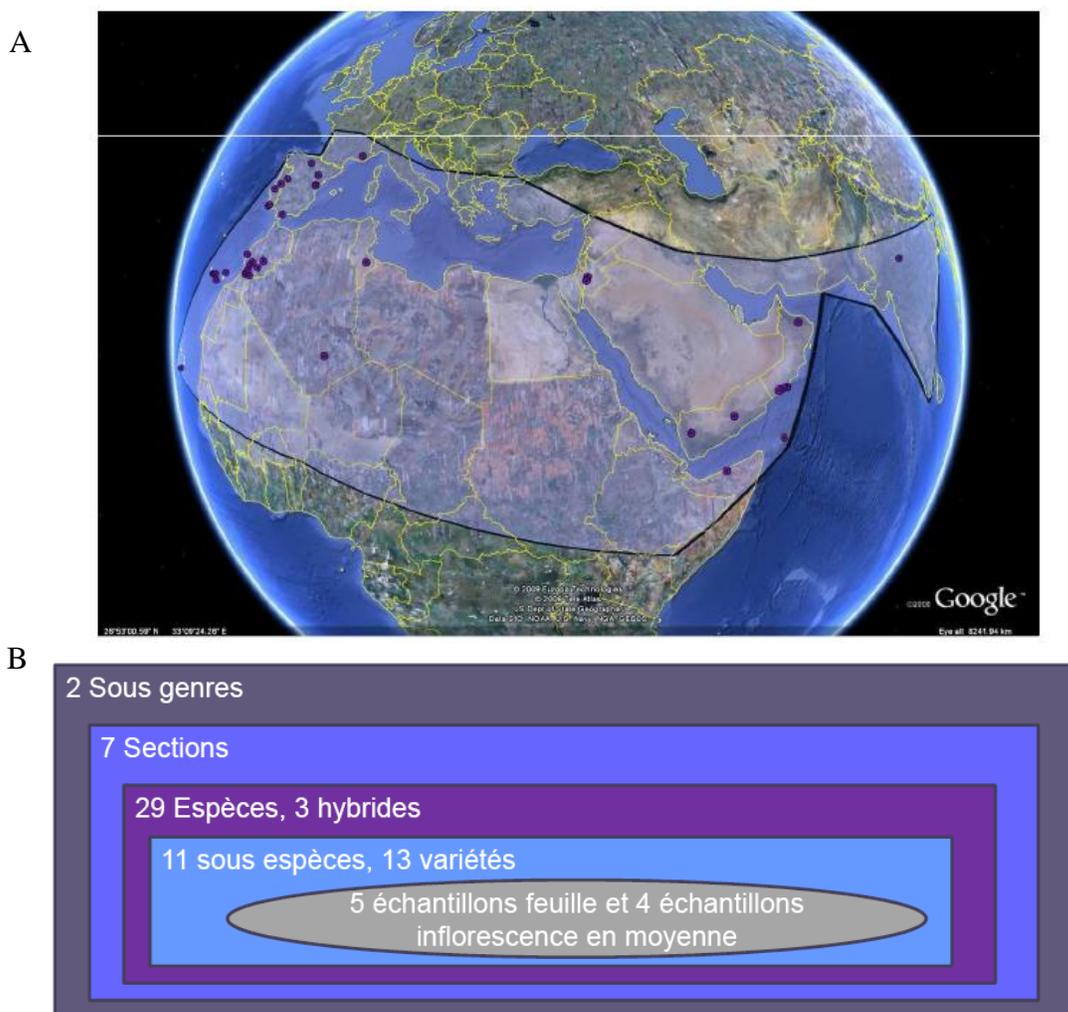


Figure 29 : Détails de l'échantillonnage de lavandes botanique.

A. Origine des échantillons de lavandes botaniques, les points violets indique les lieux d'origine des plantes prélevées au CNPMAI en 2006 et 2008, B. Schéma de l'échantillonnage stratifié.

I-2.2 Lavandes fines et lavandins

De 2005 à 2008, des récoltes de lavande fine et lavandin ont été effectuées chaque année au lycée horticole de Romans entre le mois de mai et le mois d'août, sur la base d'une récolte par semaine. La majorité de l'étude faite avec ces plantes porte sur l'évolution de la chimie de l'inflorescence lors du développement de celle-ci mais, selon les besoins d'analyse, d'autres organes ont été récoltés (*i.e.* fleurs individuelles à différents stades de développement ou position sur l'inflorescence, calices, bractées, corolles, tige florale et feuilles) (Voir Figure 30 pour une définition des stades de développement des fleurs et des zones de récolte des fleurs sur une inflorescence). Au total 2651 échantillons ont été récoltés pour cette étude (Tableau 7).

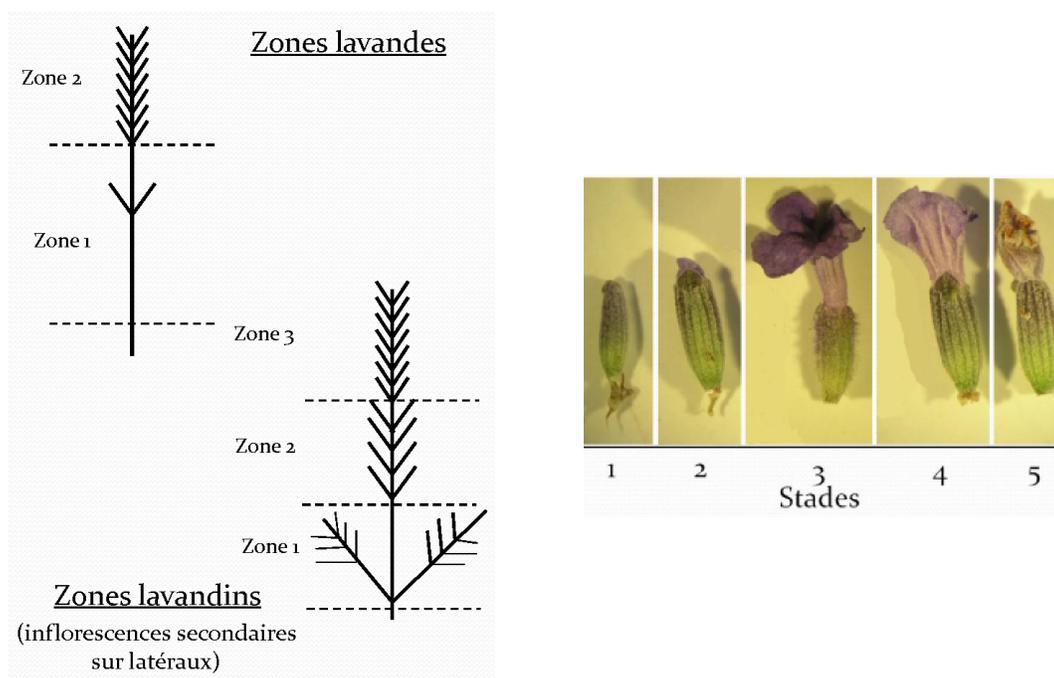


Figure 30 : Détails des zones de l'inflorescence et des stades de floraison

Matériel et Méthodes

Tableau 7 : Liste des échantillons de lavande et lavandin récoltés au lycée horticole de Romans-sur-Isère

Année	Zone*	Stade**	Test***	cultivars de lavande			cultivars de lavandin			Total		
				77-13	C15-50	Matheronne	Certitude	Grosso	Super (B110)			
2005	feuille			12	12	12	12	12	9	69		
	inflorescence			36	36	36	36	36	36	216		
	Z1	S1			3	3	3	8	8	6	31	
					S2	9	9	7	5	6	5	41
					S3	6	7	10	5	6	4	38
					S4	6	4	6	4	6	2	28
					S5	3	3	3	3	3	2	17
	Z2	S1			3	3	3	5	9	8	31	
					S2	8	10	6	11	12	12	59
					S3	7	9	11	9	11	9	56
					S4	7	7	8	9	9	7	47
					S5	3	3	3	3	3	3	18
	Z3	S1						7	8	7	22	
					S2				6	5	6	17
					S3				6	6	5	17
					S4				5	6	3	14
					S5				3	3	3	9
2006	bractée			10	10	10	10	10	10	60		
	calice		Test	5	5	5				15		
	inflorescence				60	60	60	60	60	59	359	
					Test		5	15			5	25
					Test 2				5			5
					Test 3				5			5
					Test 4				5			5
					Test 5				5			5
	petale		Test	5	5	5				15		
	tige Z1+Z2				29	30	30	30	30	30	179	
Test								5			10	
2007	feuille		non	10				10		20		
	inflorescence			129	130	127	140	140	139	805		
				Test masse			60		60		120	
				Test volume	119						119	
2008	feuille			89						89		
	inflorescence			85						85		
Nombre total d'échantillons									2651			

I-3 Conditions de culture

Les plantes échantillonnées à Milly la forêt au CNPMAI et à Romans sur isère au lycée horticole de Romans étaient cultivées en plein sol en extérieur selon les pratiques utilisées industriellement pour la culture de *L. angustifolia*.

Afin de disposer de matériel biologique frais pour certaines répétitions ou contrôles, la plupart des différents taxons ont été bouturés ou semés et cultivés en serre au laboratoire. Les plantes ont été cultivées en chambre de culture avec un cycle jour/nuit de 16 h de jour à 25°C sous une luminosité de 68 $\mu\text{mol. m}^{-2} \cdot \text{s}^{-1}$, soit environ 5000 Lux, et à l'obscurité pendant 8 h à 22°C. Le substrat de semis et de bouturage est le Substrat3 (mélange de tourbe brune et blonde, pH(H₂O)=6.0, marque Klasmann). Une fois les plants développés, ils sont transférés dans un mélange de Substrat2 (mélange de tourbe brune et blonde, pH(H₂O)=6.0, engrais enrobé NPK 14 16 18 kg/m³ marque Klasmann) de perlite et de billes d'argile (2 :1 :1 v/v).

Matériel et Méthodes

Pour *L. stoechas* le substrat est remplacé par de la terre de bruyère. Pour *L. lanata* de la litière de chat (source de calcaire) est placée en surface. L'arrosage était fait pot à pot en laissant au moins une période sèche par semaine.

Problèmes rencontrés en culture :

Fréquemment : aleurodes des serre (*Trialeurodes vaporariorum*) potentiellement vecteur de virus, développement de champignons du genre *Phytophthora* sur les racines causant un dessèchement de la plante et sa mort.

Beaucoup plus rarement : développement d'acariens et de pucerons.

II Méthodes

II-1 Microscopie électronique à balayage

Les échantillons ont été fixés pendant 4h dans une solution à 3% de glutaraldéhyde tamponnée avec 0.10 M de cacodylate de sodium (pH=7.2). Après lavage dans de la solution tampon, les échantillons étaient déshydratés, séchés au point critique puis métallisés avec une électrode de platine (Jullien, Diemer *et al.* 1998). Les observations ont été faites sur un microscope électronique à balayage (S-3000N Hitachi - Tokyo, Japon), à la Faculté de médecine de St-Etienne.

II-2 Analyse des composés organiques volatils

Une des techniques permettant l'analyse du parfum des plantes est la chromatographie en phase gazeuse (GC). Cette technique permet de séparer, de quantifier et d'identifier les composés organiques volatils constitutifs du parfum.

II-2.1 Préparation des échantillons

Juste après la récolte, les échantillons destinés à être analysés en GC étaient pesés (poids frais) et immédiatement placés dans un tube d'extraction de 16 ml contenant une solution d'hexane (Sigma) additionnée de 200 mg.l⁻¹ d'étalon interne (*cis*-3-hexenol ou *trans*-2-hexenal selon l'étude, Sigma). Le ratio poids frais/volume d'hexane était de 100 mg/ 2 ml. L'ajout d'étalon interne permet de corriger d'éventuelles variations entre échantillons liées à une évaporation de l'hexane lors des manipulations. Après une nuit d'extraction à 4°C, la solution contenant les COV est récupérée à l'aide d'une pipette pasteur en verre et placée dans un tube d'analyse GC de 1.5 ml et conservée à -20°C jusqu'à l'analyse. Le matériel végétal

Matériel et Méthodes

restant était maintenu dans le tube d'extraction et placé à l'étuve (50°C) jusqu'à stabilisation du poids (poids sec). Afin de pouvoir placer les inflorescences entières dans les tubes d'extraction celle-ci étaient coupées en tronçons d'1 cm. Les fleurs individuelles, elles, étaient détachées des inflorescences avec des ciseaux et placées par groupe de 12 dans les tubes d'extraction. Pour une étude détaillée de la production de COV par les différents organes de la fleur, nous les avons, à l'aide de pinces, disséqués en deux groupes : a/ étamines et corolle et b/ carpelle et calice (chez les lavandes, les étamines sont soudées à la corolle et les carpelles au calice). Feuilles, bractées et tiges florales étaient également récoltées avec des ciseaux et, au besoin, recoupées afin de pouvoir être placées dans les tubes d'extraction. Les extraits hexane de bractées ont été concentrés par un facteur 10 sous un flux d'air comprimé avant analyse.

II-2.2 Analyse en GC-FID

Les composés volatils ont été analysés par chromatographie en phase gazeuse et détection de flamme (GC-FID) sur un appareil Agilent 6850 équipé d'une colonne capillaire HP-Innowax (30 m de longueur, 0.25 mm de diamètre interne avec une phase stationnaire de 0.25 μm).

Les réglages étaient les suivants :

- Les températures de l'injecteur et du détecteur étaient de 250°C.
- Le gaz vecteur était l'azote (1.2 mL.min⁻¹).
- Le profil de température du four était de : 6 min à 40°C après l'injection puis augmentation de 1°C.min⁻¹ jusqu'à 60°C, puis 2°C.min⁻¹ de hausse jusqu'à 140°C et enfin 12°C.min⁻¹ jusqu'à 240°C. La température était maintenue à 240°C pendant 5 min.
- Le volume d'injection était de 2 μl avec un ratio de split de 10:1.

II-2.3 Analyse en GC-MS

Les spectres de masse des composés volatils ont été obtenus sur un spectromètre de masse à quadrupole et ionisation par bombardement électronique (Agilent 5973) après une séparation par chromatographie gazeuse sur un appareil Agilent 6850. Les conditions de chromatographie étaient les mêmes que pour l'analyse en GC-FID, sauf que le gaz vecteur était l'hélium et que le ratio de split était de 2:1. Les températures de la source d'ions et de l'interface GC-MS étaient de 230°C. Le voltage d'ionisation (ei-mode) était de 70 eV avec une mesure de 2.45 scans.s⁻¹ entre 35–350 m.z⁻¹.

II-3 Biologie moléculaire

II-3.1 Extraction de l'ADN génomique

Les ADN ont été extraits selon un protocole dérivé de Doyle et Doyle (1987). Cette méthode d'extraction utilise du CTAB et nous avons modifié le protocole initial en ajoutant deux étapes de lavage à l'éthanol après la précipitation faite avec de l'isopropanol.

En règle générale, 100 mg de tissu (poids frais feuilles/fleurs) sont broyées dans l'azote liquide, puis additionnés de 1 ml de tampon d'extraction (Tris-HCl 100 mM pH 7,5 ; NaCl 1,4 M ; EDTA 20mM ; 2% CTAB). Après 90 min d'incubation à 65°C, 450 µl d'une solution de chloroforme/alcool isoamylique (24 :1) sont ajoutés. Une centrifugation de 10 min à 3700 g est réalisée et la phase aqueuse est récupérée. L'ajout d'un volume identique d'isopropanol permet la précipitation de l'ADN. Après centrifugation à 12800 g pendant 1 min, le surnageant est éliminé et le culot est lavé avec 500 µl d'éthanol à 70%. Cette étape de lavage est répétée deux fois. Le culot d'ADN obtenu est séché sous hotte Sorbonne, puis repris dans 40 µl d'eau stérile.

L'ADNg extrait est conservé à -20°C.

II-3.2 Extraction des ARN

Lors de mes recherches deux protocoles d'extraction d'ARN ont été employés. Le premier, est réalisable avec des produits de base de laboratoire, mais est long (2 jours-Méthode de Verwoerd). Le second est rapide (3h) mais nécessite l'achat d'un kit d'extraction (NucleoSpin® RNA Plant kit de Macherey-Nagel). Pour les étapes de mise au point qui ne nécessitaient que peu d'échantillons le premier protocole a été employé. Le kit a été utilisé pour les nombreux échantillons destinés à être analysés en qPCR.

Méthode de Verwoerd

Le protocole utilisé est adapté de celui de Verwoerd et *al.* (1989). 50 mg de jeunes feuilles sont broyées dans l'azote liquide, puis additionnés de 500 µl de tampon d'extraction chauffé à 65°C [phénol / (Tris-HCl 100 mM pH 8 ; EDTA 10 mM pH 8; LiCl 100 mM ; SDS 1%), (1 :1)]. Après homogénéisation, 250 µl de chloroforme/alcool isoamylique (24 :1) sont ajoutés. Une centrifugation de 5 min à 8000 g à 4°C est réalisée, puis la phase aqueuse est récupérée et un volume égal de LiCl 4 M est ajouté. Les ARN sont précipités pendant une nuit à 4°C. Après une centrifugation pendant 30 min à 12800 g à 4°C, le culot est re-suspendu dans

Matériel et Méthodes

250 µl d'eau stérile. Les ARN sont précipités par ajout d'1/10 volume d'acétate de sodium à 3M (pH 5,2) et 2 volumes d'éthanol absolu et incubation pendant 1 h à -20°C. Le culot d'ARN obtenu après une centrifugation de 30 min à 12800 g à 4°C, est lavé avec 100 µl d'éthanol à 70%. Après 10 min de centrifugation à 12800 g à 4°C, le surnageant est éliminé et le culot est séché sous hotte Sorbonne, puis repris dans 20 µl d'eau stérile. Les ARN sont conservés à -80°C.

Kit d'extraction Malchery Nagel

Le Kit d'extraction des ARN totaux utilisé est le NucleoSpin® RNA Plant kit (Macherey-Nagel). Le kit propose deux tampons d'extraction : RAP et RA1. Les deux permettent d'extraire les ARN. RA1 a un plus faible rendement d'extraction que RAP, mais donne des extraits de meilleure qualité (moins de protéines résiduelles). RA1 a donc été utilisé pour notre étude. Le protocole contient une étape de traitement à la DNase que j'ai supprimée car les contrôles de qualité des ARN sur gel d'électrophorèse montraient une contamination en ADN même après cette étape de DNase. Une autre étape de DNase était donc nécessaire après l'éluion des ARN (cf plus bas III-3). Pour améliorer le rendement d'extraction les ARN sont élués de la colonne avec 40 µl d'eau UP puis l'éluat est repris et remplacé sur la colonne et à nouveau centrifugé.

II-3.3 Synthèse du premier brin d'ADN complémentaire

Les ARN messagers nous renseignent sur l'expression des gènes. Afin de quantifier cette expression ils doivent être transformés en ADN complémentaire (ADNc) avant de pouvoir être utilisés en Polymerase Chain Reaction (PCR). Ceci a été fait avec les protocoles suivants.

Traitement à la DNase

Malgré les précautions prises lors de l'extraction des ARN, il est possible que les extraits contiennent des traces d'ADN. Il est donc nécessaire de les traiter pour éliminer l'ADN restant. Dans un volume final de 100 µl, 10 µg d'ARN sont additionnés de 10 µl de tampon RQ1 (10X RQ1, Promega), d'1 µl de DNase (RQ1 5U/µl, Promega) et de 1 µl d'inhibiteur de RNase (RNasin 40U/µl, Promega). Le mélange réactionnel est incubé pendant 1h à 37°C. A l'issue, 100 µl d'une solution de phénol/chloroforme/alcool isoamylique (25 : 24 : 1) à pH 8 sont ajoutés, puis une centrifugation à 12800 g pendant 10 min à 4°C est réalisée et la phase aqueuse est récupérée. Cette étape, effectuée deux fois, permet l'élimination des protéines. Les ARN sont précipités avec 1/10 volume d'acétate de sodium

Matériel et Méthodes

3M (pH 5,8) et 2 volumes d'éthanol absolu, pendant 2 heures à -20°C. Après 30 min de centrifugation à 12800 g à 4°C, le surnageant est éliminé et le culot est lavé avec 100 µl d'éthanol à 70%, puis centrifugé à nouveau pendant 10 min à 12800 g à 4°C. Cette étape de lavage est réalisée deux fois. Les culots d'ARN obtenus sont séchés sous hotte Sorbonne, puis repris dans 10 µl d'eau stérile. Les ARN traités à la DNase sont conservés à -80°C.

Rétro-transcription

Une rétro-transcription des ARNm (RT) est réalisée pour synthétiser le premier brin d'ADNc qui sera utilisé comme matrice de la PCR.

Un total de 2 µg d'ARN totaux traités à la DNase (ci-dessus II-3.3) sont additionnés de 0,5 µg d'oligo (dT) (Promega) dans un volume final de 11 µl, et sont incubés 5 min à 70°C pour permettre la dénaturation des ARN, puis placés immédiatement dans la glace pendant 30 sec. La réaction de rétro-transcription est réalisée en ajoutant, pour un volume final de 25 µl, du tampon M-MLV (Promega), 1 mM de chaque dNTP, 400 U de reverse transcriptase M-MLV (Promega) et 60 U d'inhibiteur de RNase RNasin[®] (Promega). Le mélange réactionnel est incubé à 42°C pendant 1 h, puis 5 min à 95°C pour inactiver la reverse transcriptase. Les ADNc simple brin sont conservés à -20°C.

II-3.4 Dosage et contrôles de la qualité des acides nucléiques

Au spectrophotomètre

La concentration d'acides nucléiques extraits est déterminée par une mesure de la densité optique à 260 nm, en admettant que 1 unité de DO à 260 nm équivale à 40 µg/ml d'ARN ou 50 µg/ml d'ADN.

La pureté de ces acides nucléiques vis-à-vis des protéines et des polysaccharides est estimée par une mesure du ratio des DO à 230nm, 260 nm et 280 nm respectivement. Si les rapports $DO_{260\text{ nm}}/DO_{230\text{ nm}}$ et $DO_{260\text{ nm}}/DO_{280\text{ nm}}$ sont proches de 2, on considère que l'on a une bonne purification vis-à-vis des polysaccharides et des protéines.

Sur gel d'électrophorèse

La quantité et la qualité des ARN est appréciée par migration de 1 µl d'extrait d'ARN additionné de 4 µl de tampon de charge (Promega), sur gel d'agarose à 1% dans du tampon Tris-Acétate-EDTA (TAE) stérile (pH=8, 40 mM TRIS base, 40 mM Acide acétique glacial, 1 mM EDTA) contenant 0,6 µg/ml de bromure d'éthydiuim (BET).

Par « fausse » rétro- transcription

L'absence d'ADN contaminant dans les extraits d'ARN peut-être vérifiée en réalisant une « fausse » rétro-transcription (RT-). Cette procédure consiste à appliquer le protocole de rétro-transcription (ci-dessus II-3.3) sur une partie de l'extrait d'ARN en n'utilisant pas l'enzyme M-MLV (Promega). Une fois le protocole terminé, l'échantillon RT- est amplifié par PCR avec un couple d'amorce si possible spécifique de l'ADN et n'amplifiant pas l'ARN (*e.g.* défini sur un intron). Après migration sur gel d'électrophorèse et révélation aux UV il ne doit pas y avoir de bande visible. Sinon, l'extrait contient de l'ADN et doit être retiré de l'étude. On peut tester l'échantillon RT- en PCR quantitative (qPCR). Dans ce cas, il ne doit pas y avoir d'amplification ou, au pire, une détection tardive (plus de 35 cycles). L'inconvénient majeur de cette technique est qu'elle utilise une partie de l'extrait d'ARN nécessaire à l'étude et présuppose de disposer de suffisamment d'ARN pour faire une RT et une « fausse » RT.

II-3.5 Amplification par PCR

Définition des amorces

Lors de mes recherches, j'ai été amené à utiliser des couples d'amorces pour la réalisation de PCR et qPCR (Tableau 8).

Tableau 8 : Liste des amorces utilisées en PCR et qPCR ainsi que les conditions de réactions en PCR

Avec X le nombre de cycles d'amplification, X et Y le temps et la température d'hybridation, Z le temps d'élongation.

Gène ciblé	Amorces (5' Forward 3' / 5' Reverse 3')	Taille attendue de l'amplicon (bp)	Conditions PCR
La_LIM	GGCCACACAAC TAGAAATTAAGT/ TTGCACAGTCAGCTCAGCG	151	N=25 cycles X=15 sec Y=60 °C Z=45 sec
La_LIN	TTGATGAGGATGGTGATGGA/ TCTTGTGTGTTGCTTGGACA	198	N=25 cycles X=15 sec Y=60 °C Z=45 sec
La_BERG	ATTGGTGAGGCGTGTTAGC/ ATCTTTTCGGGTAGGCCAGT	235	N=25 cycles X=15 sec Y=60 °C Z=45 sec
La_Actin	TGCTCAGTGGTGGATCGACTATG/ ACCACCTTAATCTTCATGCTGCTTG	95	N=25 cycles X=15 sec Y=60 °C Z=45 sec
La_Tubulin	CCACAATCAAGACCAAGAGGACAATC/ CGCCAGGGACCACGGTTG	98	N=25 cycles X=15 sec Y=60 °C Z=45 sec
La_EFI-alpha	CCCTTCTGAGGCTCTTGAC/ GCACAGTTCCAATACCACC	106	N=25 cycles X=15 sec Y=60 °C Z=45 sec
Trnk	GGGGTTGCTAACTCAACGG/ AACTAGTCGGATGGAGTAG	2000	N=25 cycles X=30 sec Y=48 °C Z=120 sec

Pour notre étude de phylogénie, nous avons travaillé sur la région du génome codant pour l'intron du gène *trnK*. Les ADN extraits (Tableau 6) ainsi que ceux d'un plant de mélisse et de basilic fin vert ont été amplifiés par PCR en utilisant le couple d'amorces *trnK*-3914F et *trnK*-2R (Amorces notées *Trnk* dans le Tableau 8). L'amorce *trnK*-2R (C4R dans la publication de (Miikeda, Kita *et al.* 2006) a été désignée par Kelly Steele, et *trnK*-3914F dessinée par Jerry Learn a été modifiée par Soltis (Soltis et Soltis 1995).

Pour l'étude par qPCR, les amorces spécifiques des gènes LaLIMS, LaLINS et LaBERS ont été définies à partir de la séquence d'ADNc des synthèses de *L. angustifolia* publiées (Landmann, Fink *et al.* 2007) avec le logiciel Primer3 plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>, (Untergasser, Nijveen *et al.* 2007). Les séquences des gènes de référence (actine, tubuline et facteur d'élongation alpha) de lavande n'étaient pas connues au début de ma thèse. J'ai donc recherché ces gènes chez la lavande en travaillant sur les zones conservées des séquences connues de ces gènes chez les Lamiacées. Pour cela j'ai utilisé les séquences disponibles sur GenBank de sauge et de basilic. Après alignement des séquences des différents gènes, j'ai défini des couples d'amorces que j'ai testé sur des ADN de *L.angustifolia*. Les produits PCR ainsi obtenus ont été séquencés et comparés par Blast. Les

séquences obtenues sur les ADN de *L. angustifolia* correspondants aux gènes de ménage recherchés ont ensuite été utilisées pour définir des amorces spécifiques et dont les amplicons ne dépassaient pas 200 bp.

Conditions réactionnelles

Chaque réaction de polymérisation en chaîne est réalisée dans un volume final de 20 µl contenant le tampon GoTaq[®] (Promega), 0,2 mM de chaque dNTP, 5 mM d'amorces sens et antisens, 1 U d'ADN polymérase GoTaq[®] et environ 0,6 µg d'ADN. Le programme de PCR utilisé permet une dénaturation initiale (5 min à 94°C), puis N cycles d'amplification (30 sec à 94°C, X sec à Y°C, Z sec à 72°C) et une étape d'élongation finale (7 min à 72°C). Le Tableau 8 reprend les différentes conditions d'utilisation en fonction des amorces (valeurs de X, Y et Z). Il est effectué sur un thermocycleur Eppendorf Mastercycler[®].

Visualisation

Les produits PCR sont visualisés sur gel d'agarose (cf II-3.4)

II-3.6 PCR quantitative

Afin de suivre les variations des transcrits de TPS de lavande, deux gènes de TPS (*LaLIMS* and *LaLINS*) ont été étudiés par PCR quantitative. Les analyses de PCR quantitative ont été réalisées au sein du laboratoire de néphrologie de la faculté de médecine de St-Etienne sur un thermocycleur AB7500 (Applied Biosystems) avec des plaques de qPCR 96 puits. Les ADNc nécessaires à la réaction ont été obtenus après extraction des ARN avec le kit NucleoSpin[®] RNA Plant kit (Macherey-Nagel) et transcription inverse avec les protocoles décrit plus haut (II-3.3).

La réaction de PCR était faite dans un volume final de 25 µl, incluant 400 nM de chacune des amorces sens et antisens, et 5 µl d'ADNc dilué au dixième. Le cycle de qPCR était le suivant : 10 min à 95°C puis 40 cycles de dénaturation de 15 secondes à 95°C et enfin hybridation-extension de 45 secondes à 60°C. Conformément aux recommandations en matière d'analyse en qPCR des contrôles ont été effectués (Bustin, Benes *et al.* 2009). Afin de s'assurer de l'absence de contaminations par de l'ADN génomique, des « fausses » RT ont été réalisées avec le protocole décrit plus haut. La qualité du mélange réactionnel était quand à elle vérifiée par des NTC (non-template control) réalisés en remplaçant les 5 µl d'ADNc par de l'eau ultra-pure. Toutes les réactions de PCR ont été réalisées en triplicats techniques. Pour chacune des

dates testées, 6 ADNc provenant d'inflorescences différentes ont été analysées. La réaction de qPCR était conclue par la réalisation d'une courbe de fusion (de 60°C to 95°C). Cette étape permet de vérifier que le signal mesuré dans chaque puit correspond bien à un seul produit PCR.

II-3.7 Séquençage

Le séquençage était sous-traité par la société MWG (Eurofins MWG operon, Ebersberg, Allemagne). Avant l'envoi au séquençage les produits PCR étaient purifiés en utilisant le kit QIAquick PCR purification (Qiagen, Hilden, Allemagne). Lorsque les produits PCR contenaient plusieurs amplicons de tailles différentes, le produit de taille attendue était découpé sur le gel de visualisation et purifié à l'aide du kit QIAquick[®] (Qiagen, Hilden, Allemagne).

III Analyse de données

III-1 Généralités

Une des forces de mes recherches est le nombre important d'échantillons traités. Ces répétitions biologiques ou techniques permettent d'assoir les résultats avec une validation par des outils statistiques.

Pour suivre la production de COV de l'expression de gènes à l'émission des COV, j'ai été conduit à utiliser différentes techniques et outils d'analyse (Chromatographie, PCR, qPCR, séquençage...). Certains outils intègrent déjà les logiciels nécessaires à l'analyse des données (*e.g.* qPCR Applied Biosystems avec le logiciel SDS 7200, Agilent Chemsation dataAnalysis pour la GC-MS), mais d'autres nécessitent l'utilisation de logiciels externes. Dans cette partie je vous présente les méthodes d'analyse employées, les logiciels utilisés et également les outils de bioinformatique que j'ai créé afin de répondre aux besoins spécifiques de mes recherches (développés plus en détails ci-dessous dans la section Résultats).

III-2 Analyses chromatographiques

III-2.1 Identification des composés volatils

L'identification des CTV a été faite par GC-MS après une double séparation des échantillon sur une colonne polaire (HP-Innowax) et une colonne apolaire (DB-5). Le nom d'une molécule était trouvé en comparant les spectres de masse obtenus aux bases de données Wiley, NIST 05 et Adams (Adams 2007). Ces noms étaient ensuite vérifiés grâce aux indices de Kovats des COV calculés sur la colonne apolaire par comparaison avec ceux de n-alkanes (C8-C20, Fluka). Lorsque nous disposions des COV sous forme de standards commerciaux purs nous avons vérifié que le temps de rétention du composé pur correspondait bien au temps de rétention du composé à identifier (Annexe III : Liste des standards utilisés en GC). Le module MSeasy qui a été développé lors de cette thèse a été utilisé comme méthode de prétraitement du jeu de données de lavandes botaniques (600 analyses GC-MS, Figure 31) ce module a accéléré et facilité l'identification des CTV (disponible dans l'Annexe VIII : Annexes numériques).

Matériel et Méthodes

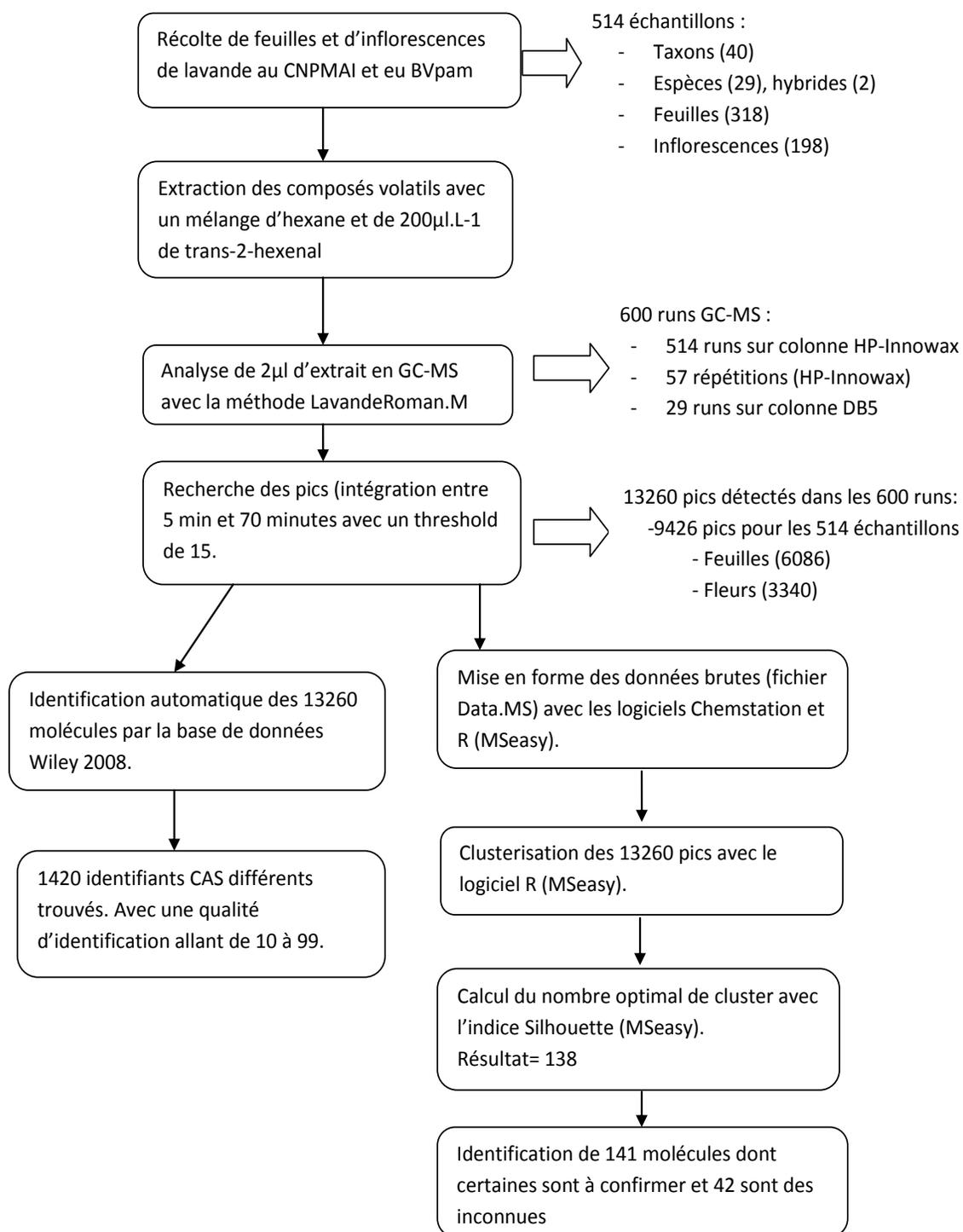


Figure 31 : Détail du nombre d'analyses GC-MS réalisées et de leur traitement par identification automatique ou par MSeasy

III-2.2 Quantification des composés volatils

La concentration de chacun des composés volatils, pour lesquels nous disposions de standards commerciaux, était calculée en comparant l'aire du pic du composé avec l'équation de la droite de régression obtenue avec différentes concentrations du composé commercial pur. Afin de tenir compte d'une éventuelle évaporation du solvant hexane, l'aire du pic était corrigée par rapport à l'aire de l'étalon interne avant le calcul de concentration.

III-2.3 Alignement des chromatogrammes

Pour mon étude sur l'accumulation des terpènes dans les feuilles et les fleurs de *L. angustifolia* et *L. Xintermedia*, je disposais d'une large bibliographie décrivant le contenu des huiles essentielles de ces espèces et je connaissais donc les molécules que j'allais rencontrer dans les chromatogrammes. Dans ce cas, il était possible de comparer les échantillons sur la base des RT des molécules. Le logiciel dataanalysis installé sur notre GC-FID par le constructeur, Agilent, ne permet pas d'aligner les milliers de chromatogrammes obtenus. J'ai donc utilisé des macros Excels développées pendant ma thèse pour permettre de reconnaître les pics de molécules spécifiques sur la base de leur RT et d'en regrouper les aires dans un tableau Excel (voir ci-dessous dans la section Résultats pour les détails sur ces macros).

Pour l'étude sur la diversité des CTV dans le genre *Lavandula*, nous nous sommes affranchi de cette étape d'alignement pour les données de GC-MS grâce au logiciel MSeasy développé sous R (voir ci-dessous dans la section Résultats pour des détails sur ce logiciel).

III-3 Analyse des PCR quantitatives

III-3.1 Choix des amorces

Pour chaque couple d'amorce (Tableau 8), l'efficacité de la PCR (E) était calculée en réalisant une gamme de dilution d'un ADNc témoin (*i.e.* issu d'un mélange d'ARN de différentes inflorescences et de différentes dates). La formule est la suivante : $E = 10^{(-1/\text{pente de la gamme étalon})}$ (Mygind, Birkelund *et al.* 2002). Seules les amorces avec une efficacité comprise entre 1.85 et 2.15 et un coefficient de détermination (R^2) de la gamme étalon supérieur ou égal à 0.98 ont été retenues pour la suite de l'étude.

III-3.2 Ratio d'expression

Les évolutions dans le temps de l'expression des gènes de la limonène synthase et de la linalol synthase ont été suivies en rapportant les valeurs obtenues à celles de la première date en utilisant la méthode du *delta-delta* du cycle threshold ($\Delta\Delta Ct$) (Livak et Schmittgen 2001). Cette méthode demande l'utilisation d'un gène de référence. Pour cela, nous avons isolé au sein du laboratoire trois gènes de *L. angustifolia* codant pour l'actine, la tubuline et le facteur d'élongation alpha (dbEST_Id 64810781, 64810782 et 64810783 respectivement). Pour nos calculs, seule l'actine a été conservée comme gène de référence car ce gène est celui dont l'expression fut la plus stable au fil du temps de développement des inflorescences de lavandes.

III-4 Analyses phylogénétiques

III-4.1 Alignement des séquences nucléotidiques

Les séquences d'ADN obtenues ont été alignées par l'algorithme ClustalW du logiciel BioEdit version 7.0.5.3. Paramétrages : options full multiple alignment et bootstrap NJ tree = 1000, et paramètre par défaut pour les Gap penalties (*i.e.* vide) (Hall 1999). Puis, pour affiner l'alignement des séquences entre elles, quelques corrections manuelles ont été apportées en utilisant la fonction Edit du logiciel BioEdit. Les nucléotides pour lesquels les chromatogrammes ne donnaient pas clairement d'identification ont été recodés par la lettre N. L'alignement de séquences a été sauvegardé au format fasta pour pouvoir être transféré vers le logiciel de phylogénie MEGA version 4 (Tamura, Dudley *et al.* 2007).

III-4.2 Calculs et visualisation des arbres de phylogénie

L'analyse phylogénétique de nos données a été faite avec le logiciel MEGA. Celui-ci permet de faire les alignements de séquences tout comme BioEdit, mais il était plus facile de visualiser les chromatogrammes des séquences et donc de corriger les alignements avec BioEdit. Par contre seul MEGA permet d'analyser les séquences en termes d'information phylogénétique. Sur les 1681 paires de bases retenues, 184 étaient variables et 89 étaient des sites informatifs. Le taux de transitions/transversions était de 1.4. Le test du χ^2 réalisé sur les proportions des différentes bases dans les séquences a montré qu'il y a proportionnellement plus de T et de A (33%) que de C et de G (17%) (ddl=93, p-value<0.05). Pour la construction d'un arbre phylogénétique nous avons fait le choix de travailler avec le

logiciel Phylip version 3.68 (Felsenstein 1995). En utilisant la fonction export data de MEGA, on génère un fichier au format phylip (option site per line = 1681 et missing data and alignement gap = include site with missing/ambiguous data and gap). Le fichier obtenu n'est pas exactement au bon format et il faut supprimer grâce à un logiciel d'édition de texte (type bloc-note de Windows® ou notepad++®) la ligne vide entre le nombre d'espèces et de paires de base et la première espèce puis supprimer les espaces entre les nucléotides. Pensez également à faire attention d'avoir 10 caractères de titre pour chaque lignes. Si besoin, ajouter des espaces pour être au bon format pour Phylip.

Les informations données sur les séquences par MEGA ont été utilisées pour paramétrer le logiciel Phylip module dnadist du programme Phylip (Phylip::dnadist). Nous avons choisi d'utiliser la correction de Felsenstein (ou maximum likelihood) (option D valeur F84) pour le calcul de la distance entre les taxons. L'avantage de cette correction est qu'elle tient compte des différences de compositions en nucléotide et également du taux de transitions/transversions (option T valeur 1.4). Le pourcentage de A, T, C et G est entré avec l'option F (entrez ces valeurs dans l'ordre suivant : 0.33 0.17 0.17 0.33 pour un pourcentage de A et T de 33% et de C et G de 17%).

La séquence d'analyse est la suivante : 1/ Phylip::seqboot avec 1000 bootstrap et une graine placée sur l'espèce 5. Ceci génère un jeu de données qui permet de calculer les valeurs de bootstrap. 2/ Puis Phylip::Dnadist avec la correction de Felsenstein (F84, T=1.4). 3/ Puis selon l'algorithme d'agglomération choisi soit Phylip::neighbor (NJ ou UPGMA), soit Phylip::dnaML (maximum de vraisemblance) ou Phylip::dnaPARS (parcimonie). 4/ Pour finir Phylip::consensus qui génère l'arbre consensus et le sauvegarde dans un fichier dont l'extention est « .phy ». Il est visualisé avec le programme de visualisation d'arbre de phylogénie ATV (<http://www.phyloXML.org> (Leebens-Mack, Vision *et al.* 2006).

III-4.3 Superposition des données de chimie et de phylogénie moléculaire

Pour superposer les données de chimie (liste des composés présents dans chaque espèce) sur l'arbre de phylogénie du genre, j'ai utilisé le logiciel TNT (Goloboff, A. Catalano *et al.* 2009). La démarche est la suivante : création manuelle d'un fichier au format TNT contenant la matrice de composés chimiques (une ligne par espèce avec autant de 0 et de 1 que de molécules absentes ou présentes) puis les information de l'arbre de phylogénie, là encore, au bon format (pour plus de détails sur le format TNT voir l'aide du logiciel). Une fois ce travail

effectué, il suffit d'ouvrir le fichier ainsi créé via le menu « File → Open Input File », puis de lire les données superposées via le menu « Optimize → Character → Character mapping ». On peut ensuite choisir de visualiser un ou plusieurs caractères (molécule) sur l'arbre de phylogénie (la touche Entrée permettant de passer d'une molécule à la suivante).

III-5 Analyses statistiques

Les analyses statistiques ont été réalisées avec les logiciels R (R Development Core Team 2007) et XLStat (Version 2008.a Fahmy 2008) et Genalex (Version 6 Peakall et Smouse 2006).

La normalité et l'homoscédasticité des données ont été testées par tests de Shapiro-Wilk et Fligner, respectivement. Afin de normaliser les données qui ne l'étaient pas, nous avons procédé à des transformations via la fonction Box cox power du logiciel R. Lorsque l'homoscédasticité et la normalité des données étaient vérifiées, des tests paramétriques (*e.g.* ANOVA) ont pu être utilisés. Lorsqu'une ANOVA indiquait qu'il y avait une différence significative entre les données, on réalisait un test de comparaisons multiples a posteriori : le test de Tukey. Lorsque les transformations ne permettaient pas de normaliser les données nous avons utilisé des tests non-paramétriques : le test de rang de Kruskal et Wallis pour les comparaisons multiples et le test de Mann et Withney pour les comparaisons d'échantillons 2 à 2.

Des analyses en composante principale (ACP) ont été faites. Pour ces ACP, les variables qualitatives étaient transformées en variables quantitatives. Les matrices de données étaient ensuite centrées et normées.

Pour analyser des corrélations 2 à 2, nous avons également utilisé les cartes de corrélations (*heatmap* en anglais), les corrélations étant calculées avec le coefficient de Spearman et la significativité des corrélations étant testée avec la correction de Bonferroni pour les comparaisons multiples. L'ordre des variables dans la carte était calculé par l'algorithme d'agglomération de Ward sur la base de la distance euclidienne entre les coefficients de corrélations.

Les analyses en coordonnées principales (ACoP) ont été réalisées sur des matrices présence/absence de COV avec le logiciel Genalex ainsi que l'analyse de variance moléculaire (AMOVA, distance de Huff).

RESULTATS

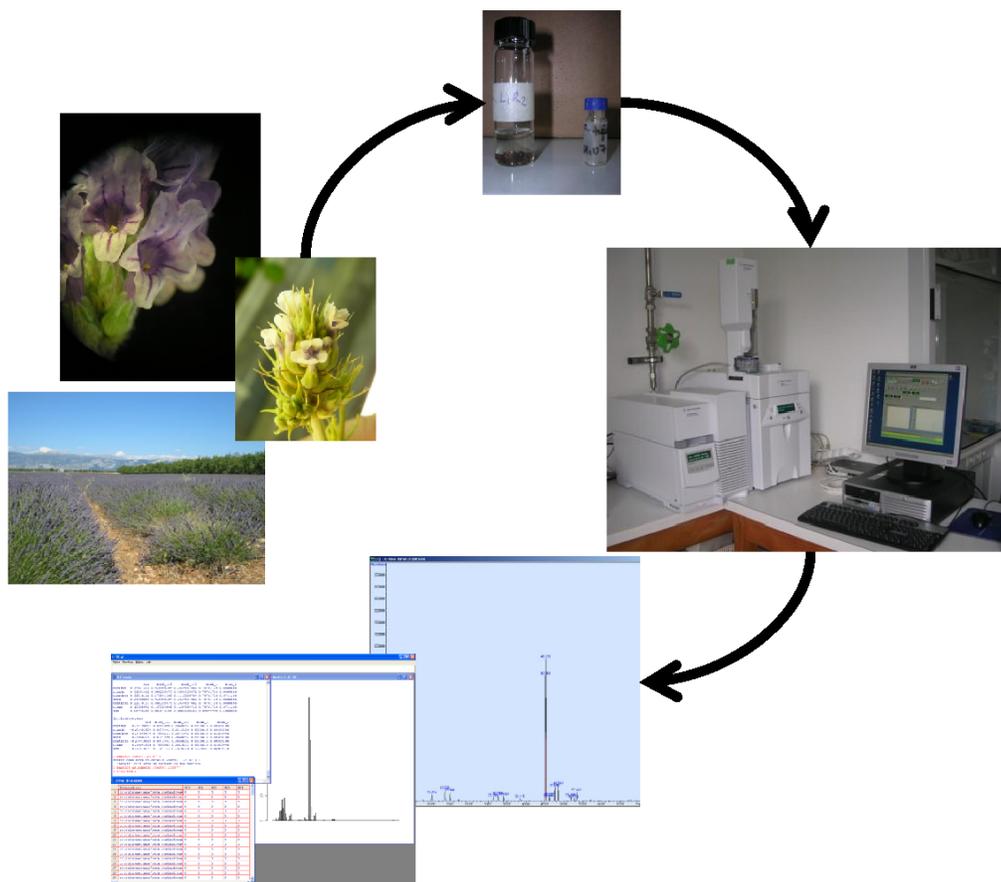
I - Création d'un logiciel de traitement des données de GC-MS (Module MSeasy pour R)

II- Evolution des composés organiques volatils accumulés par les espèces du genre *Lavandula*

III- Etude des variations spatio-temporelles de l'accumulation de terpènes chez *L. angustifolia* et *L. x intermedia*

Résultats

Création d'un logiciel de traitement des données de GC-MS (Module MSeasy pour R)



Les étapes de l'analyse des COV en GC-MS

I Création d'un logiciel de traitement des données de GC-MS (Module MSeasy pour R)

I-1 Introduction

L'analyse en chromatographie en phase gazeuse (GC) d'un grand nombre d'échantillons biologiques d'origines variées à des fins de comparaison se heurte à certaines limites. Le problème de l'alignement des chromatogrammes et de la détection des pics se posent (De Souza, Saunders *et al.* 2006; Luedemann, Strassburg *et al.* 2008). En effet, en GC-MS, GC-FID ou encore en LC-MS chaque molécule ou mélange de molécules identiques est censé être élué au même temps de rétention (RT) si le chromatogramme est obtenu sur une même colonne et dans les mêmes conditions d'analyse. La simple superposition de deux chromatogrammes devrait donc permettre de lister les molécules communes aux deux chromatogrammes. Malheureusement, ceci n'est vrai que dans une certaine mesure (Duran, Yang *et al.* 2003). De nombreux paramètres techniques peuvent, en effet, modifier de manière plus ou moins importante le RT d'une molécule entre deux analyses effectuées sur un même appareil, dans les mêmes conditions d'analyses : encrassement de la colonne, modifications de la matrice de la colonne au fil du temps, variations infimes et aléatoires des débits de gaz vecteurs ou des températures du four, maintenances sur l'appareil.... (Duran, Yang *et al.* 2003; Issaq, Van *et al.* 2009). De même, la nature biologique des échantillons peut aussi entraîner des décalages de RT. En effet, nous avons pu observer que des espèces, même proche phylogénétiquement, présentent une variation de RT pour un même composé dans des conditions d'analyse identique.

