



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Université de Lorraine

Ecole Doctorale "Sciences et Ingénierie des Ressources Naturelles"

Unité Mixte de Recherche 1121 Université de Lorraine-INRA

Laboratoire Agronomie et Environnement

## Thèse

Présentée en vue de l'obtention du titre de

**Docteur de l'Université de Lorraine en Sciences  
Agronomiques**

par

**Gianni GALATI**

**Etude de la synthèse des furocoumarines chez le  
panais par des approches d'ingénierie métabolique  
et de multi-omique**

Thèse soutenue publiquement le 17 Juillet 2019 devant un jury composé de :

Nathalie JULLIAN-PAWLICKI	(MCF, Université de Picardie)	Rapporteure
Dimitri HEINTZ	(IR, CNRS, Strasbourg)	Rapporteur
Sylvie BAUDINO	(PR, Université de St-Etienne)	Examinatrice
Claire VENEULT-FOURREY	(DR, INRA, Nancy)	Examinatrice
Romain LARBAT	(CR, INRA, Nancy)	Co-directeur de thèse
Alain HEHN	(PR, Université de Lorraine)	Directeur de thèse
Frédéric BOURGAUD	(PR en disponibilité, Université de Lorraine)	Invité

*À Régine*





## Remerciements

Tout d'abord, je remercie Madame Nathalie JULLIAN-PAWLICKI, maitre de conférences à l'université de Picardie, et M. Dimitri HEINTZ, ingénieur de recherche au CNRS de Strasbourg, pour le temps consacré à évaluer mon travail de thèse en tant que rapporteurs. Je souhaite également remercier Madame Sylvie BAUDINO, professeure à l'université de St-Etienne et Madame Claire VENEULT-FOURREY, directrice de recherche à l'INRA de Nancy, d'avoir accepté d'être les examinatrices de cette thèse.

Je tiens ensuite à remercier Christophe ROBIN pour son implication dans l'encadrement et la vie des doctorants, pour ses conseils et sa bienveillance. J'aimerais également remercier le professeur Danièle WERCK-REICHHART et le professeur Nicolas ROUHIER d'avoir accepté de participer à mon comité de thèse, et pour leurs avis éclairés.

J'aimerais ensuite exprimer toute ma gratitude à mes encadrants, le professeur Alain HEHN, Romain LARBAT, chargé de recherche à l'INRA de Nancy et le professeur Frédéric BOURGAUD, pour m'avoir fait confiance pour ce projet de thèse. Merci à Alain pour ses conseils avisés, sa pédagogie, sa bonne humeur, et pour les milliers d'heures passées à lire et corriger ce manuscrit. Merci de m'avoir constamment poussé à me remettre en question, à prendre du recul sur mon travail de thèse mais également sur tous les autres aspects de la vie de chercheur. Je remercie également Romain pour sa gentillesse, sa disponibilité, ses très bonnes idées et pour nos discussions qui très souvent m'ont remonté le moral ! Un grand merci à Frédéric également pour son implication dans mon travail de thèse, et pour avoir continué à suivre mes avancées après son départ du LAE.

J'aimerais remercier maintenant les différentes personnes que j'ai pu rencontrer dans le cadre de collaborations, notamment le Dr Sylvain LEGAY, qui m'a très gentiment accueilli au sein du LIST, grâce à qui j'ai pu me former aux méthodes de séquençage et qui nous a permis de construire une banque transcriptomique qui m'a beaucoup aidé durant cette thèse. Merci également au professeur Yves JOLIVET et à Anthony GANDIN pour leur aide et pour avoir contribué à l'élaboration d'un article de valorisation de mes travaux.

J'en arrive maintenant à remercier toutes les personnes que j'ai côtoyé au sein du LAE, et qui m'ont aidé, d'une façon ou d'une autre, à avancer dans mon projet de recherche. Tout d'abord, merci à Clément pour son efficacité et son aide précieuse tout au long de ma thèse, pour m'avoir renvoyé 100 fois les mêmes protocoles parce que je les avais perdus, pour avoir broyé et pesé des tonnes de panais et pour tous nos petits rendez-vous matinaux (voire très matinaux). A très vite pour le barbeuc ! Merci à Alex de m'avoir recueilli dans son bureau et de m'avoir très souvent aidé dans mes manip. Merci également à Aude, Julie, Thamara pour leur bonne humeur et leurs aides précieuses, merci à Jeremy pour les milliers d'échantillons analysés et pour les parties de basket qui laissent encore des séquelles chez certains ... Je remercie aussi Sissi, Cindy et Carole pour leur aide et leurs gentillesse. Même s'ils ne sont plus au labo, merci à David et Colin (et Justine et Fofette aussi ^^) pour m'avoir « obligé » à trainer à la plantation et m'empêcher de travailler en organisant des tournoi de Doodle clandestins. Merci pour tous les moments que l'on a passés ensemble, vivement les prochains ! Merci à Jeremie pour avoir été la deuxième moitié de la team Pegasus, même s'il m'a abandonné depuis... Merci à Ryosuke et Yuka de m'avoir permis d'aller jusqu'au Japon sans bouger de Nancy, pour leur aide et leurs sourires. Merci à Andreïna pour m'avoir aidé à prendre sa place d'ATER, je te souhaite le meilleur pour la suite ! Merci à Ludivine d'avoir augmenté le ratio des Nordistes au labo et pour nos (trop ?) longues pauses café.