Ainsi, une étape clef de l'interprétation des données brutes de GC-MS (et autres méthodes couplées à une masse) est donc l'alignement des pics des différents chromatogrammes ayant un RT identique, ou proche (Duran, Yang *et al.* 2003). Cette étape de prétraitement (« preprocessing ») permet l'obtention d'une liste de pics d'où découle un fichier final avec autant de lignes que d'échantillons et autant de colonnes que de pics ayant des RT différents (Figure 32 d'après De Souza, Saunders *et al.* 2006; Tohge et Fernie 2009).

Résultats

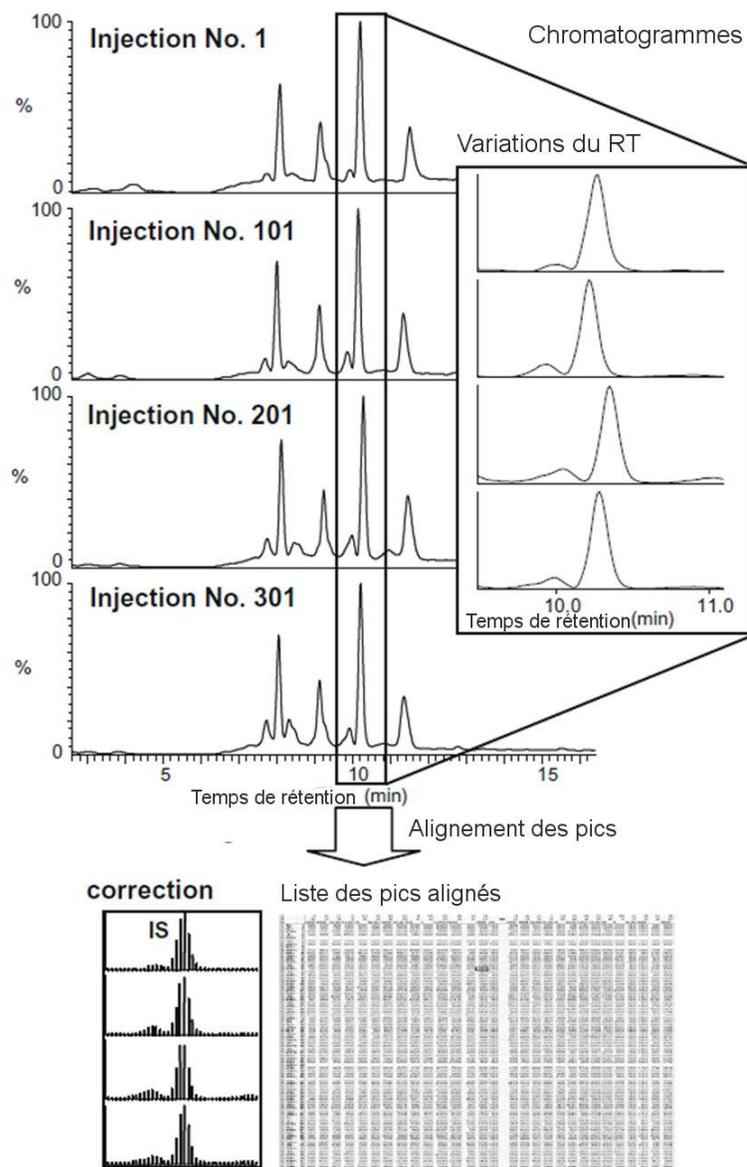


Figure 32 : Alignement de chromatogrammes sur la base du RT (modifié d'après Tohge et Fernie 2009)

Quand le nombre d'échantillons analysés est important, l'alignement manuel devient trop fastidieux voire impossible et des outils automatisés sont nécessaires. Plusieurs types d'outils ont été développés pour permettre l'alignement des chromatogrammes (Cuadros-Inostroza, Caldana *et al.* 2009; De Vos, Moco *et al.* 2007; Duran, Yang *et al.* 2003; Lisec, Schauer *et al.* 2006; Luedemann, Strassburg *et al.* 2008). Basés sur des algorithmes complexes, ils permettent généralement de regrouper les pics en se basant sur leur temps de rétention et/ou sur un indice de rétention (RI). Ces indices sont l'indice de Kovats (KI) ou l'indice arithmétique (AI) et sont obtenus en calculant la distance du RT d'un pic par rapport aux RT de pics de références (des n-alcane pour les analyses GC) (pour le calcul voir Adams 2007;

Résultats

Van den Dool et Kratz 1963). La procédure suivie par ces logiciels est généralement proche de celle-ci :

1 – récupération des données brutes de l'appareil de GC-MS, LC-MS au format constructeur ou dans un format international d'échange de données

2 – définition par l'utilisateur d'une plage acceptable de variation du temps de rétention d'une substance pour regrouper les pics de chromatogrammes différents situés dans cette même plage de RT.

3 – vérification des problèmes de collision entre pics (deux pics d'un même extrait classés dans la même colonne), ou bien création intempestive de deux colonnes dans la matrice finale pour des pics correspondants à une même molécule mais ayant des RT plus éloignés que prévu, ou encore d'échantillons aberrants présentant une plus grande variation sur ses RT que la moyenne des autres.

4 – Création d'une matrice finale pouvant être utilisée par d'autres logiciels (« postprocessing »). Cette matrice est en général un tableau dont les lignes sont les échantillons ou les individus testés, et les colonnes sont les molécules détectées.

Cette procédure constitue le prétraitement des données (avant l'établissement de la liste de pics) et peut être précédée d'une étape de nettoyage des chromatogrammes (soustraction du bruit de fond) et être suivie d'une étape d'intégration (estimation de la concentration des molécules). Le plus souvent, des logiciels fournis par le constructeur de l'appareillage effectuent le nettoyage et l'intégration ; Mais des outils développés par la communauté scientifique en métabolomique permettent aussi d'effectuer ces étapes (De Vos, Moco *et al.* 2007; Duran, Yang *et al.* 2003).

Ces différents outils de prétraitement basé sur le RT/RI peuvent cependant présenter des biais expérimentaux (De Souza, Saunders *et al.* 2006). En effet, rien n'empêche qu'un pic d'un tissu X ait le même RT qu'un pic d'un tissu Y et que les molécules associées respectivement à ces pics soient différentes chimiquement. Ce biais peut être limité si on travaille avec les indices de rétention (KI, AI) (Luedemann, Strassburg *et al.* 2008). Mais là encore, on trouve des molécules ayant des RI identiques mais des structures chimiques différentes : par exemple le *beta*-bisabolol (KI= 1675, AI 1674) et l'helifolenol A, ou encore l'*alpha*-Z-santalol, le valéranone, l'hydroxydiphényl-2-méthane (Adams 2007). De plus, les algorithmes complexes utilisés pour l'alignement des chromatogrammes conduisent souvent l'utilisateur à devoir

Résultats

déterminer des paramètres auxquels il n'a pas toujours accès ou qui ne sont pas intelligibles. L'utilisateur se retrouve alors confronté à une boîte noire sur laquelle il n'a pas le contrôle pour l'interprétation de ses données brutes.

Le couplage de la chromatographie avec un spectrophotomètre de masse permet d'ajouter à l'information du RT/RI, une information liée à la structure chimique de la molécule analysée. Cette information est constituée par le spectre de masse de la molécule. Selon la technique, ce spectre de masse est plus ou moins complexe et peut se limiter à un seul pic parent dans le cas d'une ionisation douce (souvent utilisée en LC-MS) ou bien contenir une multitude de pics issus de la fragmentation de la molécule comme, par exemple, après un bombardement électronique (souvent utilisé en GC-MS) (Peters, van Velzen *et al.* 2009). Sur la base de ces spectres de masse et du RT/RI il est possible d'identifier une molécule en utilisant des bases de données informatiques couplées à des algorithmes de comparaison (Adams 2007). Le travail d'alignement des chromatogrammes n'est ici plus nécessaire. La Figure 33 présente les différentes étapes d'une identification automatisée d'un composé par son spectre de masse via un logiciel de comparaison. En GC-MS, chaque pic représente une molécule ou un mélange de molécule en cas de co-élution. Pour chaque pic, l'appareil enregistre un certain nombre de spectres de masse par unité de temps, appelés scans (scans qui apparaissent au fur et à mesure de l'élution des molécules du chromatogramme, représentés par des fragments ionisés de masse mesurés par un rapport masse/charge (m/z) sur une période de temps). Le nombre de scans acquis par pic dépend de la fréquence de balayage du détecteur de masse et de la largeur du pic (durée de sortie d'un pic de la colonne chromatographique). Ces scans peuvent ensuite être utilisés pour l'identification. Pour comparer le spectre de masse du pic avec les spectres des bases de données, les logiciels utilisent soit le spectre de masse du scan de l'apex du pic (option par défaut sur de nombreux logiciels) soit un spectre de masse moyen issu de la moyenne de tous les scans du pic ou d'une partie des scans autour de l'apex.

Les logiciels d'identification automatisée proposent obligatoirement des noms de molécules, même si la molécule n'est pas dans la base de données. Ces noms de molécules correspondent aux spectres les plus proches du spectre du pic analysé. Un coefficient de qualité de l'identification (*match factor*) est associé à chaque proposition. Plus ce coefficient est élevé plus le spectre de masse du pic est proche de celui contenu dans la base de données.

Résultats

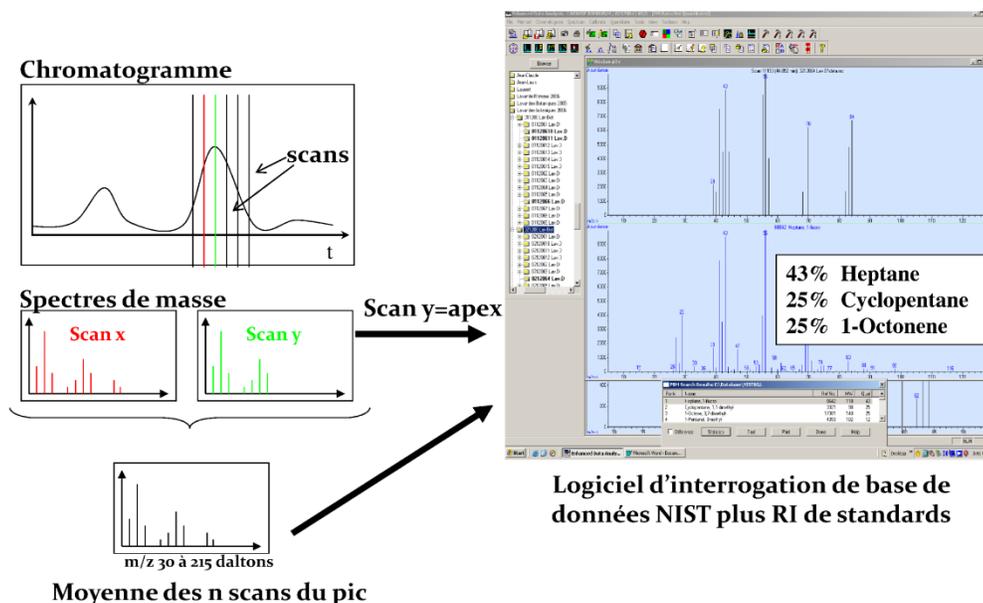


Figure 33 : Schématisation d'une identification automatique d'un composé par son spectre de masse via une base de données informatique.

En vert : Le scan y correspond au spectre de masse de l'apex du pic. En noir : le spectre de masse est le résultat de la moyenne des spectres mesurés durant la durée de sortie du pic de la colonne de chromatographie. Les 3 pourcentages indiqués sur la fenêtre du logiciel NIST correspondent aux *match factors* des 3 molécules de la base de données dont le spectre de masse est le plus proche de celui du scan y.

Cette identification automatisée est rapide et dans une certaine mesure fiable. En effet, la seule identification par le spectre de masse n'est pas suffisante, il faut pour plus de certitude vérifier que le RI est bien celui attendu pour cette molécule et si possible vérifier qu'un standard chimique pur de cette molécule analysée dans les mêmes conditions sur le même appareil a bien le même RI et le même spectre de masse (Tholl, Boland *et al.* 2006).

Au démarrage de ma thèse en novembre 2006, seuls quelques outils qui ajoutaient l'information du spectre de masse au RT pour faire l'alignement étaient disponibles. Le premier était XCMS (Smith, Want *et al.* 2006). Cet outil est un module pour le logiciel de statistique libre R (R Development Core Team 2007). Il permet un alignement de très bonne qualité mais ne peut toujours pas distinguer des pics ayant le même RT mais issus de molécules différentes. Le second était le logiciel développé par Murphy *et al* pour un appareil de spectrométrie de masse laser (Murphy, Middlebrook *et al.* 2003). Ce logiciel peut différencier des pics ayant le même RT mais présentant des spectres différents car il utilise le spectre de masse complet pour comparer les pics entre eux. Ainsi, le spectre de masse du premier pic du premier chromatogramme est comparé à l'ensemble des pics des autres chromatogrammes quelque soit son RT. La comparaison se fait par classification hiérarchique ascendante sur la base d'une matrice de distance entre spectres de masse obtenue par un calcul

Résultats

d'un « dot product » qui selon les auteurs est proche d'une distance euclidienne. Une fois la matrice de distance calculée, les deux spectres ayant la plus grande similarité (*i.e.* le plus fort « dot product ») sont moyennés pour donner un nouveau spectre de masse représentatif du groupe. Puis la matrice de distance entre les spectres de masse des pics restants et celui du groupe créé est recalculée et ainsi de suite tant que les « dot products » sont supérieurs à un seuil fixé par l'utilisateur. Au final, l'utilisateur obtient une liste de pics avec pour chacun le numéro du cluster auquel il appartient. Chaque cluster étant lié à un spectre de masse moyen issu du regroupement (*clustering*), on peut même, sous certaines conditions, identifier la molécule. Ce logiciel est donc intéressant surtout qu'il peut traiter jusqu'à 4000 pics en même temps. Mais, il a été développé pour un appareillage particulier et dans un langage informatique complexe. De plus, l'utilisateur doit faire arbitrairement le choix d'une valeur pour le seuil d'arrêt du *clustering*. Or, ce seuil, qui ne représente pas une valeur intelligible, impacte grandement le nombre de groupes formés (Marcolli, Canagaratna *et al.* 2006). Aucun indice de qualité ne permet de valider le nombre de groupes formés, qui peut alors ne pas correspondre du tout au nombre de molécules différentes réellement présentes dans les échantillons analysés. Un seuil haut, impliquant une forte identité entre les spectres de masse, pourra générer plus de groupes que de molécules. Inversement, un seuil bas risque de regrouper des molécules différentes dans un même groupe. Le dernier outil qui était disponible était celui développé pour MATLAB© par Dixon (Dixon, Brereton *et al.* 2006). Cet outil à une approche semblable à celle de Murphy (Murphy, Middlebrook *et al.* 2003).

Quelque soit l'approche, le travail d'identification manuel des pics demeure donc une étape obligatoire. Marcolli *et al* soulignent que les outils de bioinformatique peuvent aider à regrouper les molécules semblables mais qu'il sera toujours nécessaire de faire intervenir un expert en chimie pour avoir une identification certaine des molécules (Marcolli, Canagaratna *et al.* 2006).

Un des objectifs principaux de mon travail de thèse était d'analyser sans *a priori* 500 échantillons de feuilles et fleurs de différentes espèces botaniques du genre *Lavandula* pour caractériser la diversité des composés terpéniques volatiles (CTV) dans ce genre (voir chapitre de résultats suivant). Ce jeu de données analysées en GC-MS comportait quelques 50 000 pics détectables. La première étape du travail consistait donc à créer une matrice en présence/absence (« fingerprinting matrix » codée en 1/0) de la composition en COV de ces lavandes. Il fallait pour cela rechercher dans les 50 000 pics les molécules identiques (ou

Résultats

mêmes mélanges de molécules). La comparaison manuelle de ces pics permet l'identification mais nécessite beaucoup d'efforts et de temps. Dixon et al. (Dixon, Brereton *et al.* 2006) ont estimé que pour aligner et intégrer les pics d'un chromatogramme « à la main » il faut environ 5 minutes par pic (vérification du spectre de masse, du temps de rétention, correction de la ligne de base...). Ce traitement « à la main » effectué sur mon jeu de données de lavandes botaniques représenterait 4167 heures de travail, soit 104 semaines de travail à temps plein ou encore deux ans. L'automatisation de cette comparaison est dès lors indispensable. Les outils de prétraitement existants n'étant pas adaptés à nos objectifs expérimentaux, nous avons développé un outil de prétraitement adapté à nos besoins en utilisant le programme de statistique libre R (R Development Core Team 2007). Cette méthode est basée dans un premier temps sur les spectres de masse et dans un second temps, sur le RT/RI. L'ensemble des spectres de masse obtenus pour un jeu de données est analysé avec des méthodes de classification non supervisées. La validité de la classification est évaluée avec des indices de qualité et les spectres similaires sont regroupés et correspondent potentiellement à une molécule. La comparaison du spectre de masse des pics d'un même cluster avec des molécules déjà décrites est grandement facilitée. Il est possible d'intégrer des standards et/ou une bibliothèque aux analyses des échantillons, de manière à faciliter l'identification des molécules. Pour calibrer et évaluer notre outil nous avons travaillé sur trois jeux de données obtenus avec différentes méthodes, sur différents organismes. Le développement du module MSeasy est issu d'un travail collaboratif entre Florence Nicolè (MCF LBVpam), Elodie courtois (Courtois 2010) et moi-même. J'ai participé activement à la programmation du module sous R et j'ai développé l'interface conviviale MSeasytkGUI. Jérôme Mardon (Mardon 2010) et Juliana Ivanisevic (Ivanisevic) nous ont fourni des jeux de données pour tester la robustesse de notre méthode. Martine Hossaert McKey a suivi et soutenu ce projet par l'intermédiaire du GDR d'écologie chimique. Par la suite, je présente le module MSeasy (voir aussi en Annexe IV : Poster MSeasy présenté lors du congrès Metabolomic 2010 Amsterdam 27-Juin/1 Juillet 2010), mais aussi, dans un premier temps, les outils initiaux de compilation de données que j'ai développés pour le laboratoire.

I-2 Résultats : Les outils de bioinformatique créés

I-2.1 Les outils de compilation de données de GC-FID et GC-MS

Les macros Excel

A mon arrivée au laboratoire BVpam, j'ai développé des outils de bioinformatique avec le logiciel Excel de Windows® pour faciliter le traitement de nombreux échantillons. Le tableur Excel permet en effet de programmer des macros, qui sont des programmes automatisant le traitement de données. J'ai développé plusieurs macros qui permettent de faire du prétraitement convivial de données de GC-FID et de GC-MS issues du logiciel Chemstation d'Agilent. Je ne présente ici que les principales macros assemblées en un seul classeur Excel « MacroGC.xls » disponible dans l' Annexe VIII : Annexes numériques et sur demande. Le code des macros est accessible en ouvrant le classeur « MacroGC.xls » puis en ouvrant le volet Visual Basic (le mot de passe est : admin). Ces macros ont été développées pour fonctionner sur les versions 1997 et 2002 d'Excel. Avec le classeur « MacroGC.xls », il est, par exemple, possible de compiler les données brutes (Format de sortie du logiciel Chemstation d'Agilent) et de les transférer dans des classeurs Excel. Une fois les données compilées au format Excel, les macros permettent également de mettre en forme les données puis de les comparer (mise des informations nécessaires en colonne par exemple : échantillon/RT/aire du pic/spectre de masse). Les différentes possibilités offertes par ces macros sont rapidement décrites dans l'organigramme de la figure 35.

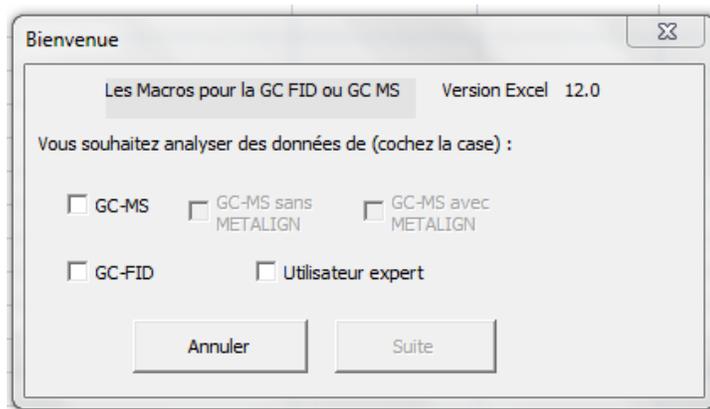


Figure 34 : Boîte de dialogue issue des macros Excel

Pour faire fonctionner les macros aisément, j'ai développé une interface graphique. La navigation entre les différentes options et l'accès aux macros, s'effectue en répondant aux

Résultats

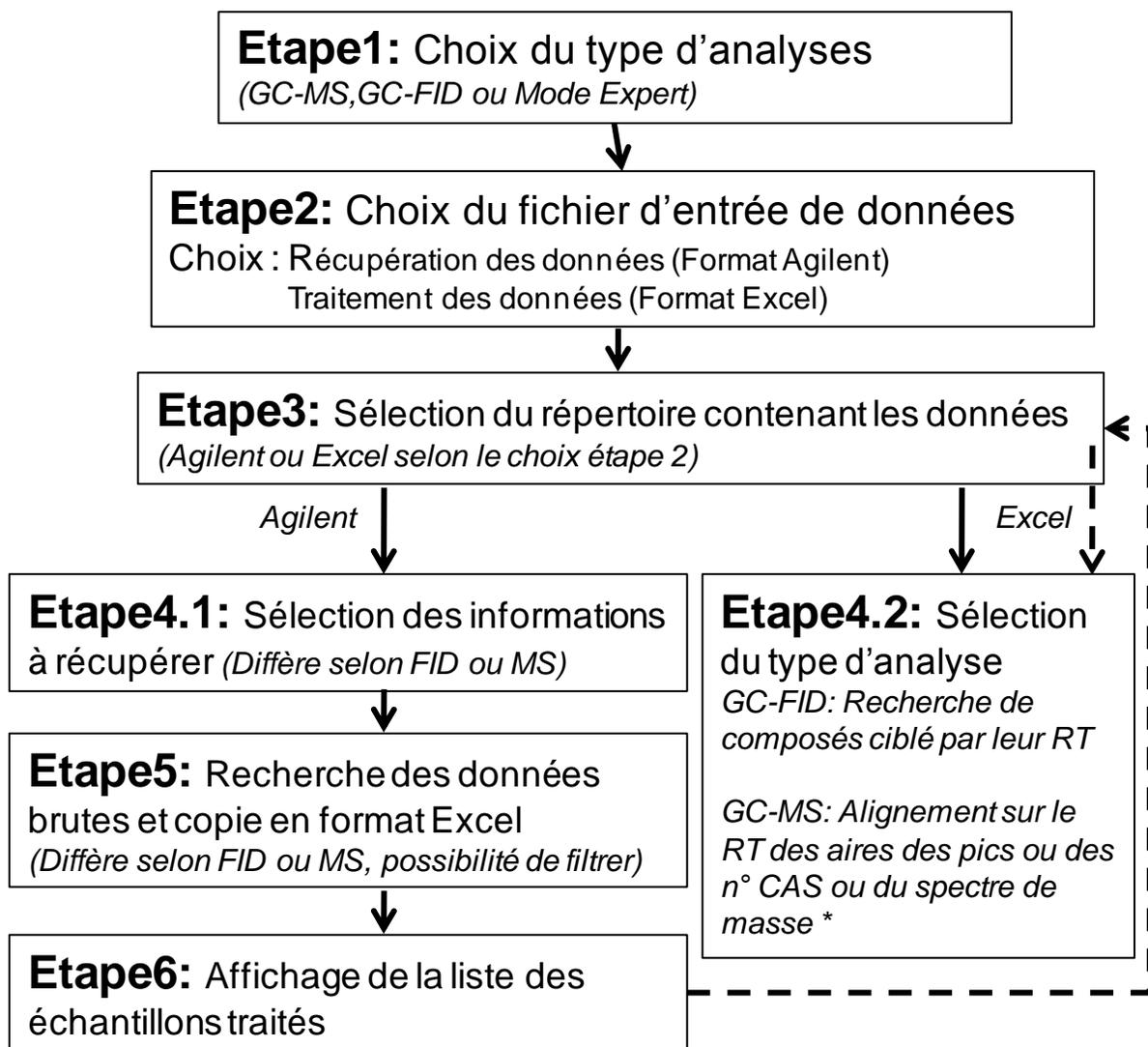


Figure 35 : Organigramme des étapes des macros Excel

Etape1 : choix entre GC-FID ou GC-MS selon le type de machine utilisée, si la réponse est GC-MS il est demandé si les données ont été traitées pour le bruit de fond par le logiciel Metalign(RIKILT/Plant-Research-International 2001). Il est également possible de passer en Mode Expert (*i.e.* sans boîtes de dialogues).

Etape2 : choix entre la récupération de données brutes (GC-MS ou GC-FID) ou bien le traitement de données déjà au format Excel (fichiers préalablement générés via les étapes 5 et 6, ou par d'autres moyens).

Etape3 : Sélection du répertoire dans lequel sont les données à travailler (données brutes sous Agilent (option par défaut) ou Excel selon le choix étape2).

Etape4.1 : choix des informations à récupérer des données brutes. Par défaut : Nom échantillon + RT + Aire de pic. Auxquelles s'ajoute si l'on souhaite : le nom de la molécule (s'il y a une identification automatique par la base de données NIST ou Wiley), le spectre de masse de chaque pic (il faut avoir créé un fichier export3D avec le logiciel Agilent Chemstation après analyse GC-MS).

Etape4.2 : Choix du type d'analyse : alignement automatique sur la base de RT (seuil de séparation des RT fixé par l'utilisateur) ou sur l'identification NIST ou Wiley (le numéro CAS), calcul du spectre de masse représentant chaque pic (différents modes de calcul possibles : moyenne de tous les scans du pic, moyenne d'un pourcentage des scans autour de l'apex du pic, spectre de l'apex du pic uniquement).

Etape5 : Recherche des données brutes répondant aux critères d'un filtre (nom de méthode, date d'analyse...) dans le répertoire sélectionné à l'étape3 et création d'une copie des données au format Excel dans un répertoire différent.

Etape6 : Affichage de la liste des échantillons trouvés et traités.

Résultats

questions posées dans les boîtes de dialogue. La première fenêtre présentée dans la Figure 34 s'affiche dès l'ouverture du classeur « MacroGC.xls » par l'utilisateur. Elle permet à l'utilisateur de commencer la récupération de ses données en indiquant sur quelle machine il travaille (GC-MS ou GC-FID). Ensuite, selon le type de données et les analyses souhaitées, l'utilisateur est amené à répondre à d'autres questions, à copier ou coller quelques lignes ou remplir certaines cases du classeur. La/les matrice(s) obtenue(s) peuvent ensuite être analysées statistiquement ; les spectres de masse peuvent être dessinés pour être comparés aux bases de données. La macro d'alignement des pics sur la base des RT sous Excel (Etape 4. Figure 35) est un peu longue lorsque le nombre de pics est important (supérieurs à 500) et est limitée par le nombre de RT différents. En effet, selon la version d'Excel, le nombre de colonnes maximum ne peut dépasser 256. Un travail sur l'efficacité des algorithmes est envisageable mais il ne résoudrait pas les limitations inhérentes aux autres logiciels (Excel ici). Afin de palier à ces problèmes de temps de création et de nombre maximum de colonnes du fichier d'alignement sur le RT et améliorer les possibilités de prétraitement de grands jeux de données de GC, j'ai développé des programmes en langage Perl.

Les programmes en Perl

Pour palier à certaines limites des macros Excel, j'ai développé des programmes en langage Perl. Ce langage est communément utilisé en bioinformatique car il est adapté au traitement de fichiers de données volumineux. Le premier programme en Perl (RTalign.pl) permet de traiter très rapidement un grand nombre d'analyses, directement à partir des données brutes en sortie du logiciel Chemstation d'Agilent (paramètres par défaut), ou sous format Excel, comme issues des macros décrites précédemment. D'autres programmes permettent d'améliorer la compilation des données de GC-MS et le rapatriement des informations des pics (RT, Aire, Identification de molécule par la base de données Wiley) (matricecomplete.pl et ses 6 sous-programmes). Ces programmes sont disponibles dans l'Annexe VIII : Annexes numériques de cette thèse et sur demande. Le code est accessible en ouvrant le fichier (*.pl) dans un éditeur de texte (Notepad de Windows® ou autre).

Les programmes en langage Perl sont efficaces pour l'alignement de grands nombres de pics mais leur utilisation nécessite l'installation du programme Perl et implique de naviguer entre le classeur Excel (ou autre tableur) et le programme Perl. Leur utilisation requière donc un certain niveau d'expertise. Ma volonté étant d'obtenir un outil convivial et efficace permettant de réaliser l'ensemble des opérations de prétraitement sur un grand nombre

Résultats

d'échantillons (compiler, mettre en forme et comparer les données de GC-MS). Cet outil existe désormais et a été formalisé sous la forme d'un module R libre et gratuit ; MSeasy. Un module, package, paquet ou bibliothèque de programmes externes, est simplement un ensemble de programmes R qui complète et permet d'augmenter les fonctionnalités de R. Un package est généralement dévolu à des méthodes particulières ou à un domaine d'applications. Il est possible de construire un package, pour usage personnel ou pour être diffusé (via le CRAN par exemple). Ce module MSeasy permet un prétraitement des données de GC/LC-MS et est détaillé dans l'article ci-dessous.

I-2.2 Le module MSeasy pour R (article en préparation pour soumission)

TITLE: “MSeasy: an R package dedicated to ecologists for the pre-processing of GC/LC-MS data

ABSTRACT

Motivation: The democratization of metabolic analyses has extended the scope of metabolomic to ecology investigations. Chemical ecology interprets the variation and diversity of chemical signals of non-model organisms in the light of species interactions. Elucidating the biological information within such complex signals, using robust statistical analyses, require a large number of replicates and dedicated bioinformatic tools.

Results: To analyse large GC/LC-MS datasets of chemical compounds, we developed an unsupervised pre-processing method. The method detects individual compounds within complex mixtures, through the clustering of mass spectra. Retention time or retention index are used in a second time to control the quality of clusters. No profile correction, migration time alignment or normalization are needed. The method was robust to the use of different types of chromatographic support and to shifts in retention times, which are common for large and/or long-term analysis due to column ageing, contamination or replacement. We validated our method on two distinct biological datasets and showed that it compares favourably to other pre-processing methods. We found that the best clustering method that groups similar mass spectra into molecules was the hierarchical clustering analysis with Euclidean distance and Ward linkage. However, we implemented a function that allows identifying if other algorithms could be more appropriate for other datasets.

Availability and implementation: an R package “MSeasy” implementing our pre-processing method is freely available on demand. For non R users a Graphical interface, called MSeasyTkGUI, was created and is also freely available at <https://dossier.univ-st-etienne.fr/lbvpam/public/MSeasy%20R%20package/>.

1-INTRODUCTION

Unveiling large metabolic profiles of biological systems (i.e. metabolomics) has become a routine process thanks to recent advances in analytical chemistry. Also, computational tools have enabled an exhaustive examination of whole metabolomes rather than specific target compounds. Mass Spectrometry (MS) is a commonly used analytical technique that detects

Résultats

and identifies metabolites in biological samples. Among the many ionisation technologies used for mass analysis, those that provide enough excess energy to lead to the parent ion fragmentation (electron bombardment for example) are particularly meaningful for small molecule analysis because they yield a mass fragmentation spectrum that is more unique to each molecule than the mass of its unfragmentated form. Due to the complex nature of biological samples, analytical separation of metabolites is often performed before MS analysis to discriminate as many metabolites as possible. The two most common separation techniques hyphenated to MS are gas chromatography (GC) and liquid chromatography (LC). Once separated (by either GC or LC), the hundreds of molecules within a sample can be chemically identified using the specific and characteristic mass spectrum obtained from the MS.

Metabolic analyses aim to identify qualitative or quantitative variations of the constituents of biological samples, using either metabolite fingerprinting or extensive profiling. Metabolite fingerprinting is used to rapidly assess differences in biological samples through the presence/absence of molecules (Fiehn 2002). This fast technique is appropriate to identify discriminating regions of chromatograms for further analyses. Metabolite profiling, on the other hand, requires molecules to be identified and quantified. It usually targets a given class of compounds or a specific metabolic pathway.

Metabolic analyses have recently spread outside of the chemistry labs and they have extended their scope to evolutionary and ecological questions (Meinwald et Eisner 2008). Metabolomics have been applied to investigate the function of chemicals in natural systems (De Vos, Van Oosten *et al.* 2005). For instance, the variation and diversity in biological chemical signals is a key aspect of many interactions between living organisms: plant-animal interactions (Johnson, Agrawal *et al.* 2009), allelopathic plant-plant interactions (Heil et Silva Bueno 2007), communication in social insects (Vander Meer, Breed *et al.* 1998), animal behavior (Penn, Oberzaucher *et al.* 2007) and sexual selection (see the special issue on chemical ecology 2008); Harborne, 1997 #945; Harborne, 2001 #936}. Unlike fields, such as medicine, agronomy or environmental science, chemical ecology usually addresses non-model organisms and investigates complex mixtures of unknown compounds using a large sample size and multiple replicates within and among individuals.

For these reasons, a need has emerged for the development and standardization of bioinformatic methods to analyse this flow of data (Fukusaki et Kobayashi 2005; Lindon, Nicholson *et al.* 2005). A consequent literature on metabolic analyses deals with the pre-processing step, which aims to obtain the fingerprinting or profiling data matrix (molecule

Résultats

presence or abundance in each sample) from raw data for statistical analyses (Goodacre, et al, 2004; Lange, et al, 2008). This first step can be challenging when large number of GC/LC-MS runs have to be treated, and consists in extracting and transforming peak information from raw data into molecule information. This involves several steps: collecting raw data from all samples, correcting chromatograms baselines, smoothing peaks, aligning and normalizing chromatograms, peak deconvolution and finally identifying the molecule based on its retention time (or retention index) and on its mass spectrum (Gullberg, Jonsson *et al.* 2004; Lisec, Schauer *et al.* 2006). These steps are too lengthy and too prone to experimentator error when conducted manually on large dataset, complex mixtures or heterogeneous samples. Also, some commonly encountered problems would need to be corrected manually. The first problem, in particular for long term analysis, is the shift in migration times due to gradual column ageing, accumulation of contaminants on the chromatographic support or by events such as peak broadening, coelution of substances and distorted peak shapes. Column changes and parallel use of different chromatographic supports also complicate data pre-processing (Issaq 2009).

Several bioinformatics' tools have already been developed to address all, or part, of these pre-processing steps: Targetsearch, XCMS and its extension for structural characterization of molecules (XCMS²) are based on the open source R interface (Benton, Wong *et al.* 2008; Cuadros-Inostroza, Caldana *et al.* 2009; Smith, Want *et al.* 2006); Metalign, TaqFinder (via Metalign), MathDAMP, MSFACTs are independent free tools (Baran, Kochi *et al.* 2006; Lommen 2009; Luedemann, Strassburg *et al.* 2008; Tikunov, Lommen *et al.* 2005);(Duran, Yang *et al.* 2003). Other authors, have developed an algorithm under commercial proprietary software (Dixon, Brereton *et al.* 2006). Other software solutions provide more global packages combining both pre and post processing steps: meltDB, metaboanalyst web server, MZmime and additional unnamed processes (Jonsson, Gullberg *et al.* 2004; Katajamaa, Miettinen *et al.* 2006; Neuweger, Albaum *et al.* 2008; Steinfath, Groth *et al.* 2008; Xia, Psychogios *et al.* 2009). In addition, manufacturers' instruments provide commercial software solutions (Agilent's genespringMS, Water's MarkerLynx, AB Sciex's MarkerView).

Most of these software solutions, however, are based on the quality of chromatograms normalization, cleansing, and alignment to select peak at a given retention time or index. Consequently, the analyses of heterogeneous samples cannot be easily streamlined e.g. when migration time shifts are important, or if several molecules share a common migration time in distinct analyses, and may potentially lead to serious mistake in data analysis. To save the

implementation of these steps, some authors proposed methods where the construction of a peak list is not needed (Bamba et Fukusaki 2006; Hansen, Andersen *et al.* 2005; Jonsson, Gullberg *et al.* 2004). Chromatogram patterns or full spectral profiles are directly analysed without deconvolution and peak identification. However, these approaches cannot resolve the specific compounds responsible for differences between samples and the achievement of a peak list becomes unavoidable.

Here, we developpe a simple method to infer a comprehensive peak list from MS analyses without prior migration time alignment. It is based on the clustering of mass spectra rather than retention times or indices. Retention times or indices are used secondarily to control the quality of the clusters. A very similar approach was successfully used for the analysis of air and water samples (Bernard et al 1986, Murphy et al 2003). This method is insensitive to shift in migration times, to the use of different columns, is exhaustive and can handle large datasets. Subsequent identification of molecules identity can then be conducted on molecules of interest by conventional techniques (RI, MS comparison with reference databases and commercial standards). Adding commercial standards or specific libraries to the datasets prior to the clustering could greatly facilitate the identification of molecules. Our method should help chemical ecologists to interpret large datasets by mining the essential information and checking for data quality. The method was implemented in an open source R package called MSeasy (R Development Core Team 2007). For non R users a user-friendly interface was also developed in a package called MSeasyTkGUI. Both packages are freely available at <https://dossier.univ-st-etienne.fr/lbvpam/public/MSeasy%20R%20package/>

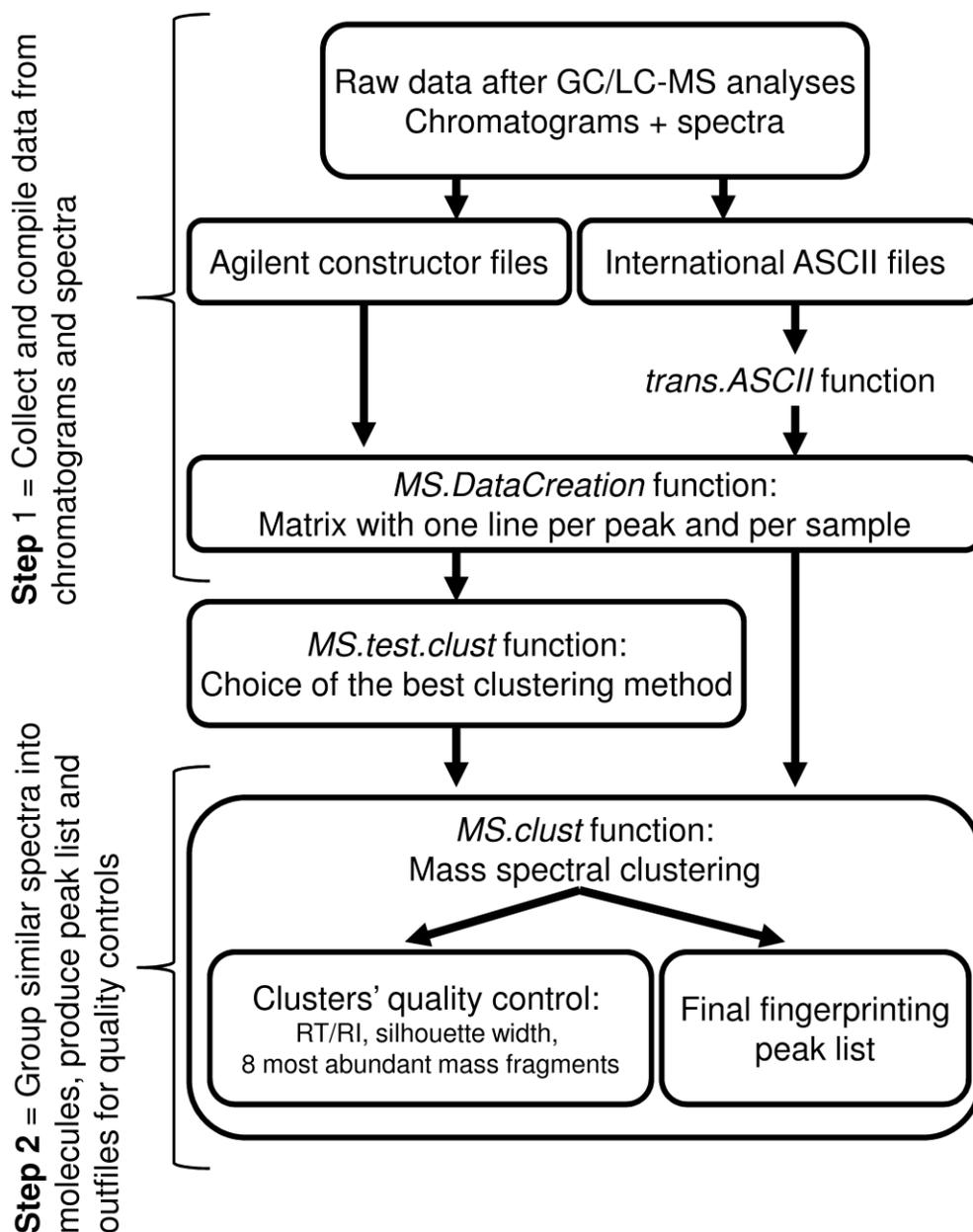
2-IMPLEMENTATION AND METHODS

2.1-PACKAGE/PROCESS DESCRIPTION

Our statistical method generates a reliable list of molecules present in a given sampling that can be used to fingerprint each sample (figure 1). Step 1, the data from all chromatograms and mass spectra are collected into a global matrix with one line for each detected chromatographic peaks. Each lines contains the information for one peak : sample name, RT/RI and relative mass spectrum. Then, prior to step 2, an optional function can be used to identify the best clustering algorithm that finds the correct structure in a test dataset where molecules are already identified. Step 2, the mass spectra from the global matrix are grouped

into homogenous clusters so as to eventually produce one cluster per molecule. The optimal number of clusters in a dataset is given graphically by plotting a cluster quality index, the Silhouette width (Rosseeuw et al). Finally several statistics are used to identify reliable clusters corresponding to a unique compound. This includes homogeneity in migration time per cluster, high Silhouette width and similarity in the 8 most abundant mass fragments.

Figure 1: General workflow of the MSeasy package



Résultats

A typical GC/LC-MS output is represented by two components:

- the chromatogram, displaying the mixture separation by the GC or LC where each peak corresponds to the elution of a distinct molecule out of the chromatographic column.
- a mass spectrum obtained by fragmentation in the MS chamber for each point of the chromatogram. A mass spectrum is represented by a histogram displaying the intensity of each fragment depending on its mass-to-charge ratio (m/z).

After a GC/LC-MS analysis, two types of data are available for each peak: the retention time in the chromatogram (i.e. when the molecule has been eluted; otherwise, retention index can be calculated), and the mass spectrum corresponding to this molecule.

Each GC/LC-MS analyser delivers the output in a specific format with various file extension (.sms for Varian, .d for Agilent, .dat for Finnigan, and so on).

In our method, two alternative input formats can be used. The first one is the international ASCII format that can be obtained from many providers directly or after transformation with translator softwares such as *GC and GCMS File Translator*TM (<http://www.chemsw.com/12149.htm>). The second input format is provided by Agilent ChemStation (MSD ChemStation, Agilent Technologies). When ASCII is the input format, data have to be compiled with the *trans.ASCII* function of the MSeasy package to obtain a global matrix for further analyses. Then the *MS.DataCreation* function smooth the chromatograms and detects peaks by the succession of 3 points with increasing intensity directly followed by 3 points of decreasing intensity (with a total intensity of the peak higher than 10 kilocounts). The default value of the filter parameter is 3, but it can be changed by increasing the *N_filt* value of the *MS.DataCreation* function (see MSeasy help files for details). When working with data from Agilent Chemstation, for each sample, the peaklist is taken from the *rteres.txt* file in the **.*D* directory and mass spectra are extracted from the text file generated with the Agilent Export3D macro from the tools menu of the Chemstation software. Then, the characteristics of the detected peaks (sample name, RT/RI and mass spectrum) are then compiled in a global matrix (space separated text file; *initial_DATA.txt*). Depending on the apex option chosen in the function *MS.DataCreation*, the mass spectrum of a peak is obtained by either averaging several mass spectra around the apex or directly extracting the mass spectrum at the peak apex. The intensity (in counts) of each mass fragment is then transformed into a relative percentage of the highest mass fragment per spectrum in

order to compare among mass spectra from different analyses with variable peak intensities. Several demonstration files (input and output files) are available at the package download page.

Test for the best clustering method (function *MS.test.clust*)

To detect similar molecules, we cluster peaks of different chromatographic runs that are characterized by similar mass spectra. The best algorithm to cluster similar mass spectra may differ depending on the species, extraction method, instrumentals, laboratory specificities and so on. We then implemented an option, the *MS.test.clust* function, to test and identify the best clustering algorithm.

Since the total number of true molecule is usually unknown in complex biological substance sampling, the use of unsupervised clustering algorithms is required. These include partitional and hierarchical algorithm (Steinbach, Ertöz *et al.* 2004). Partitional algorithms, such as kmeans or partitioning around medoids (PAM), determine all clusters at once and do not consider any hierarchical/neighborhood relations among clusters. For these algorithms, the number of clusters should be specified beforehand, but it is always possible to run the algorithm with a range of number of clusters and assess the quality of each partition with various indices of cluster validity. The optimal validity index is selected as the best partition. Unlike partitional methods, hierarchical algorithms are iterative methods for clustering datasets (hierarchical clustering analysis HCA), based on the neighborhood relations among clusters. Two types of algorithms exist: agglomerative or divisive. Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters (Halkidi, Batistakis *et al.* 2002; Zhao et Karypis 2003).

In distance-based clustering analyses, the choice of the distance metrics influences greatly the results. The p-order Minkowski distance (which includes the Euclidean distance as a special case when $p=2$) and Correlations are the most commonly used measures of respectively, dissimilarity and similarity (Jobson 1991).

Résultats

The Minkowski distance of order p between two points

$P = (x_1, x_2, \dots, x_n)$ and $Q = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ is defined as:

$$\left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}.$$

Minkowski distance is a metric distance typically used with p being 2 (Euclidean distance) or 1 (Manhattan distance). A study found that clustering large dimensions data was more efficient using p indices of Minkowski distance smaller than 1 (Aggarwal, Hinneburg *et al.* 2001; Hinneburg, Aggarwal *et al.* 2000). For that reason, we also tested two values of p indices below 1 ($p=1/2$; $p=1/3$).

The function *MS.test.clust* compares the clustering of a subset of well-known and clearly identify molecules that are present in several samples with three classical unsupervised clustering methods: partition around medoid (PAM), hierarchical divisive clustering, hierarchical agglomerative clustering, with various combinations of distance metrics and link methods. An important step in hierarchical methods is the choice of the link methods and so several options are tested (average, single, complete, ward; (Steinbach, Ertöz *et al.* 2004). The clustering algorithms are compared on their ability to correctly assign the different mass spectra to the pre-defined molecules using the matching coefficient. The results of clustering algorithms are then evaluated with two quality indices (Dunn index and Silhouette width) that assess which clustering scheme best fits the data. The matching coefficient evaluates the correct assignment of each mass spectrum to the expected molecules: when one cluster groups the different mass spectra of the same molecule, then 1 is attributed and when one cluster contains mass spectra corresponding to different molecules, then 0 is attributed. The sum is then divided by the total number of expected molecules. The value of the matching coefficient varies from 0 to 1 and 1 indicates perfect clustering results. A second cluster validity index called Silhouette width and described by Rosseeuw (Rousseeuw 1987) is commonly used to assess clustering quality (Dixon, Heinrich *et al.* 2009). This index is based on two criteria: cluster compactness and isolation. If a is the average distance of a point from the other points of the same cluster (variation intracluster / compactness), and b represents the minimum of the average distances of the point from the points of the other clusters (cluster separation), then the silhouette width s of the point is defined as: $s=(b-a)/\max\{a,b\}$. The value of the Silhouette

Résultats

width varies from -1 to 1 with -1 indicating that the object is closer to another cluster than the one it has been assigned to, 0 indicating bad clustering result and 1 a perfect clustering result.

The second validity index, the Dunn index D , is defined as: $D = [\min\{k, l - \text{numbers of clusters}\} \text{dist}(C_k, C_l)] / [\max\{m - \text{cluster number}\} \text{diam}(C_m)]$

k, l, m - numbers of clusters which come from the same partitioning, $\text{dist}(C_k, C_l)$ - inter cluster distance between clusters C_k and C_l , $\text{diam}(C_m)$ - intra cluster diameter computed for cluster C_m .

Graphics display the comparative results between algorithms on the R window and as a PDF file (for details see help files of the R package).

Step 2: Mass spectra clustering and creation of a fingerprinting matrix (function *MS.clust*)

The selected optimal unsupervised clustering method is run on the global matrix (initial_DATA.txt) to group similar mass spectra into molecules. The optimal number of clusters is identified using the Silhouette width. Running the best unsupervised clustering method with the optimal number of clusters produces different output files for facilitating the quality control of the clusters, the identification of molecules, and returns a fingerprinting matrix with homogeneous clusters.

As the total number of molecules contained in the dataset is generally unknown, the function *MS.clust* allows identifying, among a user defined range of number of clusters, the optimal number(s) of clusters using the silhouette width. For each number, a set of output files is generated. Homogeneous and inhomogeneous clusters are separated. Homogeneous clusters are expected to be a unique molecule and are defined by several quality controls:

- Low variation in migration time between samples within the cluster ($RT_{\max} - RT_{\min} < 0.1$ by default but can be user defined).
- High individual silhouette width within the cluster and high mean cluster silhouette width (high compactness and high isolation)
- Similarity of the 8 more abundant mass fragments within the cluster
- A cluster should not contain several peaks from the same chromatogram.

Résultats

Several output files are produced by the *MS.clust* routine: *output_cluster* summarizes the results at the level of the cluster and separates homogeneous from inhomogeneous clusters. This file can be used directly for subsequent molecule identifications. *Output_peak* gives the clustering outputs and the results at the level of the peak. To allow further manual investigations on homogeneous and more particularly inhomogeneous clusters, two pdf files are generated. The first displays the distribution of retention times/retention indices for each cluster and indicates the range within each cluster. The second displays the distribution of the silhouette width and indicates the mean value for each cluster. Finally, *output_matrix* is the fingerprinting matrix (0 for absence, 1 for presence) with samples in row, sample name in the first column, then migration time and homogeneous clusters.

User-friendly interface

For users that are not familiar with R, we developed a user-friendly interface that skip the writing of command lines. The MSeasyTkGUI R package contains a graphical user interface (GUI) for MSeasy. With MSeasyTkGUI only one command line - `library(«MSeasyTkGUI»); MSeasyTkGUI()` – is needed to launch the GUI (Figure 2). All MSeasy calculations can be started from the GUI menu without any command lines.

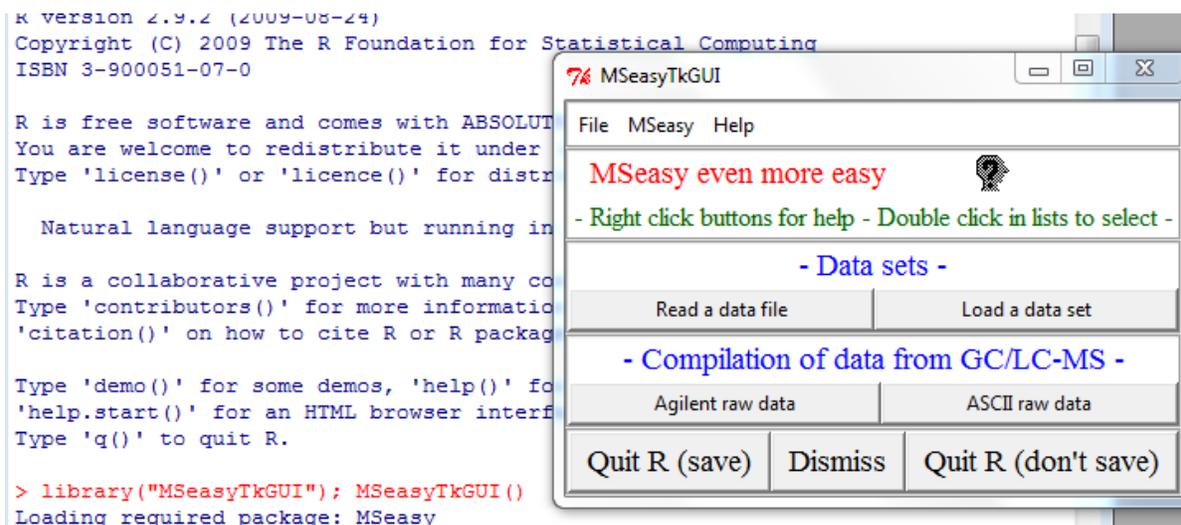


Figure 2: user-friendly graphic interface for MSeasy

Left : R commande window with in red the command line used to start the GUI. *Right* :first window of the GUI interface.

The MSeasyTkGUI package is available at the MSeasy download page.

2.2 TEST DATASETS

We calibrated and validated our method by investigating the chemical composition in two datasets obtained with different extraction methods and instrumentals. The first dataset came from a vegetal model and consisted of volatiles organic compounds (VOCs) extracted from inflorescence of *Lavandula* species by organic solvent. The second dataset came from an animal model and consisted of VOCs collected by thermal desorption of seabird feathers. Twenty individuals of the blue petrel (*Halobaena caerulea*) were analysed for their biogenically-emitted VOCs.

Lavender dataset

The Lavender data set comes from a broad study that investigates the evolution and ecological function of VOCs over the genus *Lavandula* (Guitton 2010). The dataset is composed of the VOCs from 86 inflorescences on individual plants from 16 species of two subgenera (*Lavandula* and *Fabricia*). Within each subgenus, different sections and subspecies are represented (Upton 1997). Samples were collected in one day in 2006 and 2008 from plants of similar age cultivated in the same field (CNPMAI, Milly La Forêt, France). A detailed protocol of sample preparation, VOCs extraction and analyses is described in (Guitton, Nicolè *et al.* 2010b).

The chromatographic data were generated with an Agilent GC-MS apparatus. Chromatographic and spectral data were extracted using the Agilent ChemStation (MSD ChemStation Version D.0200275, Agilent Technologies).

Bird dataset

The second dataset consisted of seabirds-emitted VOCs obtained by direct solid-phase thermal desorption of feathers. This data set was collected to investigate the role of chemical signals, and plumage scent in particular, for the social behaviors of petrel seabirds (Mardon 2010). Feathers from 20 blue petrels (*Halobaena caerulea*) were collected at a breeding colony, on a sub-Antarctic island (Ile Verte, 49°51'S, 70°05'E) from the Kerguelen archipelago. A detailed protocol of sample preparation, VOCs extraction and analyses is described in (Mardon 2010).

The Varian GC-MS used provided raw chromatographic data generated as *.sms files that were transformed into ASCII international format using a file translator software (GC and GC/MS File Translator Pro, version 4.1, ChemSW, Fairfield, CA, USA, 2003).

2.3 COMPARISON WITH DIFFERENT PRE PROCESSING METHODS

The performance of our method based on the clustering of mass spectra was compared to two other classical approaches: mass spectral library detection and manual detection. The lavender dataset was used for this analysis. The quality of detection was evaluated using the cluster validity index of silhouette width (see 2.2). For the mass spectral library detection, all peaks from all samples were submitted to the automated peak identification module integrated in the Agilent ChemStation®, using the Wiley 275 mass spectral library (McLafferty 2006). A CAS number and a match factor were attributed to each peak. Stereo isomers and redundant entries (similar molecule with different CAS numbers) were grouped after manual examination. Finally, peaks assigned to similar molecules were grouped into clusters and this partition of the data was used to calculate the silhouette indices. Manual detection was performed by comparing the retention time and mass spectrum of each peak with those of authentic standards. When standards were not available, identification were based on the mass spectral library Wiley 275 and Adams mass spectral databases (Adams, 2007) by calculating the Kovats indices using n-alkanes C8-C20 (Fluka). Peaks corresponding to similar molecules were grouped into clusters and silhouette widths were then calculated.

Quantitative and qualitative results were compared at different levels: line level (one peak of one sample), peak level (one peak identify for several samples) and molecule level (i.e. homogeneous and high quality clusters for our method; peaks with mean quality factor >80% for mass spectral library detection; and molecules identified for manual detection).

3-EXPERIMENTS AND RESULTS

3.1 IDENTIFICATION OF THE BEST CLUSTERING METHOD

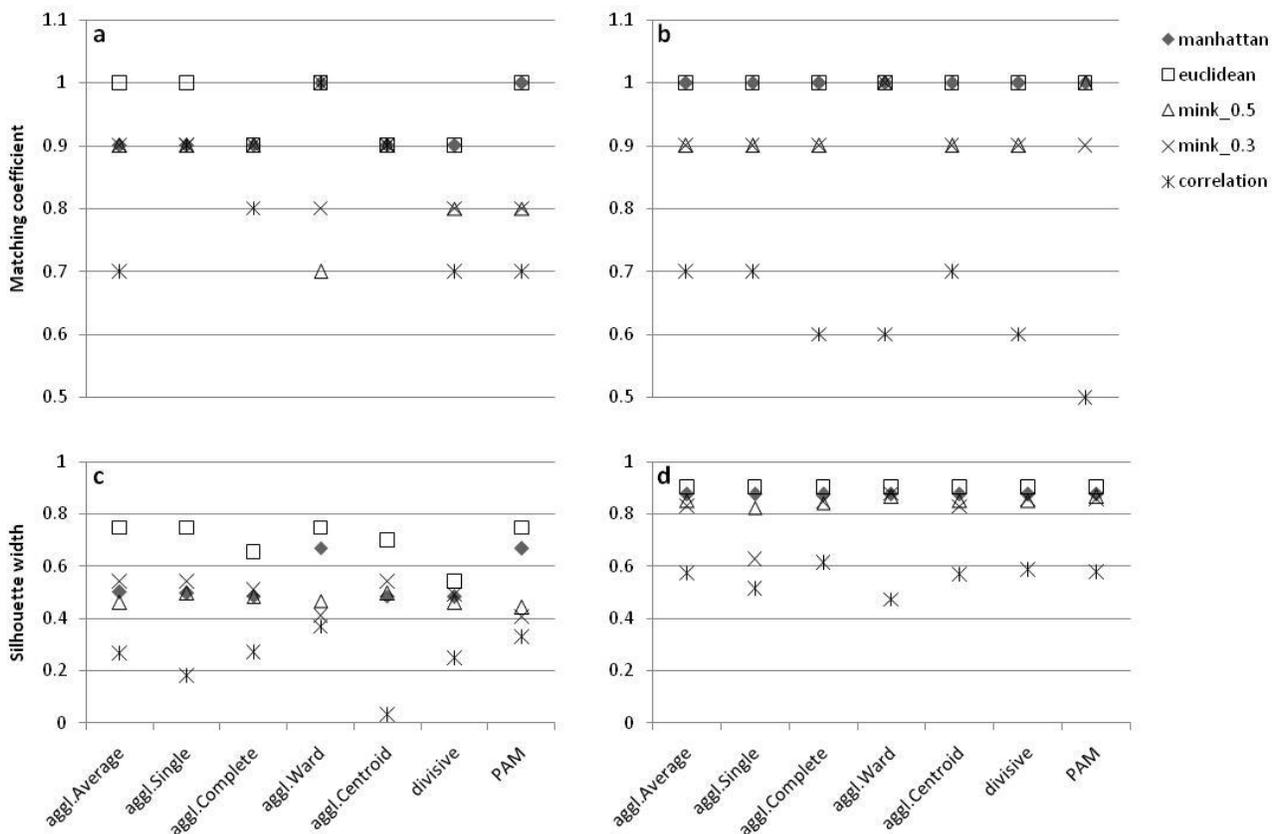
For the dataset on bird feather scent, Euclidean distance gave the best clustering results (Figure.3c) and only PAM and hierarchical agglomerative clustering with average, single or Ward linkages also provided a correct clustering (Figure.3a). For the lavender dataset, Euclidean and Manhattan distances were the best distances and gave the expected result irrespective of the clustering methods (Figure3 b and d). For both dataset, the Ward linkage method was the most efficient in grouping similar mass spectra. This was true with both apex option of the *MS.clust* routine (results not shown, see package help file for details). The

Résultats

quality and precision of the MS data influence the quality of the clustering. The bird dataset was obtained with a lower precision due to unexplained reasons compared to lavender and gave less robust clustering results (mean silhouette width over all methods respectively for bird and lavender dataset, 0.49 and 0.80).

According to these results and tests performed on other datasets (tropical trees analysed by GC-MS (Courtois, Paine *et al.* 2009) and corals analysed by LC-MS (results not shown Ivanisevic; Ivanišević, Thomas *et al.* 2010), the hierarchical agglomerative clustering with Euclidean distance and Ward linkage was chosen for the analysis of the complete datasets.

Figure 3: matching coefficient and silhouette width of several clustering methods (along the x axis) for the bird data set (a and c) and the lavender dataset (b and d). For both dataset, mass spectra were taken at the apex. PAM corresponds to partition around mediod (partitional method). Other methods were hierarchical. Several link methods were tested for agglomerative clustering (aggl.). The different symbols correspond to different distance indices.



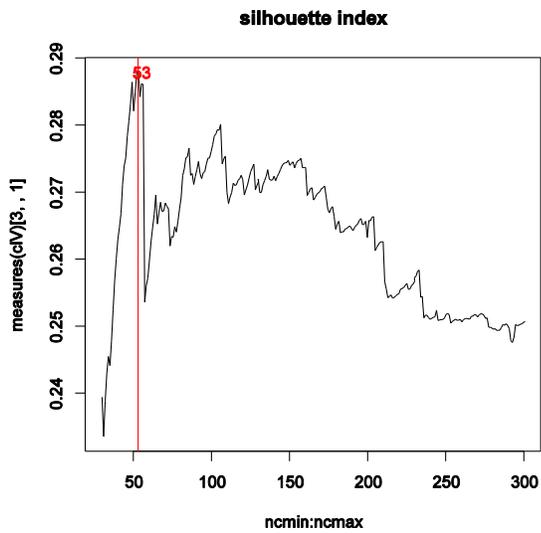
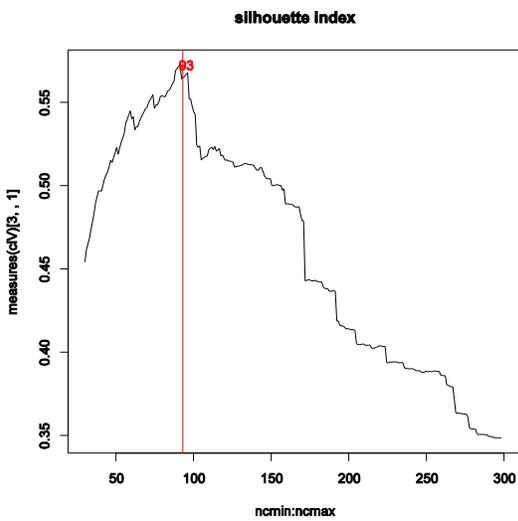
3.2 APPLICATION TO LARGER DATASETS

We applied the hierarchical agglomerative clustering with Euclidean distance and Ward linkage to the complete datasets using the function *MS.clust*.

Figure 4 : Graphical output of the function *MS.clust* representing the evolution of the silhouette width as a function of the number of clusters. The red line indicated the best number of clusters. For the two datasets, results are shown for mass spectra at the apex.

Lavender

Bird



Résultats

The search for the optimal number of clusters was applied to the Lavender dataset for a number of clusters ranging from 30 to 300 (figure 4). Here, the curve shows a unique optimal number (apex of a curve). This value should be used for further analyses. However, if the graph displays an optimal value located on a plateau, the user could perform the analyses at the optimal number of clusters and also with the minimal and maximal numbers of clusters delimiting the plateau.

Of the 1695 peaks in the lavender dataset, the optimal number of clusters was chosen equal to 93 because there was only one maximum (identical whatever the *MS.DataCreation* apex option chosen). Of these 93 clusters, 61 were directly detected as homogeneous cluster. The method was also applied to the bird dataset with a baseline correction window of 9 and for a range of clusters from 30 to 300. The optimal number of cluster depended on the *MS.DataCreation* apex option. Over the 1300 peaks, the optimum number of cluster varied from 88 when the *MS.DataCreation* option -mean mass spectra around the apex -was used to 53 when the option – only apex mass spectra - was used. On the 88 clusters, 47 homogeneous clusters could be identified directly.

After clustering, the quality of the clusters was assessed. The first and intuitive check was the variation in migration time. However, if different GC or LC columns were used to obtain the dataset, this first criterion becomes ineffective because in that case very different migration times can be obtained for the same molecule. The second criterion was based on the silhouette width that measures cluster compactness and isolation. A third criterion was based on the similarity of the 8 most abundant mass fragments of the mass spectra within a cluster.

The PDF output files generated after mass spectral clustering help interpret the cluster quality. Figure 5 presents a key to this interpretation based on migration time. Homogeneous clusters, defined by a variation of migration time lower or equal to 0.1 by default, represented two third of overall clusters in the lavender dataset. In more than 95% cases, they had a high mean silhouette width (> 0.4), a right-skewed distribution of silhouette width and correspond to a unique molecule. Inhomogeneous clusters were defined either by the presence of the same analysis twice or more within one cluster (only one cluster concerned that require a manual verification), or by a variation of migration time higher than 0.1. A total of 5.4% of the clusters were found to be a unique molecule but they presented a larger shift in migration time than 0.1. For two third of inhomogeneous clusters (22.6% of all clusters), a bimodal distribution of migration times was observed, associated with a high mean and right-skewed

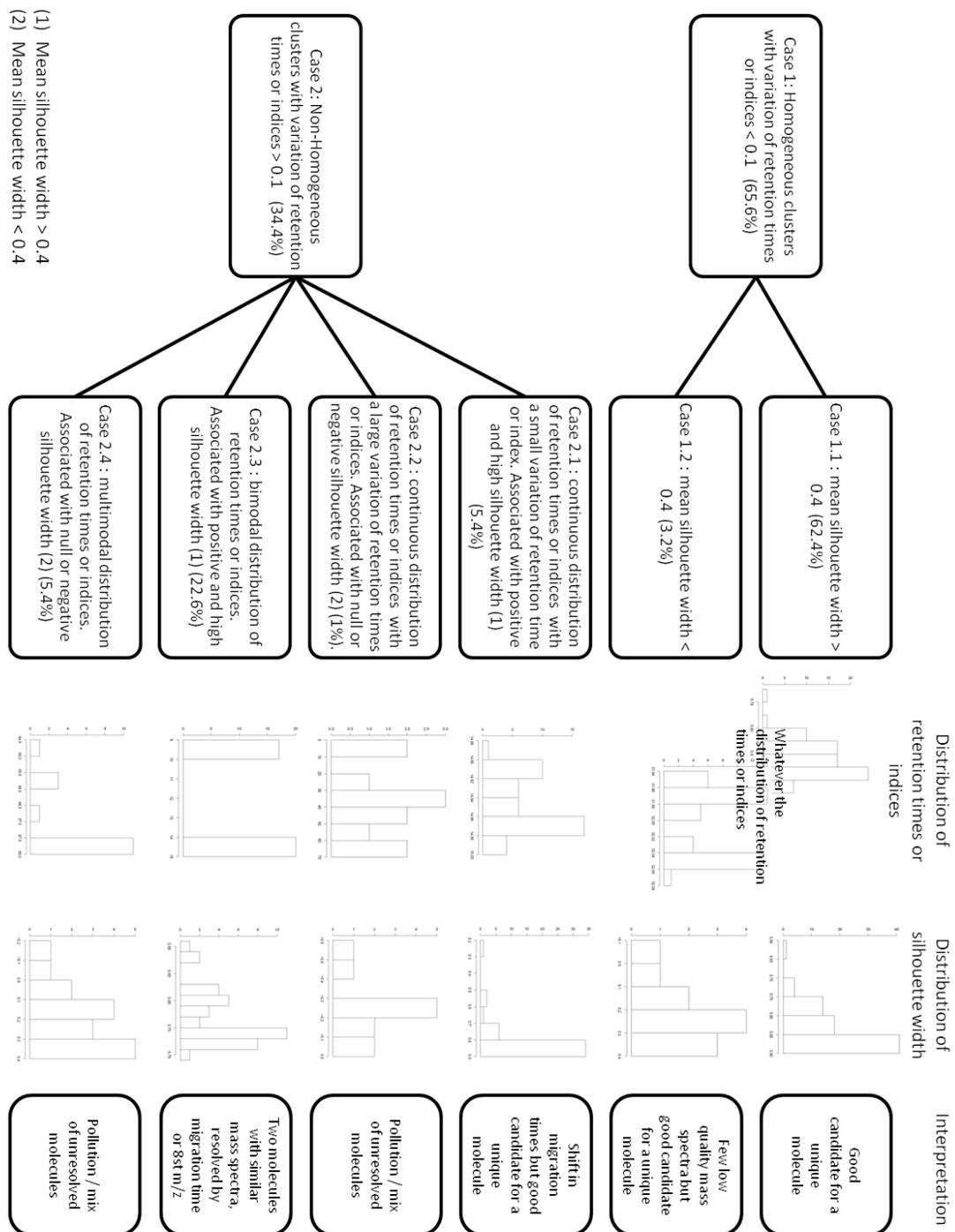
Résultats

distribution of silhouette width, indicating that two molecules could not be separated because they had very similar mass spectra. The information of the migration times and the 8 most abundant mass fragments could thus be used to distinguish the two molecules easily. Overall, only 7.4% of the clusters were identified as pollution (based on comparison with blank control analysis) or mixed of unresolved molecules that may require further extensive manual processing.

When the criterion based on migration time made no sense, the first criterion became the silhouette width. In our data, a limit of mean silhouette width equal to 0.4 (or median = 0.5) was meaningful to separate high and bad quality clusters.

Résultats

Figure 5: Key to the interpretation of cluster quality based on the criteria of variation in migration times and on the silhouette width. Main cases are represented. The percentage in parentheses gives the repartition of the different cases and was calculated on the lavender dataset.



3.3 COMPARISON WITH DIFFERENT PRE PROCESSING METHODS

Table 1 : comparative results of three different pre-processing methods.

	Mass spectra clustering method	Manual detection	Mass spectral library detection	statistics
LINE DATA				
total number of lines	1695	1695	1695	
total nb of lines with silhouette > 0.7	870 (51.3%)	881 (52%)	327 (19.3%)	KHI ² test ddl= 2 khi ² =489.8 p<0.0001
mean silhouette over lines(1)	0.641	0.610	0.071	anova ddl=3 F= 3520.6 p<0.0001
SD	0.265	0.314	0.522	
Tukey's a Posteriori Comparisons	b	b	a	
PEAK QUALITY (raw dataset, all lines confounded)				
total number of peaks	93		290	
mean silhouette per peak	0.635		-0.017	anova ddl=2 F=182.0 p<0.0001
number of peak with mean silhouette > 0.7	40 (43%)		25 (8.6%)	
Tukey's a Posteriori Comparisons	b		a	
MOLECULE QUALITY ("high quality" dataset, clusterOK (avec 8 premiers fragments OK) et groupes cas OK (>80%))				
total number of molecules	61	127	82	
mean silhouette per molecules	0.729	0.535	0.177	anova ddl=3 F=189.5 p<0.0001
number of peaks with mean silhouette > 0.7	36 (59%)	51 (40%)	20 (24.4%)	
Tukey's a Posteriori Comparisons	c	b	a	

(1) after a boxcox transformation for normalization (lambda =1.43)

Table 1 shows that the clustering method and the manual detection gave very similar results compared to the less efficient method solely based on mass spectral library detection. For the latter, only 19.3% of lines had a silhouette width higher than 0.7 while the other two methods presented approximately 51% of lines robustly assigned to a peak. At all levels, the mean silhouette width was significantly lower for the mass spectral library method than for the two other methods. The mass spectral library detection failed to group similar mass spectra which dramatically decreased the silhouette width (good cluster compactness but low cluster separation) and enlarged artificially the number of peaks (290 versus 93 for our method and 127 for the manual detection).

Compared to our method, manual detection gave a higher number of high quality molecules but with a lower mean silhouette width. However, more molecules had a higher silhouette width (51 versus 36 for mass spectra clustering). Manual detection allowed considering and assigning ambiguous mass spectra into clusters, which reduced the mean silhouette per peak. However, manual detection had better results than our method but time and efforts were different. Dixon *et al.* (Dixon, Brereton *et al.* 2006) estimated that manual alignment and

Résultats

integration of each peak lasted approximately 5 min per peak (checking mass spectra, elution times, baseline, etc.) and manual processing of a chromatographic dataset of 50 chromatograms consisting of 30 peaks would take 125 hours, i.e. around three weeks' work. At this speed, the manual treatment of the 1695 peaks of the lavender dataset would require 141 hours (3 weeks and a half). Comparatively, the time for clustering was fast and took only 15 minutes for lavender dataset and 11min for bird dataset on a laptop (CPU: 1.3 GHz, Memory: 1 Go). The 1695 original peaks were grouped in 93 clusters. This greatly shorten (reduced by 18) the amount of work needed to carry out the manual analysis. To be complete, the time to compile all GC/LC-MS analyses should be added to the time of clustering. The compilation of mass spectra and chromatogram data into a global matrix ranged from 8min for the bird dataset to 75min for the lavender dataset. Finally, overall, the treatment with our method took one to two days while the manual detection method took more than 3 weeks.

DISCUSSION

In this paper, we present a simple pre-processing method to detect molecules in GC-MS datasets. The method is implemented in the MSeasy R package and is freely available. Only one to two days were needed to treat a chemical dataset of 86 chromatograms and 1695 peaks overall. The quality of molecule detection was less effective than manual detection but comparatively, the manual approach was more labour-intensive and took approximately three weeks' work instead of few days with MSeasy.

The mass spectral clustering method, based on the clustering of mass spectra rather than retention times or indices, is simple and yields a comprehensive peak list without any extensive profile correction and alignment. Retention times or indices are also used in a second stage as a criterion for the cluster quality control.

This method is fast, exhaustive, robust to shift in migration times and to the use of different columns. It allows handling large data sets in short time and helps researcher to interpret large data set by extracting essential information and for data quality control.

Hence this method is particularly well-adapted for ecological experiments, and more generally for large and long-term experiments in non-model species. However, if retention indices are available for all samples, other methods may be more useful than mass spectral clustering.

Résultats

For molecule identification, visual exploration of the raw data is not always standardized because it is highly dependent on the chemist's ability to recognize molecule structure with a mass spectrum. As the number of peaks increases, manual peak detection and identification may become subjective and biased. As an aid, reference mass spectral libraries can be used but they are not always adapted to a given dataset, fail to group very similar molecules and may fail if the molecule is not in the database (Schauer, Steinhauser *et al.* 2005). Adding standards to the original dataset, the mass spectral clustering method greatly facilitates the identification of molecules. Spectral libraries could be created very easily and adapted to the instruments and lab specificities using standards and well-identified molecules. This library can be added very simply to the clustering analyses to facilitate the naming of clusters. Each experiment in the lab would thus enrich the spectral libraries and increase the efficiency of molecules identification. This approach is particularly powerful for labs that use the MS for a wide array of species.

Applying our method to two datasets, we found that the best clustering method for grouping similar mass spectra in molecules was the hierarchical clustering analysis with the Euclidean distance and the Ward linkage. Former studies that aimed at clustering mass spectra from air or water samples also demonstrated the greater effectiveness of Ward's method (Bernard, Van Grieken *et al.* 1986; Murphy, Middlebrook *et al.* 2003). Bernard et al (Bernard, Van Grieken *et al.* 1986) obtained similar results on the efficiency of linkage methods on electron microprobe data and found the best results from Ward's classification, moderate results with average distances measures, and the poorest results with single linkage. However, it is not excluded that other clustering algorithms could be more adapted for other particular datasets. For that reason, we recommend the use of the function *MS.test.clust* to identify the best clustering algorithm on any new dataset.

As for all methods, the mass spectral clustering method presented limits and should not be considered as an end of the data analysis. The clusters are never absolutely accurate and misclassifications occur (7.4% of bad clusters in the lavender dataset). First of all, the efficiency of clustering depends on the quality and precision of the raw dataset (scanning and analysis parameters). Even a powerful and optimized cluster algorithm will never replace the correct settings of separation analysis parameters. Unclear chromatograms with overlapping and/or saturated peaks will lead to a lower success in assigning peaks into homogeneous molecules.

Résultats

Sometimes, one spectrum cannot chemically match to a cluster even though it belongs in it for several reasons: very similar spectrum corresponding to different molecules, stereo isomers, a too low number of clusters (resulting in merging artificially different molecules). On the contrary, a spectrum can be misclassified in an unexpected cluster because it can be a mixture of several compounds, be polluted or be part of a continuum of molecules that are not easily resolved by cluster analysis.

Finally, as emphasized by Murphy (Murphy, Middlebrook *et al.* 2003), there is a vast information that an experienced scientist can bring to the interpretation of mass spectra and that cannot be easily programmed into a classification algorithm. Nevertheless, the clustering of peaks made with our mass spectral clustering method can save huge amount of time in the process of identifications of molecules.

BIBLIOGRAPHY

Aggarwal, C.C., Hinneburg, A. and Keim, D.A. (2000) On the surprising behavior of distance metrics in high dimensional spaces, *Lecture Notes in Computer Science*, 1973, 420-434.

Bamba, T. and Fukusaki, E. (2006) Technical problems and practical operations in plant metabolomics, *Journal of Pesticide Science*, 31, 300-304.

Baran, R., Kochi, H., Saito, N., Suematsu, M., Soga, T., Nishioka, T., Robert, M. and Tomita, M. (2006) MathDAMP: a package for differential analysis of metabolite profiles, *Bmc Bioinformatics*, 7, 530.

Benton, H.P., Wong, D.M., Trauger, S.A. and Siuzdak, G. (2008) XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization, *Analytical chemistry*, 80, 6382-6389.

Bernard, P.C., Van Grieken, R.E. and Eisma, D. (1986) Classification of estuarine particles using automated electron microprobe analysis and multivariate techniques, *Environmental Science & Technology*, 20, 467-473.

Cuadros-Inostroza, Á., Caldana, C., Redestig, H., Kusano, M., Lisek, J., Peña-Cortés, H., Willmitzer, L. and Hannah, M.A. (2009) TargetSearch- a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data, *BMC bioinformatics*, 10, 428.

Résultats

De Vos, M., Van Oosten, V.R., Van Poecke, R.M.P., Van Pelt, J.A., Pozo, M.J., Mueller, M.J., Buchala, A.J., Métraux, J.P., Van Loon, L.C. and Dicke, M. (2005) Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack, *Molecular Plant-Microbe Interactions*, 18, 923-937.

Dixon, S.J., Brereton, R.G., Soini, H.A., Novotny, M.V. and Penn, D.J. (2006) An automated method for peak detection and matching in large gas chromatography-mass spectrometry data sets, *Journal of Chemometrics*, 20, 325-340.

Dixon, S.J., Heinrich, N., Holmboe, M., Schaefer, M.L., Reed, R.R., Trevejo, J. and Brereton, R.G. (2009) Use of cluster separation indices and the influence of outliers: application of two new separation indices, the modified silhouette index and the overlap coefficient to simulated data and mouse urine metabolomic profiles, *Journal of Chemometrics*, 23.

Duran, A.L., Yang, J., Wang, L. and Sumner, L.W. (2003) Metabolomics spectral formatting, alignment and conversion tools (MSFACTs), *Bioinformatics*, 19, 2283.

Fiehn, O. (2002) Metabolomics--the link between genotypes and phenotypes, *Plant molecular biology*, 48, 155-171.

Fukusaki, E. and Kobayashi, A. (2005) Plant metabolomics: potential for practical operation, *Journal of bioscience and bioengineering*, 100, 347-354.

Guillon, Y. (2010). LBVPAM, Faculté des sciences. Jean Monnet, Saint Etienne.

Guillon, Y., Nicolè, F., Moja, S., Valot, N., Legrand, S., Jullien, F. and Legendre, L. (2009) Differential accumulation of volatile terpene and terpene synthase mRNAs during lavender (*Lavandula angustifolia* and *L. x intermedia*) inflorescence development, *Physiologia Plantarum*, 9999.

Gullberg, J., Jonsson, P., Nordstrom, A., Sjostrom, M. and Moritz, T. (2004) Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry, *Analytical biochemistry*, 331, 283-295.

Hansen, M.E., Andersen, B. and Smedsgaard, J. (2005) Automated and unbiased classification of chemical profiles from fungi using high performance liquid chromatography, *Journal of microbiological methods*, 61, 295-304.

Résultats

- Harborne, J.B. (1997) Recent advances in chemical ecology, *Natural product reports*, 14, 83-98.
- Harborne, J.B. (2001) Twenty-five years of chemical ecology, *Natural Product Reports*, 18, 361-379.
- Heil, M., Bueno, S. and Carlos, J. (2007) Within-plant signaling by volatiles leads to induction and priming of an indirect plant defense in nature, *Proceedings of the National Academy of Sciences*, 104, 5467.
- Hinneburg, A., Aggarwal, C.C. and Keim, D.A. (2000) What is the nearest neighbor in high dimensional spaces. *Citeseer*, 506-515.
- Jobson, J.D. (1991) *Applied multivariate data analysis: Regression and experimental design* Categorical and multivariate methods. Springer.
- Johnson, M.T.J., Agrawal, A.A., Maron, J.L. and Salminen, J.P. (2009) Heritability, covariation and natural selection on 24 traits of common evening primrose (*Oenothera biennis*) from a field experiment, *Journal of Evolutionary Biology*, 22, 1295.
- Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, M. and Moritz, T. (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS, *Analytical chemistry*, 76, 1738-1745.
- Katajamaa, M., Miettinen, J. and Oresic, M. (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, *Bioinformatics*, 22, 634.
- Lindon, J.C., Nicholson, J.K., Holmes, E., Keun, H.C., Craig, A., Pearce, J.T.M., Bruce, S.J., Hardy, N., Sansone, S.A. and Antti, H. (2005) Summary recommendations for standardization and reporting of metabolic analyses, *Nature Biotechnology*, 23, 833-838.
- Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. and Fernie, A.R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants, *Nature Protocols*, 1, 387-396.
- Lommen, A. (2009) MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing, *Analytical chemistry*.
- Luedemann, A., Strassburg, K., Erban, A. and Kopka, J. (2008) TagFinder for the quantitative analysis of gas chromatography--mass spectrometry (GC-MS)-based metabolite profiling experiments, *Bioinformatics*, 24, 732.

Résultats

Mardon, J. (2010) Olfaction chez les pétrels : caractérisation chimique d'une signature olfactive et implications évolutives. CEFE UMR 5175 Montpellier, équipe écologie comportementale. Montpellier II, Montpellier.

McLafferty, F.W. (ed) (2006) Wiley Registry of Mass Spectral Data, 8th Edition. . New York.

Meinwald, J. and Eisner, T. (2008) Chemical ecology in retrospect and prospect, Proceedings of the National Academy of Sciences, 105, 4539.

Murphy, D.M., Middlebrook, A.M. and Warshawsky, M. (2003) Cluster analysis of data from the particle analysis by laser mass spectrometry (PALMS) instrument, Aerosol Science and Technology, 37, 382-391.

Neuweger, H., Albaum, S.P., Dondrup, M., Persicke, M., Watt, T., Niehaus, K., Stoye, J. and Goesmann, A. (2008) MeltDB: a software platform for the analysis and integration of metabolomics experiment data, Bioinformatics, 24, 2726.

Penn, D.J., Oberzaucher, E., Grammer, K., Fischer, G., Soini, H.A., Wiesler, D., Novotny, M.V., Dixon, S.J., Xu, Y. and Brereton, R.G. (2007) Individual and gender fingerprints in human body odour, Journal of the Royal Society Interface, 4, 331.

RDevelopmentCoreTeam (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. In Computing, R.F.f.S. (ed). Vienna, Austria.

Rousseeuw, P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, Journal of computational and applied mathematics, 20, 53-65.

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., Lundgren, K., Roessner-Tunali, U., Forbes, M.G., Willmitzer, L., Fernie, A.R. and Kopka, J. (2005) GC-MS libraries for the rapid identification of metabolites in complex biological samples, FEBS letters, 579, 1332-1337.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem, 78, 779-787.

Steinbach, M., Ertöz, L. and Kumar, V. (2004) The challenges of clustering high dimensional data, New directions in statistical physics: econophysics, bioinformatics, and pattern recognition, 273.

Résultats

Steinfath, M., Groth, D., Lisec, J. and Selbig, J. (2008) Metabolite profile analysis: from raw data to regression and classification, *Physiologia Plantarum*, 132, 150-161.

Tikunov, Y., Lommen, A., de Vos, C.H.R., Verhoeven, H.A., Bino, R.J., Hall, R.D. and Bovy, A.G. (2005) A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles 1 [w]. *Am Soc Plant Biol*, 1125-1137.

Vander Meer, R.K. (1998) Pheromone communication in social insects: ants, wasps, bees, and termites. Westview Pr.

Xia, J., Psychogios, N., Young, N. and Wishart, D.S. (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation, *Nucleic Acids Research*.

I-3 Discussion

Les avancées rapides dans le domaine de la chimie analytique permettent d'étendre le champ d'investigation de la métabolomique aux questions écologiques. La compréhension des variations de la production de composés chimiques chez des espèces non-modèles permet de répondre à de nombreuses problématiques d'écologie. Cette démarche de recherche de signal biologique dans des systèmes biologiques complexes demande généralement l'analyse de nombreux échantillons et la répétition des expérimentations. Afin de faciliter le traitement des grands jeux de données générés par de telles études, j'ai développé (en partenariat avec Florence Nicolè et Elodie Courtois) MSeasy, un outil statistique (libre et gratuit) permettant de simplifier la détection et l'identification de pics dans des échantillons de composition inconnue analysés en GC/LC-MS. La méthode employée par MSeasy est basée sur le regroupement (*clustering*) des pics de chromatographie sur la base de leurs spectres de masse plutôt que sur celle, plus classique, du regroupement des pics selon leur temps ou indices de rétention. Nous avons montré que la meilleure méthode de *clustering* de spectres de masse est la classification hiérarchique (distance euclidienne et méthode de liaison de Ward). Ce résultat est valable pour les différents jeux de données testés (pétrels, lavandes, éponges de mer, arbres tropicaux) et des résultats similaires ont été obtenus dans des travaux antérieurs (Murphy, Middlebrook *et al.* 2003). Nous avons, cependant laissé à l'utilisateur la possibilité de tester d'autres algorithmes, d'autres distances et d'autres méthodes de liaison. En effet, il est possible qu'un autre jeu de données puisse présenter une méthode de classification optimale

Résultats

différente de celle-ci. Il est d'ailleurs recommander d'effectuer le test sur tout nouveau jeu de données.

La méthode MSeasy est rapide avec par exemple moins d'un jour pour « prétraiter » un jeu de données de 86 chromatogrammes et un total de 1695 pics. Et elle permet également, lorsque l'on souhaite identifier les molécules, de faire gagner du temps au chimiste du laboratoire en réduisant le jeu de données à l'essentiel de l'information. Ainsi, au lieu de regarder l'ensemble des spectres de masse de l'ensemble des pics pour identifier les molécules présentes celui-ci n'a plus qu'à analyser les différentes informations rendues disponibles par MSeasy pour les différents clusters (RT/RI pour chaque pic, indice Silhouette, 8 premiers fragments de masse). La procédure d'identification de molécule est certes moins efficace qu'une identification manuelle par un chimiste expérimenté mais elle a l'avantage d'être plus rapide (18 fois moins de temps pour analyser 1695 pics). MSeasy est particulièrement adaptée au prétraitement de données d'expérimentation d'écologie chimique de longue durée sur des espèces non-modèle. Cependant, si les molécules recherchées sont bien connues et que l'on dispose des RT/RI pour l'ensemble des échantillons il est évident que d'autres logiciels seront plus adaptés.

MSeasy est un outil de prétraitement et ne fournit par définition qu'une analyse préliminaire du jeu de données. Il a ces limites notamment dans le nombre de formats de données qu'il peut lire, la détection de molécules et son efficacité informatique (nombre de pics comparables, temps d'analyse...). Actuellement, MSeasy ne permet de lire que deux formats de données (Agilent et ASCII) et nous souhaiterions élargir le nombre de formats de fichier d'entrée pour que les utilisateurs d'autres types d'appareils de GC/LC-MS puissent plus aisément avoir accès à MSeasy. Ceci peut passer par l'intégration de tout ou partie du logiciel Metalign (Lommen 2009) qui permet de lire divers formats de données de divers appareils de MS ou par un couplage avec le module R XCMS² car celui-ci permet la compilation de données au format international NetCDF. XCMS² permet également une identification des spectres de masse par comparaison avec les molécules de la base de données en ligne METLIN (<http://metlin.scripps.edu/>) ou propose une structure chimique théorique si le spectre de masse n'est pas dans la base de données. MSeasy peut traiter des jeux de données de GC/LC-MS contenant jusqu'à 20000 pics en un temps relativement court (2 jours) mais que certains trouveront trop long. La qualité du *clustering* dépend grandement de la qualité des données brutes. Les meilleurs algorithmes du monde ne remplaceront jamais un paramétrage

Résultats

correct de la méthode de séparation des molécules. Ainsi, avec MSeasy, la comparaison de chromatogrammes contenant de nombreux pics saturants et/ou de nombreuses co-élutions ne pourra pas donner de résultats satisfaisants. De plus, même si un pic est mathématiquement associé à un groupe d'autres pics il est possible qu'il n'ait, en fait, pas de relations chimiques avec les autres pics du groupe. Ceci est certes une limite mais il faut rappeler que les meilleurs algorithmes ne peuvent pas totalement remplacer l'expertise d'un chimiste (Murphy, Middlebrook *et al.* 2003) mais ils peuvent lui faire gagner un temps précieux. Des développements sont d'ores et déjà envisagés pour améliorer ces performances (implémentation d'algorithmes plus efficaces (Dixon *et al.*, 2006), utilisation de nouveaux package R, programmation en C).

MSeasy a répondu à nos besoins pour notre étude de l'évolution des COV accumulés par les espèces du genre *Lavandula*. MSeasy a été présenté à diverses occasions (GDR d'écologie chimique, au colloque d'écologie 2010, au congrès de métabolomique 2010), a fait l'objet d'une formation CNRS et d'un workshop dédié. MSeasy est adapté aux besoins de nombreuses équipes travaillant sur les problématiques d'écologie chimique. C'est donc tout logiquement que notre outil bioinformatique a été utilisé avec succès par d'autres études dont les systèmes expérimentaux sont très différents du notre, notamment une étude sur l'émission de COV par 55 espèces de familles différentes d'arbres tropicaux de Guyane (Courtois, Paine *et al.* 2009) et que des collaborations ont émergé (Jérôme Mardon, Juliana Ivanisevic). Différentes personnes nous ont contactés pour tester et utiliser notre module (Marie Charpentier, CEFE Montpellier sur les Mandrills, Ramon-Perez sur les coccinelles, Gregory Roeder, FARCE de l'institut de biologie de Neuchâtel). MSeasy devrait ainsi participer à la caractérisation exhaustive de la biodiversité des substances naturelles dans des espèces non modèles.

Dans cette partie de mes travaux, j'ai également développé un ensemble d'outils de bioinformatique destinés au prétraitement de données de GC-FID ou GC-MS issues du logiciel Chemstation d'Agilent. Ces outils comprennent : des macros développées sur Excel (1997 et 2002) et en langage Perl qui ont trouvé leurs applications au sein du BVpam dna sla gestion quotidienne des données de GC-MS et de GC-FID. J'ai réussi à développer des outils conviviaux ou efficaces mais la combinaison de ces deux points restent à approfondir. Leur utilisation impose pour le moment l'intervention d'un spécialiste pour traiter les données.

Résultats

Dans ce but et suite aux remarques des utilisateurs de MSeasy j'ai développé le module MSeasyTkGUI qui permet de ne plus avoir à saisir de ligne de commande pour lancer les analyses sous R. L'interface graphique offerte par ce dernier module est donc un plus pour la distribution de notre méthodologie.

Au delà des considérations techniques liées au développement du module cette partie de mes recherches met en avant la nécessité de standardisation des formats d'acquisition de données. Lors de cette thèse j'ai passé de nombreuses heures à convertir mes données de GC-FID, GC-MS, de qPCR et de séquençage dans les différents formats informatiques nécessaires à leur analyse. Les initiatives se multiplient d'ailleurs au sein des différentes communautés scientifiques pour faire émerger des formats communs d'échange de données et/ou développer des outils de bioinformatiques modulables capables de lire l'ensemble des formats de données. La standardisation des formats de données a de nombreux avantages et assurera une plus grande efficacité et pérennité aux collaborations entre laboratoires et dans le développement des potentialités des techniques d'analyses (Brazma, Krestyaninova *et al.* 2006). Au milieu des différents formats informatiques proposés une tendance à l'utilisation du format XML (eXtensible Markup Language, [http://fr.wikipedia.org/wiki/Extensible Markup Language](http://fr.wikipedia.org/wiki/Extensible_Markup_Language)) semble se dégager (Seibel, Kruger *et al.* 2006; Strömbäck, Hall *et al.* 2007). Les outils permettant de convertir les différents formats existant en XML commencent à apparaître (avec ici une liste de quelques convertisseurs <https://wiki.nbic.nl/images/7/70/Malick.pdf>). A cela s'ajoute des initiatives de standardisation des formats de présentation des données dans les publications scientifiques. Il s'agit le plus souvent de lister les informations minimales que doivent fournir les scientifiques lorsqu'ils publient leurs travaux. Ainsi, il existe le MIQE guideline : minimum information for publication of quantitative real-time PCR experiments qui liste les informations à fournir pour la publication de travaux de RT-qPCR et qPCR (Bustin, Benes *et al.* 2009). Dans le domaine de la métabolomique ont trouvé un équivalent : le MIAMet (Bino, Hall *et al.* 2004). Ces recommandations devraient permettre de faciliter les échanges de données entre groupes de travail et entre logiciels afin d'éviter des erreurs lors de la transformation des données entre différents formats (tabulations, minuscules, ordre des champs lexicaux...) et souvent un gaspillage de temps.

Evolution de la diversité des composés organiques volatils accumulés par les espèces du genre *Lavandula*



Quelques espèces botaniques de lavande en culture au BVpam

II Evolution des composés organiques volatils accumulés par les espèces du genre *Lavandula*

II-1 Introduction

Avec un nombre total estimé à plus de 230 000 espèces, le règne végétal représente une grande part de la diversité du monde des eucaryotes (revue dans Hassan, Scholes *et al.* 2005). Depuis les travaux des premiers chercheurs sur l'évolution du vivant (revue dans Beyer 1995), la diversité du règne végétal est restée un sujet de fascination mais aussi d'étonnement.

La diversification des végétaux est manifeste sur tous les tissus et organes qui peuvent adopter une remarquable diversité de formes et de tailles, favorisant ainsi l'adaptation à des environnements écologiques des plus divers. Même les aspects intimes touchant les mécanismes reproductifs, la protection des organes reproducteurs et la dispersion des descendants affichent une large palette de variations. Toutes ces différences morpho-anatomiques ont, depuis longtemps, été utilisées pour décrire la diversité des végétaux et élaborer des systèmes de classification complexes (Grayer, Chase *et al.* 1999). La notion d'espèce est alors devenue centrale car elle constitue la plus petite unité de regroupement d'individus ayant des caractéristiques homogènes et étant isolés reproductivement des autres pour des raisons physiologiques et/ou géographiques. Ces individus sont alors caractérisés par un ensemble de traits morpho-anatomiques qui, collectivement, et sur leurs plages de variation, les distinguent des autres. Les progrès récents en Biologie Moléculaire et en analyses statistiques des matrices comparatives de séquences d'ADN ont généré de puissants outils bioinformatiques de phylogénie moléculaire. Du fait de la large congruence des résultats qu'ils ont générés avec les systèmes de classification antérieurs, ils ont solidifié notre compréhension des relations évolutives entre taxons végétaux qui peuvent désormais être reliés entre eux sur des arbres phylogénétiques (Hilu et Liang 1997).

Une autre caractéristique majeure des végétaux est leur propension à produire une myriade de petites molécules spécifiques de leur règne et non-nécessaires aux mécanismes de survie cellulaire. Avec plus de 200 000 structures chimiques connues, ces métabolites végétaux dits secondaires (Granger et Passet 1973) représentent environ 10% des molécules naturelles terrestres (Ober 2005; Wink 1988). Ils sont également très inégalement répartis dans le règne végétal à l'image de ceux impliqués dans la pigmentation et le parfum des plantes. Cette

Résultats

observation et la suggestion qu'ils dérivent de mécanismes de diversification séquentiels de voies de biosynthèse ont suggéré qu'ils pourraient être utilisés pour affiner notre compréhension des relations évolutives entre taxons végétaux, une méthodologie appelée la chémotaxonomie (Weston 2010). Une telle relation a cependant été rapidement invalidée sur des bases théoriques. En effet, si initialement les métabolites secondaires végétaux avaient été considérés comme des résidus secondaires de voies de synthèse du métabolisme énergétique, il est maintenant reconnu qu'ils sont issus de voies de synthèse spécifiques et concourent largement à l'adaptation des végétaux à leur environnement biotique et abiotique (Holopainen et Gershenzon 2010) par des mécanismes aussi divers que l'attraction des pollinisateurs, la protection directe contre des pathogènes, la protection indirecte via l'attraction d'auxiliaires carnivores consommateurs de prédateurs et l'aide à l'association symbiotique. Ils ne représentent donc pas des marqueurs neutres de l'évolution. De plus, au sein d'une famille de molécules, tous les membres sont liés par l'action séquentielle d'enzymes. Tout évènement évolutif affectant une enzyme en amont de la voie de biosynthèse se traduira par la perte de nombreuses molécules (Schwab 2003). Les métabolites secondaires végétaux ne sont donc pas des marqueurs évolutifs indépendants. Une modification de leur pattern d'expression entre les tissus d'une même plante peut également être un évènement évolutif important sans pour autant être une perte de capacité à produire une molécule par un taxon (Tholl 2006). A ces considérations théoriques s'opposent des observations pratiques. Un codage méticuleux des évènements évolutifs ayant affecté les enzymes de biosynthèse (par opposition à un codage des molécules accumulées) a généré un schéma évolutif cohérent dans le genre *Cypripedium* (Barkman 2001). Le placement de 3 familles de molécules spécialisées sur un arbre phylogénétique des Angiospermes a montré que ces molécules étaient largement conservées dans des phylla évolutifs (Wink 2003). Cependant, cette même étude révélait aussi l'existence potentielle de phénomènes de coévolution car certaines molécules avaient été détectées dans des familles végétales non-relées évolutivement. Elle concluait donc à la possibilité d'utiliser les métabolites secondaires comme marqueurs taxonomiques mais avec précaution. Hélas, toutes ces études portaient soit sur un nombre trop limité de métabolites secondaires soit sur un nombre trop limité de taxons et de niveau taxonomiques pour pouvoir situer la grande diversité des métabolomes végétaux dans un contexte évolutif de spéciation et donc pour savoir si les pressions évolutives affectent de manière équivalente toutes les molécules accumulées par un taxon.

Résultats

Pour combler cette lacune, nous avons appliqué les outils modernes de métabolomique et de phylogénie moléculaire (comparaison des séquences d'un gène chloroplastique *trnK*) sur un échantillonnage stratifié de plantes du genre *Lavandula* (niveaux : sous-genre, section, espèce, sous-espèce, intra-populationnel et inter-tissus) pour observer l'évolution des métabolites secondaires volatils (composés organiques volatils – COV : dérivés d'acides gras, mono- et sesquiterpènes qui constituent le plus grand groupe de métabolites secondaires avec plus de 10 000 structures connues) sur un cadre phylogénétique solide. Le choix du genre *Lavandula* comme genre modèle est lié à 1/ la grande quantité, et diversité, de COV produits par les différentes espèces de ce genre (éléments constitutifs de l'huile essentielle utilisée en aromathérapie pour certaines de ces espèces), 2/ le grand nombre d'espèces (39 décrites à ce jour) et la diversité des niches écologiques qu'elles occupent et 3/ l'accord entre les différentes monographies sur la classification interne de la plupart des espèces (Upson 1997).

II-2 Résultats

Relations phylogénétiques au sein du genre *Lavandula*

Le préalable indispensable à notre étude sur l'évolution des composés organiques volatils (COV) dans le genre *Lavandula* était de disposer d'une phylogénie moléculaire robuste. Elle fut construite sur 30 taxons de notre échantillonnage (26 espèces, 2 sous-espèces et 2 hybrides – Tableau 3 et (Tableau 6) Pour cela, nous avons comparé les séquences du gène chloroplastique *trnK* du fait de sa grande capacité à mettre en évidence les relations phylétiques entre espèces chez les Angiospermes (Wilson 2004) et pour compléter des résultats, préliminaires, publiés par Upson (Upson 1997) sur les séquences nucléaires d'ITS et son traitement monographique du genre avec des critères morphoanatomiques (Upson et Andrews 2004).

L'alignement des séquences de *trnK* de nos 30 taxons (Annexe VIII : Annexes numériques) a généré une matrice de 1681 caractères (gaps inclus). De ces 1681 caractères, 1461 (86%) étaient conservés, 184 (11%) étaient variables et 90 (5%) étaient phylogénétiquement informatifs. La Figure 36 présente le regroupement cladistique obtenu avec la méthode d'agglomération du maximum de vraisemblance (calcul effectué avec le logiciel Phylip::dnaML). Les cladogrammes obtenus avec les méthodes de neighbor-joining, UPGMA et de parcimonie présentaient des topologies similaires mais avec des valeurs de

Résultats

bootstraps plus faibles. La concordance entre les cladogrammes obtenus avec ces différentes méthodes est indicatrice de la robustesse des arbres obtenus avec les séquences du gène *trnK*.