J'aimerais remercier maintenant Caroline et Mariana pour leur amitié dès le début de ma thèse et qui j'espère se poursuivra bien après (même si GoT est fini...). Caroline merci pour ton aide, les vins chauds et les recettes foireuses de gâteaux aux chocolats. Mariana, merci de m'avoir supporté pendant presque quatre ans de colocation de bureau, pour ton aide, tes conseils et ton écoute (de toute façon je suis sur ton certificat de mariage ^^).

J'aimerais remercier mes amis, Adrien, Baptiste, Mattias, Alex, Mickael et Valentin (j'attends toujours la correction de mon Mat et Met ...) pour tout ce qu'il y a eu et tout ce qu'il y aura. Pour finir, merci à ma famille, ma maman, Corentin (et Rosalie !) et Vassili, pour m'avoir soutenu et aidé durant ces 4 dernières années.

Encore une fois, merci à tous !





# Sommaire

Remerciements .....	1
Sommaire .....	5
Liste des figures .....	13
Liste des tableaux.....	15
Liste des annexes.....	16
Liste des abréviations .....	17
Introduction Générale .....	20
I. Synthèse bibliographique et objectifs de thèse .....	24
1. Le métabolisme spécialisé.....	26
1.1. Introduction .....	26
1.2. Description des métabolites spécialisés.....	26
2. Evolution du métabolisme secondaire .....	30
2.1. Diversification du matériel génétique .....	31
2.1.1. Duplication tandem.....	32
2.1.2. Rétroduplication.....	33
2.1.3. Acquisitions de nouvelles fonctions .....	34
2.1.3.1. Néofonctionnalisation.....	34
2.1.3.2. Sub-fonctionnalisation.....	35
2.2. Organisation en Cluster .....	37
2.2.1. Description des clusters .....	37
2.2.2. Intérêt de l'agencement en clusters.....	41
2.2.2.1. Hérité .....	41
2.2.2.2. Co-régulation .....	41
2.2.2.3. Métabolons et Channeling .....	43
3. Les cytochromes P450 .....	44
3.1. Nomenclature des cytochromes P450 .....	45
3.2. Classification en fonction du pouvoir réducteur.....	45
3.3. Structures conservées .....	46
3.4. Rôles biologiques.....	49
3.4.1. Synthèse de métabolites spécialisés.....	49
3.4.2. Détoxification des xénobiotiques .....	50
3.4.3. Adaptation des P450 et coévolution .....	50
4. Les Furocoumarines .....	52

4.1. Structures chimiques.....	52
4.2. Répartition botanique et localisation des furocoumarines dans la plante.....	53
4.2.1. Les Apiacées.....	54
4.2.2. <i>Pastinaca sativa</i> .....	55
4.3. Rôles biologiques des furocoumarines .....	56
4.4. Voie de biosynthèse des furocoumarines .....	57
5. Objectifs de thèse .....	62
II. Résultats et Discussion.....	63
A. Etude des modifications métaboliques en condition de stress abiotique chez le panais.....	64
1. Introduction.....	64
2. Etude de la réponse métabolique du panais à deux stress environnementaux ....	66
2.1. Analyses préliminaires .....	68
2.2.....	68
2.1.1. Détermination de la dose d'ozone sub-létale chez le panais .....	68
2.1.2. Evaluation du statut oxydatif.....	68
2.2. Evolution du profil métabolique global.....	69
2.3. Identification <i>in silico</i> des SMDAs .....	73
3. Discussion et Conclusion.....	76
B. Analyse fine du profil métabolique : cas des furocoumarines.....	80
1. Comparaison inter-organes en condition contrôle .....	80
1.1. Analyse phytochimique des folioles, des tiges et des racines. ....	80
1.2. Analyse du niveau d'expression de gènes impliqués dans la voie de biosynthèse des furocoumarines en conditions contrôle.....	81
2. Etude de l'induction de la voie de biosynthèse en fonction des organes .....	83
2.1. Expression différentielle des gènes spécifiques de la voie de biosynthèse dans différents tissus en réponse à des blessures .....	83
2.2. Analyse différentielle de la production de furocoumarines dans les racines ..	85
3. Discussion et conclusion .....	86
C. Ingénierie métabolique : transposition partielle de la voie de biosynthèse des furocoumarines chez le tabac.....	90
1. Introduction.....	90
2. Construction des vecteurs GoldenBraid .....	92
2.1. La domestication .....	92
2.2. Assemblage des constructions GoldenBraid.....	94
2.2.1. Clonage des séquences d'intérêt dans pUPD .....	94