L'analyse de la comparaison des séquences du gène chloroplastique *trnK* est globalement congruente avec le découpage sectionnel du genre *Lavandula* proposé dans une monographie récente du genre sur la base d'une liste exhaustive de critères morphoanatomiques et de la comparaison des séquences des ITS présents sur ADN nucléaire (Upton 1997). Elle confirme en effet la très nette différenciation des deux sous-genres, *Lavandula* et *Fabricia* et des six sections (*i.e* *Lavandula*, *Stoechas*, *Dentatae*, *Pterostoechas*, *Subnudae* et *Chaetostachys*) que nous avons échantillonnés. Le point de non congruence dans le découpage sectionnel du genre concerne l'espèce *L. macra* qui, d'après notre analyse, est regroupée avec des espèces de la section *Pterostoechas* et non *Subnudae* comme plus largement admis d'après sa localisation géographique au nord est de l'Afrique et sa petite stature (Upton et Andrews 2004).

Les soutiens statistiques des différents nœuds sont souvent très forts sauf au sein de la section *Pterostoechas* et entre la section *Pterostoechas* et *chaetostachys*. Au sein de la section *Pterostoechas*, les valeurs bootstrap sont tout de même supérieures à celles obtenues sur la base des comparaisons des séquences des ITS (Upton 1997). Notamment, l'isolement de *L. maroccana*, *L. antinea*, *L. multifida*, et *L. pubescens* des autres espèces de la section *Pterostoechas* est mieux soutenu.

Résultats

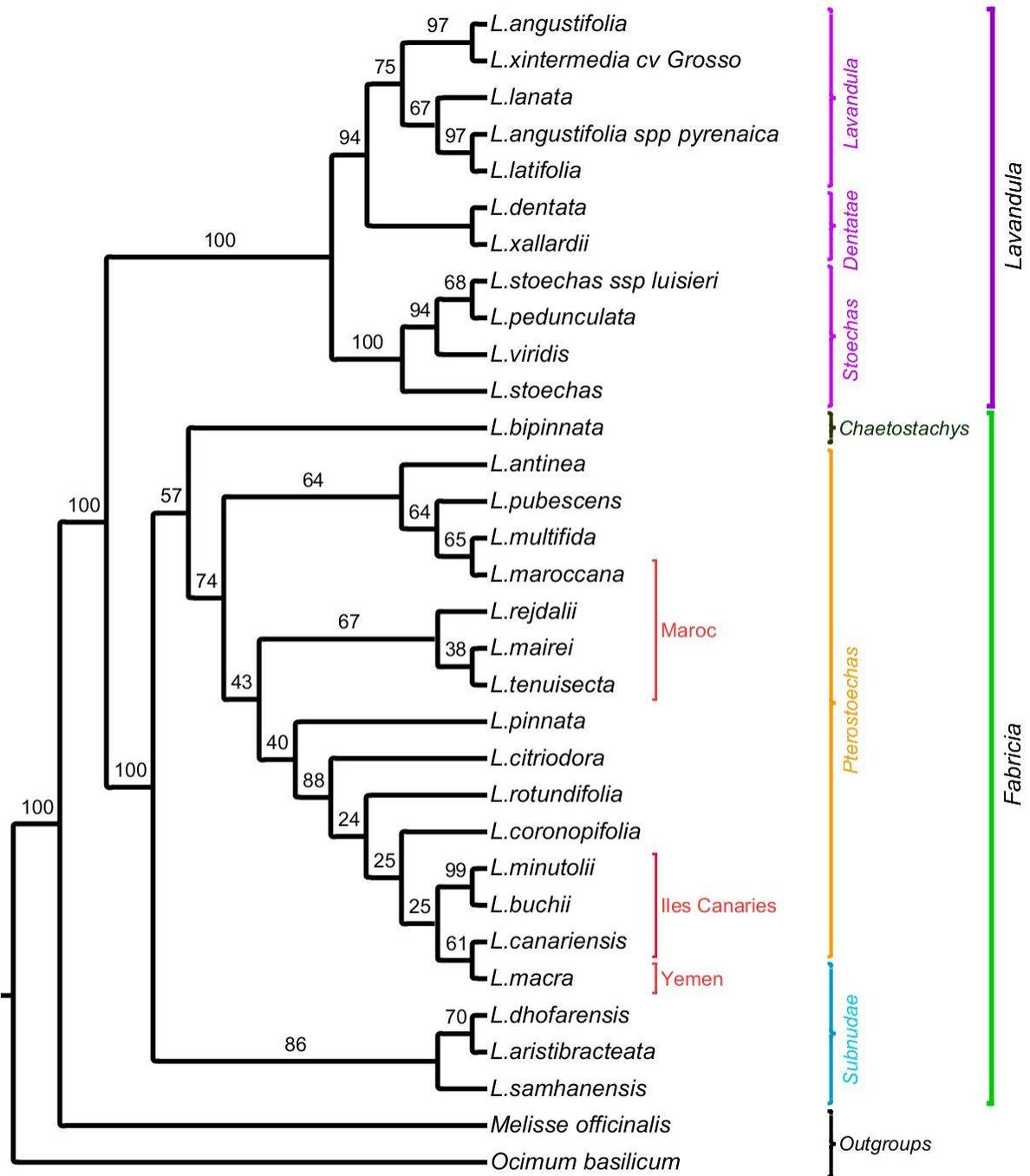


Figure 36 : Arbre phylogénétique du genre *Lavandula* basé sur la comparaison de séquences de *trnK* chez 30 taxons de lavandes.

Arbre consensus (50% majority rule) avec la méthode de maximum de vraisemblance (Phylip::dnaML). Les valeurs de bootstrap figurent au-dessus des branches (1000 bootstraps). Les 30 taxons de lavande analysés représentent 26 espèces, 2 sous-espèces et 2 hybrides. Deux des trois sous-genres sont représentés (Barre **violette** : sous-genre *Lavandula*, Barre **verte** : sous-genre *Fabricia*) ainsi que 6 des 8 sections du genre (Barre **jaune** : Section *Pterostoechas*, **Violet** : Sections *Lavandula*, *Stoechas* et *Dentatae*, **Vert foncé** : Section *Chaetostachys*, **Bleu** : Section *Subnuda*). Figure réalisée avec TreeGraph2 (Stover et Muller). La mélisse officinale et le basilic ont été utilisés pour enracer l'arbre. La localisation de certaines espèces est également indiquée en **rouge**.

Résultats

Un point majeur de non-congruence entre les phylogénies moléculaires reconstruites sur la base de comparaisons des séquences des gènes *trnK* et ITS concerne le placement de *L. bipinnata*, la seule espèce représentante de la section *Chaetostachys* dans nos analyses. Selon les arbres retenus, elle est soit à la base des sections *Pterostoechas* et *Subnudae* soit sœur de la section *Subnudae*. La place de *L. bipinnata* au sein du genre *Lavandula* sur la base de critères morphoanatomiques est cependant également sujette à débat (Upson 1997) et il faut noter que la valeur bootstrap obtenue lors de nos analyses pour le positionnement de la section *Chaetostachys* au sein du sous-genre *Fabricia* est faible (57%),

Notre analyse phylogénétique est en grande partie congruente avec la distribution géographique des espèces considérées. Les deux sous-genres séparent les espèces du nord de la méditerranée de celles du sud, la section *Lavandula* du sous-genre *Lavandula* regroupant les espèces les plus nordiques. Seules deux espèces du sous-genre *Lavandula* (*L. dentata* et *L. stoechas*) font exception en ayant une aire de distribution qui inclue les deux rives de la mer méditerranée. Les espèces du sous-genre *Fabricia* présentent aussi des regroupements sur des aires phytogéographiques communes malgré des valeurs bootstrap pas toujours très élevées. Par exemple la section *Subnudae* regroupe des espèces présentes uniquement à l'est de l'Afrique du nord alors que la section *Pterostoechas* regroupe principalement des espèces de l'Afrique de l'ouest à quelques exceptions près. Au sein de cette section, un groupe d'espèces marocaines (*L. rejdalii*, *L. mairei* et *L. tenuisecta*) se trouve séparé d'un groupe d'espèces principalement originaires des îles de la Macaronésie (*L. pinnata*, *L. canariensis*, *L. buchii*, *L. minutolii* et *L. rotundifolia*). A ces espèces Macaronésiennes, s'ajoutent *L. coronopifolia*, qui est à la fois présente sur le continent et sur les îles du Cap Vert. Les seules exceptions notables à cette présence des espèces de la section *Pterostoechas* à l'ouest de l'Afrique sont *L. citriodora* et *L. macra* qui sont endémiques des bords de la mer rouge. Plusieurs autres espèces présentent de plus une disjonction d'aire de distribution avec des populations présentes sur les deux bords du désert saharien (Tableau 3).

En conclusion, la forte congruence entre les évolutions des caractères morphoanatomiques, de la distribution géographique et des caractères moléculaires liés aux génomes chloroplastique (*trnK*) et nucléaire (ITS) suggère que l'évolution de tous ces caractères s'est faite parallèlement à la différenciation des espèces de lavande lors de l'évolution du genre. Cette situation pose ainsi un cadre idéal pour décrire et comparer les

Résultats

évolutions individuelles des COV accumulés par les taxons de ce genre lors de leurs spéciations.

Nature des COV formés par les espèces du genre *Lavandula*

La diversité des COV accumulés par la totalité de notre échantillonnage stratifié (29 espèces, 3 hybrides, 11 sous-espèces, 13 variétés et deux tissus, feuilles et fleurs, (Tableau 6) a été appréhendée par analyse GC-MS (colonne polaire HP-Innowax) d'extraits hexane. Afin de détecter d'éventuelles différences dans les COV accumulés par les membres d'un même taxon, nous avons ajouté une strate d'échantillonnage supplémentaire en analysant un minimum de 5 plantes différentes (souvent 10 plantes) par taxon. Ces plantes étaient issues de semis d'un mélange de graines prélevées sur une ou plusieurs populations sauvages. Au total, 514 échantillons ont été analysés. Afin de déterminer la reproductibilité de nos analyses, 58 échantillons ont été sélectionnés aléatoirement pour être ré-analysés de manière indépendante. La nature chimique des molécules détectées (et leur pureté dans les pics chromatographiques) a été définie en analysant par GC-MS sur une colonne de polarité inversée (colonne apolaire DB-5) 1 échantillon de chaque espèce (29 échantillons au total). Un total de 600 analyses a ainsi été réalisé. Sur les 29 espèces de notre échantillonnage, 17 n'avaient jamais été décrites en termes de production de COV (Tableau 6). Afin de réduire l'influence du bruit de fond sur les spectres de masse des COV de lavande durant les analyses, seuls les pics chromatographiques dont la hauteur maximale dépassait d'un facteur 3 l'intensité du bruit de fond (sélection d'une valeur threshold de 14 sur le logiciel chemstation d'Agilent) ont été considérés. L'analyse de nos 600 échantillons conduisit ainsi à l'identification de 13 260 pics dont le prétraitement fut effectué avec le logiciel MSeasy développé lors de mon travail de thèse (voir chapitre résultats ci-dessus dans ce manuscrit). Il put être réalisé sur un ordinateur standard (à titre d'indication, un ordinateur personnel de 2 G0 de RAM a la capacité mémoires pour « prétraiter » jusqu'à 20 000 pics). Le prétraitement de ces 13 260 pics par MSeasy a conduit à la création d'une quantité optimale de 138 clusters regroupant des pics dont les spectres de masse sont similaires (Figure 37). Une fois le nombre optimal de clusters déterminé, la fonction MS.clust a permis de finaliser le prétraitement en fournissant les fichiers PDF et texte utilisable pour créer une matrice de présence/absence de chaque cluster pour chaque échantillon.

Résultats

Chaque cluster peut cependant contenir plus d'une molécule. Ils peuvent en effet soit contenir plusieurs molécules dont les spectres de masse sont proches mais dont les indices de rétention sont éloignés soit regrouper des pics chromatographiques résultant d'un mélange non-séparé de molécules. Afin d'établir, à partir de nos 138 clusters, une liste des COV présents dans nos échantillons, nous avons 1/ créé des sous-clusters de pics ayant des temps de rétentions homogènes dans chaque cluster et 2/ supprimé tous les pics pour lesquels il était impossible de trouver un spectre de masse identique lors des analyses comparatives sur les deux colonnes de polarité inversée. L'assignation d'un nom à ces molécules s'est ensuite faite sur la base de comparaison avec des bibliothèques de données des indices de rétentions sur colonne apolaire (Adams 2007) et des spectres de masse (McLafferty 2006) Pour limiter la présence de faux-positifs, la matrice de présence/absence des COV pour chaque échantillon a été réduite en supprimant 1/ les molécules détectées dans moins de 5% des échantillons d'un taxon et 2/ les échantillons atypiques par rapport à ceux de la même espèce d'après une analyse HCA (distance = Similarité : Indice de Sokal & Sneath(1) et lien moyen) réalisée avec Xlstat (nombre d'outliers : 6 échantillons de fleur et 23 de feuille). L'analyse des réplicatas techniques a validé la reproductibilité de nos analyses chromatographiques (moins de 10% de variation de l'aire des pics entre deux duplicatas). In fine, nous avons obtenu une matrice de 141 molécules : 58 monoterpènes, 32 sesquiterpènes, 6 dérivés d'acides gras et 45 molécules inconnues dont il a parfois été possible de déterminer la classe chimique d'après la présence de fragments de masse typiques lors de l'analyse des spectres de masse (Annexe V : Liste des 141 COV accumulés dans les taxons de lavandes). Une matrice listant la présence ou absence de chacune de ces 141 molécules dans nos 514 échantillons a ensuite pu être assemblée. C'est cette matrice que nous avons utilisée pour nos analyses ultérieures (Annexe VIII : Annexes numériques).

Résultats

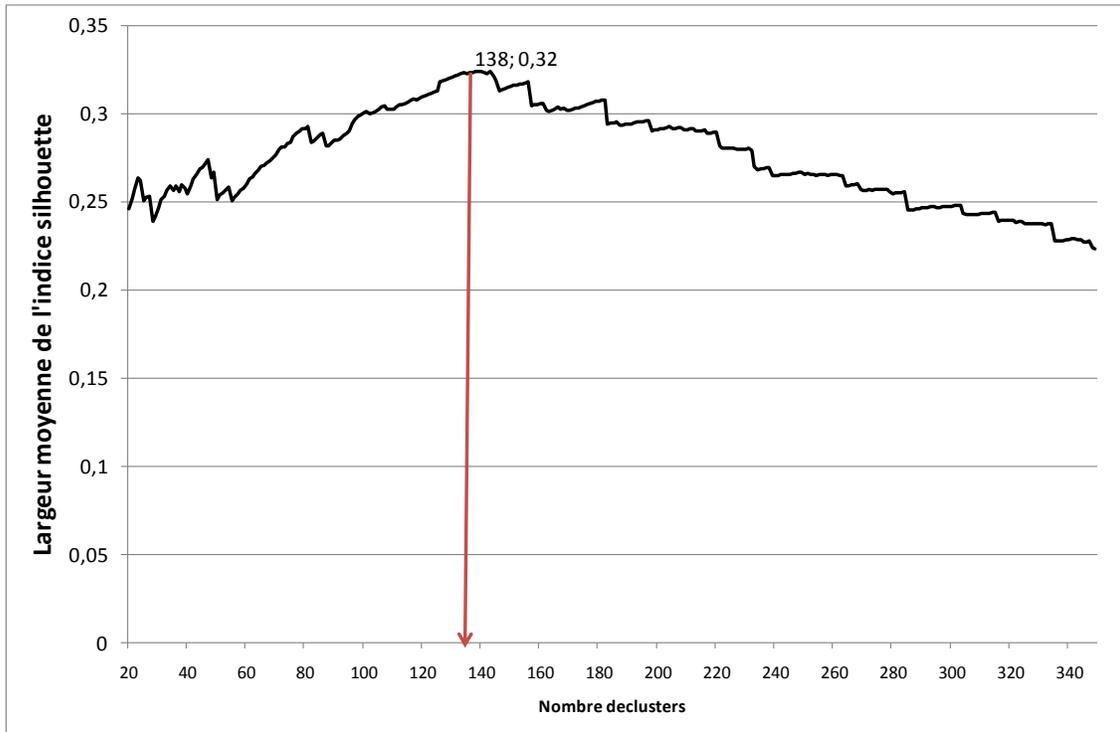


Figure 37: Variation de la largeur moyenne de l'indice silhouette en fonction du nombre de clusters

Variation de la largeur moyenne de l'indice silhouette entre 20 et 350 clusters pour le jeu de données de CTV de lavande comportant 13 260 pics. La flèche rouge indique la valeur la plus forte de largeur moyenne de l'indice silhouette et le nombre optimal de clusters. Cette figure a été obtenue en cumulant les résultats de différents fichiers *res_clValid.txt* générés par la fonction *MS.clust* du module *MSeasy*.

Résultats

Distribution des COV au sein des taxons et tissus de lavande

Les bouquets de COV accumulés par les différentes espèces de lavande étaient très différenciés. Aucun COV n'a, en effet, été détecté dans l'ensemble des taxons analysés. Une large proportion de COV n'a été détectée qu'au sein d'un même sous-genre ou d'une même section (48%, synapomorphie), 24% étaient spécifiques d'un taxon (autapomorphie) et 28% étaient homoplasiques au sein du genre (Annexe VI : Liste des synapomorphies, autapomorphies et homoplasies).

- Un caractère homoplasique, ou homoplasie est la similitude d'un caractère chez différentes espèces. Contrairement à l'homologie, celle-ci ne provient pas d'un ancêtre commun. Il existe différents types d'homoplasie : la convergence, le parallélisme et la réversion.
 - la convergence : ressemblance apparue indépendamment dans divers taxons sans lien de parenté.
 - le parallélisme : ressemblance apparue chez des taxons relativement proches phylogénétiquement.
 - la réversion : un état dérivé d'un caractère revient à un état ancestral (antérieur).
- Un caractère autapomorphique, ou autapomorphie est un caractère dérivé (ou apomorphique), propre à un taxon. Les caractères autapomorphiques permettent de définir les caractéristiques propres d'un taxon, voire d'une espèce.
- Un caractère synapomorphique, ou synapomorphie est un caractère dérivé (ou apomorphique), partagé par deux ou plusieurs taxons phylogénétiquement proches.

Pour chaque tissu (fleur et feuille), nous avons calculé le nombre total de COV détectés dans les échantillons de chaque sous-genre ou chaque section (Figure 38). Une analyse ANOVA a révélé qu'il y avait significativement plus de COV détectables dans nos échantillons de fleurs et de feuilles du sous-genre *Lavandula* (ANOVA sur l'effet sous-genre ddl=2 F=126 P-value<0.0001 et ddl=2 F=747 P-value<0.0001 respectivement). Une conclusion similaire pouvait être tirée en comparant les sections des deux sous-genres entre elles, et ce malgré le fait que notre échantillonnage contenait un nombre plus limité de taxons dans le sous-genre *Lavandula* que dans le sous-genre *Fabricia* (Test de Tuckey après ANOVA sur l'effet section

Résultats

ddl=4 F=20 P-value<0.0001 et ddl=5 F=34 P-value<0.0001 respectivement). En conclusion, les espèces du sous-genre *Lavandula* se distinguent des espèces du sous-genre *Fabricia* par leur capacité à accumuler à des niveaux détectables une palette plus diversifiée de COV. Toutes les familles majeures de composés chimiques avaient, cependant, des représentants dans les deux sous-genres.

Toutes nos extractions et analyses de COV ont été réalisées à poids de tissu constant pour permettre une comparaison entre taxons de leurs capacités d'accumulation d'une molécule. La non-détection d'une molécule dans un taxon ne signifie donc pas que cette molécule n'est pas produite par ce taxon (l'analyse d'un extrait concentré l'aurait peut-être mise en évidence) mais indique que cette molécule y est beaucoup moins abondante (voire absente) que dans les taxons où elle a été détectée et qu'elle ne pourra pas y assurer son rôle biologique à la même intensité.

Résultats

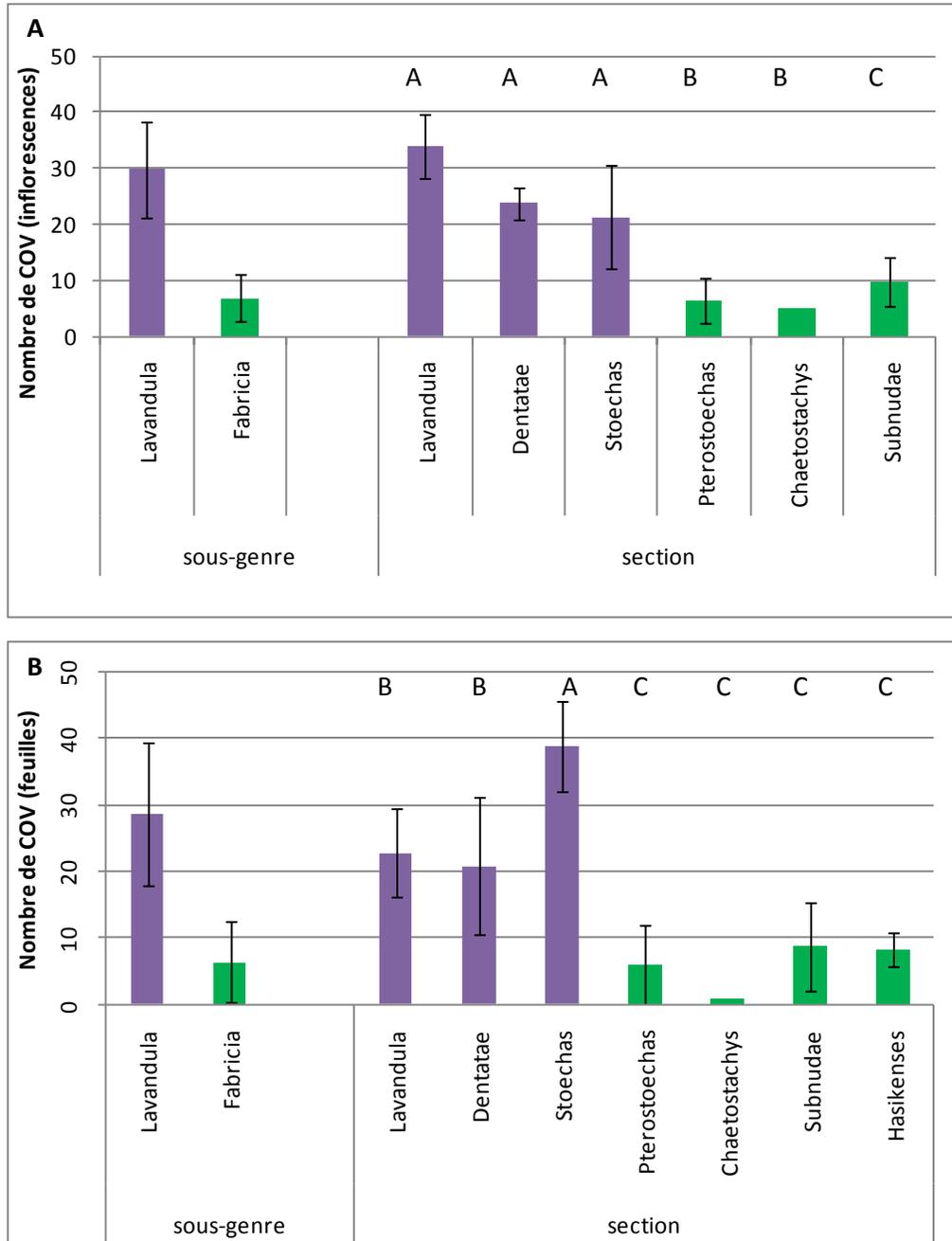


Figure 38 : Comparaison des nombres de COV détectés dans les feuilles et dans les fleurs de lavande à deux niveaux taxonomiques (sous-genre et section). Les barres violettes correspondent au sous-genre *Lavandula* et à ses sections alors que les barres vertes correspondent au sous-genre *Fabricia* et à ses sections.

A : comparaison du nombre de COV détectés dans les inflorescences de lavande pour les sous-genres *Lavandula* et *Fabricia* ainsi qu'au niveau des sections. Les barres d'erreur représentent l'erreur standard sur la moyenne par comparaison des différents taxons.

B : comparaison du nombre de COV détectés dans les feuilles de lavande pour les sous-genres *Lavandula* et *Fabricia* ainsi qu'au niveau des sections. Les barres d'erreur représentent l'erreur standard sur la moyenne par comparaison des différents taxons. Les lettres au-dessus des histogrammes indiquent les différences significatives (Test de Tuckey). Les analyses ont été réalisées avec le logiciel XLStat.

Résultats

Une analyse en coordonnées principale (ACoP) de la variabilité des COV accumulés par nos échantillons de lavande a révélé que les échantillons des deux sous-genres de lavande constituent deux sous-groupes distincts (Figure 39). Ceci est lié au fait que 47 des 141 molécules étudiées présentaient une distribution synapomorphique entre les deux sous-genres (Annexe VI : Liste des synapomorphies, autapomorphies et homoplasies). La présence de COV synapomorphiques des sections *Lavandula*, *Stoechas* et *Pterostoechas* (20 molécules au total) explique également la formation de groupes distincts par les échantillons de ces sections par rapport aux autres échantillons des autres sections de leurs sous-genres sur la représentation graphique de l'ACoP (Figure 39).

L'analyse ACoP de la composition en COV de nos échantillons de lavande révèle également que les échantillons du sous-genre *Lavandula* ne contiennent pas seulement plus de COV que ceux du sous-genre *Fabricia* comme déterminé ci-dessus (Figure 38), ils sont aussi plus différents entre eux quant à leur composition en COV (distribution des points-échantillons plus large sur la représentation graphique de l'analyse ACoP Figure 39). Une analyse de la répartition de la variance de la distribution des COV au sein du genre *Lavandula* a révélé que seulement 25% et 19% de la variance portaient respectivement sur les niveaux sous-genre et sectionnel et que la majorité de la variance (56%) se situait entre échantillons au sein d'une même section (Figure 40 A). Cette part de variance était majoritairement due à des différences entre taxa (44%), la part de variabilité entre individus d'un même taxon étant cependant relativement élevée (31%) et presque aussi importante que la variabilité inter-taxa (probablement dû au fait que certaines molécules peu abondantes se sont trouvées sous le seuil de détection dans certains échantillons –Figure 40 B).

Même si l'évolution de la production de COV par les végétaux supérieurs peut se faire via des modifications du pattern d'expression tissulaire, seulement 25% de la variabilité intra-sectionnelle est due à des différences d'expression entre feuilles et fleurs (Figure 40 B). Une analyse espèce par espèce de la différenciation des COV accumulés par les feuilles et les fleurs a, en effet, révélé que la majorité des COV accumulés par une espèce sont détectables dans ces deux tissus (Tableau 9). Cependant, toutes les espèces de notre échantillonnage différenciaient leurs feuilles et leurs inflorescences en y accumulant spécifiquement un petit nombre de COV. De manière surprenante, cette analyse comparative a révélé que les différentes espèces diffèrent par la proportion relative de COV qu'elles accumulent spécifiquement dans leurs feuilles ou fleurs. La distribution de ce caractère semble très variable et ne corrèle pas avec les relations phylogénétiques entre espèces mis à part dans le sous-genre *Lavandula* où les

Résultats

espèces de la section *Stoechas* accumulent toutes une palette de COV plus diversifiée dans leurs feuilles que leurs fleurs et celles de la section *Lavandula* accumulent toutes une plus grande richesse de COV dans leurs inflorescences que dans leurs feuilles.

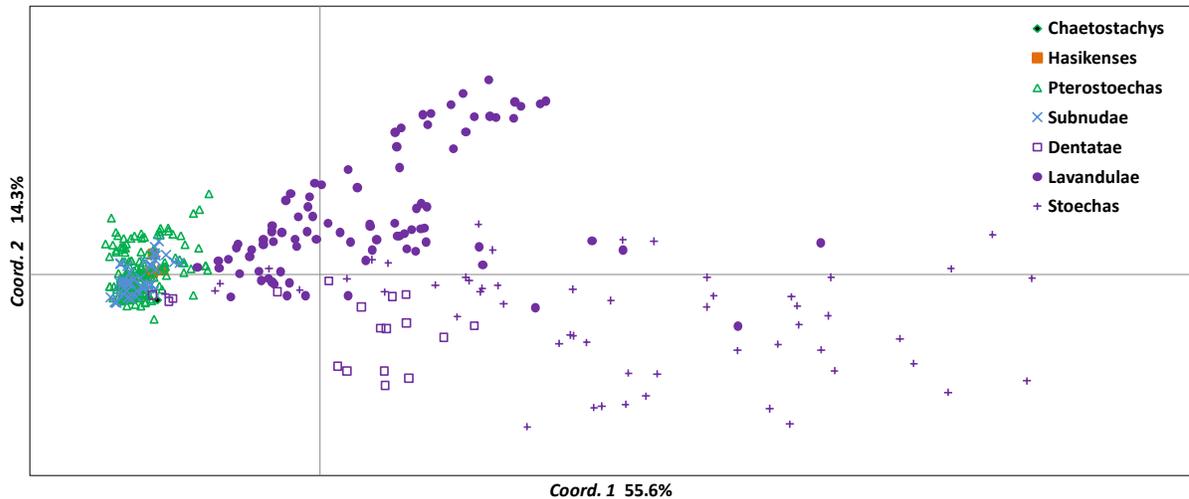


Figure 39 : Analyse en coordonnées principales (ACoP) de la composition en COV des échantillons de lavandes. L'ACoP a été réalisée avec le logiciel Genalex version 6. Les échantillons sont regroupés par sections. Les échantillons des sections du sous-genre *Lavandula* ont des teintes violettes et ceux des sections du sous-genre *Fabricia* ont des teintes vertes ou orange.

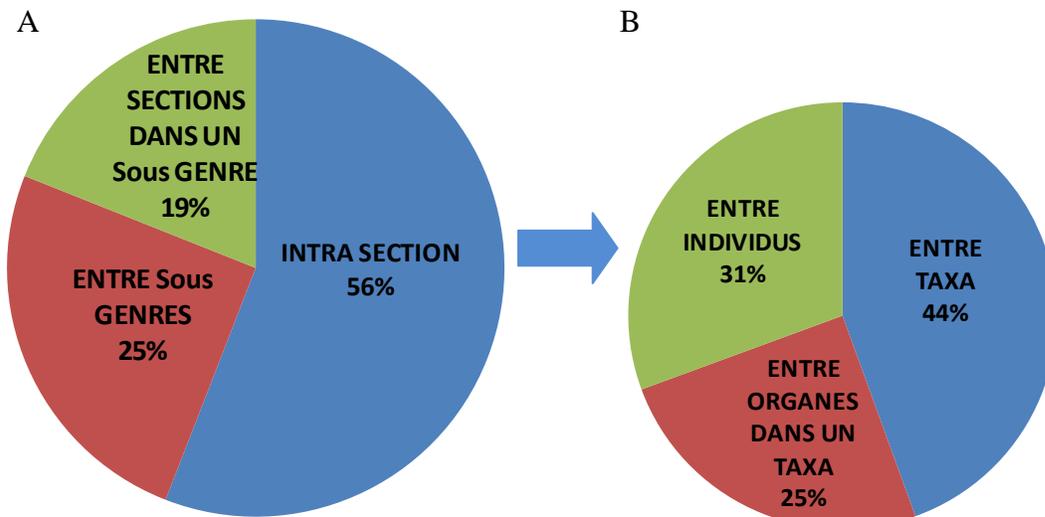


Figure 40 : Distribution de la variance de la composition en COV des échantillons de lavandes entre strates de notre échantillonnage. Analyse réalisée avec le logiciel Genalex.

A : Distribution entre sous-genre et section, B : Distribution intra section entre individus, taxa et entre organes dans un taxa

Résultats

Tableau 9 : Différentiation des COV accumulés par les feuilles et les fleurs de 23 espèces de lavandes
Le tissu ayant le plus de COV spécifique est mis en valeur en orange

Sous-genre	Espèce	Fleur	Commun	Feuille
<i>Lavandula</i>	L. angustifolia	14	56	9
	L. latifolia	19	39	8
	L. lanata	19	54	7
	L. dentata	8	42	18
	L. pedunculata	1	56	60
	L. viridis	0	13	77
	L. stoecha	1	49	47
<i>Fabricia</i>	L. antinea	9	13	3
	L. multifida	3	12	22
	L. pubescens	1	13	14
	L. marocana	2	9	9
	L. mairei	1	22	20
	L. rejdalii	4	9	7
	L. tenuisecta	5	7	2
	L. pinata	9	1	1
	L. coronopifolia	3	7	9
	L. citriodora	1	15	8
	L. dhofarensis	1	1	15
	L. aristibracteata	12	20	1
	L. canariensis	8	21	2
	L. macra	2	13	16
	L. minutolii	7	28	6
	L. buchii	8	2	2

Résultats

Relations entre la présence d'un COV, son origine biosynthétique, les relations phylétiques entre taxon et l'environnement écologique des taxons

Une répartition des COV en catégories chimiques a révélé de surprenantes coïncidences de relation entre la conservation d'un COV au sein d'un phylum évolutif du genre *Lavandula* et son appartenance à une catégorie chimique (Tableau 11). La quasi-totalité substances linéaires simples n'ayant subi aucune modification chimique après synthèse de leur schéma carboné de base (oxydation sous forme d'aldéhyde ou cétone, acétylation) ou ne provenant pas d'une synthèse atypique (attaque tête-tête de deux DMAPP, isomérisation *trans-cis*, ajout d'un cycle époxy) avaient en effet une distribution homoplasiques et ce, quelle que soit la catégorie de composé considérée, dérivés d'acides gras, mono- et sesquiterpènes.

Une autre remarque surprenante est que même si les terpènes synthases sont capables de fabriquer des composés cycliques aliphatiques ou hydroxylés à partir du GPP (monoterpènes) ou FPP (sesquiterpènes), les composés cycliques hydroxylés sont beaucoup moins homoplasiques que les composés aliphatiques (14% contre 69% pour les composés monoterpéniques et 8% contre 38% pour les composés sesquiterpéniques). Les composés bi- ou tri-cycliques constituaient la majorité des composés synapomorphiques monoterpéniques, le bornéol, le camphre et le camphène étant même les seules substances à présenter des synapomorphies strictes sur un sous-genre. Pareillement, les composés sesquiterpéniques cycliques aliphatiques qui présentent une plus grande complexité (et donc unicité) de structure que les composés monoterpéniques cycliques aliphatiques, avaient, en parallèle, plus fréquemment une distribution synapomorphique (38% contre 23%).

Résultats

Tableau 10 : Distribution des COV par catégorie chimique au sein du genre *Lavandula*.

Les synapomorphies retenues sont au niveau des sous-genres ou des sections et les informations ont été compilées à partir des données présentes dans la matrice des COV détectés dans les échantillons (Annexe Numérique).

Famille de molécule	Synapomorphies	Homoplasies	Autapomorphies	Total
Dérivé AG alcool	0	3	1	4
Dérivé AG acétylé	0	1	1	2
Total dérivés AG				6
Monoterpène cyclique aliphatique	3	9	1	13
Monoterpène cyclique alcool*	10	2	2	14
Monoterpène cyclique acétylé	1	0	1	2
Monoterpène cyclique aldéhyde, cétone	6	1	2	9
Monoterpène cyclique benzénique	1	2	1	4
Monoterpène linéaire aliphatique	0	4	0	4
Monoterpène linéaire alcool	1	2	0	3
Monoterpène linéaire acétylé	1	2	0	3
Monoterpène linéaire aldéhyde	2	0	0	2
Monoterpène atypique	3	0	1	4
Total monoterpènes				58
Sesquiterpène cycliques aliphatique	8	8	5	21
Sesquiterpène cycliques alcool	4	1	2	7
Sesquiterpène atypique	1	1	0	2
Sesquiterpène linéaire aliphatique	0	2	0	2
Total sesquiterpènes				32
Inconnu	26	2	17	45
Total	67	40	34	141

* inclus 1,8-cinéol

Résultats

Tableau 11 : Distribution des couples COV acétylé/oxydés et substrats potentiels au sein du genre *Lavandula*.

n.p. non-prédictible (voies de synthèse inconnue ou substrat potentiel non-déTECTÉ dans nos échantillons). * origine biosynthétique soumise à débat ou inconnue. Synapomorphie et autapomorphie portaient respectivement sur les mêmes sections ou taxon sauf pour a *L. stoechas subsp. luisieri* ; b *L. stoechas subsp. luisieri* et *L. pedunculata* ; c sous-genre *Lavandula* ; d section *Pterostoechas*

COV	Distribution	Substrat potentiel	Distribution substrat potentiel
COV acétylé			
3-acétate d'octyle	homoplasique	3-octanol	homoplasique
1-acétate d'octyle	autapomorphie	n.p.	-
acétate de <i>trans</i> - α -nécrodyde	autapomorphie ^a	<i>trans</i> - α -nécrodol*	autapomorphie ^b
acétate de bornyle	synapomorphie	bornéol	synapomorphie
acétate de linalyle	homoplasie	linalol	homoplasie
acétate de géranyle (<i>trans</i>)	homoplasie	trans-géraniol	homoplasie
acétate de néryle (<i>cis</i>)	synapomorphie ^c	<i>cis</i> -géraniol*	synapomorphie ^d
COV oxydé			
myrténal	synapomorphie	myrténol	synapomorphie
α -campholéal	synapomorphie	n.p.	-
<i>cis</i> -citral	synapomorphie	n.p.	-
<i>trans</i> -citral	synapomorphie	n.p.	-
camphre	synapomorphie	bornéol	synapomorphie
fenchone	homoplasie	fenchol	homoplasie
pipéritone	autapomorphie	n.p.	-
carvenone	autapomorphie	n.p.	-
pinocarvone	synapomorphie	pinocarvéol	synapomorphie
verbenone	synapomorphie	n.p.	-
carvone	synapomorphie	<i>trans</i> -carvéol	synapomorphie

Afin de souligner d'éventuelles associations biosynthétiques entre COV pouvant expliquer leur co-présence dans un taxon, nous avons réalisé une analyse de corrélation de présence/absence des COV par paires sur l'ensemble de notre échantillonnage (coefficient de Pearson et agglomération de Ward - Figure 42 - (Tikunov, Lommen *et al.* 2005). Un premier carré de corrélations concerne des monoterpènes cycliques (camphène, camphre et bornéol) qui partagent tous le même intermédiaire carbocationique (cation camphyle) au niveau d'une étape enzymatique réalisée par une terpène synthase (carré C - Jarvis 2000). De nombreuses terpènes synthases ayant été décrites pour générer plusieurs produits à partir d'un même substrat (Degenhardt, Köllner *et al.* 2009), il est possible que le camphène et le bornéol soient produits par une même enzyme mais une telle relation n'a encore jamais été décrite chez ces plantes. Le camphre étant, quant à lui, une forme oxydée du bornéol, il est possible qu'il en dérive par l'action d'une enzyme oxydante qui reste à déterminer. Un deuxième carré regroupe

Résultats

le sabinène, sabinène hydrate et le *delta*- et l' *alpha*- terpinéol et suggère des relations similaires car elles partagent le carbocation terpinyle comme intermédiaire réactionnel au niveau d'une étape enzymatique catalysée par une terpène synthase (carré E - Jarvis 2000). Bien que leur voie de synthèse soit sujette à débat, le *cis*-géraniol, le *cis*-citral et le *trans*-citral (carré F) étaient co-présents dans deux espèces (*L. coronopifolia* et *L. citriodora*), suggérant ainsi une origine biosynthétique commune bien que, comme pour les carrés précédents, les démonstrations expérimentales restent à réaliser. De même, les sesquiterpènes hydroxylés cubébol et viridiflorol étaient co-présents dans de nombreuses espèces communes (carré B) sans que les terpènes synthases qui les produisent n'aient pas déjà été décrites. Les autres groupements réunissent des substances sur le simple fait qu'elles sont autapomorphiques dans une même espèce. Dans le cas des acétates de *cis*- et *trans-alpha*-nécrodol qui sont co-présents, et uniquement décrits dans le règne végétal, chez *L. luisieri* (carré A - Garcia-Vallejo, Garcia-Vallejo *et al.* 1994), il est possible qu'ils soient issus d'une voie de synthèse commune. Mais dans d'autres cas, ces regroupements ne sont pas liés à une origine biosynthétique commune (carré D - *alpha*-fenchène et érémosphylène présents uniquement dans les échantillons de feuilles de *L. pedunculata* et regroupés malgré le fait que le premier soit un monoterpène et le deuxième un sesquiterpène).

Les substances dont on peut prédire avec le plus de certitude la molécule précurseur concernent les dérivés acétylés ou oxydés (forme cétonique ou aldéhydique) car ces étapes sont souvent conduites par des acétyl transférases et des Cyt P₄₅₀ (Jarvis 2000). De nombreuses peuvent être détectées dans les COV de lavande (Tableau 11). Même si un seul couple substance modifiée-molécule précurseur potentiel s'est trouvé statistiquement associé dans de nombreuses espèces (couple camphre-bornéol cité ci-dessus –Figure 42), tous les couples ainsi formés regroupent tout de même des substances accumulées par les espèces d'une même section botanique ou étant toutes deux homoplasiques pour le genre (Tableau 11). Malgré une distribution globalement similaire des deux molécules couple, il existe de nombreuses espèces qui n'accumulent que l'un des deux partenaires. Dans la plupart des cas, c'est le partenaire le plus modifié chimiquement (le plus terminal sur une voie de synthèse) qui fut trouvé dans un nombre plus restreint de taxons. Ce fut le cas par exemple de l'acétate de linalyle par rapport au linalol (Figure 41 C,D) ou du myrténal par rapport au myrténol (Figure 41 A,B). Dans de rares cas, le partenaire le moins modifié était absent (vraisemblablement dû à sa trop grande consommation lors du processus biosynthétique qui

Résultats

l'avait rendu non-délectable). Par exemple, le *trans*-géraniol et l'acétate de *trans*-géranyle étaient présents chez les mêmes espèces à l'exception près que la forme acétylée avait également été détectée chez *L. samhanensis*, *L. latifolia* et *L. angustifolia*.

Afin de déterminer si la présence d'un COV dans une espèce peut être liée à l'adaptation de cette espèce à un paramètre environnemental, des calculs de régression logistique ont été effectués entre la présence de chaque COV et l'altitude ou la latitude de l'aire sauvage de distribution des taxons la produisant. Ces analyses ont révélé une corrélation positive significative ($R^2=0.6$, $p<0,001$) entre l'accumulation de bicyclogermacrène et l'altitude ainsi qu'entre la latitude et la présence d'*alpha*-gurjunène, *beta*-bisabolène, *beta*-pinène, carvone, lédol, myrténol, pinocarvone, terpinène-4-ol, transpinocarvéol, *trans*-carvéol, verbénone et viridiflorol. Ces premiers résultats demandent à être approfondis. En effet, pour l'altitude deux espèces sont concernées *L. mairei* et *L. tenuisecta*. Il faudrait avoir accès à des populations de ces espèces issuent d'altitudes plus variées et prendre en compte la teneur en bicyclogermacrène plutôt qu'uniquement la présence/absence. Pour la relation avec la latitude les résultats sont peut-être liés au fait que les espèces produisant quantitativement plus d'HE sont au nord de l'aire de distribution et au fait que bon nombre de nos échantillons proviennent d'Espagne et du Maroc, il faudrait avoir plus d'analyses de taxons du sud-est de l'aire de répartition et peut-être concentrer ces extraits.

Résultats

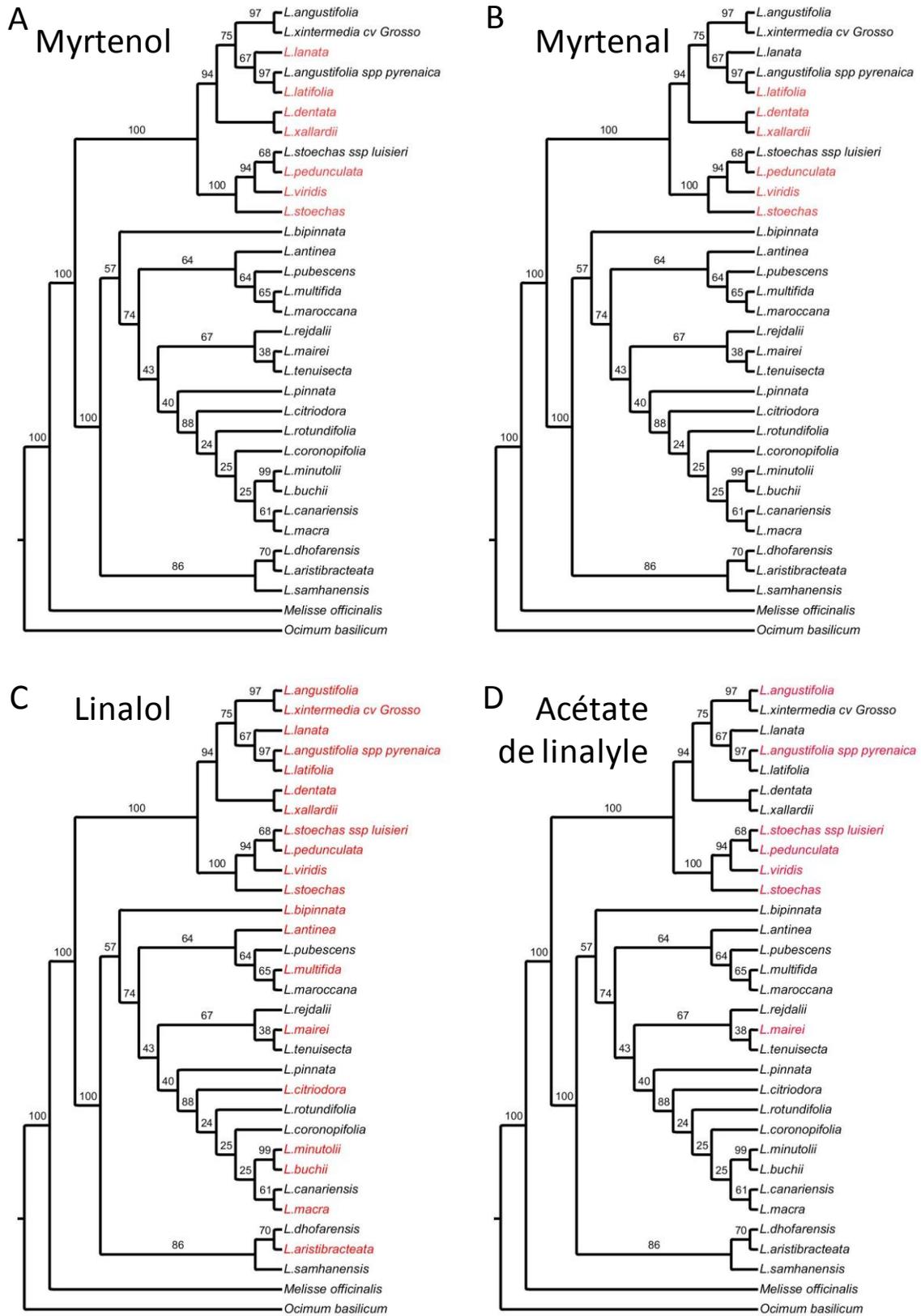


Figure 41 : Comparaison des taxons accumulant le myrtenol, le myrtenal, le linalol et son acétate
En rouge les espèce accumulant le COV.

Résultats

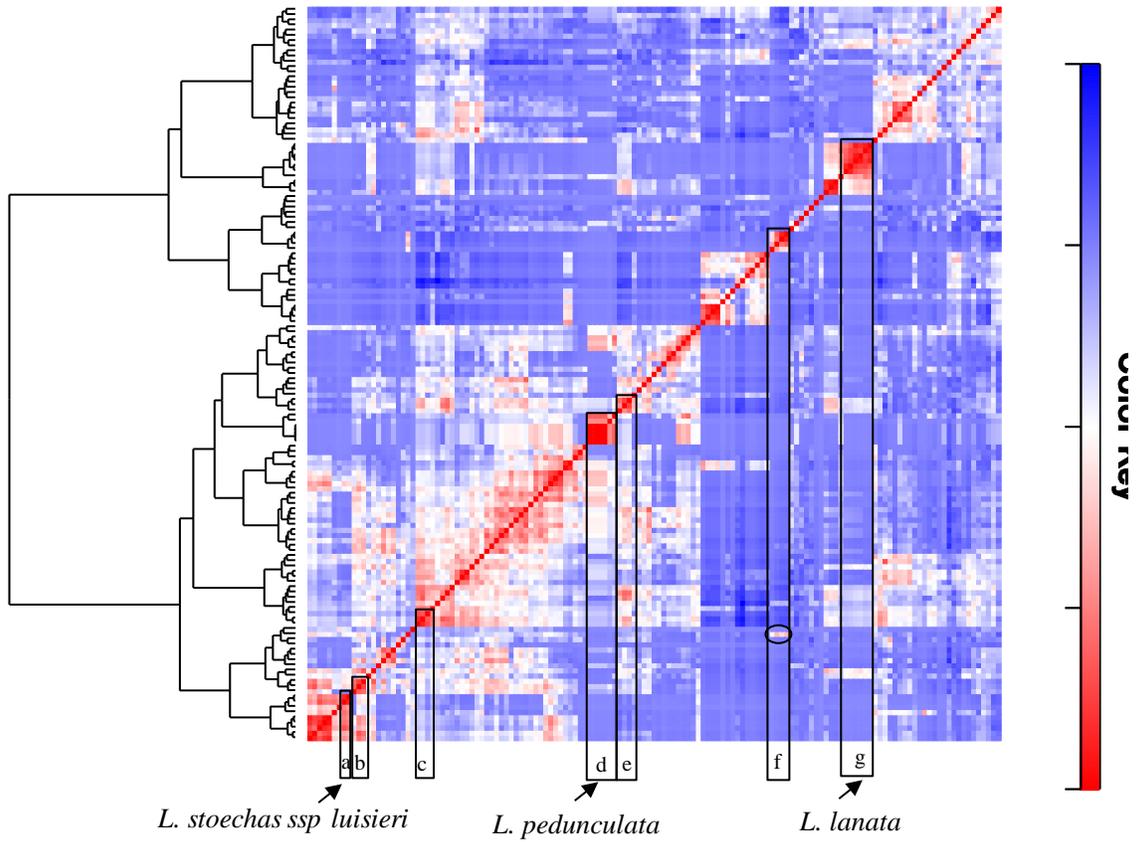


Figure 42: Carte de corrélation des 141 COV de lavande

Carte de corrélation des 141 COV de lavande (coefficient de Pearson et agglomération de Ward). L'échelle de couleur indique les valeurs de corrélation. Les encadrés noirs soulignent les zones de fortes corrélations. L'encadré *a* : présente des COV uniquement présents chez *L.stoechas ssp. luisieri* (l'acétate de *trans-alpha-necrodol* et le *cis-alpha-necrodol*), *b* : présente le cubébol, le viridiflorol et l'inconnu 31, *c* : camphène, camphre et bornéol, *d* : COV autapomorphes de *L. pedunculata*, *e* : sabinène, sabinène hydrate, *delta*-terpinéol, *alpha*-terpinéol, *f* : *trans*-citral, *cis*-citral, *cis*-géraniol, citronellol et dans le cercle noir le *trans*-géraniol, *g* : COV autapomorphe de *L.lanata*.

II-3 Discussion

Durant cette étude, nous avons tout d'abord établi une phylogénie moléculaire du genre *Lavandula* sur la base de comparaison de séquences du gène chloroplastique *trnK*. Les valeurs bootstrap étaient élevées pour la plupart des nœuds et les arbres consensus obtenus avec diverses méthodes cladistiques (maximum de vraisemblance, UPMAG ou parcimonie) étaient similaires. Nos résultats cladistiques étaient congruents avec les traitements monographiques du genre basés sur la comparaison des séquences d'un gène nucléaire, ITS, et de critères morphoanatomiques. Nous avons ensuite analysé par GC-MS la composition des mélanges de COV accumulés par les plantes de notre échantillonnage. Un total de 141 molécules ont ainsi pu être identifiées dans des classes chimiques diverses, 6 dérivés d'acide gras, 56 monoterpènes et 32 sesquiterpènes (Annexe V : Liste des 141 COV accumulés dans les taxons de lavandes). La structure chimique de 45 molécules n'a pas pu être déterminée car absentes des bases de données publiques. La robustesse des relations phylogénétiques entre espèces de lavandes a alors constitué un cadre idéal pour décrire l'évolution au sein du genre des 141 COV accumulés par les 26 espèces de notre échantillonnage.

Une première observation est que toutes les espèces de lavande de notre échantillonnage accumulent des mélanges de COV différents. Aucun COV n'a de plus été détecté dans l'ensemble de nos taxons. Par contre, de nombreux COV (67 sur 141) n'ont été détectés que dans les espèces d'une section botanique du genre ou dans un des deux sous-genres. Avec seulement 40 COV présentant une distribution homoplasique et 34 étant autapomorphiques d'un taxon, la distribution globale des COV était congruente avec les relations phylogénétiques entre espèces de lavandes. Les mélanges de COV accumulés par les espèces de lavande se sont donc diversifiés parallèlement à l'évolution du genre, et donc parallèlement au phénomène de spéciation dans ce genre, au même titre que les caractères morphologiques. Cette conclusion rejoint celle faite par Wink (Wink 2003) après l'analyse de la distribution de 3 classes de composés spécialisés chez les Angiospermes.

La modification d'un mélange de COV lors des processus évolutifs requiert la perte d'accumulation de certaines substances et l'acquisition de nouvelles molécules. L'analyse comparative de génomes entièrement séquencés, de transcriptomes et d'enzymes recombinantes mutées sur des acides aminés ponctuels a suggéré qu'un certain nombre de mécanismes évolutifs sont à l'œuvre chez les Angiospermes pour faire évoluer les structures chimiques (Field et Osbourn 2008; Ober 2005). Ceux-ci incluent 1/ la duplication de gènes

Résultats

suivie de leur divergence, 2/ l'évolution convergente, 3/ l'évolution répétée et 4/ l'altération de l'expression de gènes pour des raisons variées suivie de modifications dans la spécificité des enzymes. Ces évolutions sont d'autant plus rapides qu'un petit nombre de mutations dans la séquence codante des gènes de la plupart des enzymes de biosynthèse peut conduire à en modifier la spécificité au niveau des substrats mais aussi des produits formés (Schwab 2003). Plusieurs de ces mécanismes sont compatibles avec les observations faites sur les COV chez la lavande. Les espèces de la section *Lavandula* sont par exemple celles qui ont le plus grand nombre de chromosomes ($2n=36$ et $2n=75$ – Upson et Andrews 2004), une étude caryologique du genre ayant même suggéré que ces espèces résultent d'anciens évènements de polypléidisation suivi de périodes de spéciation (Upson et Andrews 2004). Leur plus grand nombre de COV pourrait donc résulter d'un élargissement des familles de gènes impliqués dans les étapes de biosynthèse suivi d'évènements de neofonctionnalisation des gènes dupliqués. Les espèces de la section *Stoechas* (également du sous-genre *Lavandula*) sont aussi caractérisées par la présence d'un plus grand nombre de COV. Leur compte chromosomique n'est cependant pas supérieur à celui des autres espèces du genre ($2n=30$). Dans leur cas, il est possible qu'un certain nombre d'enzymes biosynthétiques aient une spécificité de produit plus faible, conduisant ainsi à la formation d'un grand nombre de COV comme déjà décrit par exemple pour une sesquiterpène synthase (TPS4-B73) qui était capable de synthétiser divers sesquiterpènes (le 7-epi-sesquithujène, le *beta*-bisabolène et 12 sesquiterpènes) chez *zea mays* (Kollner, Schnee *et al.* 2004). Le fait que seulement 1-3 COV aient des teneurs supérieures à 0,1% chez ces espèces soutient une telle hypothèse. La distribution de certaines molécules dans un petit nombre d'espèces distantes évolutivement suggère que des processus d'évolution concertée sont aussi vraisemblablement à l'œuvre. Additionnellement, toutes les espèces que nous avons analysées différenciaient les COV qu'elles accumulaient entre leurs fleurs et feuilles, suggérant que ce phénomène est un important trait adaptatif. Par contre, nos données ne nous permettent pas de savoir si la modification de la composition en COV d'un tissu résulte d'une modification d'expression tissulaire d'un gène ou de la modification du fonctionnement d'un gène préexistant. Il serait maintenant intéressant de découvrir les mécanismes moléculaires qui soutiennent la biosynthèse de COV chez la lavande pour en découvrir les bases évolutives et les comparer à celles décrites dans d'autres espèces modèles.

Qu'ils dérivent d'un acide gras ou qu'ils soient issus de la voie de synthèse des mono- et sesquiterpènes, les COV de lavande sont issus d'une ou deux étapes enzymatiques à partir

Résultats

de substrats communs. Des études portant sur des espèces modèles telles la tomate et l'arabette ont ainsi suggéré que le *cis*-3-hexénol résulte de la dégradation séquentielle d'acides gras polyinsaturés (C18:2 acide linoléique et C18:3 acide linoléique) par une lipoxygénase et une hydroperoxyde lyase (Matsui 2006). Acétylation et oxydation sont ensuite respectivement effectuées par une acétyl transférase ou alcool deshydrogénase (D'Auria, Pichersky *et al.* 2007). Les voies biosynthétiques des autres dérivés d'acides gras de notre échantillonnage sont hélas moins connues. La biosynthèse des terpènes a, par contre, fait l'objet de nombreuses études mécanistiques et fait appel à une terpène synthase suivie éventuellement d'une acétyl transférase ou CytP₄₅₀ qui agissent sur les produits hydroxylés (Figure 43). De manière surprenante, à chaque substance terpénique acétylée ou oxydée ayant une distribution synapomorphique dans notre échantillonnage de plants de lavande, correspondait une forme hydroxylée au schéma carboné similaire présentant une distribution restreinte à la même section ou sous-genre. Même si les formes acétylées ou oxydées avaient des fois des distributions plus restreintes que leurs précurseurs potentiels, les espèces les accumulant ne constituaient pas une sous-clade génétique. Les enzymes réalisant ces modifications sur les terpènes hydroxylés sont connues pour avoir de faibles spécificités de substrat (Schwab 2003) et nos résultats suggèrent que, chez les lavandes, ces enzymes doivent avoir eu des spécificités de substrat suffisamment faibles pour avoir pu systématiquement accommoder de nouveaux substrats de manière impromptue durant les processus évolutifs. Leurs niveaux de fonctionnement semblent cependant très fluctuants d'une espèce à l'autre et ainsi ne garantissent pas de trouver systématiquement les deux molécules couple simultanément. Ces résultats suggèrent cependant que l'étape enzymatique précédant l'ajout de modifications est celle qui a été mise à contribution dans le genre *Lavandula* pour différencier les mélanges de COV entre les sections et sous-genres.

Résultats

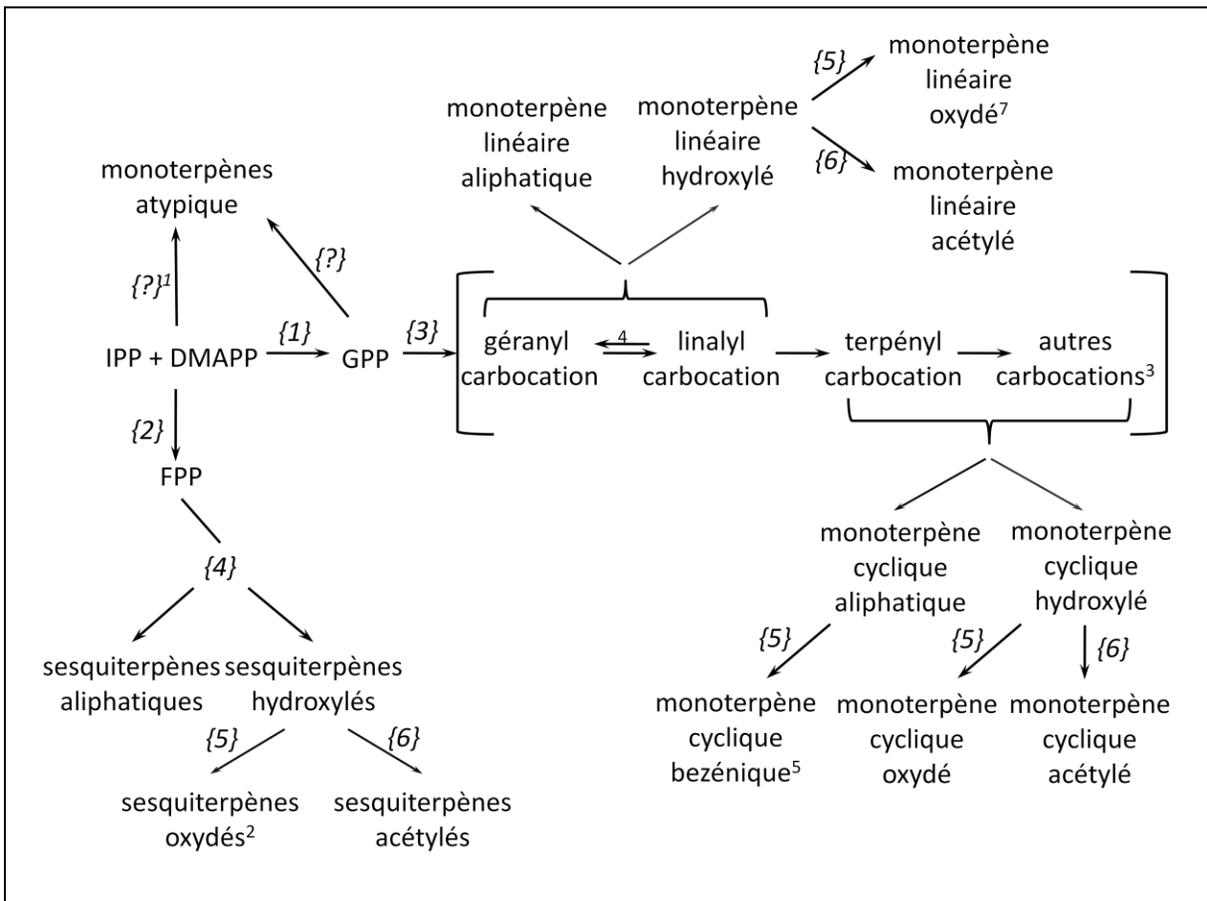


Figure 43 : Représentation schématique de la voie de synthèse des mono- et sesquiterpènes.

IPP : isopentényl diphosphate ; DMAPP : isopentényl diphosphate ; GPP : géranyl diphosphate ; FPP : farnésyldiphosphate ; {1} GPP synthase ; {2} FPP synthase ; {3} monoterpène synthase ; {4} sesquiterpène synthase ; {5} CytP450 type enzyme ; {6} acétyl transférase ; 1 synthèse lavandulol et citronellol ; 2 dont oxide de caryophyllène ; 3 pinyll, bornyl et saninyl cations ; 4 isomérisation allylique réversible pouvant conduire à une modification de l'isomérie de la double liaison en position 1 ; 5 formation de p-cymène et dérivés (modifié de Degenhardt et al., 2009)

Résultats

Mais un des résultats les plus surprenants de cette étude est que la probabilité qu'un COV non-acétylé ou oxydé ne s'accumule que dans les espèces d'une section botanique ou un sous-genre est très parallèle à la structure physico-chimique du COV. Tous les COV linéaires, qu'ils soient hydroxylés ou aliphatiques, qu'ils dérivent du catabolisme des acides gras ou de la voie des mono- ou sesquiterpènes (voies de synthèse non-atypiques), présentaient presque tous une distribution homoplasique dans le genre *Lavandula*. Pareillement, les terpènes cycliques aliphatiques avaient plus souvent une distribution homoplasique que leurs homologues hydroxylés bien que les deux groupes de substances soient générés par la même classe de terpène synthase (Degenhardt, Köllner *et al.* 2009; Pichersky, Noel *et al.* 2006; Tholl 2006). La propension à avoir une distribution homoplasique diminuait également avec l'augmentation de la complexité du schéma carboné de base. Dans le cas des monoterpènes, les composés bicycliques étaient ainsi moins homoplasiques que les dérivés monocycliques. Ceci correspond à la création de substances dont le contour physico-chimique est plus unique mais aussi à des terpènes synthases plus élaborées dans le sens où elles sont capables d'abaisser le niveau énergétique d'un plus grand nombre d'intermédiaires réactionnels carbocationiques (Figure 43). Les composés terpéniques dits atypiques car ils possèdent un schéma carboné peu courant et qui résultent de mécanismes biosynthétiques rares sont également très souvent synapomorphes d'une section ou sous-genre.

Si les connaissances actuelles permettent d'expliquer comment une plante individuelle peut, par le biais de modifications de son génome, modifier les COV qu'elle accumule, elles n'expliquent pas comment une telle modification peut être conservée dans les générations qui suivent pour finalement devenir un trait commun à l'intégralité d'une espèce botanique, et encore plus difficilement un trait commun à l'intégralité des espèces d'une section ou sous-genre. Ces processus font vraisemblablement appel à une sélection positive basée sur le gain de fonction associé à la modification du bouquet de COV produits. Ils sont mal connus du fait des faibles connaissances sur les valeurs adaptatives des COV individuellement, mais surtout en mélange avec ceux produits par la même plante, ceux produits par les plantes environnantes et en relation avec les autres traits exprimés par la plante et ayant une fonction écologique similaire.

Le fait que les COV linéaires soient homoplasiques pourrait suggérer qu'ils sont dénués de valeur adaptative favorisant leur sélection positive. Ce n'est vraisemblablement pas le cas car plusieurs d'entre eux ont des fonctions biologiques bien établies comme les

Résultats

farnésènes (Arimura, Dezen *et al.* 2004; Markovic, Norris *et al.* 1996) ou les ocimènes (Takabayashi et Dicke 1996). Il en va de même pour les terpènes cycliques aliphatiques comme le caryophyllène (Gershenzon et Dudareva 2007). Une première hypothèse plus vraisemblable est que leur accumulation soit en fait induite par un stress biotique ou abiotique, un niveau d'induction qui aurait été variable dans notre échantillonnage même si tous les végétaux échantillonnés étaient cultivés au même endroit et récoltés le même jour. Alternativement, le fait que ces substances aient un contour physicochimique simple et similaire signifie peut être qu'elles jouent des rôles écologiques similaires, et donc interchangeables. Ces substances dont la formation nécessite le plus petit niveau de spécialisation de leurs enzymes biosynthétiques (et donc de pression de sélection - surtout vrai des terpènes synthases – Figure 43) assumeraient ainsi un certain nombre de fonctions écologiques de base et communes à une vaste majorité d'espèces végétales. Le fait que la majorité de ces substances comme le *cis*-3-hexénol, le linalol, le limonène, le géraniol, les ocimènes, le caryophyllène, les farnésènes soient connues pour être présentes dans une très grande majorité des taxons dont les COV ont été analysés soutient une telle hypothèse (Knudsen, Eriksson *et al.* 2006). La majorité des COV connus pour être induits par des stress biotiques et abiotiques font aussi partie de ces catégories de COV, en appui de la première hypothèse (Holopainen et Gershenzon 2010). En contre partie, les COV dont les structures physico-chimiques sont plus uniques et dont, en parallèle, la biosynthèse fait appel à des événements de modification des génomes moins probables (terpènes atypiques, bi- ou tricycliques) apparaîtraient de manière plus ponctuelle lors de l'évolution du règne végétal et seraient conservés pour favoriser la différenciation d'un groupe de végétaux et son adaptation à une situation écologique qui lui serait plus spécifique. En soutien d'une telle hypothèse, des composés tels que le lavandulol, le fenchol ont été assez rarement observés dans un nombre restreint d'espèces végétales. Les *alpha*- et *beta*-pinènes, des composés monoterpéniques bicycliques, constituent le principal contre-exemple à cette hypothèse car ils présentaient une distribution homoplasique au sein du genre *Lavandula* et ont été détectés dans une vaste majorité d'espèces végétales (Knudsen, Eriksson *et al.* 2006). Cette hypothèse n'explique pas non-plus pourquoi les composés terpéniques hydroxylés sont plus fréquemment synapomorphiques que leur homologues aliphatiques. Ceci peut être lié au fait qu'ils peuvent être modifiés chimiquement et donc générés des substances aux propriétés écologiques plus uniques et dont la sélection positive imposerait la sélection des intermédiaires biosynthétiques hydroxylés. Alternativement, ceci peut être lié à leur fonction biologique. Trop peu d'études

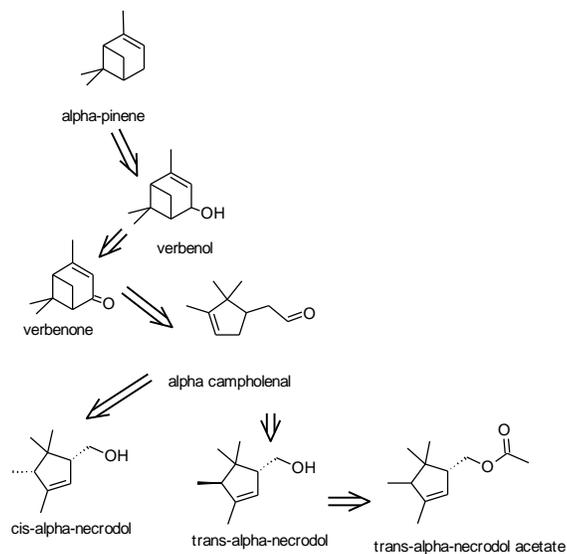
Résultats

comparatives des valeurs adaptatives des terpènes sont hélas disponibles pour trancher en faveur de l'une ou l'autre de ces deux hypothèses. Pareillement, l'absence d'étude similaire à la notre couvrant de manière relativement exhaustive un genre végétal sur des individus cultivés et analysés en parallèle ne nous permet pas de savoir si nos observations sur le genre *Lavandula* constituent une règle ou une exception chez les végétaux. Il est espéré que le développement d'outils bioinformatiques de prétraitement des données chromatographique, tels que notre module MSeasy, encouragera de nombreuses autres équipes de recherche à réaliser des travaux similaires sur leur genre d'intérêt.

Les COV faisant partie de la famille chimique des nécrodanes sont les molécules de lavande les plus spécifiques dans le règne végétal puisqu'elles ne sont connues que de chez *L. stoechas* ssp. *luisieri* (Garcia-Vallejo 1992; Garcia-Vallejo, Garcia-Vallejo *et al.* 1994). En dehors du règne végétal, elles sont connues pour être produites par plusieurs insectes (Garcia-Vallejo 1992; Garcia-Vallejo, Garcia-Vallejo *et al.* 1994). Avec un cycle à 5 carbones, ces substances ont un squelette carboné unique pour les métabolites secondaires végétaux et leur voie de biosynthèse est inconnue. La société Firmenich SA a récemment publié dans un brevet la synthèse chimique des *cis*- et *trans-alpha*-nécrodols ainsi que la forme acétylée de ce dernier (Figure 44 A- (Chapuis, Barthe *et al.* 2006). Cette voie de synthèse organique débute avec l'*alpha*-pinène. Il est possible qu'un schéma biosynthétique similaire existe chez ces espèces de lavande car les autres espèces de leur section botanique contenaient de l'*alpha*-pinène et les autres substances intermédiaires potentielles, le campholénal et la verbénone. De manière intéressante, la probabilité de présence des nécrodols ou de l'acétate de nécrodol était inversement proportionnelle à la présence d'*alpha*-pinène, campholénal et verbénone dans cette section botanique (Figure 44 B). Un raisonnement similaire pourrait être conduit sur de nombreux composés inconnus (32% des molécules identifiées dans notre échantillonnage) et qui sont co-exprimés avec d'autres substances connues (Figure 42). La découverte de leur structure chimique faciliterait la prédiction, et la découverte, des voies de biosynthèse des COV dans le genre *Lavandula*.

Résultats

A



B

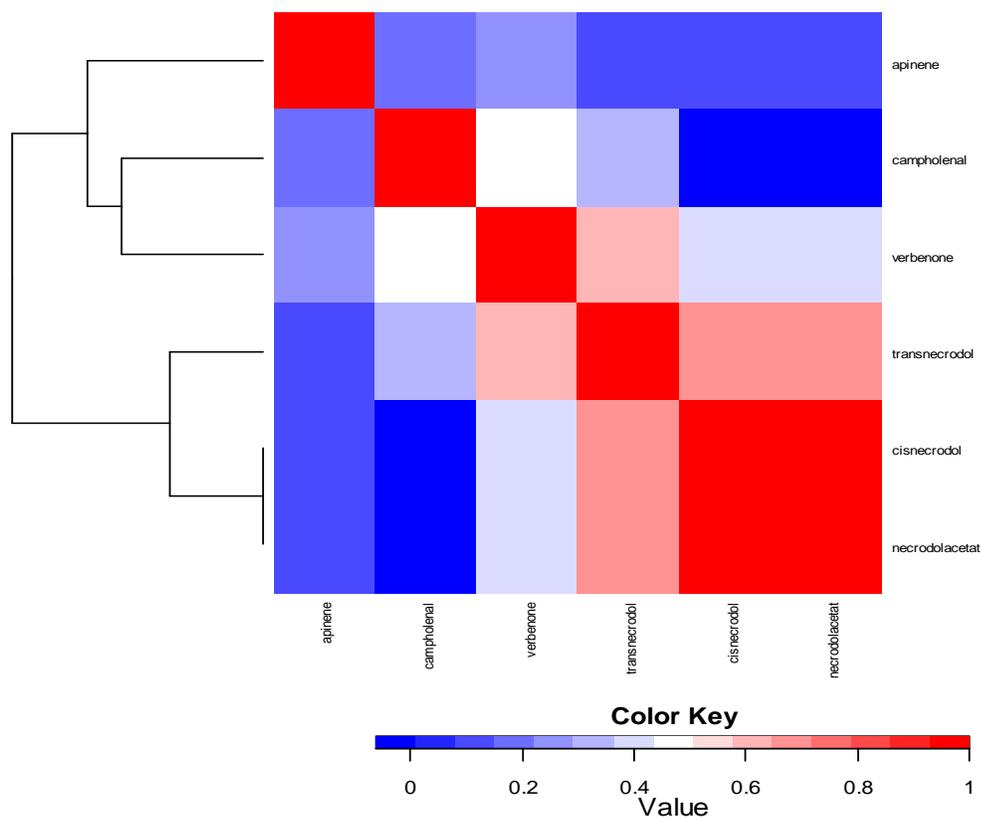


Figure 44 : Les nécrodanes voie de synthèse organique et carte de corrélation

A : Synthèse organique des nécrodanes (modifié d'après Chapuis, Barthe *et al.* 2006).

Voie de synthèse chimique simplifiée (sans les intermédiaires et les catalyseurs) des nécrodanes présents chez *L. stoechas* ssp. *luisieri*.

B: Carte de corrélation des COV impliqués dans la synthèse des nécrodanes au sein du genre *Lavandula*.

Résultats

Etude des variations spatio-temporelles de
l'accumulation de terpènes chez *L. angustifolia* et *L. x*
intermedia



Inflorescences de *L. angustifolia* cv 77-13 au lycée horticole de Romans sur Isère

III Etude des variations spatio-temporelles de l'accumulation de terpènes chez *L. angustifolia* et *L. x intermedia*

III-1 La problématique

L'expression des voies du métabolisme secondaire chez les plantes s'inscrit dans le schéma développemental des tissus et organes végétaux. La production de chaque terpène peut ainsi être différentielle entre tissus ou lors du développement de ceux-ci, l'intensité des productions pouvant être modulées de manière post-développementale par des stimulations biotiques ou abiotiques en provenance de l'environnement immédiat de la plante. La production de certains terpènes peut ainsi être limitée à certains organes comme par exemple les pétales des fleurs de rose au moment précis de l'ouverture des boutons floraux (Bergougnoux, Caissard *et al.* 2007), les feuilles (McConkey, Gershenzon *et al.* 2000) ou les racines (Chen, Ro *et al.* 2004). Bien que ces phénomènes soient insuffisamment décrits scientifiquement, l'inflorescence de lavande n'échappe pas à cette règle. Car les lavandiculteurs ont depuis longtemps observé des différences de composition terpénique des huiles essentielles de lavande entre des récoltes effectuées précocément et d'autres effectuées tardivement. Cependant, ces observations sont basées sur des récoltes faites sur des sites différents. Le degré de développement des inflorescences était laissé à la discrétion du lavandiculteur et, surtout, ces observations portent sur des terpènes accumulés dans la plante (dans des poches spécialisées à la surface des têtes de trichomes spécialisés) et ne permettent donc pas de savoir si ces accumulations différentielles sont liées à des différences d'expression des gènes de biosynthèse lors d'un processus développemental ou si elles sont liées à des évaporations différentielles des substances sur la plante. Notre problématique a donc été de décrire les variations de terpènes au cours de la saison estivale dans les feuilles et les inflorescences de lavande sur différentes années et plusieurs cultivars. Ceci nous a permis de modéliser ces variations et de les corrélérer avec des étapes du développement de la plante. Elles ont ensuite été corrélées avec des expressions différentielles des gènes de terpène synthase. Cette étude constitue un pré-requis pour utiliser la lavande comme plante modèle pour la famille des Lamiacées afin de caractériser dans le futur les rôles écologiques de ces terpènes et les mécanismes moléculaires de la régulation de leur biosynthèse au cours du développement de la plante.

Note : Bien qu'il s'agisse d'un abus de langage le terme inflorescence remplacera le mot thyrses dans le reste du document (voir Figure 15°)

III-2 Résultats

III-2.1 Articles

Les résultats de cette partie de mes recherches ont été publiés dans le volume 138 de février 2010 de la revue *Physiologia Plantarum* (Guitton, Nicolè *et al.* 2010b) et dans *Plant Signaling & Behaviors* en tant qu'*addendum* (Guitton, Nicolè *et al.* 2010a). Ce premier article présente les résultats principaux. Ceux-ci sont mis en perspectives dans le second. A la suite de ces deux articles, j'ai également ajouté quelques résultats complémentaires non-inclus dans ces articles et qui soulignent les implications agronomiques de mon travail.

Résumé de l'article publié dans la revue *Physiologia Plantarum*

Malgré l'importance économique des huiles essentielles (HE) de fleurs de lavande fine (*Lavandula angustifolia* Mill.) et de lavandin (*L. x intermedia* Emeric ex Loisel), aucune information n'est disponible sur les variations éventuelles de la composition en composés organiques volatils (COV) accumulés par ces plantes durant le développement de leurs inflorescences. Chez les espèces étudiées, comme chez de nombreuses Lamiacées, les COV s'accumulent majoritairement à la surface du calice. Les 20 COV les plus abondants ont pu être classés en 3 sous-groupes sur la base des variations temporelles de leurs concentrations. Les trois sous-groupes dominent ainsi successivement le bouquet floral. La transition entre le 1^{er} et le 2^{ème} groupe a lieu au moment de l'ouverture de la première fleur de l'inflorescence, la transition suivante, entre les VOC des sous-groupes 2 et 3, ayant lieu après la formation des premières graines matures. Nos résultats montrent également que les variations d'accumulation de VOC dans les calices d'une inflorescence est sous l'influence de deux facteurs, l'un lié à la date de récolte qui tend à synchroniser la composition en COV de tous les calices et l'autre lié au stade de développement de fleurs individuelles d'une inflorescence (très hétérogènes dans les inflorescences de lavande) qui fait que des fleurs moins matures sont plus retardées que les autres dans les évolutions globales de la composition en COV de l'inflorescence. Les feuilles accumulent un nombre plus restreint de COV qui correspondent à une partie des COV accumulés dans les inflorescences avant la floraison (sous-groupe 1). Les COV dans les feuilles restent qualitativement et quantitativement constants durant toute la période de floraison. L'expression par PCR quantitative de 2 gènes (*LaLIMS* et *LaLINS*)

Résultats

codants pour des terpènes synthases (TPS) responsables de la biosynthèse de composés des sous-groupes 1 et 2-3 respectivement, démontre une tendance similaire entre les profils d'expression des gènes et l'accumulation de leurs produits de biosynthèse au cours du développement de l'inflorescence. Ces données suggèrent que des changements dans l'expression des TPS ont lieu lors du développement des inflorescences et de lavande et qu'ils affectent leur composition en HE.

Résultats

*L'article publié dans *Physiologia Plantarum* (DOI: 10.1111/j.1399-3054.2009.01315.x)*

[Lien vers l'article en ligne](#)

*L'article publié dans *Plant Signaling & Behaviours**

[Lien vers l'article en ligne](#)

Résultats

NOTE : Les tableaux supplémentaires sont donnés ci-dessous

Résultats

Supplemental Table 1. Essential oil composition in *L. angustifolia* and *L. × intermedia* organs. RT: retention time in minutes. AI: calculated arithmetic index on a DB5 column (comparison with n-alkanes C8-C21). AI.ref: AI from Adams (2007). MS: identification by comparison of the experimental mass spectrum with Wiley's database (Wiley, 2007). STD: identification on the basis of similar retention behaviour as authentic commercial standards. Commercial standards were from Acros, Aldrich, Fluka or Payan & Bertrand (Grasse, France). * detected in few samples (<5%). Content in mg gFW⁻¹, N=10.

peak	identity	HP Innowax			DB-5			Identification methods		<i>L. angustifolia</i>						<i>L. × intermedia</i>						
		RT		AI	RT		AI	AI.ref	methods		Leaves		inflorescences		Leaves		inflorescences		Leaves		inflorescences	
		min	max		min	max		min	max	min	max	min	max	min	max	min	max	min	max	min	max	
1	a-pinene	8.60		930	11.36	932	MS, STD		0												0.01 - 0.12	
2	camphene	10.67		945	12.20	946	MS		0												0 - 0.10	
3	b-pinene	13.11		973	13.70	974	MS, STD		0												0.05 - 0.53	
4	d-3-carene	16.20		1005	15.42	1008	MS, STD		0												0 - 0.47	
5	myrcene	18.07		988	14.53	988	MS, STD		*												0.04 - 0.69	
6	limonene	20.40		1026	16.54	1024	MS, STD		*												0.03 - 0.58	
7	1,8-cineol	21.10		1029	16.72	1026	MS, STD		0.17 - 0.25												0.39 - 2.22	
8	trans-beta-ocimene	24.90		1067	17.66	1065	MS		0												0.09 - 1.20	
9	cis-beta-ocimene	26.50		1047	17.06	1044	MS, STD		0												0 - 0.64	
10	camphre	45.20		1142	22.67	1141	MS, STD		*												0.48 - 1.50	
11	linalool	48.90		109	20.47	1095	MS, STD		0												0.38 - 5.04	
12	linalyl acetate	49.33		1253	28.08	1254	MS, STD		0												0.22 - 2.54	
13	bornyl acetate	49.42		1281	29.40	1287	MS, STD		0												0 - 0.04	
14	caryophyllene	50.60		1414	35.36	1417	MS, STD		0												0 - 0.11	
15	terpinen-4-ol	51.50		1177	24.42	1178	MS, STD		0												0 - 0.55	
16	lavandulyl acetate	52.30		1284	29.55	1288	MS, STD		0												0.06 - 0.19	
17	germacrene D	54.60		1475	37.92	1484	MS, STD		0												0 - 0.22	
18	trans-beta-farnesene	55.60		1450	36.88	1454	MS		0												0 - 0.07	
19	lavandulol	56.20		1162	23.68	1165	MS, STD		0												0 - 0.06	
20	borneol	57.10		1168	23.98	1165	MS, STD		0.06 - 0.21												0 - 0.59	

Supplemental Table 2. Terpene synthase genes specific primer pairs used in this work for real time quantitative PCR

<i>Gene abbreviation</i>	<i>Blastn best match</i>	<i>E-value</i>	<i>% Identity</i>	<i>Primer Pair (5' Forward 3' / 5' Reverse 3')</i>	<i>Amplicon length (bp) / efficiency</i>	<i>r²</i>
La_LIM	DQ263740.1	9.00E-44	99%	GCGCCACACAACACTAGAAATTAAGT/ TTGCACAGTCAGCTCAGCG	151/2.06	0.99
La_LIN	DQ263741.1	2.00E-77	100%	TTGATGAGGATGGTGATGGA/ TCTTGTGTGTTGCTTGAGCA	198/1.82	0.99
La_BERG	DQ263742.1 ***	1.00E-64	96%	ATTGGTGAGGCGGTTTAGC/ ATCTTTCGGGTAGGCCAGT	***	

Closest homolog was identified using BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi> with Blastn options)

*** only detected on genomic DNA despite the fact that primers were designed on cDNA sequence

Supplemental Table 3. Reference genes specific primer pairs used in this work for real time quantitative PCR

Gene abbreviation	Sequence Identity (GenBank)	Arabidopsis ortholog locus	Arabidopsis locus description	Primer Pair (5' Forward 3' / 5' Reverse 3')	Amplicon length (bp) / efficiency	r ²	CV
La_Actin	dbEST_id 64810781	AT5G09810.1	Actin 7(ACT7)/actin 2 other hits include ACT3, ACT11, ACT1, ACT2, ACT8, ATC4, ACT12, ACT9	TGCTCAGTGGTGGATCGACTATG/ ACCACCTTAATCTTCATGCTGCTTG	95/1.95	0.99	4.32
La_Tubulin	dbEST_id 64810782	AT4G14960.2	Tubulin alpha 6 chain (TUA6) other hits TUA4, TUA2, TUA3, TUA5, TUA1	CCACAATCAAGACCAAGAGGACAATC/ CGCCAGGGACCACCGGTTG	98/1.93	0.99	5.08
La_EFI-alpha	dbEST_id 64810783	AT5G60390.3	Elongation factor 1-alpha other hits AT5G60390.3, AT5G60390.2, AT5G60390.1, AT1G07940.2, AT1G07940.1, AT1G07930.1, AT1G07920.1, AT1G07930.2, AT1G18070.2, AT1G18070.1, AT5G10630.1	CCCTTCTTGAGGCTCTTGAC/ GCACAGTTCCAATACCACC	106/2.05	0.93	6.84

Closest Arabidopsis homologs were identified using TAIR BLAST 2.2.8

Efficiency : $E=10^{-(1/\text{slope})}$ was calculated using the slope of the standard curve given by the Applied Biosystems SDS 7500 software ver 1.2.3

Résultats

Résultats

L'article publié dans Plant Signaling & Behaviours

Plant Signaling & Behavior 5:6, 1-3; June 2010; © 2010 Landes Bioscience

ARTICLE ADDENDUM

Résultats

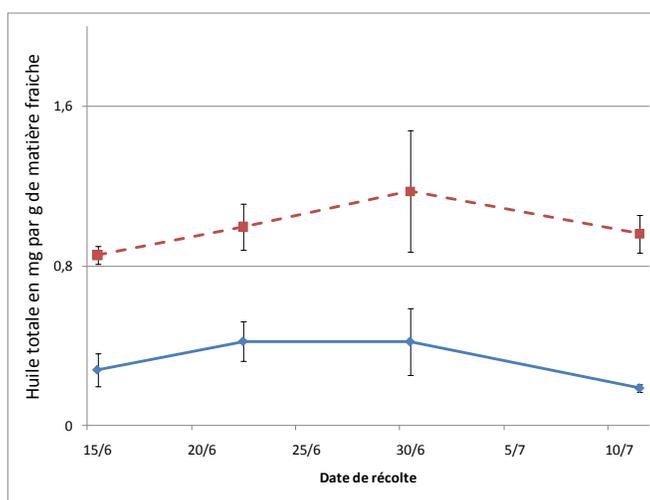
Résultats

III-2.2 Résultats complémentaires

Variation de la production de COV dans les feuilles de *L. angustifolia* et *L. x intermedia*

Les lavandes accumulent des COV dans leurs différents organes aériens, et principalement dans les calices des inflorescences. Nous avons focalisé l'étude de la variabilité quantitative des COV dans l'article de *Physiologia Plantarum* sur les inflorescences de lavande car ces organes présentent la plus grande quantité et diversité de COV de la plante. Ce sont également les COV des inflorescences qui constituent la majorité de l'HE essentielle produite industriellement à partir de ces plantes. La Figure 45 A et le Tableau 12 présentent, respectivement, les variations quantitatives et qualitatives de la production de COV par les feuilles de lavande et lavandin durant la période de floraison. Les variations quantitatives d'HE observées dans les feuilles ne sont pas significativement différentes entre les dates de récolte, ni entre les cultivars, comme le montre le test de Kruskal-Wallis présenté dans la Figure 45 B. Aucune variation significative n'a non plus été observée dans la composition en COV des feuilles sur la période de récolte et ce, tant chez les cultivars de lavande que chez ceux de lavandin (Tests de Kruskal-Wallis présentés dans le Tableau 12).

A



B

Test de Kruskal et Wallis	K valeur obs	ddl	p-value (seuil alpha =0.001)	Différence significative
Quantité d'Huile effet date (lavande)	6,400	3	0,094	non
Quantité d'Huile effet date (lavandin)	6,198	3	0,102	non
Quantité d'Huile effet cultivar (lavande)	32,835	11	0,001	non
Quantité d'Huile effet cultivar (lavandin)	15,216	10	0,124	non

Figure 45 : Variation de la production de COV dans les feuilles de *L. angustifolia* et *L. x intermedia* en mg.g⁻¹

A : Evolution de la quantité d'huile dans les feuilles de *L. angustifolia* (bleu) et *L. x intermedia* (rouge, pointillés). Les barres d'erreur représentent l'erreur standard sur la moyenne. B : Statistiques associées aux variations de la quantité d'HE en fonction de la date et du cultivar (Kruskal et Wallis).

Résultats

Tableau 12 : Variation qualitative de la composition des principaux COV dans les feuilles de *L. angustifolia* et de *L. x intermedia* au cours de la période de floraison

Test de Kruskal et Wallis Feuille de lavandin	K valeur obs	ddl	p-value (seuil alpha =0.001)	Différence significative
Quantité de 1,8-cinéole effet date	3,173	3	0,366	non
Quantité de Camphre effet date	6,873	3	0,076	non
Quantité de Bornéol effet date	6,436	3	0,092	non

Test de Kruskal et Wallis Feuille de lavande	K valeur obs	ddl	p-value (seuil alpha =0.001)	Différence significative
Quantité de 1,8-cinéole effet date	0,894	3	0,827	non
Quantité de Camphre effet date	9,524	3	0,023	non
Quantité de Bornéol effet date	8,166	3	0,043	non

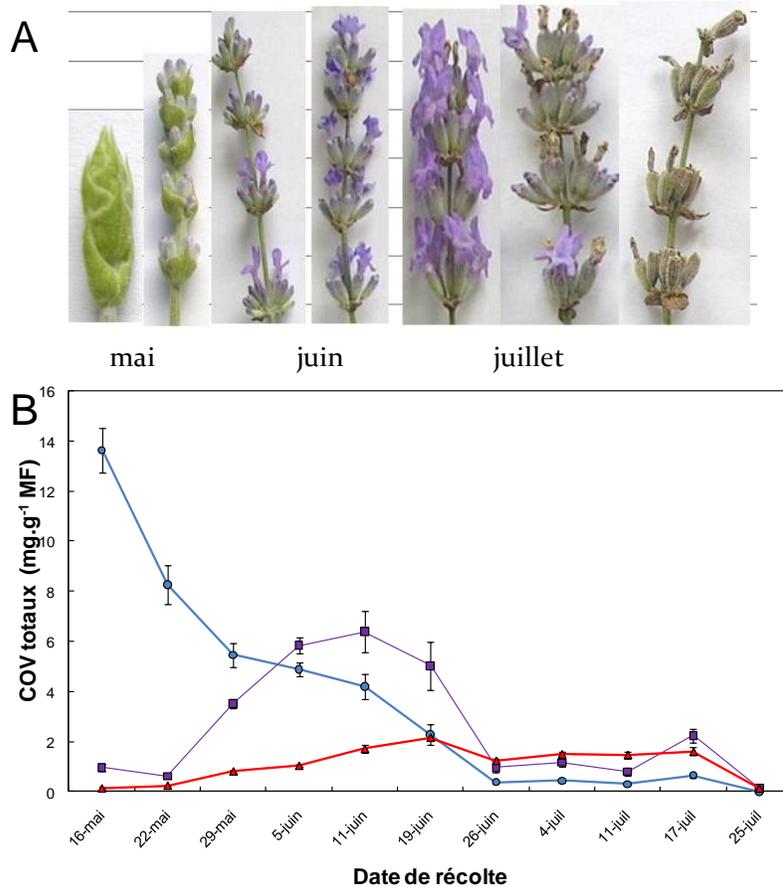


Figure 46 : Evolution au cours de la période de floraison de la morphologie des inflorescences et des COV accumulés chez la lavande.

A : Evolution de la morphologie des inflorescences de *L. angustifolia* cv 77-13 de la mi-mai à la fin juillet. B Evolution de la quantité de COV dans les inflorescences de *L. angustifolia* cv 77-13 de la mi-mai à la fin juillet. Les COV ont été divisés en 3 groupes sur la base des corrélations de leur accumulation. Les quantités moyennes des COV d'un même groupe ont été sommées et reportés sur le graphique. Groupe 1 (bleu) Groupe 2 (violet) Groupe 3 (rouge). La barre sur chaque point représente la somme des erreurs standards sur les moyennes.

Résultats

Variations temporelles de l'accumulation des COV dans les inflorescences de lavandes et lavandins au cours de la floraison

Les variations différentielles, au cours du temps du contenu en COV des inflorescences de 3 cultivars de lavandes nous ont conduit à définir 3 groupes de composés dont la production est variable au cours du développement de l'inflorescence. Les COV du groupe 1 (*delta*-3-carène, limonène, myrcène, bornéol, acétate de bornyl, 1,8-cinéol et le *trans-beta*-ocimène) sont tous des monoterpènes. Les molécules du groupe 2 sont soit des sesquiterpènes (*trans*-farnésène, *beta*--caryophyllène, germacrène D) soit des dérivés acétylés de monoterpènes (acétate de linalyl et acétate de lavandulyl). Le groupe 3 contient le linalol et le terpinène-4-ol. Ces groupes ont été définis par une comparaison de l'évolution de chaque COV au cours du temps avec celle de tous les autres à l'aide d'un test de corrélation de Spearman (Figure 5 de l'article Guitton, Nicolè *et al.* 2010b). La définition de ces 3 groupes de composés a permis de tracer l'évolution de ces composés en fonction du temps. La Figure 46 représentée aussi dans l'article est complétée ici par la superposition des changements morphologiques de l'inflorescence. Ce document supplémentaire permet de voir clairement que le groupe 1 de COV peut être associé à des inflorescences très jeunes où la corolle des fleurs n'est pratiquement pas visible. Le groupe 2 est associé à des inflorescences portant majoritairement des fleurs épanouies. Le groupe 3 devient majeur dans les inflorescences portant majoritairement des fleurs fanées.

Un travail identique a été réalisé sur l'évolution qualitative et quantitative des COV au cours de la floraison chez le lavandin. Les données non explicitées dans les deux publications sont rapportées dans ce paragraphe. Une analyse des correspondances multiples (Figure 47) présente l'évolution temporelle des COV dans les inflorescences de 3 cultivars de lavandin. L'axe 1 porte 75.9% de la variance et sépare les premières dates de récoltes (jusqu'au jour 34) associées à l'absence de composés (COV notés avec un 0) des récoltes plus tardives associées à la présence de nombreux composés (COV notés avec un 1). L'ACM permet de différencier certains cultivars sur la base de leur profil d'accumulation de COV au fil de la floraison. En effet, le barycentre des observations des inflorescences du cultivar Super (B110) est plus éloigné de ceux des cultivars Certitude et Grosso dont les barycentres de distribution de données se superposent sur l'ACM de la Figure 47. Ceci démontre une évolution similaire du contenu en COV des

Résultats

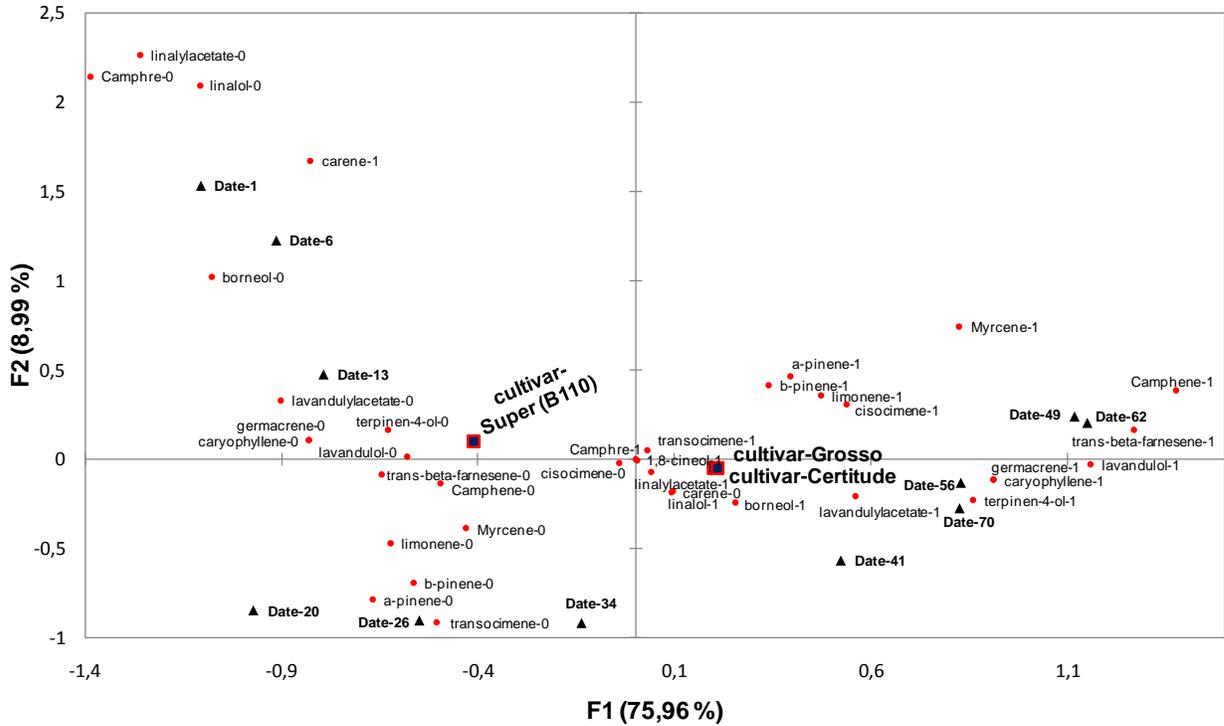


Figure 47: Analyse des correspondances multiples de l'évolution temporelle des COV des inflorescences de 3 cultivars de lavandin (année 2007)

Comparaison de la composition en COV d'inflorescences (présence/absence) de 3 cultivars de lavandin (Super (B110), Grosso et Certitude). Les points rouges indiquent la présence ou l'absence de COV (présence par un -1 et absence par un -0 à la suite du nom du COV). Les carrés bleus représentent les barycentres des observations. Les triangles noirs représentent les dates de récoltes (le numéro indique le nombre de jours depuis le début de campagne de récolte).

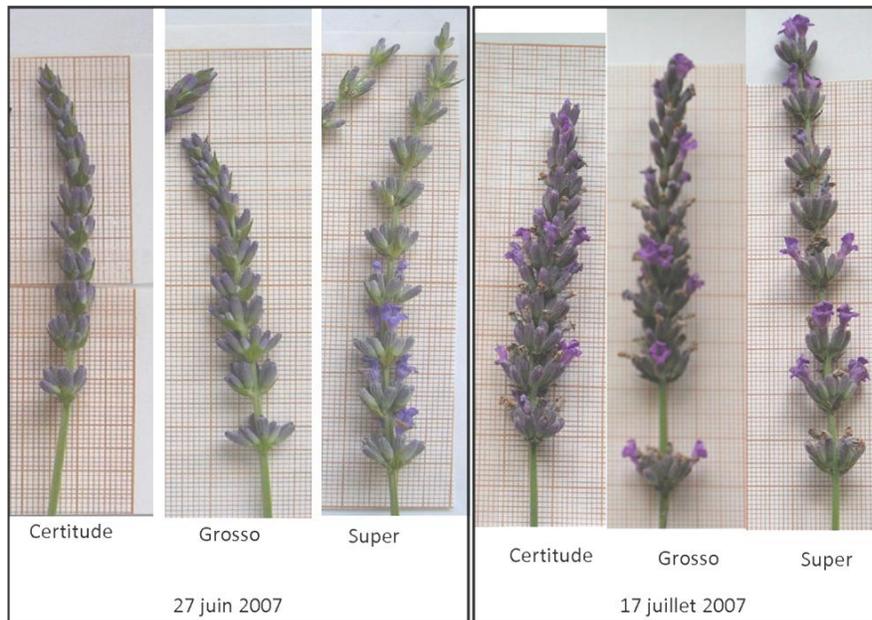


Figure 48 : Evolution comparative des inflorescences de 3 cultivars de lavandin au cours de la floraison (année 2007)

Résultats

inflorescences des lavandins Grosso et Certitude. Le cultivar Super (B110) diffère morphologiquement et génétiquement des 2 autres. En effet, l'inflorescence du cultivar Super (B110) est moins compacte que celle des deux autres car l'espacement entre deux verticilles y est plus grand (Figure 48). Ce cultivar est également plus précoce. Certitude est quant à lui un somaclone de Grosso (Source CRIEPPAM). Le cultivar Super (B110) est associé aux dates de récoltes précoces ce qui correspond bien avec sa courte durée de floraison comparée avec celle des cultivars Grosso ou Certitude.