2.2.2. Construction des unités transcriptionnelles .....	96
2.2.3. Assemblage des unités transcriptionnelles.....	98
3. Expression transitoire des unités transcriptionnelles .....	99
3.1. Test préliminaire.....	100
3.2. Expression transitoire de PsDiox et de PsPT1 dans des feuilles de <i>N. benthamiana</i> .....	102
3.2.1. Analyse de l'expression des gènes .....	103
3.2.2. Analyses phytochimiques .....	104
4. Construction de tabacs mutants, producteurs de furocoumarines : résultats préliminaires .....	106
4.1. Introduction .....	106
4.2. Transformation stable de <i>Nicotiana tabacum</i> .....	107
4.2.1. Obtention de plantes transgéniques .....	107
4.2.2. Analyse phénotypique .....	107
5. Discussion et conclusion .....	108
D. Identification de nouveaux gènes et caractérisation fonctionnelle .....	115
1. Introduction.....	115
2. Recherche de gènes candidats .....	115
2.1. Approche transcriptomique .....	115
2.1.1. Analyse de banque RNAseq (séquençage 454) de panais .....	116
2.1.2. Mesures quantitatives de l'expression des séquences candidates.....	119
2.1.2.1. Mise au point méthodologique : évaluation de l'efficacité des amorces de PCR quantitative .....	119
2.1.2.2. Analyse du profil d'expression des séquences candidates .....	121
2.1.2.3. Affiliation des candidats à des familles de P450 : vers une prédiction fonctionnelle ?.....	122
2.1.3. Recherche de séquences codantes complètes .....	125
2.1.3.1. Construction de la banque, séquençage et assemblage de la banque .....	125
2.1.3.2. Analyse de la banque RNAseq de panais (Illumina) .....	126
2.1.3.2.1. Analyse globale .....	126
2.1.3.2.2. Détermination de la séquence complète des contigs candidats	127
2.2. Approche génomique .....	129
2.2.1. Contexte et identification de <i>CYP71AT</i> .....	129
2.2.2. Analyse du profil d'expression .....	130
3. Clonage et caractérisation fonctionnelle des gènes candidats .....	131

3.1. Clonage dans le plasmide pYeDP60_GW <sup>®</sup> .....	132
3.1.1. Clonage de CYP71-C10. ....	132
3.1.2. Clonage de <i>CYP71AT</i> .....	132
3.2. Expression des protéines dans la levure.....	133
3.2.1. Analyse de l'expression .....	133
3.2.2. Criblage métabolique : Etude de l'activité métabolique des gènes candidats .....	134
4. Discussion et conclusion .....	136
Conclusion générale et perspectives .....	141
III. Matériel et Méthodes.....	147
1. Matériel.....	148
1.1. Matériel végétal.....	148
1.1.1. <i>Pastinaca sativa</i> .....	148
1.1.2. <i>Nicotiana benthamiana</i> .....	148
1.1.3. <i>Nicotiana tabacum</i> var. Samsung.....	148
1.2. Souches de bactéries.....	148
1.2.1. <i>Escherichia coli</i> MC1022 .....	148
1.2.2. <i>Agrobacterium tumefaciens</i> EHA105.....	148
1.3. Souches de levures <i>Saccharomyces cerevisiae</i> WAT11 .....	149
1.4. Vecteurs et plasmides .....	149
1.4.1. pCR <sup>™</sup> 8/GW/TOPO <sup>™</sup> .....	149
1.4.2. pYeDP60_GW <sup>®</sup> .....	150
1.4.3. pEAQ-HT-DEST1 .....	151
1.4.4. pSoup .....	152
1.4.5. pUPD .....	153
1.4.5.1. pUPD-35s .....	154
1.4.5.2. pUPD-tNos .....	154
1.4.6. pDBG1_α.....	154
1.4.7. pDGB1_α1:35s- <i>GFP</i> -tNos.....	155
1.4.8. pDGB1_Ω.....	156
1.5. Conditions de culture .....	156
1.5.1. Milieux de culture de bactéries .....	156
1.5.1.1. Milieu LB .....	156
1.5.1.2. Milieu YEB .....	156
1.5.1.3. Antibiotiques utilisés .....	157



1.5.2. Milieux de culture de levures .....	158
1.5.2.1. Milieu YPGA.....	158
1.5.2.2. Milieu SGI .....	158
1.5.2.3. Milieu YPGE.....	158
1.5.2.4. Milieu YPL.....	158
1.5.3. Culture de plantes en terre .....	159
1.5.4. Culture de plante <i>in vitro</i> .....	160
1.5.4.1. Stérilisation des graines de tabac .....	160
1.5.4.2. Semis des graines et culture.....	160
1.5.4.3. Conditions de culture des plantes <i>in vitro</i> .....	160
1.6. Induction de la synthèse des furocoumarines chez le panais .....	161
1.6.1. Elicitation mécanique.....	161
1.6.1.1. À partir de folioles, de tiges et de racines séparées.....	161
1.6.1.2. À partir de folioles de la plante entière .....	162
1.6.2. Elicitation à l'ozone .....	162
1.6.2.1. Détermination de la concentration d'ozone sub-létale .....	162
1.6.2.2. Etude de l'évolution de la concentration en furocoumarines chez le panais en fonction du temps .....	163
1.7. Outils de bio-informatique .....	163
1.7.1. Logiciels en ligne .....	163
1.7.1.1. MBCF Oligo Calculator .....	163
1.7.1.2. Neb Cutter V2.0 .....	163
1.7.1.3. Multalin.....	164
1.7.1.4. XCMS.....	164
1.7.1.5. MetaboAnalyst .....	164
1.7.2. Logiciels à télécharger.....	164
1.7.2.1. BioEdit.....	164
1.7.2.2. Gene Studio Professional Edition .....	164
1.7.2.3. PrimerExpress .....	165
1.7.2.4. R .....	165
1.7.2.5. StepOne plus v2.3 .....	165
1.7.2.6. LabSolution .....	165
1.7.2.7. Xcalibur.....	165
1.7.2.8. MS-Dial .....	165
1.7.2.9. MsFinder .....	166

1.7.2.10. CLC Genomics Workbench v6.5 .....	166
1.7.2.11. Blast2Go Pro.....	166
2. Méthodes utilisées .....	166
2.1. Outils de biologie moléculaire .....	166
2.1.1. Extraction d'ADN génomique de plante .....	166
2.1.2. Extraction d'ARN totaux de plante .....	167
2.1.3. Synthèse d'ADN complémentaire .....	167
2.1.4. Amplification d'un fragment d'ADN .....	167
2.1.4.1. Par PCR en point final.....	167
2.1.4.1.1. EmeraldAmp® PCR Master Mix 2X.....	168
2.1.4.1.2. SapphireAmp® Fast PCR Master Mix 2X.....	168
2.1.4.1.3. PrimeSTAR® Max DNA Polymerase 2X .....	169
2.1.4.2. Par PCR quantitative .....	169
2.1.6. Electrophorèse sur gel d'agarose .....	170
2.1.7. Extraction d'ADN à partir d'un gel d'agarose .....	171
2.1.8. Méthodes de clonage .....	171
2.1.8.1. Clonage d'un fragment de PCR dans le plasmide pCR™8.....	171
2.1.8.2. Recombinaison dans le plasmide pYeDP60_GW® et pEAQ-HT-DEST1 .....	171
2.1.8.3. Par la méthode GoldenBraid.....	172
2.1.8.3.1. Domestication.....	172
2.1.8.3.2. Ligation dans le plasmide pUPD .....	173
2.1.8.3.3. Construction des unités transcriptionnelles simples .....	174
2.1.8.3.4. Construction des unités transcriptionnelles multiples .....	175
2.1.9. Préparation de cellules compétentes.....	175
2.1.9.1. Bactéries électrocompétentes .....	175
2.1.9.1.1. Escherichia coli MC1022 .....	176
2.1.9.1.2. <i>Agrobacterium tumefaciens</i> .....	176
2.1.9.2. Levures chimiocompétentes .....	177
2.1.10. Transformation de cellules compétentes .....	177
2.1.10.1. Bactéries électrocompétentes.....	177
2.1.10.2. Levures chimiocompétentes .....	178
2.1.11. Extraction d'ADN plasmidiques .....	178
2.1.12. Quantification par spectrophotométrie.....	179
2.1.13. Western Blot.....	179