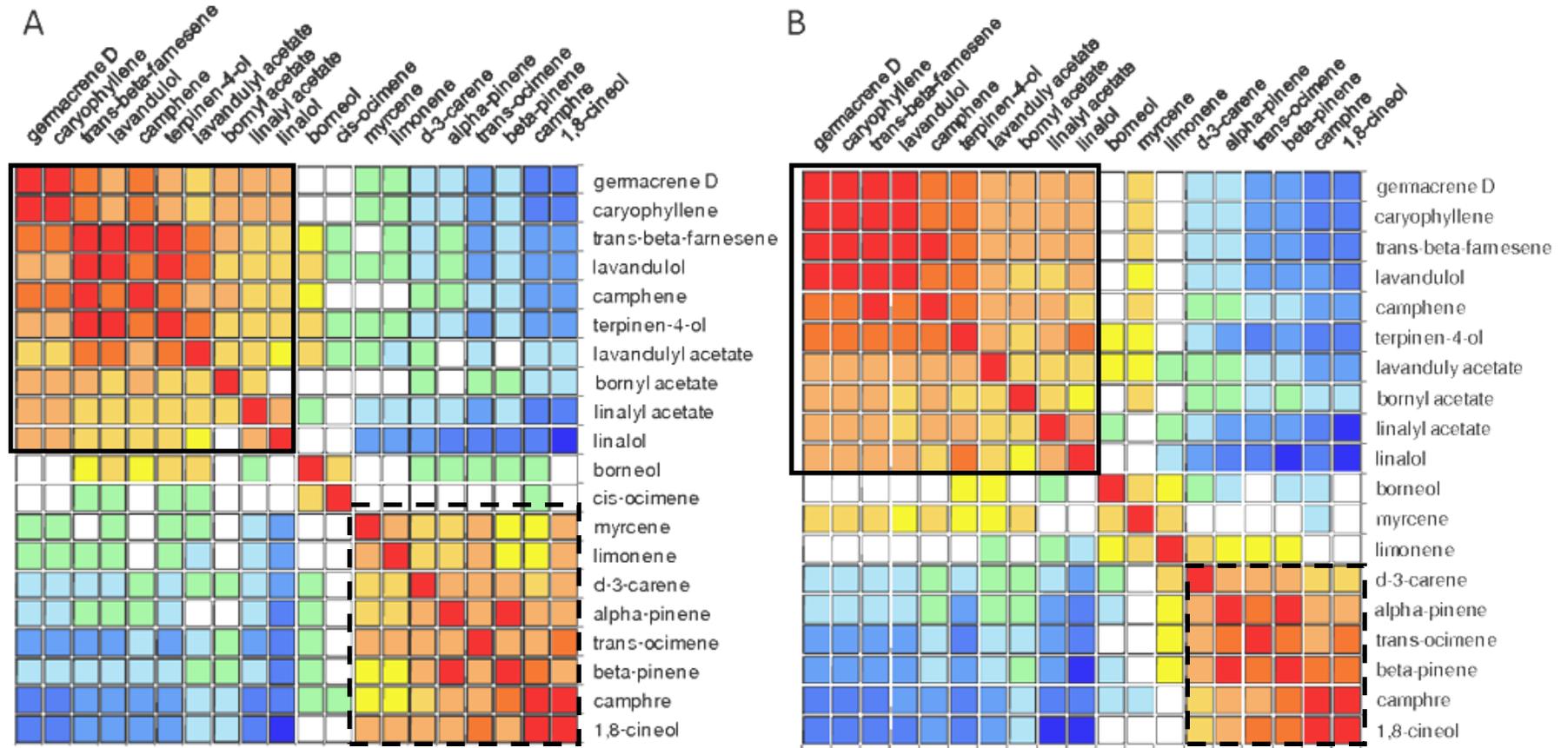
Lorsqu'on réalise le test de corrélation de Spearman sur les COV des échantillons des 3 cultivars de lavandins (415 échantillons, Figure 49 A), on distingue 2 groupes de COV. Le premier regroupe (carré noir en trait pointillé sur la Figure 49) comme chez les lavandes, le δ -3-carène, le limonène, le myrcène, le 1,8-cinéol et le *trans*-ocimène auxquels s'ajoutent le camphre, l'*alpha* et le *beta*-pinène. Le bornéol et son acétate ne sont pas associés à ce premier groupe contrairement à ce qu'on a pu observer pour les lavandes. Le bornéol est cependant positivement associé à quelques COV du groupe 1 (R^2 inférieurs à 0.2). Ceci s'explique par la production de camphre par cette espèce qui rend moins détectables le bornéol et l'acétate de bornyl, la biosynthèse du camphre étant faite au dépend du bornéol et donc aussi de son acétate (Figure 51). Le deuxième groupe de COV est composé du germacrène D, du caryophyllène, du *trans-beta*-farnésène, de l'acétate de lavandulyl, de l'acétate de linalyl comme chez la lavande. A ceux-ci s'ajoutent le camphène, le lavandulol, l'acétate de bornyl ainsi que les composés du groupe 3 des lavandes le linalol et le terpinène-4-ol. Comme dans le cas de la lavande le *cis*-ocimène n'est clairement associé à aucun des deux groupes. La figure 49 B représente les corrélations entre les COV des inflorescences des cultivars Certitude et Grosso (280 échantillons), le cultivar Super (B110) ayant été retiré du fait de son évolution de chimie différente des deux autres cultivars (Figure 47). Les corrélations entre les COV d'un même groupe s'en trouvent nettement améliorées (certains R^2 augmentant de 0.2 et plus par exemple pour le lavandulol par rapport au germacrène D et au caryophyllène). La Figure 50 représente l'évolution des groupes de COV dans les inflorescences de lavandins (Cultivars Certitude et Grosso) au fil de la floraison. Pour chaque groupe les moyennes des COV accumulés dans chaque cultivar à chaque date de récolte sont additionnées. La teneur en COV du groupe 1 dans les inflorescences chute brutalement jusqu'au 5 juin puis ces molécules restent présentes jusqu'à la fin de la floraison. Les COV du groupe 2 s'accumulent progressivement et dépassent les COV du groupe 1 vers le 5 juin et atteignent leur maximum

Résultats

vers le 20 juillet. Les premières fleurs de lavandin se sont épanouies vers la fin juin en 2007, avec une floraison plus précoce des Super (B110) visible sur la figure 48 (les premières corolles épanouies étant visibles dès le 27 juin). Cette dernière figure montre également que les premières fleurs fanées sont visibles à partir du 17 juillet ce qui correspond avec le début de la baisse des COV du groupe 2. Cette baisse est principalement liée à la diminution du linalol et de son acétate qui masquent même une hausse de l'accumulation de terpinène-4-ol autour du 17 juillet, la quantité de COV ne chutant brutalement que vers début septembre avec la chute massive des calices. En conclusion, contrairement à la lavande :

- Les COV du groupe 1 restent relativement présents tout au long de la période de floraison ; ceci est lié à la formation continue de nouveaux calices (cymes plus longues et nombre de verticilles plus élevé et moins synchronisés que chez la lavande) et donc à la présence de fleurs du stade 1 sur les inflorescences de lavandin tout au long de notre période d'investigation.
- La persistance des COV jusqu'au début du mois de septembre qui correspond aussi au fait que l'inflorescence de lavandin est plus compacte et longue que celle des lavandes. De fait les fleurs sont plus longtemps pérennes.
- L'absence de groupe 3 qui est vraisemblablement masqué par la présence continue de grandes quantités de fleurs non-fanées sur les inflorescences au cours de l'étude.

Résultats



Sur la carte de corrélation, chaque carré coloré indique une corrélation significative. Le signe et la valeur de la corrélation sont donnés par le code couleur suivant. Une corrélation positive est représentée par : [0.8 ; 1] [red], [0.6 ; 0.8] [orange], [0.4 ; 0.6] [light orange], [0.2 ; 0.4] [yellow], [0.1 ; 0.2] [light yellow]. Une corrélation négative : [0.1 et 0.2] [green], [0.2 ; 0.4] [light blue], [0.4 ; 0.6] [blue], [0.6 ; 0.8] [dark blue]. Les corrélations non significatives après correction de Bonferroni du facteur alpha pour des comparaisons multiple est un carré blanc.

Figure 49 : Cartes des corrélations de Spearman des évolutions des COV dans les inflorescences de lavandins.

Le contenu en COV des inflorescences est calculé sur les inflorescences entières de lavandin (10 extractions par cultivars et par semaine entre la mi-mai et début octobre 2007). Le contenu en COV est exprimé en mg.g^{-1} de poids frais avant analyse statistique. Les COV présents dans moins de 5% des échantillons ont été retirés des analyses. Les corrélations entre les changements temporels en contenu de COV ont été calculées par le coefficient de corrélation de Spearman. La significativité des corrélations deux à deux a été testée après correction de Bonferonni du facteur alpha.

A : corrélations entre les COV des inflorescences de 3 cultivars de lavandin (Grosso, Certitude et Super (B110)). B : corrélations entre les COV des inflorescences de 2 cultivars de lavandin (Grosso et Certitude).

Résultats

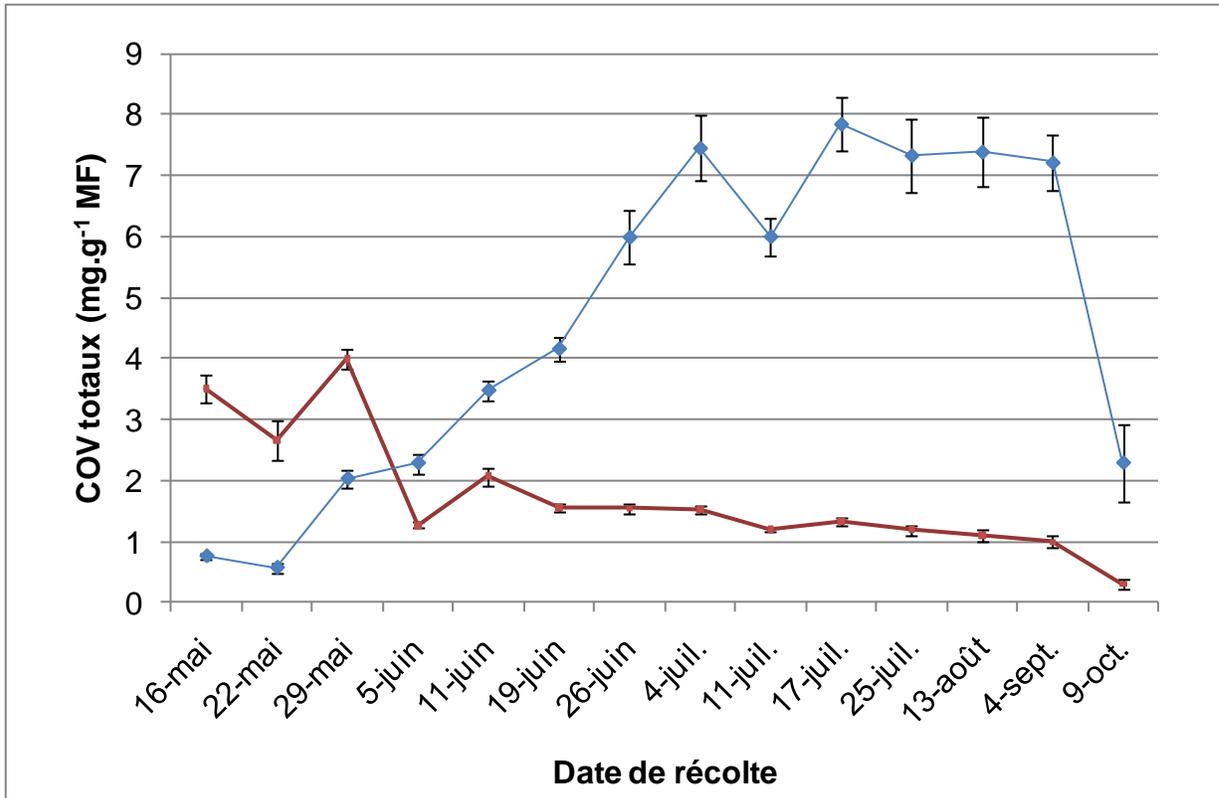


Figure 50 : Evolution au cours de la période de floraison des COV accumulés par les lavandins.

Evolution de la quantité de COV dans les inflorescences de 2 cultivars de lavandins (Grosso et Certitude) de la mi-mai au début octobre 2007. Les COV ont été divisés en 2 groupes sur la base des corrélations de leur accumulation. Les quantités moyennes des COV d'un même groupe ont été sommées et reportés sur le graphique. Groupe 1 (rouge) Groupe 2 (bleu). La barre sur chaque point représente la somme des erreurs standards sur les moyennes.

Résultats

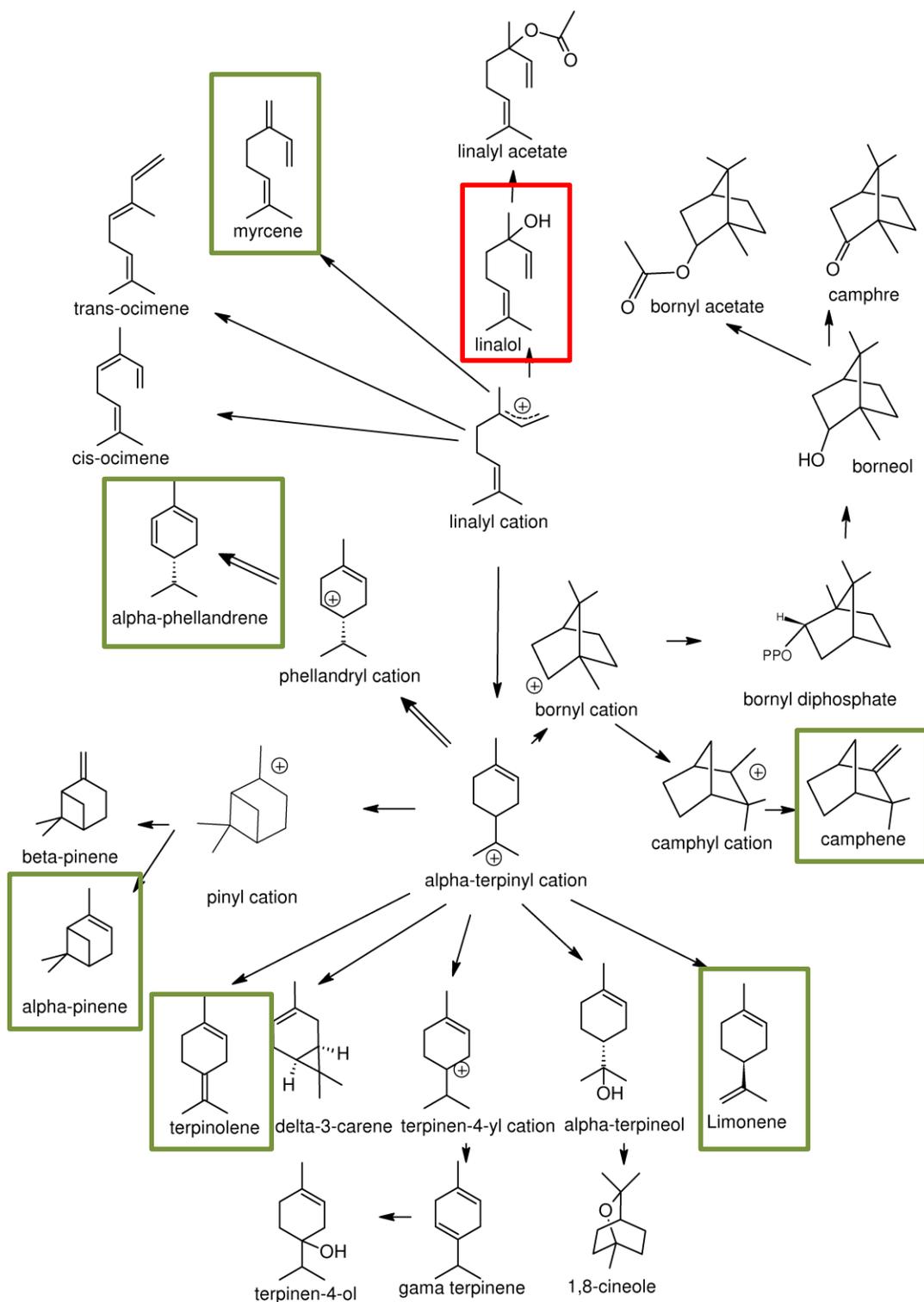


Figure 51 : Voies de synthèse des principaux monoterpènes des lavandes et lavandins

Les molécules encadrées en vert sont celles formées par la LaLIMS et en rouge le linalol formé par la LaLINS (Landmann, Fink *et al.* 2007).

III-3 Discussion

Dans cette étude, nous avons montré pour la première fois que les 20 composés majeurs des HE de *L. angustifolia* et *L. x intermedia* s'accumulent différemment dans les calices des fleurs lors du développement de l'inflorescence. Les analyses des CTV des inflorescences entières au long de la période de floraison montrent que ces CTV peuvent être groupés en 3 groupes chez *L. angustifolia* et 2 groupes chez *L. x intermedia*. Ces groupes correspondent à des phases différentes de la floraison. Ces résultats ont été confirmés sur les différents cultivars de chaque espèce et ce, quatre années de suite. Pour *L. angustifolia* les CTV du groupe 1 sont : δ -3-carène, limonène, myrcène, bornéol, acétate de bornyle, 1,8-cinéol et *trans-beta*-ocimène sont tous des monoterpènes. Les molécules du groupe 2 sont soit des sesquiterpènes (*i.e.* *trans*-farnésène, *beta*-caryophyllène, germacrène D) soit des dérivés acétylés de monoterpènes (*i.e.* acétate de linalyle ou acétate de lavandulyle). Le groupe 3 contient le linalol et le terpinène-4-ol. Chez *L. x intermedia* on ne distingue que deux groupes de CTV. Le groupe 1 est formé, comme chez la lavande, de δ -3-carène, limonène, myrcène, 1,8-cinéol et *trans*-ocimène auxquels s'ajoutent le camphre, l'*alpha* et le *beta*-pinène. Ces trois derniers étant en fait présents chez les lavandes mais en faible quantité et n'ont pas été utilisés dans l'analyse finale des corrélations chez les lavandes. Contrairement aux lavandes le bornéol et son acétate ne sont pas associés au groupe 1. Le bornéol est cependant positivement associé à quelques CTV du groupe 1 (R^2 inférieurs à 0.2). Ceci s'explique par la présence de camphre dans le premier groupe, la biosynthèse de ce dernier étant faite au dépend du bornéol et donc de son acétate (Figure 51). Le deuxième groupe de CTV est composé, comme chez la lavande, de germacrène D, de caryophyllène, de *trans-beta*-farnésène, de l'acétate de lavandulyle, de l'acétate de linalyle. A ceux-ci s'ajoutent le camphène, le lavandulol, l'acétate de bornyle ainsi que les composés du groupe 3 des lavandes (*i.e.* le linalol et le terpinène-4-ol). Chez les lavandes, le camphène et le lavandulol étaient selon les cultivars associés soit au groupe 2 soit au groupe 3 et étaient présents dans peu d'échantillons. Les lavandins produisant de plus grandes quantités d'HE ces composés sont plus souvent détectés dans nos extraits de lavandins ce qui explique qu'on ait pu les associer clairement à un groupe. Le *cis*-ocimène n'est clairement associé à aucun groupe chez la lavande ou le lavandin. Il semble aussi ne pas y avoir de troisième groupe chez les lavandins. Cependant on observe un décalage dans l'accumulation du terpinène-4-ol par rapport aux autres composés du groupe 2 celui-ci

Résultats

atteignant son maximum après que les premières fleurs soient fanées (ceci correspond à la formation des premières graines).

Une étude approfondie des CTV accumulés par des fleurs individuelles de 5 stades différents a démontré que les fleurs individuelles accumulent des CTV différents selon leur stade de maturité. Les stades précoces accumulant plutôt les CTV du groupe 1 et les stades plus tardifs accumulant les CTV des groupes 2 et 3 (chez la lavande et le lavandin). La même étude a également mis en évidence que les CTV accumulés par les calices de fleurs présentant un même stade de développement évoluent au cours des dates de prélèvement. Ceci suggère un certain degré de synchronisation des variations de la composition en CTV des différentes fleurs d'une inflorescence au cours de l'évolution de la saison, celui-ci étant modulé par le stade de développement des fleurs individuelles. L'HE des inflorescences entières représentant un mélange des CTV accumulés par les différentes fleurs présentes au même instant sur une même inflorescence, elle sera affectée par la structure de l'inflorescence (nombre et synchronisation des verticilles, longueur des cymes). Ce type de profil d'accumulation de CTV est plus complexe que ceux précédemment décrits chez des plantes modèle comme *Anthriscum*, *Arabidopsis* ou encore la rose chez qui l'ensemble des COV floraux sont émis de manière concomitante au moment de l'ouverture de la fleur (Dudareva, Cseke *et al.* 1996; Dudareva, Martin *et al.* 2003; Nagegowda, Gutensohn *et al.* 2008). Cependant, dans ces plantes les pétales sont responsables de l'émission des COV alors que chez les lavandes et lavandins les pétales participent peu à l'émission globale de CTV par l'inflorescence, les CTV s'accumulant principalement dans les calices. Les pétales des lavandes et lavandins sont rapidement fanés et accumulent 10 fois moins de CTV que les calices. Le fait que les calices de lavandes et lavandins aient une durée de vie plus longue que les pétales est peut-être responsable de cette variation séquentielle des CTV accumulés du fait du rôle écologique variable des calices en fonction de l'avancement du développement de la fleur. Ceci s'opposerait au rôle écologique unique, et limité dans le temps, des pétales (restreint à l'attraction des pollinisateurs). Ainsi, les terpènes accumulés par les calices pourraient avoir un rôle de protection contre les ravageurs, les composés du groupe 1, dans les fleurs, immatures (d'où peut-être leur similitude avec ceux trouvés dans les feuilles). Ensuite les CTV du groupe 2 initiés avec l'ouverture des corolles pourraient être impliqués dans l'attraction de pollinisateurs. Les composés du groupe 3 n'étant présents que pendant la phase de maturation des graines pourraient assurer un rôle de protection. Cette hypothèse est soutenue par les

Résultats

démonstrations des rôles attractifs pour des pollinisateurs généralistes (impliqués dans la pollinisation des lavandes) joués par les CTV du groupe 2, et les rôles insecticides joué par le linalol et plus fortement par le terpinène-4-ol (<http://www.pherobase.com> Kordali, Kesdek *et al.* 2007; Ma, Han *et al.* 2008). Une étude directe des rôles écologiques des différents mélanges de CTV des différents groupes serait nécessaire pour valider cette hypothèse.

L'activité des TPS de plantes étudiées est régulée au niveau transcriptionnel dans de divers modèles expérimentaux tels que durant le développement des pétales (Nagegowda, Gutensohn *et al.* 2008), la maturation des fruits (Sharon-Asa, Shalit *et al.* 2003), le développement des feuilles (McConkey, Gershenzon *et al.* 2000) ou encore en réponse à un traitement par un éliciteur (Herde, Gartner *et al.* 2008). De la même façon, nos données suggèrent que l'accumulation différentielle des CTV sur les calices est, au moins en partie, expliquée par une régulation de la transcription des gènes de TPS. LaLIMS et les CTV qu'elle produit montrant une décroissance lors du développement de l'inflorescence. LaLINS et les CTV associés ont le même profil d'accumulation en cloche durant les 3 semaines que dure la floraison.

Durant nos quatre années de prélèvement et d'analyses, nous avons observé de temps en temps quelques quantités de CTV accumulés qui étaient 20-50% inférieures à la courbe de tendance temporelle (voir prélèvement du 11 juillet Figure 50 pour un exemple). Tous les CTV des extraits étaient affectés similairement. Les 6 cultivars que nous analysions en parallèle subissaient aussi les mêmes évolutions. Comme nos plantes étaient cultivées en plein champs, il est probable que des évènements climatiques (pluie, vent, froid) aient été à l'origine de ces baisses momentanées des teneurs en CTV. Le stockage des CTV dans les poches d'HE des calices est en effet connue pour présenter des fuites même en conditions climatiques stables puisque des mesures de CTV émis par les inflorescences grâce à des sondes SPME a montré que le bouquet de CTV émis est identique au mélange de CTV stockés (An, Haig *et al.* 2001). Le fait que les prélèvements aux dates suivantes révélaient des taux de CTV en phase avec l'évolution temporelle globale suggère que les calices ont la capacité de rectifier des modifications apportées aléatoirement par l'environnement abiotique. L'hypothèse la plus vraisemblable est que la synthèse des CTV reste active durant toute la durée de vie du calice et qu'elle est régulée en fonction des besoins, ou capacité, de stockage de CTV. Les facteurs environnementaux responsables de la baisse des teneurs en CTV mériteraient d'être définis ainsi que les mécanismes moléculaires mis en jeu pour en corriger l'impact car ils ont un réel impact sur la capacité de production d'une lavanderaie.

Résultats

Des variations dans l'activité transcriptomique de certains gènes ont déjà été démontrées dans d'autres genres végétaux comme les *Physalis* lors du développement des calices. Dans ce genre, l'expression du facteur de transcription MPF2 est induite après la pollinisation de la fleur et conduit au développement de la cage protectrice autour du fruit qui donne cette forme typique au fruit de l'amour en cage (He et Saedler 2005). Les sépales de nombreuses autres espèces de plantes montrent des changements de phénotypes en réponse à la pollinisation. Ces changements sont généralement des dépigmentations (perte des anthocyanes) comme décrit chez *Helleborus* (Herrera 2005). Certaines orchidées modifient l'odeur de leur périanthe après la pollinisation (Ayasse, Schiestl *et al.* 2000; Schiestl, Ayasse *et al.* 1997). La pollinisation peut donc agir comme source de signal pour la modification des odeurs des fleurs de certaines espèces. Qu'en est-il pour les lavandes et les lavandins ? Dans une étude menée sur la production de nectar par les lavandes et lavandin du sud-est de la France (Barbier 1963), il a été démontré que la germination du pollen sur le stigmate de lavande est responsable de l'arrêt de la formation de nectar. Dans cette étude, il est dit que ce sont probablement des sécrétions émises par des abeilles lors du butinage qui déclenchent ce phénomène (Barbier 1963). Cet arrêt de la production de nectar est, chez la lavande et le lavandin, simultanément associé à une perte de turgescence et de la couleur bleue des pétales qui conduisent au flétrissement des pétales (Barbier 1963). L'impact de la visitation des fleurs par les abeilles sur les CTV produits et accumulés dans les calices mériterait donc d'être étudié.

Lors de notre étude précédente (voir chapitre précédent) sur les COV accumulés par 30 taxons de lavande nous avons pu observer que la majorité des espèces de lavande expriment quelques CTV différemment dans leurs feuilles et fleurs. Cependant, la capacité d'exprimer différents gènes de TPS et d'accumuler différents CTV au cours du développement des inflorescences est certainement un événement évolutif restreint à la section botanique *Lavandula*. En effet, une étude conduite sur *L. latifolia* appartenant à la même section que *L. angustifolia* suggère un comportement identique dans l'émission des CTV (Salido, Altarejos *et al.* 2004). Chez cette espèce, le linalol et le terpinène-4-ol étaient majoritairement présents dans les HE d'inflorescences fanées alors que les jeunes inflorescences contenaient de l'acétate de linalyle comme composé dominant (Salido, Altarejos *et al.* 2004). Inversement, *L. stoechas* appartenant au même sous-genre mais à une section botanique différente ne montre aucun changement dans la composition de son HE au cours du développement de son inflorescence (Angioni, Barra *et al.* 2006). Par ailleurs les CTV accumulés dans les calices

Résultats

étaient les même que ceux des feuilles. Ces composés étaient dans une large mesure identiques à ceux appartenant au groupe 1 identifié chez *L. angustifolia*.

Le genre *Lavandula* est donc un modèle intéressant pour l'étude de la régulation de l'expression des gènes de TPS et de l'accumulation de CTV lors du développement des plantes. Le développement des recherches sur le promoteur potentiel de *LaLINS* récemment découvert (Biswas, Foster *et al.* 2009) et l'étude approfondie des banques d'EST obtenues dernièrement (Lane, Boecklemann *et al.*) seront certainement cruciaux pour la suite de ces recherches.

La connaissance sur l'évolution séquentielle de l'apparition des CTV dans les inflorescences de lavande apportée par mes recherches pourrait, en théorie, servir à rationaliser la date de récolte optimale, celle-ci ayant, d'après nos données une influence majeure sur la composition des HE récoltées. En surveillant par exemple l'apparition du terpinène-4-ol on pourrait par exemple éviter une récolte trop tardive. La floraison sur les champs de variétés clonales étant synchronisée cette surveillance pourrait se faire sur quelques pieds choisis au hasard dans chaque parcelle. Mais, cette surveillance sera plus complexe dans les champs de lavandes de population où les différentes plantes présentent de grandes divergences de précocité. Par contre, le coût afférant à la venue, pendant plusieurs semaines d'un technicien sur chaque parcelle et aux analyses GC de tous les échantillons prélevés, constitue un frein économique au développement de cette pratique à grande échelle. Une mesure plus traditionnelle en agronomie du nombre de degrés-jours depuis l'ouverture de la première fleur serait, en revanche, peut être une alternative viable pour définir plus précisément la date optimale de récolte des différents champs. Un suivi de l'évolution des différents CTV au cours du développement des inflorescences peut cependant permettre d'améliorer la comparaison entre variétés dans le cadre d'essais de développement variétal réalisés par la filière car ceux-ci sont réalisés sur la même parcelle, souvent sur une même date (ou au mieux à quelques jours d'intervalle), et ne tiennent donc pas compte de l'influence de la précocité sur les qualité des HE analysées. La connaissance apportée par nos travaux peut en tout cas aider à prédire l'influence de l'architecture de l'inflorescence (un critère facilement détectable) sur la composition en CTV des HE. Dans les plantes ayant un grand nombre de verticilles et de grandes cymes on trouvera une plus grande proportion de composés du groupe 1 (1,8-cinéole et limonène principalement).

CONCLUSIONS ET PERSPECTIVES

Pour ces recherches sur la diversité des composés terpénique volatils accumulés dans les espèces du genre *Lavandula*, j'ai été amené à développer des outils de bioinformatique afin de pouvoir analyser les données obtenues en GC-FID ou GC-MS. En effet, au démarrage de cette thèse, aucun des outils de traitement de données de GC-FID ou de GC-MS ne permettaient de réaliser les comparaisons nécessaires sur nos nombreuses analyses. Ceci s'explique, d'une part par le grand nombre d'analyses réalisées (600 en GC-MS, *cf.* chapitre 2 des résultats) et (2651 en GC-FID, *cf.* chapitre 3 des résultats) et d'autre part par le fait que nous travaillons sur des espèces non-modèles. Les outils de bioinformatique créés m'ont permis d'analyser mes données et ont été utiles à d'autres travaux de recherche au sein du BVpam ainsi que pour des recherches menées dans d'autres laboratoires. Le module MSeasy en particulier a été utilisé pour une étude des COV de 55 espèces de familles différentes d'arbres tropicaux de Guyane (Courtois, Paine *et al.* 2009) et pour une étude par LC-MS du métabolisme secondaire des éponges *Homoscleromorpha* (Ivanisevic). Suite à la présentation du package MSeasy lors du congrès de métabolomique d'Amsterdam (27/06-1/07/2010) nous avons eu des demandes de la part de chercheurs en écologie-chimique souhaitant l'utiliser pour leurs recherches. Comme je l'indique dans la discussion du 1^{er} chapitre de résultats, le module MSeasy est efficace sur des jeux de données relativement grands (jusqu'à 20 000 pics) et pourrait encore être amélioré pour pouvoir traiter plus de pics.

Grâce à MSeasy et à la construction d'un arbre phylogénétique à partir de marqueurs chloroplastiques (*trnK*) nous avons construit une phylogénie moléculaire du genre *Lavandula* sur laquelle nous avons superposé les 141 COV détectés dans 30 taxons du genre *Lavandula*. Les 141 molécules détectées font partie des principales classes de COV (6 dérivés d'acide gras, 56 monoterpènes et 32 sesquiterpènes) auxquelles s'ajoutent l'acétate de crésol et le cryptone. Les structures de 45 de ces molécules restent indéterminées. Cela nous a permis de caractériser la diversité des composés terpéniques volatils accumulés dans le genre *Lavandula* et d'appréhender les liens évolutifs entre des mélanges de COV et les espèces de lavande. La diversité intra-spécifique de certaines espèces comme *L. pedunculata* méritent par ailleurs des études phylogénétiques plus précises afin de vérifier si ces phénomènes de spéciation s'accompagnent de variabilité terpénique. De même, chez *L. stoechas* la sous-espèce *luisieri* possède un profil terpénique atypique avec notamment la production de nécrodanes

Conclusions et perspectives

contrairement à *L. stoechas* ssp. *stoechas*. L'aire de répartition de la première sous-espèce est limitée au centre de l'Espagne alors que *L. stoechas* ssp. *stoechas* se retrouve sur l'ensemble du pourtour méditerranéen. La description précise des terpènes de chacune des espèces encore non étudiée constitue donc un large champ d'investigation phytochimique. De telles études nécessiteront une quantité de matériel végétal conséquente permettant d'effectuer une distillation à partir de laquelle les composés mineurs pourront être plus facilement caractérisés. Certaines espèces botaniques pourront servir de matériel végétal de choix pour la caractérisation de TPS originales encore non-décrites ainsi que l'élucidation des voies de synthèse correspondantes. Par exemple, la voie de biosynthèse des nécrodanes, constituant majeur de *L. stoechas* ssp. *luiseri*, est actuellement inconnue. De même peu de TPS conduisant à des terpènes irréguliers ont été caractérisés. Le lavandulol, un terpène irrégulier que l'on trouve comme composés mineurs chez *L. angustifolia* est produit de manière abondante dans les fleurs de *L. lanata*. Cette espèce est donc un matériel de choix pour caractériser l'enzyme (ou les) qui catalyse la synthèse du lavandulol. Enfin, certaines espèces comme *L. dentata* occupent des zones refuges disjointes (Figure 22). La présence de populations de *L. dentata* au sud de l'Arabie Saoudite ou en Ethiopie est particulièrement intéressante d'un point de vue phytogéographique. Une étude phylogénétique précise de différentes populations de cette espèce pourrait expliquer comment elle a colonisé ces différentes zones géographiques. Par ailleurs, une étude de la diversité terpénique de ces mêmes populations pourrait permettre aussi de mesurer si ces isolements génétiques se sont accompagnés d'évolutions divergentes dans la production de composés volatils organiques.

Les recherches sur l'accumulation des COV dans le genre a été complétée par le suivi de l'accumulation temporelle des composés au cours du développement de l'inflorescence de lavande cultivée (*L. angustifolia*) et du lavandin (*L. x intermedia*). Ce travail montre que le calice des fleurs est capable d'accumuler des composés terpéniques différents suivant sa maturité (Guitton, Nicolè *et al.* 2010b). Sachant que ces composés terpéniques sont synthétisés dans des glandes sécrétrices il faut donc émettre l'hypothèse qu'une même glande sécrétrice est capable de produire des terpènes différents où qu'il existe une mise en place continue sur le calice de glandes au cours du développement de la fleur et que celles-ci synthétisent des composés différents. Si la première hypothèse était vérifiée, cela constituerait un résultat original car il n'existe aucun exemple actuellement de glandes sécrétrices où l'expression de différentes terpènes synthases se succèdent dans le temps. La différenciation des trichomes

Conclusions et perspectives

sécréteurs et leur fonctionnement biosynthétique ont été particulièrement étudiés chez une autre Lamiacée, la menthe. Ces études montrent que les trichomes glandulaires sont différenciés très tôt lors du développement des organes et que l'expression des terpènes synthases est réduite dans le temps aux premiers stades de développement des trichomes. Seule l'expression différentielle de *cytP₄₅₀* au cours du vieillissement des structures sécrétrices a été remarquée. Celle-ci permet la transformation du limonène en pulégone, menthone et enfin menthol. La seconde hypothèse qui conduirait à penser qu'au cours du développement du calice de nouvelles structures sécrétrices sont différenciées reste à vérifier. Cependant, l'observation de jeunes calices met en évidence des glandes sécrétrices très rapprochées ne laissant que peu de place pour l'ajout de trichomes supplémentaires. Une étude histologique de la différenciation des trichomes au cours du développement du calice sera nécessaire pour répondre à ces interrogations. Récemment, une banque de 5000 ESTs a été publiée sur la lavande permettant de caractériser de manière potentielle certaines TPS (Lane, Boeckleemann *et al.*). Un séquençage à haut débit (pyroséquençage 454) des transcrits de fleurs, feuilles et racines de *L. angustifolia* est disponible au laboratoire BVpam depuis octobre 2010. Avec ces outils il faut ajouter la collection de lavandes botaniques dont dispose le laboratoire. En effet, à partir des taxons de cette collection il est envisageable de suivre l'évolution du parfum et de traits morphoanatomiques dans des conditions maîtrisées au travers de croisements et du suivi des descendants. Ces outils permettront la caractérisation d'autres TPS impliquées dans la biosynthèse des terpènes au cours du développement de l'inflorescence. Ces outils ouvrent de nombreuses perspectives dans l'étude de l'expression d'autres TPS mais aussi dans la recherche de facteurs de transcription potentiellement impliqués dans la régulation des TPS. Ces outils permettront aussi de rechercher d'autres gènes impliqués dans les modifications secondaires subies par les terpènes comme par exemple ceux codant pour les hydroxylases ou/et acétyltransferases. Ces familles de gènes sont très intéressantes car de nombreux composés majeurs de l'HE de lavande ou du lavandin sont des terpènes ayant subi des modifications tels que l'acétate de linalyle ou le camphre. Ainsi, l'étude de l'expression des gènes impliqués dans les modifications secondaires pourrait contribuer aussi à l'amélioration de la qualité des HE de lavande.

Conclusions et perspectives

BIBLIOGRAPHIE

-A-

- Adams, R. P. (2007). Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry (4th edition). Carol Stream, Allured Publ. Corporation.
- AFNOR (1996). Huiles essentielles, recueil de normes françaises. AFNOR, AFNOR, Paris La Défense, France.
- Aggarwal, C., A. Hinneburg, et al. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. Database Theory — ICDT 2001, Springer Berlin / Heidelberg. **1973**: 420-434.
- Aharoni, A., A. P. Giri, et al. (2004). "Gain and Loss of Fruit Flavor Compounds Produced by Wild and Cultivated Strawberry Species." Plant Cell **16**(11): 3110-3131.
- Al-Nafie, A. H. (2008). "Phytogeography of Saudi Arabia." Saudi Journal of Biological Sciences **15**(1): 159-176.
- Allan, G. J., J. Francisco-Ortega, et al. (2004). "Molecular phylogenetic evidence for the geographic origin and classification of Canary Island Lotus (Fabaceae: Loteae)." Molecular Phylogenetics and Evolution **32**(1): 123-38.
- Alonso, W. R., J. I. Rajaonarivony, et al. (1992). "Purification of 4S-limonene synthase, a monoterpene cyclase from the glandular trichomes of peppermint (*Mentha x piperita*) and spearmint (*Mentha spicata*)." Journal of Biological Chemistry **267**(11): 7582-7587.
- An, M., T. Haig, et al. (2001). "On-site field sampling and analysis of fragrance from living Lavender (*Lavandula angustifolia* L.) flowers by solid-phase microextraction coupled to gas chromatography and ion-trap mass spectrometry." Journal of Chromatography A **917**(1-2): 245-250.
- Angioni, A., A. Barra, et al. (2006). "Chemical Composition, Seasonal Variability, and Antifungal Activity of *Lavandula stoechas* L. ssp. *stoechas* Essential Oils from Stem/Leaves and Flowers." Journal of Agriculture and Food Chemistry. **54**(12): 4364-4370.
- Arigoni, D., S. Sagner, et al. (1997). "Terpenoid biosynthesis from 1-deoxy-d-xylulose in higher plants by intramolecular skeletal rearrangement." Proceedings of the National Academy of Sciences of the United States of America **94**(20): 10600-10605.
- Arimura, G. I., P. W. H. Dezen, et al. (2004). "Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of terpenoid volatiles in hybrid poplar (*Populus trichocarpa* x *deltoides*): cDNA cloning, functional characterization, and patterns of gene expression of (-)-germacrene D synthase, PtdTPS1." The Plant Journal **37**(4): 603-616.
- Atti-Santos, A. C., M. R. Pansera, et al. (2004). "Seasonal variation of essential oil yield and composition of *Thymus vulgaris* L.(Lamiaceae) from South Brazil." Journal of Essential Oil Research **16**(4): 294-295.
- Aubourg, S. Aubourg, et al. (2002). "Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*." Molecular Genetics and Genomics **267**(6): 730-745.
- Ayasse, M., F. P. Schiestl, et al. (2000). "Evolution of reproductive strategies in the sexually deceptive orchid *Ophrys sphegodes*: how does flower-specific variation of odor signals influence reproductive success?" Evolution **54**(6): 1995-2006.

Bibliographie

Azuma, H., L. B. Thien, et al. (1999). "Molecular Phylogeny of Magnolia (Magnoliaceae) Inferred from cpDNA Sequences and Evolutionary Divergence of the Floral Scents." Journal of Plant Research **112**(3): 291-306.

-B-

- Back, K. and J. Chappell (1995). "Cloning and Bacterial Expression of a Sesquiterpene Cyclase from *Hyoscyamus muticus* and Its Molecular Comparison to Related Terpene Cyclases." Journal of Biological Chemistry **270**(13): 7375-7381.
- Baldovini, N., S. Lavoine-Hanneguelle, et al. (2005). "Necrodane monoterpenoids from *Lavandula luisieri*." Phytochemistry **66**(14): 1651-5.
- Baldovini, N., A. Muselli, et al. (1998). "Variabilite chimique de *Lavandula Stoechas* L. ssp. *stoechas* de Corse." Rivista Italiana EPPOS **9**: 773-780.
- Baldwin, I. T., R. Halitschke, et al. (2006). "Volatile Signaling in Plant-Plant Interactions: "Talking Trees" in the Genomics Era." Science **311**(5762): 812-815.
- Bamba, T. and E. Fukusaki (2006). "Technical problems and practical operations in plant metabolomics." Journal of Pesticide Science **31**(3): 300-304.
- Baran, R., H. Kochi, et al. (2006). "MathDAMP: a package for differential analysis of metabolite profiles." BMC Bioinformatics **7**(1): 530.
- Barbier, E. (1963). "Les lavandes et l'apiculture dans le sud-est de la France." Annales de l'abeille **6**(2): 85-159.
- Barkman, T. J. (2001). "Character coding of secondary chemical variation for use in phylogenetic analyses." Biochemical Systematics and Ecology **29**(1): 1-20.
- Bellakhdar, J. (1997). Contribution à l'étude de la pharmacopée traditionnelle au Maroc : la situation actuelle, les produits, les sources du savoir (enquête ethnopharmacologique de terrain réalisée de 1969 à 1992). Centre des sciences de l'environnement. Metz, Université de Metz. **PhD**: 1157.
- Bellakhdar, J., M. Berrada, et al. (1985). "Etude chimique comparative des huiles essentielles de 10 populations de *Lavandula multifida* L. ." Al Biruniya **tome 1**(2): 95-106.
- Benton, H. P., D. M. Wong, et al. (2008). "XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization." Analytical Chemistry **80**(16): 6382-9.
- Bergougnot, V., J.-C. Caissard, et al. (2007). "Both the adaxial and abaxial epidermal layers of the rose petal emit volatile scent compounds." Planta **226**(4): 853-866.
- Bernard, P. C., R. E. Van Grieken, et al. (1986). "Classification of estuarine particles using automated electron microprobe analysis and multivariate techniques." Environmental Science & Technology **20**(5): 467-473.
- Besser, K., A. Harper, et al. (2009). "Divergent Regulation of Terpenoid Metabolism in the Trichomes of Wild and Cultivated Tomato Species." Plant Physiology **149**(1): 499-514.
- Beyer, H.-G. (1995). "Toward a Theory of Evolution Strategies: Self-Adaptation." Evolutionary Computation **3**(3): 311-347.
- Bino, R. J., R. D. Hall, et al. (2004). "Potential of metabolomics as a functional genomics tool." Trends in Plant Science **9**(9): 418-425.
- Birse, A. C. R., W. F. Bott, et al. (1997). "The Mesozoic and Early Tertiary tectonic evolution of the Socotra area, eastern Gulf of Aden, Yemen." Marine and Petroleum Geology **14**(6): 675-684.

Bibliographie

- Biswas, K., A. Foster, et al. (2009). "Essential oil production: relationship with abundance of glandular trichomes in aerial surface of plants." Acta Physiologiae Plantarum(31): 6.
- Boeckelmann, A. (2008). Monoterpene production and regulation in lavenders (*Lavandula angustifolia* and *Lavandula x intermedia*). Okanaga, University of British Columbia. **Master of Science**.
- Boelens, M. H. (1986). "The essential oil of spike lavender *Lavandula latifolia* Vill.(*L. spica* D. C.)." Perfumer & flavorist **11**(5): 43-63.
- Bohlmann, J., D. Martin, et al. (2000). "Terpenoid Secondary Metabolism in *Arabidopsis thaliana*: cDNA Cloning, Characterization, and Functional Expression of a Myrcene/(E)-[beta]-Ocimene Synthase." Archives of Biochemistry and Biophysics **375**(2): 261-269.
- Bohlmann, J., G. Meyer-Gauen, et al. (1998). "Plant terpenoid synthases: Molecular biology and phylogenetic analysis." Proceedings of the National Academy of Sciences of the United States of America **95**(8): 4126-4133.
- Bohlmann, J., C. L. Steele, et al. (1997). "Monoterpene synthases from Grand Fir (*Abies grandis*). cDNA isolation, characterization, and functional expression of myrcene synthase, (-)-(4s)-limonene synthase, and (-)-(1s,5s)-pinene synthase." Journal of Biological Chemistry **272**(35): 21784-21792.
- Bousmaha, L., J. B. Boti, et al. (2006). "Intraspecific chemical variability of the essential oil of *Lavandula dentata* L. from Algeria." Flavour and Fragrance Journal **21**(2): 368-372.
- Brazma, A., M. Krestyaninova, et al. (2006). "Standards for systems biology." Nature Reviews Genetics **7**(8): 593-605.
- Breitmaier, E. (2006). Terpenes: flavors, fragrances, pharmaca, pheromones. Weinheim, Wiley-VCH.
- Brown, J. T., P. K. Hegarty, et al. (1987). "The toxicity of monoterpenes to plant cell cultures." Plant Science **48**(3): 195-201.
- Burke, C. C., M. R. Wildung, et al. (1999). "Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer." Proceedings of the National Academy of Sciences of the United States of America **96**(23): 13062-13067.
- Bustin, S. A., V. Benes, et al. (2009). "The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments." Clinical Chemistry **55**(4): 611-22.

-C-

- Cane, D. E., Q. Xue, et al. (1996). "Trichodiene Synthase. Probing the Role of the Highly Conserved Aspartate-Rich Region by Site-Directed Mutagenesis" Biochemistry **35**(38): 12369-12376.
- Castle, A. L., O. Fiehn, et al. (2006). "Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results." Briefings in Bioinformatics **7**(2): 159-165.
- Catchpole, G. S., M. Beckmann, et al. (2005). "Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops." Proceedings of the National Academy of Sciences of the United States of America **102**(40): 14458-14462.
- Cavanagh, H. M. A. and J. M. Wilkinson (2002). "Biological activities of Lavender essential oil." Phytotherapy Research **16**(4): 301-308.

Bibliographie

- Chaisse, E. and M. Fontaine (2009). Horizon bleu. CRIEPPAM. Manosque, CRIEPPAM: 2.
- Chapuis, C., M. Barthe, et al. (2006). "Analogues of alpha-Campholenal (= (1R)-2,2,3-Trimethylcyclopent-3-ene-1-acetaldehyde) as Building Blocks for (+)-beta-Necrodol (= (1S,3S)-2,2,3-Trimethyl-4-methylenecyclopentanemethanol) and Sandalwood-like Alcohols13." Helvetica Chimica Acta **89**(11): 2638-2653.
- Charles, D. J. and E. N. C. Renaudand (2002). Comparative study of essential oil quantity and composition from ten cultivars of organically grown lavender and lavandin. Lavender: The Genus Lavandula: 232-242.
- Chatzopoulou, P. S. and A. H. Goliaris (2003). "Contribution to the Analysis of the Volatile Constituents from some Lavender and Lavandin Cultivars Grown in Greece." Scientia Pharmaceutica (Sci. Pharm.) **71**: 229-234.
- Chen, F., D.-K. Ro, et al. (2004). "Characterization of a Root-Specific Arabidopsis Terpene Synthase Responsible for the Formation of the Volatile Monoterpene 1,8-Cineole." Plant Physiology **135**(4): 1956-1966.
- Chograni, H., C. Messaoud, et al. (2008). "Genetic diversity and population structure in Tunisian *Lavandula stoechas* L. and *Lavandula multifida* L. (Lamiaceae)." Biochemical Systematics and Ecology **36**(5-6): 349-359.
- Christianson, D. W. (2006). "Structural Biology and Chemistry of the Terpenoid Cyclases." Chemical Reviews (Washington, D. C.) **37**(44).
- Christianson, D. W. (2008). "Unearthing the roots of the terpenome." Current Opinion in Chemical Biology **12**(2): 141-150.
- Clark, R. J. and R. C. Menary (1980). "Environmental Effects on Peppermint (*Mentha piperita* L.). I. Effect of Daylength, Photon Flux Density, Night Temperature and Day Temperature on the Yield and Composition of Peppermint Oil." Functional Plant Biology **7**(6): 685-692.
- Colby, S. M., W. R. Alonso, et al. (1993). "4S-limonene synthase from the oil glands of spearmint (*Mentha spicata*). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase." Journal of Biological Chemistry **268**(31): 23016-23024.
- Connolly, J. D. and R. A. Hill (1991). Dictionnaire of terpenoids. London, Chapman and Hall.
- Coste, H. and C. Flahault (1998). Flore descriptive et illustrée de la France, de la Corse et des contrées limitrophes. Paris, Librairie scientifique et technique A. Blanchard.
- Costeur, L. and S. Legendre (2008). "Spatial and temporal variation in European Neogene large mammals diversity." Palaeogeography, Palaeoclimatology, Palaeoecology **261**(1-2): 127-144.
- Couderc-Le-Vaillant, M., N. Segur-Fantino, et al. (1990). "Etude phytodermatologique de *Lavandula angustifolia* Mill." Revue de Cytologie et de Biologie Vegetales le Botaniste **13**: 75-88.
- Courtois, E. (2010). Le parfum de la défense des plantes en forêt amazonienne. Laboratoire Evolution et Diversité Biologique. Toulouse, Université Toulouse III – Paul Sabatier. **PhD**.
- Courtois, E., C. Paine, et al. (2009). "Diversity of the Volatile Organic Compounds Emitted by 55 Species of Tropical Trees: a Survey in French Guiana." Journal of Chemical Ecology **35**(11): 1349-1362.
- Croteau, R., M. Felton, et al. (1981). "Relationship of Camphor Biosynthesis to Leaf Development in Sage (*Salvia officinalis*)." Plant Physiology **67**(4): 820-824.
- Cuadros-Inostroza, A., C. Caldana, et al. (2009). "TargetSearch- a Bioconductor package for the efficient pre-processing of GC-MS metabolite profiling data." BMC Bioinformatics **10**(1): 428.