2.1.13.1. Electrophorèse sur gel de polyacrylamide en condition dénaturante .....	179
2.1.13.2. Transfert des protéines séparées sur une membrane PVDF .....	180
2.1.13.3. Immunodétection .....	181
2.1.14. Construction d'une banque transcriptomique de panais par séquençage Illumina .....	182
2.1.14.1. Préparation de la librairie et séquençage .....	182
2.1.14.2. Assemblage et annotation de la banque .....	182
2.2. Production de protéines en systèmes hétérologues.....	183
2.2.1. Dans <i>Saccharomyces cerevisiae</i> WAT11.....	183
2.2.2. Dans <i>Nicotiana benthamiana</i> par expression transitoire .....	185
2.2.2.1. Préparation de la suspension bactérienne à inoculer .....	185
2.2.2.2. Agro infiltration des feuilles de <i>Nicotiana benthamiana</i> .....	185
2.2.3. Dans <i>Nicotiana tabacum</i> var. Samsung par expression stable .....	186
2.2.3.1. Préparation de la suspension bactérienne .....	186
2.2.3.2. Infection d'explants foliaires .....	187
2.2.3.3. Sélection et régénération des plantes transformées .....	187
2.3. Etude de l'activité des protéines recombinantes .....	188
2.3.1. À partir de microsomes de levures .....	188
2.3.1.1. Criblage métabolique .....	188
2.3.1.2. Extraction des composés phénoliques à l'acétate d'éthyle .....	188
2.3.2. À partir de <i>Nicotiana benthamiana</i> .....	189
2.3.2.1. Infiltration des substrats .....	189
2.3.2.2. Traitement des feuilles transformées .....	189
2.4. Analyses métaboliques .....	190
2.4.1. Extraction des composés à analyser .....	190
2.4.2. Analyse du contenu métabolique.....	190
2.4.2.1. Analyse ciblée des polyphénols .....	190
2.4.2.1.1. UPLC-MS .....	190
2.4.2.1.2. Orbitrap-LTQ .....	192
2.4.2.1.3. Identification et quantification relative des composés phénoliques ciblés .....	193
2.4.2.2. Analyse métabolomique non ciblée .....	193
Références Bibliographiques.....	196
Annexes.....	219



## Liste des figures

FIGURE 1 : CARACTERISTIQUES DES TROIS SOUS-FAMILLES D'ALCALOÏDES.....	27
FIGURE 2 : STRUCTURE CHIMIQUE DE QUELQUES METABOLITES SPECIALISES ACCUMULES CHEZ LES PLANTES	28
FIGURE 3 : EXEMPLES DE COULEURS DE PETALES INDUITES PAR LA PRESENCE D'ANTHOCYANES.....	29
FIGURE 4 : REPRESENTATION SCHEMATIQUE DE LA SYNTHESE DE METABOLITES SPECIALISES ORGANISEE EN VOIE METABOLIQUE .....	31
FIGURE 5: REPRESENTATION SIMPLIFIEE D'UN ENJAMBEMENT ET D'UN ENJAMBEMENT INEGAL ENTRE DEUX BRAS DE CHROMOSOMES. ....	32
FIGURE 6: REPRESENTATION SCHEMATIQUE DU MECANISME DE RETRODUPLICATION. ....	33
FIGURE 7: REPRESENTATION SCHEMATIQUE DU MECANISME DE NEO-FONCTIONNALISATION.....	35
FIGURE 8 : REPRESENTATION SCHEMATIQUE DU MECANISME DE LA SUB-FONCTIONNALISATION.....	36
FIGURE 9: ORGANISATION GENOMIQUE DE L'OPERON LACTOSE CHEZ ESCHERICHIA COLI. ....	37
FIGURE 10: ORGANISATION GENOMIQUE DE CERTAINS CLUSTERS DE CYTOCHROME P450 IMPLIQUES DANS LA SYNTHESE DE METABOLITES SECONDAIRES.....	40
FIGURE 11: REPRESENTATION SCHEMATIQUE DE LA STRUCTURE D'UN CYTOCHROME P450 DE CLASSE II. ....	46
FIGURE 12: REPRESENTATION SCHEMATIQUE DE LA STRUCTURE PRIMAIRE D'UN P450 MEMBRANAIRE EUCARYOTE. ....	47
FIGURE 13: ALIGNEMENT DE LA SEQUENCE PROTEIQUES DE 19 CYTOCHROMES P450 APPARTENANT A LA SOUS- FAMILLE CYP71AJ.....	48
FIGURE 14: REPRESENTATION SCHEMATIQUE DE LA STRUCTURE DE BASE DES FUROCOUMARINES LINEAIRES (GAUCHE : PSORALENE) ET ANGULAIRES ( DROITE : ANGELICINE).....	53
FIGURE 15: EXEMPLES D'APIACEES. GAUCHE : FLEUR DE CAROTTE (DAUCUS CAROTA), MILIEU : FLEUR DE PANAIS, DROITE : RACINE DE PANAI.....	55
FIGURE 16 : REPRESENTATION DE LA VOIE DE BIOSYNTHESE DES FUROCOUMARINES CHEZ LE PANAI.....	60
FIGURE 18: MECANISME GLOBAL SIMPLIFIE DE LA REPOSE DE LA PLANTE A UN STIMULUS EXTERIEUR.....	65
FIGURE 19 : APPARITION DE CHLOROSIS (FIGURE (B) A (E) APRES APPLICATION D'OZONE SUR DES FEUILLES DE CORNILLE (VIGNA UNGUICULATA, FABACEES). ....	67
FIGURE 20: MESURE DE LA CONCENTRATION EN H <sub>2</sub> O <sub>2</sub> DANS LES FEUILLES DE PANAI.....	69
FIGURE 21: ANALYSE EN COMPOSANTE PRINCIPALE A PARTIR DE DONNEES METABOLIQUES NON CIBLEES ISSUES DU TRAITEMENT DE FOLIOLES DE PANAI DANS DIFFERENTES CONDITIONS.....	70
FIGURE 22 : REPRESENTATION EN HEATMAP DE 646 SMDAS. ....	71
FIGURE 23: MESURE DE LA CONCENTRATION EN FUROCOUMARINES DANS LES FOLIOLES (GRIS CLAIR), TIGES (GRIS) ET RACINES (GRIS FONCE) DE PANAI EN CONDITION CONTROLE.....	81
FIGURE 24: ANALYSE DU NIVEAU D'EXPRESSION DES GENES PSDIOX, PSPT1, PSPT2, CYP71AJ3 ET CYP71AJ4 DANS LES FOLIOLES, TIGES ET RACINES DE PANAI EN CONDITION CONTROLE PAR PCR QUANTITATIVE. ....	82
FIGURE 25: ANALYSE EN PCR QUANTITATIVE DE L'EXPRESSION DE DES GENES PSDIOX, PSPT1, PSPT2, CYP71AJ3 ET CYP71AJ4 DANS LES FOLIOLES, TIGES ET RACINES DE PANAI EN CONDITION CONTROLE (RESPECTIVEMENT FNI, TNI ET RNI, GRIS) ET APRES L'APPLICATION DE BLESSURE MECANIQUE (RESPECTIVEMENT FI, TI ET RI, NOIR). ....	84
FIGURE 26: MESURES DE LA CONCENTRATION EN FUROCOUMARINES DANS DES RACINES DE PANAI EN CONDITION CONTROLE (GRIS) ET APRES L'APPLICATION DE BLESSURES MECANIQUES (NOIR).....	85
FIGURE 27: PRODUCTION DE DEMETHYLSUBEROSINE A PARTIR DE P-COUMAROYL-COA PAR L'INTERVENTION DE PSDIOX ET PSPT1. ....	90
FIGURE 28: REPRESENTATION SCHEMATIQUE DE LA CONSTRUCTION D'UNE DOUBLE UNITE TRANSCRIPTIONNELLE PAR LA METHODE GOLDENBRAID .....	91
FIGURE 29: DOMESTICATION DES SEQUENCES CODANTES DE PSDIOX ET PSPT1. ....	93
FIGURE 30: CLONAGE DE PSDIOX ET PSPT1 DANS LE PLASMIDE PUPD. ....	95
FIGURE 31: DIGESTION DES CONSTRUCTIONS PUPD : PSDIOX ET PUPD : PSPT1 PAR L'ENZYME BSAI.....	96
FIGURE 32: ASSEMBLAGE DU PROMOTEUR 35S, DE LA SEQUENCE CODANTE PSDIOX ET DU TERMINATEUR TNOS EN UNITE TRANSCRIPTIONNELLE DANS LE PLASMIDE PDBG1A. ....	97