Bibliographie

-D-

- D'Auria, J. C., E. Pichersky, et al. (2007). "Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in *Arabidopsis thaliana*." *The Plant Journal* **49**(2): 194-207.
- Davis, P. H. and I. C. Hedge (1971). "Floristic links between NW Africa and SW Asia." *Annalen des Naturhistorischen Museums in Wien* **75**: 43-57.
- De Pascual-Teresa, J. and G. Machin (1983). "Constituents of the essential oil of *Lavandula latifolia*." *Phytochemistry* **22**(4): 1033-1034.
- De Souza, D. P., E. C. Saunders, et al. (2006). "Progressive peak clustering in GC-MS Metabolomic experiments applied to *Leishmania* parasites." *Bioinformatics* **22**(11): 1391-1396.
- De Vos, M., V. R. Van Oosten, et al. (2005). "Signal Signature and Transcriptome Changes of *Arabidopsis* During Pathogen and Insect Attack." *Molecular Plant-Microbe Interactions* **18**(9): 923-937.
- De Vos, R. C. H., S. Moco, et al. (2007). "Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry." *Nature Protocols* **2**(4): 778-791.
- Degenhardt, J., T. G. Köllner, et al. (2009). "Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants." *Phytochemistry* **70**(15-16): 1621-1637.
- Dettmer, K., P. A. Aronov, et al. (2007). "Mass spectrometry-based metabolomics." *Mass Spectrometry Reviews* **26**(1): 51-78.
- Dewick, P. M. (2002). "The biosynthesis of C 5–C 25 terpenoid compounds." *Natural Product Reports* **19**(2): 181-222.
- Dixon, S. J., R. G. Brereton, et al. (2006). "An automated method for peak detection and matching in large gas chromatography-mass spectrometry data sets." *Journal of Chemometrics* **20**(8-10): 325-340.
- Dixon, S. J., N. Heinrich, et al. (2009). "Use of cluster separation indices and the influence of outliers: application of two new separation indices, the modified silhouette index and the overlap coefficient to simulated data and mouse urine metabolomic profiles." *Journal of Chemometrics* **23**(1): 19-31.
- Dob, T., D. Dahmane, et al. (2006). "Essential Oil Composition of *Lavandula stoechas* from Algeria." *Pharmaceutical Biology (Formerly International Journal of Pharmacognosy)* **44**(1): 60-64.
- Dob, T., D. Dahmane, et al. (2005). "Chemical composition of the essential oil of *Lavandula dentata* L. from Algeria." *International Journal of Aromatherapy* **15**(2): 110-114.
- Dudareva, N., L. Cseke, et al. (1996). "Evolution of Floral Scent in *Clarkia*: Novel Patterns of S-Linalool Synthase Gene Expression in the *C. breweri* Flower." *Plant Cell* **8**(7): 1137-1148.
- Dudareva, N., D. Martin, et al. (2003). "(E)-{beta}-Ocimene and Myrcene Synthase Genes of Floral Scent Biosynthesis in Snapdragon: Function and Expression of Three Terpene Synthase Genes of a New Terpene Synthase Subfamily." *Plant Cell* **15**(5): 1227-1241.
- Dudareva, N. and F. Negre (2005). "Practical applications of research into the regulation of plant volatile emission." *Current Opinion in Plant Biology* **8**(1): 113-118.
- Dudareva, N., F. Negre, et al. (2006). "Plant Volatiles: Recent Advances and Future Perspectives." *Critical Reviews in Plant Sciences* **25**(5): 417-440.
- Duran, A. L., J. Yang, et al. (2003). "Metabolomics spectral formatting, alignment and conversion tools (MSFACTs)." *Bioinformatics* **19**(17): 2283-2293.

Bibliographie

Duriyaprapan, S., E. J. Britten, et al. (1986). "The Effect of Temperature on Growth, Oil Yield and Oil Quality of Japanese Mint." Annals of Botany **58**(5): 729-736.

-E-

El-Hilaly, J., M. Hmammouchi, et al. (2003). "Ethnobotanical studies and economic evaluation of medicinal plants in Taounate province (Northern Morocco)." Journal of Ethnopharmacology **86**(2-3): 149-158.

Emberger, L., H. Gaussen, et al. (1963). Carte bioclimatique de la zone mediterraneenne (Etude ecologique de la zone mediterraneenne), UNESCO-FAO, Paris.

-F-

Fahlén, A., Welander, M. and Wennersten, R. (1997). "Effects of Light-Temperature Regimes on Plant Growth and Essential Oil Yield of Selected Aromatic Plants." Journal of the Science of Food and Agriculture **73**(1): 111-119.

Fahmy, T. (2008). XLStat. Paris, Addinsoft.

Fahn, A. (1988). "Secretory tissues in vascular plants." New Phytologist **108**(3): 229-257.

Fakhari, A. R., P. Salehi, et al. (2005). "Hydrodistillation-headspace solvent microextraction, a new method for analysis of the essential oil components of *Lavandula angustifolia* Mill." Journal of Chromatography A **1098**(1): 14-18.

Felsenstein, J. (1995). PHYLIP (phylogeny inference package). Seattle, University of Washington.

Fiehn, O. (2002). "Metabolomics – the link between genotypes and phenotypes." Plant Molecular Biology **48**(1): 155-171.

Fiehn, O., G. Wohlgemuth, et al. (2008). "Quality control for plant metabolomics: reporting MSI-compliant studies." The Plant Journal **53**(4): 691.

Field, B. and A. E. Osbourn (2008). "Metabolic Diversification--Independent Assembly of Operon-Like Gene Clusters in Different Plants." Science **320**(5875): 543-547.

Figueiredo, A. C., J. G. Barroso, et al. (1995). "Composition of the Essential Oil of *Lavandula pinnata* L. fil. var. *pinnata* grown on madeira." Flavour and Fragrance Journal **10**(2): 93-96.

Flores, G., G. P. Blanch, et al. (2005). "Enantiomeric composition studies in *Lavandula* species using supercritical fluids." Journal of Separation Science **28**(17): 2333-2338.

Fraga, B. M. (2007). "Natural sesquiterpenoids." Natural Product Reports **24**(6): 1350-1381.

Freitas, H. and A. Brehm (2001). "Genetic Diversity of the Macaronesian Leafy Liverwort *Porella canariensis* Inferred From RAPD Markers." Journal of Heredity **92**(4): 339-345.

Fukusaki, E. and A. Kobayashi (2005). "Plant metabolomics: potential for practical operation." Journal of Bioscience and Bioengineering **100**(4): 347-354.

-G-

Gamez, M. J., J. Jimenez, et al. (1990). "Study of the essential oil of *Lavandula dentata* L." Pharmazie **45**(1): 69-70.

Gamez, M. J., J. Jimenez, et al. (1987). "Hypoglycemic activity in various species of the genus *Lavandula*. Part 1: *Lavandula stoechas* L. and *Lavandula multifida* L." Pharmazie **42**(10): 706-7.

Bibliographie

- Gamez, M. J., A. Zarzuelo, et al. (1988). "Hypoglycemic activity in various species of the genus *Lavandula*. Part 2: *Lavandula dentata* and *Lavandula latifolia*." Pharmazie **43**(6): 441-2.
- Garcia-Vallejo, M. I. (1992). Aceites esenciales de las Lavandulas ibéricas: ensayo de la quimiotaxonomía. Departamento de biología vegetal I. Madrid, Universidad Complutense de Madrid. **PhD**: 498.
- Garcia-Vallejo, M. I., M. C. Garcia-Vallejo, et al. (1994). "Necrodane(1, 2, 2, 3, 4-pentamethylcyclopentane) derivatives in *Lavandula luisieri*, new compounds to the plant Kingdom." Phytochemistry **36**(1): 43-45.
- Gershenzon, J. and N. Dudareva (2007). "The function of terpene natural products in the natural world." Nat Chem Biol **3**(7): 408-414.
- Gershenzon, J., M. E. McConkey, et al. (2000). "Regulation of Monoterpene Accumulation in Leaves of Peppermint." Plant Physiology **122**(1): 205-214.
- Gingins de la Sarraz, F. C. J. (1826). Histoire naturelle des lavandes. Paris, Charles -Béchet.
- Goloboff, P. A., S. A. Catalano, et al. (2009). "Phylogenetic analysis of 73 060 taxa corroborates major eukaryotic groups." Cladistics **25**: 1-20.
- Goodacre, R., S. Vaidyanathan, et al. (2004). "Metabolomics by numbers: acquiring and understanding global metabolite data." Trends in Biotechnology **22**(5): 245-252.
- Gotsiou, P., G. Naxakis, et al. (2002). "Diversity in the composition of monoterpenoids of *Origanum microphyllum* (Labiatae)." Biochemical Systematics and Ecology **30**(9): 865-879.
- Gouinguene, S. P. and T. C. J. Turlings (2002). "The Effects of Abiotic Factors on Induced Volatile Emissions in Corn Plants." Plant Physiology **129**(3): 1296-1307.
- Granger, R. and J. Passet (1973). "Thymus vulgaris spontane de France: Races chimiques et chemotaxonomie." Phytochemistry **12**(7): 1683-1691.
- Grayer, R. J., M. W. Chase, et al. (1999). "A comparison between chemical and molecular characters for the determination of phylogenetic relationships among plant families: An appreciation of Hegnauer's "Chemotaxonomie der Pflanzen"." Biochemical Systematics and Ecology **27**(4): 369-393.
- Grayer, R. J., G. C. Kite, et al. (1996). "Intraspecific taxonomy and essential oil chemotypes in sweet basil, *Ocimum basilicum*." Phytochemistry **43**(5): 1033-1039.
- Greenhagen, B. and J. Chappell (2001). "Molecular scaffolds for chemical wizardry: Learning nature's rules for terpene cyclases." Proceedings of the National Academy of Sciences of the United States of America **98**(24): 13479-13481.
- Greenhagen, B. T., P. E. O'Maille, et al. (2006). "Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases." Proceedings of the National Academy of Sciences **103**(26): 9826-9831.
- Guitton, Y. (2010). Diversité des composés terpéniques volatils au sein du genre *Lavandula* : aspects évolutifs et physiologiques. St-Etienne, Université Jean Monnet: 224.
- Guitton, Y., F. Nicolè, et al. (2010a). "Lavender inflorescence: A model to study regulation of terpenes synthesis." Plant Signaling & Behavior **5**(6): 1-3.
- Guitton, Y., F. Nicolè, et al. (2010b). "Differential accumulation of volatile terpene and terpene synthase mRNAs during lavender (*Lavandula angustifolia* and *L. x intermedia*) inflorescence development." Physiologia Plantarum **138**: 150-163.
- Gullberg, J., P. Jonsson, et al. (2004). "Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry." Analytical Biochemistry **331**(2): 283-95.

Bibliographie

-H-

- Halkidi, M., Y. Batistakis, et al. (2002). "Clustering validity checking methods: part II." SIGMOD Record **31**(3): 19-27.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.
- Hallahan, D. L. (2000). "Monoterpenoid biosynthesis in glandular trichomes of Labiate plants." Advances in Botanical Research **31**: 77-120.
- Hanamanthagouda, M. S., S. B. Kakkalameli, et al. (2009). "Essential oils of *Lavandula bipinnata* and their antimicrobial activities." Food Chemistry **In Press, Accepted Manuscript**.
- Hansen, M. E., B. Andersen, et al. (2005). "Automated and unbiased classification of chemical profiles from fungi using high performance liquid chromatography." Journal of Microbiological Methods **61**(3): 295-304.
- Harley, R. M., S. Atkins, et al. (2004). Labiatae. Flowering Plants. Dicotyledons Lamiales (except Acanthaceae including Avicenniaceae). J. W. Kadereit. Berlin, Springer-Verlag. **7**: 167–275.
- Hassan, R. M., R. Scholes, et al. (2005). Ecosystems and human well-being: current state and trends: findings of the Condition and Trends Working Group of the Millennium Ecosystem Assessment. Washington, DC, Island Press.
- He, C. and H. Saedler (2005). "Heterotopic expression of MPF2 is the key to the evolution of the Chinese lantern of *Physalis*, a morphological novelty in Solanaceae." Proceedings of the National Academy of Sciences of the United States of America **102**(16): 5779-5784.
- Heil, M. and J. C. Silva Bueno (2007). "Within-plant signaling by volatiles leads to induction and priming of an indirect plant defense in nature." Proceedings of the National Academy of Sciences **104**(13): 5467-5472.
- Hemmerlin, A., S. B. Rivera, et al. (2003). "Enzymes Encoded by the Farnesyl Diphosphate Synthase Gene Family in the Big Sagebrush *Artemisia tridentata* ssp. *spiciformis*." Journal of Biological Chemistry **278**(34): 32132-32140.
- Hemmerlin, A. a., J.-F. o. Hoefler, et al. (2003). "Cross-talk between the Cytosolic Mevalonate and the Plastidial Methylerythritol Phosphate Pathways in Tobacco Bright Yellow-2 Cells." Journal of Biological Chemistry **278**(29): 26666-26676.
- Hendel-Rahmanim, K., T. Masci, et al. (2007). "Diurnal regulation of scent emission in rose flowers." Planta **226**(6): 1491-1499.
- Herde, M., K. Gartner, et al. (2008). "Identification and Regulation of TPS04/GES, an Arabidopsis Geranylinalool Synthase Catalyzing the First Step in the Formation of the Insect-Induced Volatile C16-Homoterpene TMTT." Plant Cell **20**(4): 1152-1168.
- Herrera, C. M. (2005). "Post-floral perianth functionality: contribution of persistent sepals to seed development in *Helleborus foetidus* (Ranunculaceae)." American Journal of Botany **92**(9): 1486-1491.
- Herrera, C. M. and P. Bazaga (2007). "Adding a third dimension to the edge of a species' range: altitude and genetic structuring in mountainous landscapes." Heredity.
- Herrera, J. (1997). "The role of colored accessory bracts in the reproductive biology of *lavandula stoechas*." Ecology **78**(2): 494-504.
- Hilu, K. W. and H. Liang (1997). "The matK gene: sequence variation and application in plant systematics." American Journal of Botany **84**(6): 830-.

Bibliographie

- Hinneburg, A., C. C. Aggarwal, et al. (2000). What is the nearest neighbor in high dimensional spaces. 26th VLDB Conference
Cairo, Egypt.
- Holm, Y. I., I. Laakso, et al. (1997). "Variation in the essential oil composition of *Artemisia annua* L. of different origin cultivated in Finland." Flavour and Fragrance Journal **12**(4): 241-246.
- Holopainen, J. K. and J. Gershenzon (2010). "Multiple stress factors and the emission of plant VOCs." Trends in Plant Science **15**(3): 176-184.
- Huang, S.-S., B. K. Kirchoff, et al. (2008). "The capitate and peltate glandular trichomes of *Lavandula pinnata* L. (Lamiaceae): histochemistry, ultrastructure, and secretion1." The Journal of the Torrey Botanical Society **135**(2): 155-167.
- Hyatt, D. C., B. Youn, et al. (2007). "Structure of limonene synthase, a simple model for terpenoid cyclase catalysis." Proceedings of the National Academy of Sciences of the United States of America **104**(13): 5360-5365.

-I-

- Iijima, Y., R. Davidovich-Rikanati, et al. (2004). "The Biochemical and Molecular Basis for the Divergent Patterns in the Biosynthesis of Terpenes and Phenylpropenes in the Peltate Glands of Three Cultivars of Basil." Plant Physiology **136**(3): 3724-3736.
- Iijima, Y., D. R. Gang, et al. (2004). "Characterization of Geraniol Synthase from the Peltate Glands of Sweet Basil." Plant Physiology **134**(1): 370-379.
- Imai, H., K. Osawa, et al. (2001). "Inhibition by the essential oils of peppermint and spearmint of the growth of pathogenic bacteria." Microbios **106**: 31-39.
- Iriti, M., G. Colnaghi, et al. (2006). "Histo-cytochemistry and scanning electron microscopy of lavender glandular trichomes following conventional and microwave-assisted hydrodistillation of essential oils: a comparative study." Flavour and Fragrance Journal **21**(4): 704-712.
- Issaq, H. J., Q. N. Van, et al. (2009). "Analytical and statistical approaches to metabolomics research." Journal of Separation Science **32**(13): 2183-2199.
- Ivanišević, J. Métabolisme secondaire des éponges Homoscleromorpha: diversité et fluctuation de son expression en fonction de facteurs biotiques et abiotiques. Nice, Faculté de Sciences, Université de Nice Sophia-Antipolis – CNRS UMR 6001 LCMBA.
- Ivanišević, J., O. Thomas, et al. (2010). "Metabolic fingerprinting as an indicator of biodiversity: towards understanding inter-specific relationships among Homoscleromorpha sponges." Metabolomics: 1-16.

-J-

- Jarvis, B. (2000). "Volume 2. Isoprenoids Including Carotenoids and Steroids Edited by David E. Cane (Brown University). xxxviii + 400 pp." Journal of Natural Products **63**(2): 288-288.
- Jobson, J. D. (1991). Applied multivariate data analysis: Regression and experimental design
Categorical and multivariate methods, Springer.

Bibliographie

- Johnson, C. B., A. Kazantzis, et al. (2004). "Seasonal, populational and ontogenic variation in the volatile oil content and composition of individuals of *Origanum vulgare* subsp. *Hirtum*, assessed by GC headspace analysis and by SPME sampling of individual oil glands." Phytochemical Analysis **15**(5): 286-292.
- Johnson, M. T. J., A. A. Agrawal, et al. (2009). "Heritability, covariation and natural selection on 24 traits of common evening primrose (*Oenothera biennis*) from a field experiment." Journal of Evolutionary Biology **22**(6): 1295-1307.
- Jonsson, P., J. Gullberg, et al. (2004). "A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS." Analytical Chemistry **76**(6): 1738-45.
- Juan, C., B. C. Emerson, et al. (2000). "Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands." Trends in Ecology & Evolution **15**(3): 104-109.
- Jullien, F., F. Diemer, et al. (1998). "An optimising protocol for protoplast regeneration of three peppermint cultivars (*Mentha x piperita*)." Plant Cell, Tissue and Organ Culture **54**(3): 153-159.
- Jung, J., S. Sewenig, et al. (2005). "Comprehensive authenticity assessment of lavender oils using multielement/multicomponent isotope ratio mass spectrometry analysis and enantioselective multidimensional gas chromatography–mass spectrometry." European Food Research and Technology **220**(2): 232-237.

-K-

- Kampranis, S. C., D. Ioannidis, et al. (2007). "Rational Conversion of Substrate and Product Specificity in a Salvia Monoterpene Synthase: Structural Insights into the Evolution of Terpene Synthase Function." Plant Cell **19**(6): 1994-2005.
- Kaplan, F., J. Kopka, et al. (2004). "Exploring the Temperature-Stress Metabolome of Arabidopsis." Plant Physiology **136**(4): 4159-4168.
- Katajamaa, M., J. Miettinen, et al. (2006). "MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data." Bioinformatics **22**(5): 634-636.
- Katoh, S., D. Hyatt, et al. (2004). "Altering product outcome in *Abies grandis* (-)-limonene synthase and (-)-limonene/(-)-[alpha]-pinene synthase by domain swapping and directed mutagenesis." Archives of Biochemistry and Biophysics **425**(1): 65-76.
- Kaufmann, M. and M. Wink (1994). "Molecular systematics of the Nepetoideae (family Labiatae): phylogenetic implications from rbcL gene sequences." Zeitschrift fur Naturforschung C-Journal of Biosciences **49**(9): 635-645.
- Kearsley, A. J., W. E. Wallace, et al. (2005). "A numerical method for mass spectral data analysis." Applied Mathematics Letters **18**(12): 1412-1417.
- Kim, N.-S. and D.-S. Lee (2002). "Comparison of different extraction methods for the analysis of fragrances from *Lavandula* species by gas chromatography-mass spectrometry." Journal of Chromatography A **982**(1): 31-47.
- Knudsen, J., R. Eriksson, et al. (2006). "Diversity and distribution of floral scent." The Botanical Review **72**(1): 1-120.
- Kokkalou, E. (1988). "The Constituents of the Essential Oil from *Lavandula stoechas* Growing Wild in Greece." Planta Medica **54**(1): 58-59.
- Kollner, T. G., C. Schnee, et al. (2004). "The Variability of Sesquiterpenes Emitted from Two *Zea mays* Cultivars Is Controlled by Allelic Variation of Two Terpene Synthase Genes Encoding Stereoselective Multiple Product Enzymes." Plant Cell **16**(5): 1115-1131.

Bibliographie

- Kordali, S., M. Kesdek, et al. (2007). "Toxicity of monoterpenes against larvae and adults of Colorado potato beetle, *Leptinotarsa decemlineata* Say (Coleoptera: Chrysomelidae)." Industrial Crops and Products **26**(3): 278-297.
- Krausz, M. and E. Chaisse (2009). Situation de la production 2009. L'essentiel n°58. **58**: 1.
- Kumar, C. and M. Mann (2009). "Bioinformatics analysis of mass spectrometry-based proteomics data sets." FEBS Letters **583**(11): 1703-12.

-L-

- L'Agriculture drômoise (2007). Lavandiculture drômoise. L'agriculture drômoise. Valence. **1820**.
- Landmann, C., B. Fink, et al. (2007). "Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*)." Archives of Biochemistry and Biophysics **465**(2): 417-429.
- Lane, A., A. Boecklemann, et al. "A genomics resource for investigating regulation of essential oil production in *Lavandula angustifolia*." Planta.
- Lavoine-Hanneguelle, S. and H. Casabianca (2004). "New Compounds from the Essential Oil and Absolute of *Lavandula luisieri* L." Journal of Essential Oil Research **16**: 445-448.
- Leebens-Mack, J., T. Vision, et al. (2006). "Taking the First Steps towards a Standard for Reporting on Phylogenies: Minimum Information about a Phylogenetic Analysis (MIAPA)." OMICS: A Journal of Integrative Biology **10**(2): 231-237.
- Lesburg, C. A., G. Zhai, et al. (1997). "Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology." Science **277**(5333): 1820-1824.
- Levin, R. A., L. A. McDade, et al. (2003). "The Systematic Utility of Floral and Vegetative Fragrance in Two Genera of Nyctaginaceae." Systematic Biology **52**(3): 334-351.
- Lichtenthaler, H. K. (1999). "The 1-deoxy-d-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants." Annual Review of Plant Physiology and Plant Molecular Biology **50**(1): 47-65.
- Lindon, J. C., J. K. Nicholson, et al. (2005). "Summary recommendations for standardization and reporting of metabolic analyses." Nature Biotechnology **23**(7): 833-838.
- Lis-Balchin, M. (2002). Lavender: The Genus *Lavandula*, Taylor & Francis.
- Lis-Balchin, M. and S. Hart (1999). "Studies on the mode of action of the essential oil of Lavender *Lavandula angustifolia* P. Miller." Phytotherapy Research **13**(6): 540-542.
- Lisec, J., N. Schauer, et al. (2006). "Gas chromatography mass spectrometry-based metabolite profiling in plants." Nature Protocols **1**(1): 387-396.
- Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**: 402 - 8.
- Lommen, A. (2009). "MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing." Analytical Chemistry **81**(8): 3079-3086.
- Loreto, F., C. Barta, et al. (2006). "On the induction of volatile organic compound emissions by plants as consequence of wounding or fluctuations of light and temperature." Plant, Cell and Environment **29**(9): 1820-1828.
- Luedemann, A., K. Strassburg, et al. (2008). "TagFinder for the quantitative analysis of gas chromatography--mass spectrometry (GC-MS)-based metabolite profiling experiments." Bioinformatics **24**(5): 732-737.

Bibliographie

-M-

- Ma, Z.-q., X.-l. Han, et al. (2008). "Effects of Terpinen-4-ol on Four Metabolic Enzymes and Polyphenol Oxidase (PPO) in *Mythimna separata* Walker." *Agricultural Sciences in China* **7**(6): 726-730.
- Marculli, C., M. R. Canagaratna, et al. (2006). "Cluster analysis of the organic peaks in bulk mass spectra obtained during the 2002 New England Air Quality Study with an Aerodyne aerosol mass spectrometer." *Atmospheric Chemistry and Physics* **6**(12): 5649-5666.
- Mardon, J. (2010). Chemical communication in petrel seabirds. School of Biomedical, Biomolecular and Chemical Sciences, UWA, Perth, Australia
Centre d'Écologie Fonctionnelle et Évolutive, UMR 5175 CNRS, Montpellier, France.
Montpellier, Université Montpellier II. **PhD**.
- Markovic, I., D. M. Norris, et al. (1996). "Volatiles Involved in the Nonhost Rejection of *Fraxinus pennsylvanica* by *Lymantria dispar* Larvae." *Journal of Agriculture and Food Chemistry*. **44**(3): 929-935.
- Martín-Bravo, S., H. Meimberg, et al. (2007). "Molecular systematics and biogeography of Resedaceae based on ITS and trnL-F sequences." *Molecular Phylogenetics and Evolution* **44**(3): 1105-1120.
- Martin, D., D. Tholl, et al. (2002). "Methyl Jasmonate Induces Traumatic Resin Ducts, Terpenoid Resin Biosynthesis, and Terpenoid Accumulation in Developing Xylem of Norway Spruce Stems." *Plant Physiology* **129**(3): 1003-1018.
- Matsui, K. (2006). "Green leaf volatiles: hydroperoxide lyase pathway of oxylipin metabolism." *Current Opinion in Plant Biology* **9**(3): 274-280.
- McConkey, M. E., J. Gershenzon, et al. (2000). "Developmental Regulation of Monoterpene Biosynthesis in the Glandular Trichomes of Peppermint." *Plant Physiology* **122**(1): 215-224.
- McGarvey, D. J. and R. Croteau (1995). "Terpenoid Metabolism." *Plant Cell* **7**(7): 1015-1026.
- McGeady, P. and R. Croteau (1995). "Isolation and Characterization of an Active-Site Peptide from a Monoterpene Cyclase Labeled with a Mechanism-Based Inhibitor." *Archives of Biochemistry and Biophysics* **317**(1): 149-155.
- McLafferty, F. W. (2006). Wiley Registry of Mass Spectral Data, 8th Edition. J. W. a. S. Inc. New York, John Wiley and Sons Inc.
- Meimberg, H., T. Abele, et al. (2006). "Molecular evidence for adaptive radiation of *Micromeria* Benth. (Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data." *Molecular Phylogenetics and Evolution* **41**(3): 566-78.
- Meinwald, J. and T. Eisner (2008). "Chemical ecology in retrospect and prospect." *Proceedings of the National Academy of Sciences* **105**(12): 4539-4540.
- Messerli, G., V. Partovi Nia, et al. (2007). "Rapid Classification of Phenotypic Mutants of *Arabidopsis* via Metabolite Fingerprinting." *Plant Physiology* **143**(4): 1484-1492.
- Meulenkamp, J. E. and W. Sissingh (2003). "Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African-Eurasian convergent plate boundary zone." *Palaeogeography, Palaeoclimatology, Palaeoecology* **196**(1-2): 209-228.

Bibliographie

- Miikeda, O., K. Kita, et al. (2006). "Phylogenetic relationships of *Clematis* (Ranunculaceae) based on chloroplast and nuclear DNA sequences." Botanical Journal of the Linnean Society **152**(2): 153-168.
- Millar, J., S. Midland, et al. (2005). "(2,3,4,4-Tetramethylcyclopentyl)Methyl Acetate, a Sex Pheromone from the Obscure Mealybug: First Example of a New Structural Class of Monoterpenes." Journal of Chemical Ecology **31**(12): 2999-3005.
- Muller, C. H. (1966). "The role of chemical inhibition (allelopathy) in vegetational composition." Bulletin of the Torrey Botanical Club **93**(5): 332-351.
- Munoz-Bertomeu, J., I. Arrillaga, et al. (2006). "Up-regulation of 1-deoxy-D-xylulose-5-phosphate synthase enhances production of essential oils in transgenic spike lavender." Plant Physiology **142**(3): 890-900.
- Munoz-Bertomeu, J., I. Arrillaga, et al. (2007). "Essential oil variation within and among natural populations of *Lavandula latifolia* and its relation to their ecological areas." Biochemical Systematics and Ecology **35**(8): 479-488.
- Muñoz-Bertomeu, J., R. Ros, et al. (2008). "Expression of spearmint limonene synthase in transgenic spike lavender results in an altered monoterpene composition in developing leaves." Metabolic Engineering **10**(3-4): 166-177.
- Muñoz-Bertomeu, J., E. Sales, et al. (2007). "Up-regulation of an N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase enhances production of essential oils and sterols in transgenic *Lavandula latifolia*." Plant Biotechnology Journal **5**(6): 746-758.
- Murphy, D. M., A. M. Middlebrook, et al. (2003). "Cluster analysis of data from the particle analysis by laser mass spectrometry (PALMS) instrument." Aerosol Science and Technology **37**(4): 382-391.
- Mygind, T., S. Birkelund, et al. (2002). "Determination of PCR efficiency in chelex-100 purified clinical samples and comparison of real-time quantitative PCR and conventional PCR for detection of *Chlamydia pneumoniae*." BMC Microbiology **2**(1): 17.

-N-

- Nagegowda, D. A., M. Gutensohn, et al. (2008). "Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers." The Plant Journal **55**(2): 224-239.
- Negre, F., C. M. Kish, et al. (2003). "Regulation of Methylbenzoate Emission after Pollination in Snapdragon and Petunia Flowers." Plant Cell **15**(12): 2992-3006.
- Neuweger, H., S. P. Albaum, et al. (2008). "MeltDB: a software platform for the analysis and integration of metabolomics experiment data." Bioinformatics **24**(23): 2726-2732.
- Nicolas, J. P., J. M. Fernandez-Palacios, et al. (1989). "Inter-island floristic similarities in the Macaronesian region." Plant Ecology **84**(2): 117-125.
- Nogueira, J. M. F. and A. Romano (2002). "Essential oils from micropropagated plants of *Lavandula viridis*." Phytochemical Analysis **13**(1): 4-7.

-O-

- Ober, D. (2005). "Seeing double: gene duplication and diversification in plant secondary metabolism." Trends in Plant Science **10**(9): 444-449.
- Oliver, S. G., M. K. Winson, et al. (1998). "Systematic functional analysis of the yeast genome." Trends in Biotechnology **16**(9): 373-378.
- ONIPPAM (2006). Rapport d'Activité.

Bibliographie

Owen, S. M., C. Boissard, et al. (2001). "Volatile organic compounds (VOCs) emitted from 40 Mediterranean plant species: : VOC speciation and extrapolation to habitat scale." Atmospheric Environment **35**(32): 5393-5409.

-P-

- Pala-Paul, J., J. J. Brophy, et al. (2004). "Analysis of the volatile components of *Lavandula canariensis* (L.) Mill., a Canary Islands endemic species, growing in Australia." Biochemical Systematics and Ecology **32**(1): 55-62.
- Paton, A. J., D. Springate, et al. (2004). "Phylogeny and evolution of basil and allies (Ocimeae, Labiatae) based on three plastid DNA regions." Molecular Phylogenetics and Evolution **31**(1): 277-299.
- Peakall, R. and P. E. Smouse (2006). "genalex 6: genetic analysis in Excel. Population genetic software for teaching and research." Molecular Ecology Notes **6**(1): 288-295.
- Penn, D. J., E. Oberzaucher, et al. (2007). "Individual and gender fingerprints in human body odour." Journal of The Royal Society Interface **4**(13): 331-340.
- Peñuelas, J. and S. Munné-Bosch (2005). "Isoprenoids: an evolutionary pool for photoprotection." Trends in Plant Science **10**(4): 166-169.
- Perrin, A. and M. Colson (1986). "L'appareil sécréteur des lavandes et des lavandins." Parfums, cosmétiques, arômes: 61-63.
- Peters, S., E. van Velzen, et al. (2009). "Parameter selection for peak alignment in chromatographic sample profiling: objective quality indicators and use of control samples." Analytical and Bioanalytical Chemistry.
- Phillips, M. A. and R. B. Croteau (1999). "Resin-based defenses in conifers." Trends in Plant Science **4**(5): 184-190.
- Pichersky, E. and D. R. Gang (2000). "Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective." Trends in Plant Science **5**(10): 439-445.
- Pichersky, E. and J. Gershenzon (2002). "The formation and function of plant volatiles: perfumes for pollinator attraction and defense." Current Opinion in Plant Biology **5**(3): 237-243.
- Pichersky, E., J. P. Noel, et al. (2006). "Biosynthesis of Plant Volatiles: Nature's Diversity and Ingenuity." Science **311**(5762): 808-811.
- Politi, M., N. De Tommasi, et al. (2002). "Structure and Absolute Configuration of New Diterpenes from *Lavandula multifida*." Journal of Natural Products **65**(11): 1742-1745.
- Prosser, I., I. G. Altug, et al. (2004). "Enantiospecific (+)- and (-)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis aspartate-rich motif." Archives of Biochemistry and Biophysics **432**(2): 136-144.

-R-

- R Development Core Team (2007). R: A language and environment for statistical computing. , R Foundation for Statistical Computing, Vienna, Austria. .
- Rasmann, S., T. G. Kollner, et al. (2005). "Recruitment of entomopathogenic nematodes by insect-damaged maize roots." Nature **434**(7034): 732-737.
- Rasmann, S. and T. C. J. Turlings (2008). "First insights into specificity of belowground tritrophic interactions." Oikos **117**: 362-369.
- RIKILT/Plant-Research-International (2001). metAlign software. Wageningen, The Netherlands **2008**.

Bibliographie

- Ristorcelli, D., F. Tomi, et al. (1998). "C-NMR as a tool for identification and enantiomeric differentiation of major terpenes exemplified by the essential oil of *Lavandula stoechas* L. ssp. *stoechas*." Flavour and Fragrance Journal **13**(3): 154-158.
- Rivera, S. B., B. D. Swedlund, et al. (2001). "Chrysanthemyl diphosphate synthase: Isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*." Proceedings of the National Academy of Sciences of the United States of America **98**(8): 4373-4378.
- Rodriguez-Concepcion, M., O. Fores, et al. (2004). "Distinct Light-Mediated Pathways Regulate the Biosynthesis and Exchange of Isoprenoid Precursors during Arabidopsis Seedling Development." Plant Cell **16**(1): 144-156.
- Rohmer, M. (2007). "Diversity in isoprene unit biosynthesis: The methylerythritol phosphate pathway in bacteria and plastids." Pure and Applied Chemistry **79**(4): 739-752.
- Rohmer, M., M. Knani, et al. (1993). "Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate." Biochemistry Journal **295**(2): 517-524.
- Rohmer, M. and M. Rohmer (1999). "The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants." Natural Product Reports **16**(5): 565-574.
- Rousseuw, P. J. (1987). "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics **20**(1): 53-65.

-S-

- Salido, S., J. Altarejos, et al. (2004). "Chemical composition and seasonal variations of spike lavender oil from Southern Spain." Journal of Essential Oil Research **16**(3): 206-210.
- Sallaud, C., D. Rontein, et al. (2009). "A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato *Solanum habrochaites*." Plant Cell **21**(1): 301-317.
- Sangwan, N. S., A. H. A. Farooqi, et al. (2001). "Regulation of essential oil production in plants." Plant Growth Regulation **34**(1): 3-21.
- Santoyo, S., S. Cavero, et al. (2005). "Chemical Composition and Antimicrobial Activity of *Rosmarinus officinalis* L. Essential Oil Obtained via Supercritical Fluid Extraction." Journal of Food Protection **68**: 790-795.
- Sanz, J., A. C. Soria, et al. (2004). "Analysis of volatile components of *Lavandula luisieri* L. by direct thermal desorption-gas chromatography-mass spectrometry." Journal of Chromatography A **1024**(1-2): 139-146.
- Schauer, N., D. Steinhauser, et al. (2005). "GC-MS libraries for the rapid identification of metabolites in complex biological samples." FEBS Letters **579**(6): 1332-1337.
- Schiestl, F. P., M. Ayasse, et al. (1997). "Variation of Floral Scent Emission and Postpollination Changes in Individual Flowers of *Ophrys sphegodes* Subsp. *sphogodes*." Journal of Chemical Ecology **23**(12): 2881-2895.
- Schmiderer, C., P. Grassi, et al. (2008). "Diversity of essential oil glands of clary sage (*Salvia sclarea* L., Lamiaceae)." Plant Biology **10**(4): 433-440.
- Schnee, C., T. G. Kollner, et al. (2002). "The Maize Gene terpene synthase 1 Encodes a Sesquiterpene Synthase Catalyzing the Formation of (E)-beta -Farnesene, (E)-Nerolidol, and (E,E)-Farnesol after Herbivore Damage." Plant Physiology **130**(4): 2049-2060.

Bibliographie

- Schwab, W. (2003). "Metabolome diversity: too few genes, too many metabolites?" Phytochemistry **62**(6): 837-849.
- Segur-Fantino, N., G. Bahuaud, et al. (1988). "Etude de la variabilité génétique au sein d'une population de lavande (*Lavandula angustifolia* P. Pill.)." Parfums, cosmétiques, arômes(80): 116-120.
- Seibel, P., J. Kruger, et al. (2006). "XML schemas for common bioinformatic data types and their application in workflow systems." BMC Bioinformatics **7**(1): 490.
- Sharon-Asa, L., M. Shalit, et al. (2003). "Citrus fruit flavor and aroma biosynthesis: isolation, functional characterization, and developmental regulation of *Cstps1*, a key gene in the production of the sesquiterpene aroma compound valencene." The Plant Journal **36**(5): 664-674.
- Shellie, R., P. Marriott, et al. (2000). "Characterization and Comparison of Tea Tree and Lavender Oils by Using Comprehensive Gas Chromatography." Journal of High Resolution Chromatography **23**(9): 554-560.
- Shellie, R., L. Mondello, et al. (2002). "Characterisation of lavender essential oils by using gas chromatography-mass spectrometry with correlation of linear retention indices and comparison with comprehensive two-dimensional gas chromatography." Journal of Chromatography A **970**(1-2): 225-234.
- Singh, N., R. Luthra, et al. (1989). "Effect of Leaf Position and Age on the Essential Oil Quantity and Quality in Lemongrass (*Cymbopogon flexuosus*)1." Planta Medica **55**(3): 254-6.
- Skoula, M., C. Abidi, et al. (1996). "Essential oil variation of *Lavandula stoechas* L. ssp. *stoechas* growing wild in crete (Greece)." Biochemical Systematics and Ecology **24**(3): 255-260.
- Slaughter, J. C. (1999). "The naturally occurring furanones: formation and function from pheromone to food." Biological Reviews **74**(3): 259-276.
- Smilde, A., J. Westerhuis, et al. "Dynamic metabolomic data analysis: a tutorial review." Metabolomics.
- Smith, C. A., E. J. Want, et al. (2006). "XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification." Analytical Chemistry **78**(3): 779-787.
- Soltis, P. S. and D. E. Soltis (1995). Plant molecular systematics: inferences of phylogeny and evolutionary processes. Evolutionary biology Vol. 28. **28**: 139-194.
- Sosa, S., G. Altinier, et al. (2005). "Extracts and constituents of *Lavandula multifida* with topical anti-inflammatory activity." Phytomedicine **12**(4): 271-277.
- Starks, C. M., K. Back, et al. (1997). "Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase." Science **277**(5333): 1815-1820.
- Steele, C. L., J. Crock, et al. (1998). "Sesquiterpene Synthases from Grand Fir (*Abies grandis*). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase." Journal of Biological Chemistry **273**(4): 2078-2089.
- Steinbach, M., L. Ertöz, et al. (2004). "The challenges of clustering high dimensional data." New directions in statistical physics: econophysics, bioinformatics, and pattern recognition: 273.
- Steinfath, M., D. Groth, et al. (2008). "Metabolite profile analysis: from raw data to regression and classification." Physiologia Plantarum **132**(2): 150-161.
- Stover, B. and K. Muller (2010). "TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses." BMC Bioinformatics **11**(1): 7.

Bibliographie

- Strömbäck, L., D. Hall, et al. (2007). "A review of standards for data exchange within systems biology." Proteomics **7**(6): 857-867.
- Suarez-Cervera, M. and J. A. Seoane-Camba (1986). Taxonomía numérica de algunas especies de Lavandula L., basada en caracteres morfológicos, cariológicos y palinológicos.
- Sudriá, C., M. T. Piñol, 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**(3): 177-184.
- Sumner, L. W., P. Mendes, et al. (2003). "Plant metabolomics: large-scale phytochemistry in the functional genomics era." Phytochemistry **62**(6): 817-836.

-T-

- Takabayashi, J. and M. Dicke (1996). "Plant--carnivore mutualism through herbivore-induced carnivore attractants." Trends in Plant Science **1**(4): 109-113.
- Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0." Molecular Biology and Evolution **24**(8): 1596-1599.
- Theis, N. and M. Lerchau (2003). "The Evolution of Function in Plant Secondary Metabolites." International Journal of Plant Sciences **164**(s3): S93-S102.
- Tholl, D. (2006). "Terpene synthases and the regulation, diversity and biological roles of terpene metabolism." Current Opinion in Plant Biology **9**(3): 297-304.
- Tholl, D., W. Boland, et al. (2006). "Practical approaches to plant volatile analysis." The Plant Journal **45**(4): 540-560.
- Tholl, D., F. Chen, et al. (2005). "Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers." The Plant Journal **42**(5): 757-71.
- Thulasiram, H. V., H. K. Erickson, et al. (2008). "A Common Mechanism for Branching, Cyclopropanation, and Cyclobutanation Reactions in the Isoprenoid Biosynthetic Pathway." Journal of the American Chemical Society **130**(6): 1966-1971.
- Tikunov, Y., A. Lommen, et al. (2005). "A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles." Plant Physiology **139**(3): 1125-1137.
- Tohge, T. and A. R. Fernie (2009). "Web-based resources for mass-spectrometry-based metabolomics: A user's guide." Phytochemistry **70**(4): 450-456.
- Ton, J., M. D'Alessandro, et al. (2007). "Priming by airborne signals boosts direct and indirect resistance in maize." The Plant Journal **49**(1): 16-26.
- Topcu, G., M. N. Ayril, et al. (2001). "Triterpenoids of the roots of *Lavandula stoechas* ssp. *stoechas*." Pharmazie **56**(11): 892-5.
- Trapp, S. and R. Croteau (2001a). "Defensive resin biosynthesis in conifers." Annual Review of Plant Physiology and Plant Molecular Biology **52**(1): 689-724.
- Trapp, S. C. and R. B. Croteau (2001b). "Genomic Organization of Plant Terpene Synthases and Molecular Evolutionary Implications." Genetics **158**(2): 811-832.
- Trusty, J. L., R. G. Olmstead, et al. (2004). "Using molecular data to test a biogeographic connection of the Macaronesian genus *Bystropogon* (Lamiaceae) to the New World: a case of conflicting phylogenies." Systematic Botany **29**(3): 702-715.
- Tsuro, M., M. Inoue, et al. (2001). "Variation in essential oil components in regenerated lavender (*Lavandula vera* DC) plants." Scientia Horticulturae **88**(4): 309-317.
- Tufféry, S. (2005). Data mining et statistique décisionnelle: L'intelligence dans les bases de données. Paris, Editions Technip.

Bibliographie

Tutin, T. G., V. H. Heywood, et al. (1972). Flora Europaea. Vol. 3. Cambridge, Cambridge University Press.

-U-

Ulubelen, A., N. Goren, et al. (1988). "Longipinene derivatives from *Lavandula stoechas* subs. *stoechas*." Phytochemistry **27**(12): 3966-3967.

Unsicker, S. B., G. Kunert, et al. (2009). "Protective perfumes: the role of vegetative volatiles in plant defense against herbivores." Current Opinion in Plant Biology **12**(4): 479-485.

Untergasser, A., H. Nijveen, et al. (2007). "Primer3Plus, an enhanced web interface to Primer3." Nucleic Acids Research **35**(suppl_2): W71-74.

Upson, T. (2009). The Monograph, an overview: new classification and future prospects. First international Lavender conference, Cambridge university botanic garden and Clare college.

Upson, T. and S. Andrews (2004). The Genus Lavandula. Portland, Timber Press.

Upson, T. M. (1997). Systematics of the genus *Lavandula* L. (Lamiaceae), University of Reading: 446.

Upson, T. M. and S. L. Jury (2002). "A revision of native Moroccan species of *Lavandula* L. section *Pterostoechas* Ging. (Lamiaceae)." Taxon **51**(2): 309-327.

Urwin, N., J. Horsnell, et al. (2007). "Generation and characterisation of colchicine-induced autotetraploid *Lavandula angustifolia*." Euphytica **156**(1): 257-266.

-V-

Valentini, G., N. Arnold, et al. (1993). "Etude chimique comparative des huiles essentielles de quatre populations de *Lavandula stoechas* L. subsp. *stoechas* spontanées de Chypre." Plantes médicinales et phytothérapie **26**(4): 289-299.

Van Dam, N. M., J. A. Harvey, et al. (2003). "Interactions between aboveground and belowground induced responses against phytophages." Basic and Applied Ecology **4**(1): 63-77.

Van Dam, N. M. and G. M. Poppy (2008). "Why plant volatile analysis needs bioinformatics; detecting signal from noise in increasingly complex profiles." Plant Biology **10**(1): 29-37.

Van den Dool, H. and P. D. Kratz (1963). "A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography." Journal of Chromatography A **11**: 463-471.

Van Schie, C., M. Haring, et al. (2007). "Tomato linalool synthase is induced in trichomes by jasmonic acid." Plant Molecular Biology **64**(3): 251-263.

Vander Meer, R. K., M. D. Breed, et al. (1998). Pheromone communication in social insects, Westview Press Boulder, CO.

Venskutonis, P. R., A. Dapkevicius, et al. (1997). "Composition of the essential oil of Lavender (*Lavandula angustifolia* Mill.) from Lithuania." Journal of Essential Oil Research **9**: 107-110.

Verlaque, R., F. Médail, et al. (1997). "Endémisme végétal et paléogéographie dans le Bassin Méditerranéen." Geobios **30**(Supplement 2): 159-166.

Vickers, C. E., J. Gershenzon, et al. (2009). "A unified mechanism of action for volatile isoprenoids in plant abiotic stress." Nature Chemical Biology **5**(5): 283-291.

Bibliographie

Voirin, B. and C. Bayet (1996). "Developmental changes in the monoterpene composition of *Mentha x piperita* leaves from individual peltate trichomes." Phytochemistry **43**(3): 573-580.

-W-

Wagner, G. J. (1991). "Secreting Glandular Trichomes: More than Just Hairs." Plant Physiology **96**(3): 675-679.

Werker, E. (1993). "Function of essential oil-secreting glandular hairs in aromatic plants of Lamiaceae - a review." Flavour and Fragrance Journal **8**(5): 249-255.

Weston, R. J. (2010). "Chemosystematic comparison of essential oils from eight species of four genera of the New Zealand Apiaceae." Biochemical Systematics and Ecology **In Press, Corrected Proof**.

Whittington, D. A., M. L. Wise, et al. (2002). "Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase." Proceedings of the National Academy of Sciences of the United States of America **99**(24): 15375-15380.

Wiesefeld, E. (1997, 23/12/1999). "Aroma profiles of various *Lavandula* species. ." Retrieved 29/07, 2007, from <http://www.sisweb.com/referenc/applnote/noville.htm>.

Wild, D. J. and C. J. Blankley (2000). "Comparison of 2D fingerprint types and hierarchy level selection methods for structural grouping using Ward's clustering." Journal of Chemical Information and Computer Science **40**(1): 155-162.

Williams, D. C., D. J. McGarvey, et al. (1998). "Truncation of Limonene Synthase Preprotein Provides a Fully Active 'Pseudomature' Form of This Monoterpene Cyclase and Reveals the Function of the Amino-Terminal Arginine Pair." Biochemistry **37**(35): 12213-12220.

Wilson, C. A. (2004). "Phylogeny of *Iris* based on chloroplast matK gene and trnK intron sequence data." Molecular Phylogenetics and Evolution **33**(2): 402-412.

Wink, M. (1988). "Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores." TAG Theoretical and Applied Genetics **75**(2): 225-233.

Wink, M. (2003). "Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective." Phytochemistry **64**(1): 3-19.

Wise, M. L., T. J. Savage, et al. (1998). "Monoterpene Synthases from Common Sage (*Salvia officinalis*). cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase." Journal of Biological Chemistry **273**(24): 14891-14899.

-X-

Xia, J., N. Psychogios, et al. (2009). "MetaboAnalyst: a web server for metabolomic data analysis and interpretation." Nucleic Acids Research **37**(suppl_2): W652-660.

Xu, Y.-H., J.-W. Wang, et al. (2004). "Characterization of GaWRKY1, a Cotton Transcription Factor That Regulates the Sesquiterpene Synthase Gene (+)- δ -Cadinene Synthase-A." Plant Physiology **135**(1): 507-515.

Bibliographie

-Y-

- Yanhe, L., L. E. Craker, et al. (1995). Effect of light levels on essential oil production of sage (*S. officinalis*) and thyme (*T. vulgaris*). International Symposium on Medicinal and Aromatic Plants, Amherst, Massachusetts, USA.
- Yassaa, N. and J. Williams (2005). "Analysis of enantiomeric and non-enantiomeric monoterpenes in plant emissions using portable dynamic air sampling/solid-phase microextraction (PDAS-SPME) and chiral gas chromatography/mass spectrometry." Atmospheric Environment **39**(27): 4875-4884.