FIGURE 33: INSERTION DES UNITES TRANSCRIPTIONNELLES 35S-PSDIOX-TNOS ET 35S-PSPT1-TNOS DANS LE PLASMIDE PDGB1Ω.....	98
FIGURE 34: INFILTRATION DE SUSPENSIONS D'AGROBACTERIES DANS DES FEUILLES DE NICOTIANA BENTHAMIANA.....	100
FIGURE 35: OBSERVATION DE LA FLUORESCENCE DE LA GFP DANS DES FEUILLES DE NICOTIANA BENTHAMIANA A LA LOUPE BINOCULAIRE (X100).....	101
FIGURE 36: RT-PCR REALISEES SUR DE L'ARN EXTRAITS DE FEUILLES DE TABAC APRES 4 JOURS D'INCUBATION.....	104
FIGURE 37: ANALYSE DES PRODUITS DE METABOLISATION DANS LE TABAC DES DIFFERENTES UNITES TRANSCRIPTIONNELLES GOLDENBRAID PAR UPLC/MS.....	105
FIGURE 38: PLANTULES DE NICOTIANA TABACUM REGENEREES A PARTIR D'EXPLANTS TRANSFORMES PAR LA CONSTRUCTION GOLDENBRAID [35S-PSDIOX-TNOS : 35S-PSPT1-TNOS].....	108
FIGURE 39: SCHEMA REPRESENTANT LES DIFFERENTS CAS DE FIGURE DANS LA DEMARCHE D'ASSEMBLAGE DE SEQUENCES RNASEQ.....	118
FIGURE 40: QUANTIFICATION RELATIVE DES DIFFERENTES SEQUENCES CANDIDATES DANS LES RACINES DE PANAI ELICITEES MECANIQUEMENT (GRIS FONCE) OU NON ELICITEES (GRIS CLAIR).....	121
FIGURE 41: COMPARAISON DES MECANISMES REACTIONNELS DE LA MENTHOFURANE SYNTHASE (HAUT) ET DE LA MARMESINE SYNTHASE (BAS).....	124
FIGURE 42 : ALIGNEMENT DE LA SEQUENCE DU CONTIG C10 AVEC LA SEQUENCE CORRESPONDANTE ISSUE DE LA BANQUE TRANSCRIPTOMIQUE DE PANAI GENEREES AVEC LA METHODE DE SEQUENÇAGE ILLUMINA.....	128
FIGURE 43: LOCALISATION DES GENES CYP71AJ3, CYP71AJ4, PSDIOX , PSPT1 ET CYP71AT SUR LE GENOME DU PANAI.....	130
FIGURE 44: PROFIL D'EXPRESSION DE CYP71AT EN PCR QUANTITATIVE A PARTIR D'ARN EXTRAITS DE RACINE DE PANAI CONTROLE (RT) ET ELICITEES MECANIQUEMENT (RE).....	131
FIGURE 45 : RECONSTRUCTION DE LA SEQUENCE DU GENE CYP71AT A PARTIR D'ADN GENOMIQUE DE PANAI.....	132
FIGURE 46: IMMUNODETECTION DES PROTEINES CYP71AT ET MFS (C10).....	134
FIGURE 47:ANALYSE DES PRODUITS DE METABOLISATION DE CYP71AT PAR UPLC.....	135
FIGURE 48: REPRESENTATION SCHEMATIQUE DE LA PRODUCTION PROTEINES RECOMBINANTES DANS LA LEVURE.....	139
FIGURE 49: REPRESENTATION SCHEMATIQUE DU PLASMIDE PCR™8/GW/TOPO™.....	150
FIGURE 50: REPRESENTATION SCHEMATIQUE DU PLASMIDE PYEDP60_GW.....	151
FIGURE 51: REPRESENTATION SCHEMATIQUE DU PLASMIDE PEAQ-HT-DEST1.....	152
FIGURE 52: REPRESENTATION SCHEMATIQUE DU PLASMIDE PSOUP.....	153
FIGURE 53: REPRESENTATION SCHEMATIQUE DU PLASMIDE PUPD.....	154
FIGURE 54: REPRESENTATION SCHEMATIQUE DU PLASMIDE PDGB1_A.....	155
FIGURE 55: REPRESENTATION DU MODE D'ACTION DES ENZYMES DE TYPE IIS.....	172
FIGURE 56: LIGATION D'UNE SOUS-UNITE DE CASSETTE D'EXPRESSION DANS LE PLASMIDE PUPD PAR LA METHODE GOLDENBRAID.....	174
FIGURE 57: ASSEMBLAGE D'UNE UNITE TRANSCRIPTIONNELLE DANS UN PLASMIDE PDGB1A PAR LA METHODE DU GOLDENBRAID.....	175
FIGURE 58: REPRESENTATION SCHEMATIQUE DU DISPOSITIF EXPERIMENTAL DE TRANSFERT DES PROTEINES DU GEL DE SEPARATION VERS LA MEMBRANE PVDF.....	180
FIGURE 59: EXEMPLE D'UN « BUBBLE SIZE » DE 31 NUCLEOTIDES LORS DE L'ALIGNEMENTS DE 2 SEQUENCES NUCLEOTIDIQUES.....	183
FIGURE 60: DISPOSITIF EXPERIMENTAL DE L'INFILTRATION DE SUBSTRATS DANS LES FEUILLES DE TABAC TRANSFORMEES.....	189
FIGURE 61: : GRADIENT DE LA PHASE MOBILE UTILISE POUR LA SEPARATION DES COMPOSES PHENOLIQUES EN UPLC.....	191
FIGURE 62: GRADIENT DE LA PHASE MOBILE UTILISE POUR LA SEPARATION DES COMPOSES PHENOLIQUES EN ORBITRAP.....	<b>ERREUR ! SIGNET NON DEFINI.</b>