-Z-

- Zhao, P., K. Inoue, et al. (2003). "Characterization of Leachianone G 2" - Dimethylallyltransferase, a Novel Prenyl Side-Chain Elongation Enzyme for the Formation of the Lavandulyl Group of Sophoraflavanone G in *Sophora flavescens* Ait. Cell Suspension Cultures." Plant Physiology **133**(3): 1306-1313.
- Zhao, Y. and G. Karypis (2003). "Clustering in life sciences." Methods in molecular biology **224**: 183-218.
- Zhao, Y. and G. Karypis (2005). "Data clustering in life sciences." Molecular Biotechnology **31**(1): 55-80.
- Zrira, S. and B. Benjilali (2003). "The constituents of the oils of *Lavandula stoechas* L. ssp. *atlantica* Br.-Bl. and *L. stoechas* ssp. *stoechas* from Morocco." Journal of Essential Oil Research **15**(2): 68-69.

ANNEXES

Annexe I : Table des illustrations

Annexe II : Références produits

Annexe III : Liste des standards utilisés en GC

Annexe IV : Poster MSeasy présenté lors du congrès Metabolomic 2010 Amsterdam 27-Juin/1
Juillet 2010

Annexe V : Liste des 141 COV accumulés dans les taxons de lavandes

Annexe VI : Liste des synapomorphies, autapomorphies et homoplasies

Annexe VII : Coordonnées GPS et altitudes des taxons

Annexe VIII : Annexes numériques

Annexes

Annexe I : Table des illustrations

Liste des tableaux :

Tableau 1 : Historique des principales évolutions de la classification du genre <i>Lavandula</i> (modifié d'après Gingins de la Sarraz 1826; Upson 1997)	37
Tableau 2 : Taxonomie du genre <i>Lavandula</i> (d'après Upson et Andrews 2004).....	38
Tableau 3 : Liste des origines géographiques des espèces du genre <i>Lavandula</i>	51
Tableau 4 : Les huiles essentielles des espèces du genre <i>Lavandula</i> dans la littérature	58
Tableau 5 : Résumé de la classification chémotaxonomique des espèces espagnoles de <i>Lavandula</i> (modifié d'après Garcia-Vallejo 1992)	60
Tableau 6 : Listes des espèces de lavandes utilisées dans les expérimentations de GC-MS, phylogénie et cytométrie en flux en 2006 et 2008.....	73
Tableau 7 : Liste des échantillons de lavande et lavandin récoltés au lycée horticole de Romans-sur-Isère.....	77
Tableau 8 : Liste des amorces utilisées en PCR et qPCR ainsi que les conditions de réactions en PCR.....	84
Tableau 9 : Différentiation des COV accumulés par les feuilles et les fleurs de 23 espèces de lavandes	150
Tableau 10 : Distribution des COV par catégorie chimique au sein du genre <i>Lavandula</i>	152
Tableau 11 : Distribution des couples COV acétylé/oxydés et substrats potentiel au sein du genre <i>Lavandula</i>	153
Tableau 12 : Variation qualitative de la composition des principaux COV dans les feuilles de <i>L. angustifolia</i> et de <i>L. x intermedia</i> au cours de la période de floraison	194

Liste des figures :

Figure 1 : Schéma des interactions des plantes avec leur environnement via les COV (modifié d'après Dudareva, Negre <i>et al.</i> 2006).....	16
Figure 2: Glandes capitées sur des feuilles de <i>L. multifida</i> (microscopie optique).....	19
Figure 3 : Trois types de trichomes sur calice de <i>L. angustifolia</i> (MEB x500 12kV).....	19

Annexes

Figure 4: Détail des 8 cellules sécrétrices d'une glande peltée sur calice de <i>L. lanata</i> (MEB x500 8kV).....	20
Figure 5 : Schéma d'un trichome pelté de lavande (d'après Iriti, Colnaghi <i>et al.</i> 2006)	20
Figure 6 : Schéma de l'IPP et du DMAPP avec la numérotation des carbones	21
Figure 7 : Mécanisme de réaction des prényltransférases illustrant la liaison faite par la GPP synthase, la FPP synthase et la GGPP synthase d'après (Burke, Wildung <i>et al.</i> 1999).....	22
Figure 8 : Principales réaction de liaison à l'origine de la diversité des squelettes carbonés des terpènes (modifié d'après Thulasiram, Erickson <i>et al.</i> 2008).....	23
Figure 9 : Schémas des principales formes de squelettes carbonés des mon et sesquiterpènes (modifié d'après Dudareva, Negre <i>et al.</i> 2006).....	24
Figure 10 : Voies de synthèses des isoprénoïdes dans les plantes (modifié d'après Rodriguez-Concepcion, Fores <i>et al.</i> 2004).....	26
Figure 11 : Principales zones de production d'HE de lavande et lavandin en France (source CPPARM).....	33
Figure 12 : Arbre de classification des espèces du genre <i>Lavandula</i> basé sur les séquences nucléaire ITS (modifié d'après Upson et Andrews 2004).....	39
Figure 13: Variations de la morphologie du calice dans certaines espèces du genre <i>Lavandula</i> (modifié d'après Tim Upson dans Lis-Balchin 2002).....	40
Figure 14 : Variation de la morphologie des feuilles des espèces du genre <i>Lavandula</i>	41
Figure 15 : Schémas des inflorescences du genre <i>Lavandula</i> (d'après Upson 1997).	42
Figure 16 : Photographies de thyrses représentatives de 8 sections du genre <i>Lavandula</i>	43
Figure 17 : Variations de la morphologie des bractées dans les espèces du genre <i>Lavandula</i> (modifié d'après Tim Upson dans Lis-Balchin 2002).....	44
Figure 18: Calice et trichomes de <i>L. angustifolia</i> (MEB x 100 15 kV)	45
Figure 19 : Les trichomes de <i>L. angustifolia</i> (modifié d'après Couderc-Le-Vaillant, Segur-Fantino <i>et al.</i> 1990).....	46
Figure 20 : Variations de la morphologie des fleurs des espèces du genre <i>Lavandula</i> (D'après Upson dans Lis-Balchin 2002)	47

Annexes

Figure 21: Détails de la corolle de <i>L. stoechas</i> et <i>L. pedunculata</i>	48
Figure 22: Schématisation de l'aire de répartition des sections du genre <i>Lavandula</i>	49
Figure 23: Carte des îles de la Macaronésie et de la répartition des espèces de lavande (modifié d'après Allan, Francisco-Ortega <i>et al.</i> 2004)	50
Figure 24 : Reconstruction de la zone d'origine probable du genre <i>Lavandula</i> à la fin du Jurassique (modifié d'après Birse, Bott <i>et al.</i> 1997)	52
Figure 25: Carte détaillée de la répartition des espèces de lavandes sur les îles Canaries et au Maroc (modifié d'après Allan, Francisco-Ortega <i>et al.</i> 2004)	53
Figure 26 : Grandes régions phytogéographiques (modifié d'après Al-Nafie 2008)	54
Figure 27: La cascade des "omiques" (modifié d'après Dettmer, Aronov <i>et al.</i> 2007)	63
Figure 28: Stratégies d'investigation en métabolomique (modifié d'après Dettmer, Aronov <i>et al.</i> 2007).....	65
Figure 29 : Détails de l'échantillonnage de lavandes botanique.	75
Figure 30: Détails des zones de l'inflorescence et des stades de floraison	76
Figure 31: Détail du nombre d'analyses GC-MS réalisées et de leur traitement par identification automatique ou par MSeasy	88
Figure 32: Alignement de chromatogrammes sur la base du RT (modifié d'après Tohge et Fernie 2009).....	97
Figure 33: Schématisation d'une identification automatique d'un composé par son spectre de masse via une base de données informatique.	100
Figure 34: Boîte de dialogue issue des macros Excel	103
Figure 35: Organigramme des étapes des macros Excel	104
Figure 36 : Arbre phylogénétique du genre <i>Lavandula</i> basé sur la comparaison de séquences de <i>trnK</i> chez 30 taxons de lavandes.	140
Figure 37: Variation de la largeur moyenne de l'indice silhouette en fonction du nombre de clusters	144
Figure 38 : Comparaison des nombres de COV détectés dans les feuilles et dans les fleurs de lavande à deux niveaux taxonomiques (sous-genre et section). Les barres violettes	

Annexes

correspondent au sous-genre <i>Lavandula</i> et à ses sections alors que les barres vertes correspondent au sous-genre <i>Fabricia</i> et à ses sections.	147
Figure 39 : Analyse en coordonnées principales (ACoP) de la composition en COV des échantillons de lavandes.	149
Figure 40 : Distribution de la variance de la composition en COV des échantillons de lavandes entre strates de notre échantillonnage. Analyse réalisée avec le logiciel Genalex.	149
Figure 41 : Comparaison des taxons accumulants le myrtenol, le myrtenal, le linalol et son acétate	156
Figure 42: Carte de corrélation des 141 COV de lavande	157
Figure 43: Représentation schématique de la voie de synthèse des mono- et sesquiterpènes.	161
Figure 44 : Les nécrodanes voie de synthèse organique et carte de corrélation.....	165
Figure 45 : Variation de la production de COV dans les feuilles de <i>L. angustifolia</i> et <i>L. x intermedia</i> en mg.g ⁻¹	193
Figure 46 : Evolution au cours de la période de floraison de la morphologie des inflorescences et des COV accumulés chez la lavande.	194
Figure 47: Analyse des correspondances multiples de l'évolution temporelle des COV des inflorescences de 3 cultivars de lavandin (année 2007)	196
Figure 48 : Evolution comparative des inflorescences de 3 cultivars de lavandin au cours de la floraison (année 2007)	196
Figure 49 : Cartes des corrélations de Spearman des évolutions des COV dans les inflorescences de lavandins.	199
Figure 50 : Evolution au cours de la période de floraison des COV accumulés par les lavandins.	200
Figure 51 : Voies de synthèse des principaux monoterpènes des lavandes et lavandins	201

Annexes

Annexe II : Références produits

	Matériel (nombre de pièces)	Fournisseur	Référence	Estimation du coût unitaire Euros HT
Chromatographie	GC-MS	Agilent	5973	50 000
	GC-FID	Agilent	6850	25 000
	Colonne HP-Innowax	Agilent	1909N-133E	400
	Colonne DB-5	Agilent	122-5032E	400
	Capuchon à vis avec septa PTFE (6000)	Agilent	5185-5820	928
	Flacon 1.5 ml pour analyse GC (1000)	Sodipro	LPP11090500	75
	Hexane (2.5L)	Sodipro	SIS322932500	20
	Réducteurs pour tube GC 1.5ml (1000)	Sodipro		305
	Tube 16 ml (800)	Dutcher		306
	Bouchons noir teflon pour tubes 16ml (1400)	Dutcher		894
	Pipette pasteur verre (1500)	Dutcher		
Biologie Moléculaire	GoTaq® DNA Polymérase 2500u			
	Tampon de charge coloré fourni	Promega	E2.33 TAQ EN8	437
	Oligo dT (15 primer) 20 µg	Promega	E2.34 CES CES	48
	M-MLV Reverse transcriptase 10000u	Promega	E2.33 MOD RTV	54
	RQ1 Rnase-free Dnase 1000u	Promega	E2.33 MOD NCA	43
	Rnasin® Ribonuclease inhibitor 10000u	Promega	E2.33 RSN RSN	247
	dATP 100mM	Promega	E2.34 LAB NCL	35
	dGTP 100mM	Promega	E2.34 LAB NCL	35
	dCTP 100mM	Promega	E2.34 LAB NCL	35
	dTTP 100mM	Promega	E2.34 LAB NCL	35
	1 kb DNA ladder 500 µl	Promega	E2.31 NUC LAA	81
	100 bp DNA ladder 250 µl	Promega	E2.31 NUC LAD	67
	Agarose, LE, Analytical Grade 100 g	Promega	E2.25 CHE AGA	83
	Microtube 1.5 ml Treff (1000 pièces)	Dutcher	034013	17
	Microtube PCR Abgène 0.2ml(barrette de 8)	Dutcher	016000	114
	Tubes à centri Falcon 15ml (500)	Dutcher	352097	88
	Tubes à centri Falcon 50ml (200)	Dutcher	352098	86
	Amorces (environ 20 bp) 100µM	Eurofins-MWG		8
	Séquençage (1 Run)	Eurofins-MWG		8
	Kit d'extraction ARN 250 prep	Malchery-Nagel		750
	Kit purification ADN Quia-Quick	Qiagen		
	Plaque 96 puits qPCR (20)	Applied Biosystems	4306737	148
	Film pour plaque 96 (100)	Applied Biosystems	4311971	215
	Tube pour PCR (barrette de 8 par 250)	Dutcher		114
	Thermocycleur	BioRad		5 000
	qPCR 7500	Applied Biosystems		40 000
	SYBR Green Master Mix (400 réactions)	Applied Biosystems	4309155	664
	Kit RACE (5 banques) + enzyme	Ozyme		830 + 350
	Cônes à filtre différents volumes (960)	Starlab		50
	Cônes différents volumes (250)	Dutcher		3
	Bromure d'ethidium (10 ml)			20

Annexes

Annexes

Annexe III : Liste des standards utilisés en GC

Boîte	Composé	Fournisseur	Référence	n° CAS	Toxicité	Quantité
1	Aldehyde C-9	Payan & Bertrand		?	?	1 mL
	Alkane standard C ₂₁ -C ₄₀	Fluka	04071	?	F - X _n	2 mL
	Alkane standard C ₈ -C ₂₀	Fluka	04070	?	F	2 mL
	Benzoate de benzyl	Payan & Bertrand		120-51-4	X _n	5 mL
	Benzyl alcool	Acros	148390050	100-51-6	X _n	5 mL
	(1S)-endo(-)-Borneol	Aldrich	13,911-4	464-45-9	F	25 g
	Bornyl acetate	Aldrich	B55203	5655-61-8		5 mL
	Camphre	Fluka	21293	464-48-2	F - X _n	1 g
	3-Carene	Aldrich	11,557-6	13466-78-9	X _n - N	25 mL
	L(-)Carvone	Acros	154590050	6485-40-1	X _n	5 mL
2	Carvophyllene	Payan & Bertrand		87-44-5	?	5 mL
	Cineole	Acros	110340050	470-82-6		5 mL
	Citral	Payan & Bertrand		5392-40-5	X _i	5 mL
	Citronellal	Payan & Bertrand		106-23-0	X _i	1 mL
	Citronellol	Payan & Bertrand		106-22-9	X _i	5 mL
	β-Citronellol	Aldrich	30,346-1	1117-61-9	X _i	5 mL
	Citronellyl acetate	Payan & Bertrand		150-84-5	?	5 mL
	Damascenone	Payan & Bertrand		23696-85-7	F	1 mL
	α-Damascone	Payan & Bertrand		?	?	1 mL
	β-Damascone	Payan & Bertrand		?	?	1 mL
	Eugenol	Payan & Bertrand		97-53-0	X _n	5 mL
	methyl-Eugenol	Payan & Bertrand		93-15-2	X _n	5 mL
	Geraniol	Sigma	G5135	106-24-1	X _i	10 mL
	Geranyl acetate	Payan & Bertrand		105-87-3		5 mL
3	Germacrene D	RC Treatt	Sa 136310	?	?	10 mL
	1-Heptadecene	Acros	338330010	6765-39-5	?	1 mL
	Hexanal	Sigma	H9008	66-25-1	X _i	100 mL*
	trans-2-Hexenal	Aldrich	13,265-9	6728-26-3	X _n	10 mL
	cis-3-Hexenol	Acros	120770100	928-96-1	?	10 mL
	trans-2-Hexenyl acetate	Acros	310530500	2497-18-9	X _i	50 mL*
	Hexyl formate	Aldrich	W25,700-1	629-33-4	X _i	10 mL
	β-Ionol	Fluka	00468	?		100 µL
	α-Ionone	?	?	127-41-3	?	100 µL
	β-Ionone	?	?	14901-07-6	?	100 µL
	(+/-)-Lavandulol	Fluka	61735	58461-27-1		10 mL
	Lavandulol acetate	Fluka	61736	25905-14-0		5 mL
	(+)Limonene	Acros	179390050	5989-27-5	X _n - N	5 mL
	Linalol	Payan & Bertrand		78-70-6	X _i	5 mL
	Linalyl acetate	Acros	232300050	115-95-7	X _i	5 mL
4	Menthofuran	Roth	5188.1	?	?	1 g
	Menthol	Fluka	63660	89-78-1	X _i	100 g*
	iso-Menthone	Roth	6458.1	491-07-6	?	1 g
	L-Menthone	Acros	203610250	14073-97-3		25 mL
	Menthyl ester	Roth	5189.1	?	?	1 g
	Myrcene	Payan & Bertrand		123-35-3	X _i	5 mL
	Nerol	Payan & Bertrand		106-25-2	?	5 mL
	Nerolidol	Payan & Bertrand		40716-66-3	X _i	1 mL
5	Ocimene	Fluka	74730	3338-55-4		1 mL
	Origan Vert					1 mL
	Oxyde de Rose	Payan & Bertrand				1 mL
	Phenyl acetaldehyde	Acros	148181000	122-78-1	X _n	100 mL*
	Phenyl ethylique alcool	Payan & Bertrand		?	?	5 mL
	α-Pinene	Payan & Bertrand		80-56-8	X _n - N	5 mL
	β-Pinene	Payan & Bertrand		127-91-3	?	5 mL
	Pulegone	Acros	418500050	89-82-7		5 mL
	Saccocalyx saturoïdes					5 mL
	Sclareol	Fluka	74327	515-03-7		1 g
	Terpènes d'Orange	Payan & Bertrand				5 mL
	Terpinen-1-ol-4	Payan & Bertrand		562-74-3	X _n	5 mL
	α-Terpineol	Payan & Bertrand		98-55-5	X _i	5 mL
	Thymol	Acros	150331000	89-83-8	C - N	100 g*
1,3,5-Tri-methoxy-benzene	Acros	160550500	621-23-8	X _n	50 g	

Annexes

MSeasy: an R package for pre-processing of GC/LC-MS data adapted to chemical ecology

Yann Guitton^a, Florence Nicolè^a, Elodie Courtois^b, Jérôme Mardon^c, Martine Hossaert-Mckey^c, Laurent Legendre^a

Contact: yann.guitton@univ-st-etienne.fr

^a Université de Lyon, F-69603, Lyon, France; Université de Saint-Etienne, F-42000, Saint-Etienne, France; BPam, EA 3061: 23 rue du Dr Paul Michelon, F-42000, Saint-Etienne, France.

^b Laboratoire Evolution & Diversité Biologique UMR CNRS 5174 Bâtiment 4R3 Université Paul Sabatier 118, route de Narbonne 31062 Toulouse cedex 4, France.

^c Behavioural Ecology Group, Centre d'Ecologie Fonctionnelle et Evolutive UMR CNRS 5175, 1919 route de Mende, F-34293 Montpellier, cedex 3, France.

Abstract

The democratization of metabolic analyses has extended the scope of metabolomics to ecological questions. Chemical ecology interprets the variation and diversity of chemical signals of non-model organisms in the light of species interactions. Elucidating the biological information within such complex signals, using robust statistical analyses, requires a large number of replicates.

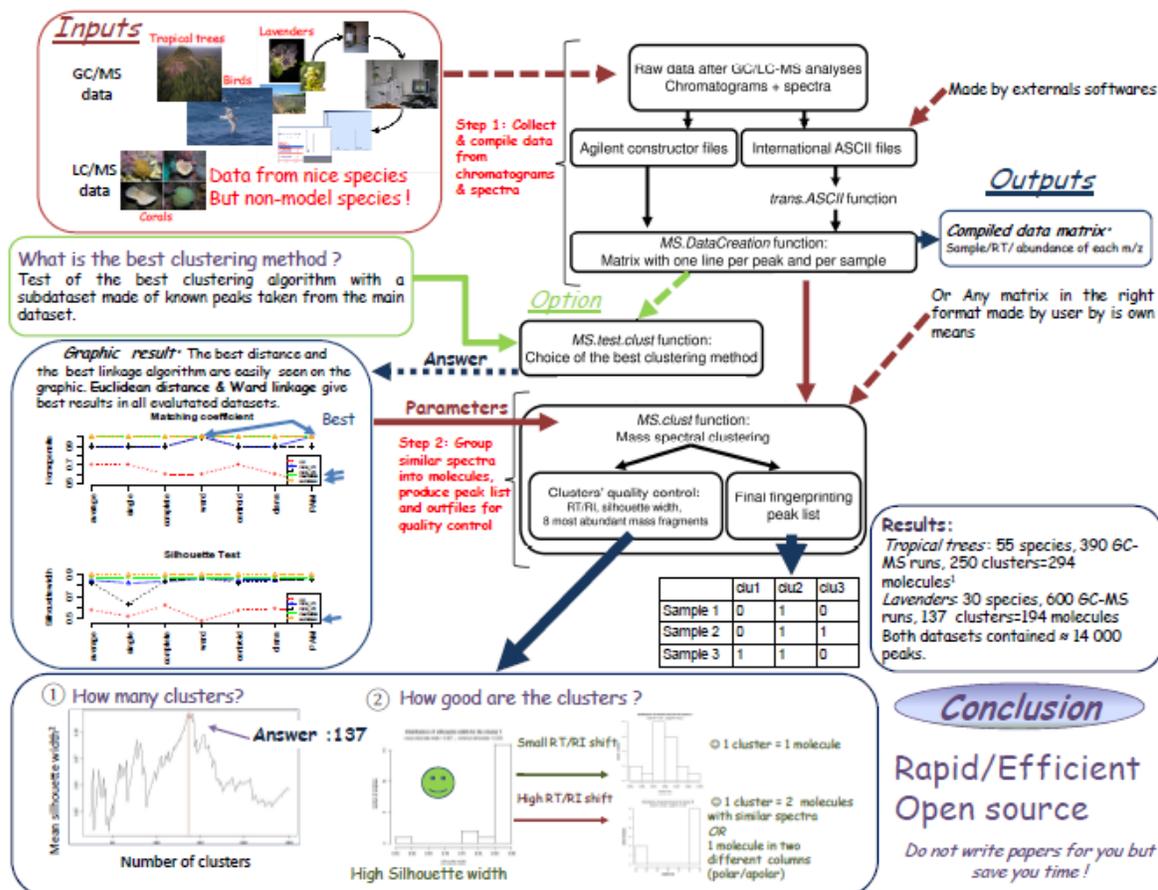
We developed an unsupervised pre-processing method, that generates a fingerprinting peak list from large GC/LC-MS dataset. The method is based on the clustering of mass spectra and does not require any profile correction, retention time alignment or normalization. It is robust to the use of different types of columns and to shifts in retention times particularly common for large/long-term experiments. On the two datasets, used for validation we found that the best clustering method for grouping similar mass spectra was the hierarchical clustering analysis with the Euclidean distance and the Ward linkage. However, it is not excluded that other clustering algorithms could be more adapted for other datasets. For that reason, we have developed a function to identify the best clustering algorithm for each dataset.

Availability : The R package "MSeasy" implementing our pre-processing method is freely available. For non R users a graphical interface was created.

Process description

Objective: unsupervised pre-processing of GC /LC-MS data in chemical ecology investigations

Workflow of the MSeasy package



Metabolomic 2010 congress
Amsterdam 27 June - 1st July

¹ Guitton, Y., Nicolè, F., et al. (2010). "Diversity of the volatile organic compound profiles of Tropical Trees: a Survey in French Guiana." *Journal of Chemical Ecology* 36(12): 1549-1560.
² Guitton, Y. (2007). "MSeasy: a graphical tool for the interpretation and validation of cluster analysis." *Journal of computational and applied mathematics* 202(2): 83-90.

Annexes

Annexes

Annexe V : Liste des 141 COV accumulés dans les taxons de lavandes

	Nom	RI calculé	RI selon Adams	Famille de molécule	Fonctionnalisation
1	<i>cis</i> -3-hexénol	844	844	dérivé AG	alcool
2	hexanol	862	863	dérivé AG	alcool
3	tricyclène	918	921	monoterpène	aliphatique
4	<i>alpha</i> -pinène	930	932	monoterpène	aliphatique
5	<i>alpha</i> -fenchène	944	945	monoterpène	aliphatique
6	camphène	945	946	monoterpène	aliphatique
7	sabinène	969	969	monoterpène	aliphatique
8	<i>beta</i> -pinène	973	974	monoterpène	aliphatique
9	1-octène-3-ol	979	974	dérivé AG	alcool
10	Inconnu 1 'dérivé d'acide gras'	984		dérivé AG	x
11	<i>beta</i> -myrcène	988	988	monoterpène	aliphatique
12	3-octanol	996	988	dérivé AG	alcool
13	<i>delta</i> -3-carène	1005	1008	monoterpène	aliphatique
14	<i>alpha</i> -terpinène	1013	1014	monoterpène	aliphatique
15	<i>p</i> -cymene	1021	1022	monoterpène	réduit
16	limonène	1026	1024	monoterpène	aliphatique
17	<i>beta</i> -phéllandrène	1028	1025	monoterpène	aliphatique
18	1,8-cinéole	1028	1026	monoterpène	alcool
19	Inconnu 2	1032		x	x
20	<i>cis-beta</i> -ocimène	1035	1032	monoterpène linéaire	aliphatique
21	<i>trans-beta</i> -ocimène	1045	1044	monoterpène linéaire	aliphatique
22	Inconnu 3 'monoterpène'	1055		monoterpène	x
23	<i>cis</i> -sabinène hydrate	1067	1095	monoterpène	alcool
24	oxyde de linalol (<i>cis</i>)	1068	1067	monoterpène	alcool
25	terpinolène	1081	1086	monoterpène	aliphatique
26	fenchone	1084	1083	monoterpène	cétone
27	oxyde de linalol (<i>trans</i>)	1084	1084	monoterpène	alcool
28	<i>m</i> -cyménène	1086	1082	monoterpène	aliphatique
29	linalol	1098	1095	monoterpène	alcool
30	Inconnu 4	1105		x	x
31	fenchol	1115	1114	monoterpène	alcool
32	3-acétate d'octyle	1119	1120	dérivé AG	acétylé
33	Inconnu 5 'monoterpène'	1121		monoterpène	x
34	<i>alpha</i> -élémyène	1123	1122	sesquiterpène	aliphatique
35	<i>allo</i> -ocimène	1126	1128	monoterpène linéaire	aliphatique
36	acétate de crésol	1136	1134	x	x
37	<i>trans</i> -pinocarvéol	1136	1135	monoterpène	alcool
38	camphre	1142	1141	monoterpène	cetone
39	<i>trans-alpha</i> -nécredol	1145	1144	monoterpène	alcool
40	pinocarvone	1157	1160	monoterpène	cétone
41	lavandulol	1162	1165	atypique monoterpène	alcool
42	<i>delta</i> -terpinéol	1167	1162	monoterpène	alcool
43	bornéol	1168	1165	monoterpène	alcool
44	<i>cis-alpha</i> -nécredole	1171	1169	monoterpène	alcool
45	Inconnu 6	1172		x	x
46	Inconnu 7 'monoterpène'	1172		monoterpène	alcool
47	terpinène-4-ol	1176	1174	monoterpène	alcool
48	<i>p</i> -cymèn-8-ol	1178	1179	monoterpène	réduit, alcool
49	cryptone	1182	1183	x	x

Annexes

Annexe V suite:

	Nom	RI calculé	RI selon Adams	Famille de molécule	Fonctionnalisation
50	Inconnu 8	1182		x	x
51	Inconnu 9	1186		x	x
52	Inconnu 10 'dérivé d'acide gras'	1191		dérivé AG	x
53	<i>alpha</i> -amorphène	1192	1186	sesquiterpène	aliphatique
54	<i>alpha</i> -campholénal	1192	1186	monoterpène	aldéhyde
55	<i>alpha</i> -terpinéol	1192	1186	monoterpène	alcool
56	myrténol	1192	1194	monoterpène	alcool
57	myrténal	1194	1195	monoterpène	aldéhyde
58	Inconnu 11	1196		x	x
59	verbénone	1202	1204	monoterpène	cétone
60	1-acétate d'octyle	1210	1211	dérivé AG	acétylé
61	dihydrocarvéol	1214	1212	monoterpène	alcool
62	Inconnu 12 'monoterpène'	1215		monoterpène	aldéhyde
63	<i>trans</i> -carvéol	1215	1215	monoterpène	alcool
64	<i>cis</i> -géraniol	1221	1227	monoterpène linéaire	alcool
65	citronellol	1225	1223	monoterpène linéaire	réduit
66	Inconnu 13 'monoterpène'	1235		monoterpène	x
67	Inconnu 14	1235		x	x
68	<i>cis</i> -citral	1236	1235	monoterpène linéaire	aldéhyde
69	cuminaldéhyde	1238	1238	monoterpène	réduit, aldéhyde
70	carvone	1239	1239	monoterpène	cétone
71	pipéritone	1246	1249	monoterpène	cétone
72	<i>trans</i> -géraniol	1248	1249	monoterpène linéaire	alcool
73	carvénone	1253	1255	monoterpène	cétone
74	acétate de linalyl	1253	1254	monoterpène linéaire	acétylé
75	<i>trans</i> -citral	1266	1264	monoterpène linéaire	aldéhyde
76	acétate de bornyle	1281	1287	monoterpène	acétylé
77	acétate de <i>alpha</i> -necrodyle (<i>trans</i>)	1281	1282	monoterpène	acétylé
78	Inconnu 15 'monoterpène'	1282		monoterpène	x
79	p-cymène-7-ol	1287	1289	monoterpène	réduit, alcool
80	Inconnu 16	1293		x	x
81	p-mentha-1-ène-9-ol	1294	1294	monoterpène	alcool
82	Inconnu 17	1306		x	x
83	Inconnu 18 'monoterpène'	1310		monoterpène	cétone
84	Inconnu 19	1327		x	x
85	acétate de néryle (<i>cis</i>)	1356	1359	monoterpène linéaire	acétylé
86	Inconnu 20 'sesquiterpène'	1370		sesquiterpène	x
87	acétate de géranyle (<i>trans</i>)	1376	1379	monoterpène linéaire	acétylé
88	Inconnu 21 'sesquiterpène'	1389		sesquiterpène	x
89	<i>beta</i> -cubébène	1389	1394	sesquiterpène	aliphatique
90	<i>alpha</i> -gurjunène	1400	1404	sesquiterpène	aliphatique
91	<i>alpha</i> -santalène	1414	1416	sesquiterpène	aliphatique
92	<i>beta</i> -caryophyllène	1414	1417	sesquiterpène	aliphatique
93	<i>gamma</i> -éléémène	1424	1434	sesquiterpène	aliphatique
94	<i>trans-alpha</i> -bergamotène	1428	1432	sesquiterpène	aliphatique
95	Inconnu 22 'sesquiterpène'	1439		sesquiterpène	x
96	<i>alpha</i> -humulène	1449	1452	sesquiterpène	aliphatique
97	<i>cis-beta</i> -farnésène	1450	1454	sesquiterpène linéaire	aliphatique
98	<i>allo</i> -aromadendrène	1453	1458	sesquiterpène	aliphatique
99	Inconnu 23 'sesquiterpène'	1454		sesquiterpène	x

Annexes

Annexe V suite:

	Nom	RI calculé	RI selon Adams	Famille de molécule	Fonctionnalisation
100	germacrène D	1475	1484	sesquiterpène	aliphatique
101	phényléthyl-2-méthylbutanoate	1481	1486	x	x
102	<i>beta</i> -sélinène	1482	1489	sesquiterpène	aliphatique
103	Inconnu 24	1482		x	x
104	Inconnu 25 'sesquiterpène'	1486		sesquiterpène	x
105	bicyclogermacrène	1488	1500	sesquiterpène	aliphatique
106	Inconnu 26 'dérivé d'acide gras'	1489		dérivé AG	x
107	<i>delta</i> -sélinène	1489	1492	sesquiterpène	aliphatique
108	<i>alpha</i> -phéllandène	1493	1500	monoterpène	aliphatique
109	Inconnu 27 'sesquiterpène'	1493		sesquiterpène	x
110	Inconnu 28 'sesquiterpène'	1496		sesquiterpène	x
111	<i>alpha</i> -bisabolène	1497	1506	sesquiterpène	aliphatique
112	E,E- <i>alpha</i> -farnésène	1501	1505	sesquiterpène linéaire	aliphatique
113	<i>beta</i> -bisabolène	1503	1505	sesquiterpène	aliphatique
114	Inconnu 29 'sesquiterpène'	1503		sesquiterpène	x
115	<i>alpha</i> -chamigrène	1504	1503	sesquiterpène	aliphatique
116	cubébol	1509	1514	sesquiterpène	alcool
117	Inconnu 30 'sesquiterpène'	1513		sesquiterpène	x
118	Inconnu 31 'sesquiterpène'	1513		sesquiterpène	x
119	<i>trans</i> -calamène	1513	1521	sesquiterpène	aliphatique
120	Inconnu 32 'sesquiterpène'	1518		sesquiterpène	x
121	Inconnu 33 'sesquiterpène'	1526		sesquiterpène	x
122	Inconnu 34 'sesquiterpène'	1526		sesquiterpène	x
123	<i>cis-gamma</i> -bisabolène	1536	1529	sesquiterpène	aliphatique
124	séлина-3,7(11)-diène	1536	1542	sesquiterpène	aliphatique
125	Inconnu 35 'monoterpène'	1547		monoterpène	x
126	germacrène B	1551	1559	sesquiterpène	aliphatique
127	Inconnu 36 'sesquiterpène'	1557		sesquiterpène	x
128	<i>cis-alpha</i> -muuro-5-ène-4-ol	1564	1559	sesquiterpène	alcool
129	germacrène D-4-ol	1570	1574	sesquiterpène	alcool
130	oxyde de caryophyllène	1574	1582	sesquiterpène	époxy
131	Inconnu 37 'monoterpène'	1576		monoterpène	x
132	viridiflorol	1587	1592	sesquiterpène	alcool
133	lédol	1596	1602	sesquiterpène	alcool
134	1,10-di- <i>epi</i> -cubénol	1611	1618	sesquiterpène	alcool
135	Inconnu 38 'sesquiterpène'	1623		sesquiterpène	x
136	Inconnu 39	1630		x	x
137	<i>epi-alpha</i> -cadinol	1635	1638	sesquiterpène	alcool
138	Inconnu 40 'sesquiterpène'	1650		sesquiterpène	x
139	Inconnu 41	1651		x	x
140	Inconnu 42	1597		x	x
141	éremophilène	1492	1490*	sesquiterpène	aliphatique

NB: les molécules inconnues sont caractérisées par un MS et RI

NB: prédiction de la famille chimique des composés inconnus sur la base de fragments de masse spécifiques

RI*: référence externe au livre de Adams

Annexes

Annexes

Annexe VI : Liste des synapomorphies, autapomorphies et homoplasies

	Nom	Synapomorphie	Homoplasie intraclade	Homoplasie sur le genre	Autapomorphie
1	<i>cis</i> -3-hexénol	non	non	oui	non
2	hexanol	non	non	non	oui
3	tricyclène	s gen lavandula	oui	non	non
4	<i>alpha</i> -pinène	non	non	oui	non
5	<i>alpha</i> -fenchène	non	non	non	oui
6	camphène	s gen lavandula	non	non	non
7	sabinène	non	non	oui	non
8	<i>beta</i> -pinène	non	non	oui	non
9	1-octène-3-ol	non	non	oui	non
10	Inconnu 1 'dérivé d'acide gras'	non	non	oui	non
11	<i>beta</i> -myrcène	non	non	oui	non
12	3-octanol	non	non	oui	non
13	<i>delta</i> -3-carène	non	non	oui	non
14	<i>alpha</i> -terpinène	non	non	oui	non
15	p-cymene	non	non	oui	non
16	limonène	non	non	oui	non
17	<i>beta</i> -phéllandrène	non	non	oui	non
18	1,8-cinéole	s gen lavandula	oui	non	non
19	Inconnu 2	s gen lavandula	oui	non	non
20	<i>cis-beta</i> -ocimène	non	non	oui	non
21	<i>trans-beta</i> -ocimène	non	non	oui	non
22	Inconnu 3 'monoterpène'	non	non	non	oui
23	<i>cis</i> -sabinène hydrate	s gen lavandula	oui	non	non
24	oxyde de linalol (<i>cis</i>)	s gen lavandula	oui	non	non
25	terpinolène	non	non	oui	non
26	fenchone	non	non	oui	non
27	oxyde de linalol (<i>trans</i>)	s gen lavandula	oui	non	non
28	<i>m</i> -cyménène	sec pterostoechas	oui	non	non
29	linalol	non	non	oui	non
30	Inconnu 4	s gen lavandula	oui	non	non
31	fenchol	non	non	oui	non
32	3-acétate d'octyle	non	non	oui	non
33	Inconnu 5 'monoterpène'	non	non	non	oui
34	<i>alpha</i> -élémente	s gen lavandula	oui	non	non
35	<i>allo</i> -ocimène	non	non	oui	non
36	acétate de crésol	non	non	non	oui
37	<i>trans</i> -pinocarvéol	s gen lavandula	oui	non	non
38	camphre	s gen lavandula	non	non	non
39	<i>trans-alpha</i> -nécrodol	sec stoechas	oui	non	non
40	pinocarvone	s gen lavandula	oui	non	non
41	lavandulol	s gen lavandula	oui	non	non
42	<i>delta</i> -terpinéol	s gen lavandula	oui	non	non
43	bornéol	s gen lavandula	non	non	non
44	<i>cis-alpha</i> -nécrodole	non	non	non	oui
45	Inconnu 6	non	non	non	oui
46	Inconnu 7 'monoterpène'	s gen lavandula	oui	non	non
47	terpinène-4-ol	s gen lavandula	oui	non	non
48	p-cymèn-8-ol	non	non	oui	non
49	cryptone	s gen lavandula	oui	non	non

Annexes

Annexe VI suite:

	Nom	Synapomorphie	Homoplasie intraclade	Homoplasie sur le genre	Autapomorphie
50	Inconnu 8	s gen lavandula	oui	non	non
51	Inconnu 9	s gen lavandula	oui	non	non
52	Inconnu 10 'dérivé d'acide gras'	non	non	non	oui
53	<i>alpha</i> -amorphène	non	non	oui	non
54	<i>alpha</i> -campholénal	sec stoechas	oui	non	non
55	<i>alpha</i> -terpinéol	non	non	oui	non
56	myrténol	s gen lavandula	oui	non	non
57	myrténal	s gen lavandula	oui	non	non
58	Inconnu 11	s gen fabricia	oui	non	non
59	verbénone	s gen lavandula	oui	non	non
60	1-acétate d'octyle	non	non	non	oui
61	dihydrocarvéol	non	non	non	oui
62	Inconnu 12 'monoterpène'	s gen fabricia	oui	non	non
63	<i>trans</i> -carvéol	s gen lavandula	oui	non	non
64	<i>cis</i> -géraniol	sec pterostoechas	oui	non	non
65	citronellol	non	non	non	oui
66	Inconnu 13 'monoterpène'	sec pterostoechas	oui	non	non
67	Inconnu 14	s gen fabricia	oui	non	non
68	<i>cis</i> -citral	s gen fabricia	oui	non	non
69	cuminaldéhyde	non	non	non	oui
70	carvone	s gen lavandula	oui	non	non
71	pipéritone	non	non	non	oui
72	<i>trans</i> -géraniol	non	non	oui	non
73	carvénone	non	non	non	oui
74	acétate de linalyl	non	non	oui	non
75	<i>trans</i> -citral	s gen fabricia	oui	non	non
76	acétate de bornyle	s gen lavandula	oui	non	non
77	acétate de <i>alpha</i> -necrodyle (<i>trans</i>)	non	non	non	oui
78	Inconnu 15 'monoterpène'	s gen lavandula	oui	non	non
79	p-cymène-7-ol	s gen lavandula	oui	non	non
80	Inconnu 16	s gen lavandula	oui	non	non
81	p-mentha-1-ène-9-ol	s gen fabricia	oui	non	non
82	Inconnu 17	non	non	oui	non
83	Inconnu 18	s gen fabricia	oui	non	non
84	Inconnu 19	non	non	non	oui
85	acétate de néryle (<i>cis</i>)	s gen lavandula	oui	non	non
86	Inconnu 20 'sesquiterpène'	sec stoechas	oui	non	non
87	acétate de géranyle (<i>trans</i>)	non	non	oui	non
88	Inconnu 21 'sesquiterpène'	non	non	non	oui
89	<i>beta</i> -cubébène	non	non	non	oui
90	<i>alpha</i> -gurjunène	s gen lavandula	oui	non	non
91	<i>alpha</i> -santalène	non	non	non	oui
92	<i>beta</i> -caryophyllène	non	non	oui	non
93	<i>gamma</i> -élémente	non	non	oui	non
94	<i>trans</i> - <i>alpha</i> -bergamotène	s gen lavandula	oui	non	non
95	Inconnu 22 'sesquiterpène'	s gen lavandula	oui	non	non
96	<i>alpha</i> -humulène	non	non	oui	non
97	<i>cis</i> - <i>beta</i> -farnésène	non	non	oui	non
98	<i>allo</i> -aromadendrène	sec stoechas	oui	non	non
99	Inconnu 23 'sesquiterpène'	s gen lavandula	oui	non	non

Annexes

Annexe VI suite:

	Nom	Synapomorphie	Homoplasie intraclade	Homoplasie sur le genre	Autapomorphie
100	germacrène D	non	non	oui	non
101	phényléthyl-2-méthylbutanoate	non	non	non	oui
102	<i>beta</i> -sélinène	non	non	non	oui
103	Inconnu 24	non	non	non	oui
104	Inconnu 25 'sesquiterpène'	sec stoechas	non	non	non
105	bicyclgermacrène	non	non	oui	non
106	Inconnu 26 'dérivé d'acide gras'	sec stoechas	oui	non	non
107	<i>delta</i> -sélinène	sec stoechas	non	non	non
108	<i>alpha</i> -phéllandrène	non	non	oui	non
109	Inconnu 27 'sesquiterpène'	sec stoechas	oui	non	non
110	Inconnu 28 'sesquiterpène'	s gen lavandula	oui	non	non
111	<i>alpha</i> -bisabolène	sec stoechas	oui	non	non
112	E,E- <i>alpha</i> -farnésène	non	non	oui	non
113	<i>beta</i> -bisabolène	non	non	oui	non
114	Inconnu 29 'sesquiterpène'	non	non	non	oui
115	<i>alpha</i> -chamigrène	non	non	non	oui
116	cubébol	sec stoechas	oui	non	non
117	Inconnu 30 'sesquiterpène'	sec lavandula	oui	non	non
118	Inconnu 31 'sesquiterpène'	sec stoechas	oui	non	non
119	<i>trans</i> -calamène	s gen lavandula	oui	non	non
120	Inconnu 32 'sesquiterpène'	sec lavandula	oui	non	non
121	Inconnu 33 'sesquiterpène'	non	non	non	oui
122	Inconnu 34 'sesquiterpène'	sec stoechas	oui	non	non
123	<i>cis-gamma</i> -bisabolène	s gen lavandula	oui	non	non
124	séline-3,7(11)-diène	sec stoechas	non	non	non
125	Inconnu 35 'monoterpène'	non	non	non	oui
126	germacrène B	non	non	oui	non
127	Inconnu 36 'sesquiterpène'	sec stoechas	oui	non	non
128	<i>cis-alpha</i> -muuro-5-ène-4-ol	non	non	non	oui
129	germacrène D-4-ol	non	non	oui	non
130	oxyde de caryophyllène	non	non	oui	non
131	Inconnu 37 'monoterpène'	non	non	non	oui
132	viridiflorol	s gen lavandula	oui	non	non
133	lédol	s gen lavandula	oui	non	non
134	1,10-di- <i>epi</i> -cubébol	s gen lavandula	oui	non	non
135	Inconnu 38 'sesquiterpène'	non	non	non	oui
136	Inconnu 39	non	non	non	oui
137	<i>epi-alpha</i> -cadinol	non	non	non	oui
138	Inconnu 40 'sesquiterpène'	non	non	non	oui
139	Inconnu 41	non	non	non	oui
140	Inconnu 42	sec stoechas	non	non	non
141	érémophilène	non	non	non	oui

NB: les molécules inconnues sont caractérisées par un MS et RI

NB: prédiction de la famille chimique des composés inconnus sur la base de fragments de masse spécifiques

RI*: référence externe au livre de Adams

Annexes

Annexes

Annexe VII : Coordonnées GPS et altitudes des taxons

Espece	pays	altitude	Coordonnées GPS	
<i>Lavandula angustifolia</i>	culture	n.d	N 45 50,669	E 5 58,399
<i>Lavandula angustifolia</i> Mill. ssp. <i>pyrenaica</i> (DC.) Guinea var. <i>pyrenaica</i>	Espagne	n.d	N 41 25,617	WO 0 42,244
<i>Lavandula angustifolia</i> Mill. ssp. <i>pyrenaica</i> (DC.) Guinea var. <i>turolenis</i> (Pau) O. Bolos & J. Vigo	Espagne	n.d	N 40 17,701	WO 0 30,506
<i>Lavandula antheae</i> Maire ssp. <i>antheae</i>	Algérie*	n.d	N 23 26,440	E 6 11,653
<i>Lavandula aristibracteata</i> A.G. Mill.	Somalie	n.d	N 10 48,000	E 47 22,000
<i>Lavandula buchii</i> Webb & Berthel.	Iles Canaries	155 m	N 28 31,409	WO 16 11,607
<i>Lavandula buchii</i> Webb & Berthel. (gracile) vert	Iles Canaries	192 m	N 28 21,934	WO 16 52,984
<i>Lavandula CANARENSIS</i> VAR <i>CANARIAE</i>	Iles Canaries	123 m	N 28 23,730	WO 16 39,380
<i>Lavandula citriodora</i> A.G. Miller	Yémen	n.d	N 15 17,371	E 44 09,476
<i>Lavandula coronopifolia</i> Poir. VAR OCCIDENTALIS	Maroc	706 m	N 29 44,341	WO 07 55,018
<i>Lavandula coronopifolia</i> Poir. VAR ORIENTALIS	Jordanie	386 m	N 30 38,825	E 35 31,256
<i>Lavandula dentata</i> L. var. <i>CANDICANS</i>	Maroc	70 m	N 31 47,504	WO 09 34,079
<i>Lavandula dentata</i> L. var. <i>dentata</i>	sud du Maroc	n.d	N 30 19,564	WO 08 33,889
<i>Lavandula DHOFARENSIS</i> SSP <i>DHOFARENSIS</i>	Oman	n.d	N 17 13,342	E 54 4,249
<i>Lavandula lanata</i> Bois.	Espagne	777 m	N 36 45,064	WO 04 55,776
<i>Lavandula lanata</i> Medik.	Espagne	531 m	N 42 39,767	WO 03 01,890
<i>Lavandula macra</i> Baker	SE du Yémen	n.d	N 15 58,168	E 48 45,890
<i>Lavandula mairei</i> Humbert var. <i>atlantica</i> Maire	Maroc	1564 m	N 29 44,612	WO 08 49,529
<i>Lavandula mairei</i> Humbert var. <i>mairei</i>	Maroc	1362 m	N 30 47,584	WO 06 43,338
<i>Lavandula maroccana</i> Murb.	Maroc	484 m	N 31 12,742	WO 08 04,417
<i>Lavandula minutifolia</i> Bolle var. <i>minutifolia</i>	Iles Canaries	640 m	N 27 55,213	WO 15 33,374
<i>Lavandula multifida</i> L.	Sud Tunisie	n.d	N 32 53,394	E 10 16,589
<i>Lavandula pedunculata</i> (Mill.) Cav. ssp. <i>ATLANTICA</i>	Maroc	1919 m	N 30 51,393	WO 08 23,262
<i>Lavandula pedunculata</i> (Mill.) Cav. ssp. <i>Iustanica</i>	Portugal	54 m	N 39 16,526	WO 08 33,117
<i>Lavandula pedunculata</i> (Mill.) Cav. ssp. <i>pedunculata</i> (Mill.) Sampa. ex Rozeira	Espagne	631 m	N 40 35,418	WO 06 40,396
<i>Lavandula pedunculata</i> (Mill.) Cav. ssp. <i>sampalana</i> Rozeira	Portugal	313 m	N 39 55,797	WO 07 38,261
<i>Lavandula pinata</i> L. f.	Iles Canaries	370 m	N 29 07,992	WO 13 30,311
<i>Lavandula pubescens</i> Decne.	Jordanie	469 m	N 30 39,267	E 35 32,490
<i>Lavandula rejdalii</i> Upson	Maroc	1239 m	N 29 51,192	WO 09 00,333
<i>Lavandula rotundifolia</i> Benth. Var. " <i>flora rosea</i> "	Cap Vert	n.d	N 14 55,485	WO 24 21,793
<i>Lavandula rotundifolia</i> Benth. Var. " <i>flora violacea</i> "	Cap Vert	n.d	N 14 55,485	WO 24 21,793
<i>Lavandula stoechas</i> L. ssp. <i>LUSIERI</i>	Portugal	187 m	N 37 29,912	WO 08 16,330
<i>Lavandula tenuisecta</i> Coss. ex Ball.	Maroc	1777 m	N 31 34,521	WO 05 53,418
<i>Lavandula viridis</i> L'Hér.	Portugal	440 m	N 37 19,279	WO 08 32,062
<i>Lavandula x intermedia</i> Emeric ex Loisel.	culture	n.d	N 45 50,669	E 5 58,399

Annexes

Annexes

Annexe VIII : Annexes numériques

MacroGC.xls : fichier Excel pour la compilation et le prétraitement de données de GC-FID ou GC-MS issues du logiciel Chemstation d'Agilent. Pour éditer le code le mot de passe est : admin.

RTAlign.pl : programme en Perl destiné à aligner les pics de différents chromatogrammes (format Agilent) selon leur temps de rétention. Ouverture possible avec un éditeur de texte.

MSeasy et MSeasyTkGUI : packages R pour le prétraitement de données de GC-MS (format Agilent et ASCII)

Alignement_trnk_Lavandula.FAS : fichier au format fasta contenant l'alignement des séquences trnK de 30 taxons de lavande, de la mélisse et du basilic. Ouverture possible avec BioEdit ou des logiciels permettant la lecture du format fasta.

MatriceCOV.xls : Matrice des COV détectés en GC-MS dans les échantillons de lavande du chapitre 2 au format Excel.

Ces documents sont consultables à l'adresse internet suivante : <https://dossier.univ-st-etienne.fr/lbvpam/public/AnnexesNumeriques/>

En cas de difficultés vous pouvez contacter le laboratoire BVPAM par email lbvpam@univ-st-etienne.fr