## Liste des tableaux

TABLEAU 1 : NOMBRE DE SMDAS EN CONDITION « OZONE » ET « BLESSURE MECANIQUE » PAR RAPPORT AU CONTROLE. ....	73
TABLEAU 2 : RESUME DES METABOLITES PUTATIVEMENT ANNOTES DIFFERENTIELLEMENT ACCUMULES EN FONCTION DES TRAITEMENTS ET DU TEMPS .....	74
TABLEAU 3: RECAPITULATIF DE LA FREQUENCE D'APPARITION DES CODONS CHEZ LE TABAC .....	94
TABLEAU 4: RECAPITULATIFS DES RESULTATS DE L'INCUBATION DES FEUILLES DE TABAC TRANSFORMEES AVEC LES CONSTRUCTIONS GOLDENBRAID EN PRESENCE DU SUBSTRAT ASSOCIES. ....	106
TABLEAU 5: RECAPITULATIF DES DIFFERENTES ETAPES DE TRI DE LA BANQUE RNASEQ DE PANAI GENERE AVEC LA METHODE DE SEQUENÇAGE 454.....	117
TABLEAU 6 : RECAPITULATIFS DES RESULTATS D'AMPLIFICATION PAR PCR DES DIFFERENTS CONTIGS CANDIDATS A PARTIR D'ADN GENOMIQUE DE PANAI .....	119
TABLEAU 7: RESULTATS DE L'ANALYSE DE L'EFFICACITE DES COUPLES D'AMORCES SPECIFIQUES DES SEQUENCES CANDIDATES PAR PCR QUANTITATIVE A PARTIR DE 100NG D'ARN DE PANAI. ....	120
TABLEAU 8: RESULTATS DE L'ANALYSE DES SEQUENCES NUCLEOTIDIQUES DES CONTIGS CANDIDATS PAR L'ALGORITHME BLAST. ....	123
TABLEAU 9 : SYNTHESE DES ETAPES DE SELECTION DES SEQUENCES CANDIDATES A PARTIR DES SEQUENCES VALIDEES PAR PCR.....	129
TABLEAU 10: COMPOSITION DU MILIEU YEB.....	157
TABLEAU 11: LISTE DES ANTIBIOTIQUES UTILISES POUR LA CULTURE DE BACTERIES.....	157
TABLEAU 12 :COMPOSITION DE 1L DES DIFFERENTS MILIEUX UTILISES POUR LA CULTURE DE LEVURE.....	159
TABLEAU 13: COMPOSITION DU MILIEU DE CULTURE DE NICOTIANA TABACUM IN VITRO .....	161
TABLEAU 14: COMPOSITION DES DIFFERENTS TAMPONS UTILISES POUR LA GENERATION DE MICROSOMES DE LEVURES.....	184
TABLEAU 15 : COMPOSITION DES DIFFERENTS MILIEU UTILISES POUR LA TRANSFORMATION STABLE DE TABAC PAR <i>AGROBACTERIUM TUMEFACIENS</i> .....	187

## Liste des annexes

ANNEXE 1 : GALATI <i>ET AL</i> , 2019 (EN COURS DE REVISION).....	221
ANNEXE 2 : ROSELLI <i>ET AL</i> , 2017 .....	235
ANNEXE 3 : KRIEGER <i>ET AL</i> , 2018 .....	249
ANNEXE 4 : SEQUENCES CODANTES DES DIFFERENTS GENES CONNUS DE LA VOIE DE BIOSYNTHESE DES FUROCOUMARINES CHEZ LE PANAI.....	264
ANNEXE 5: LISTE DES AMORCES UTILISES EN RTQPCR .....	267
ANNEXE 6: ALIGNEMENTS DES SEQUENCES DU GENE <i>PSDIOX</i> (A) ET <i>PSPT1</i> (B) NATIVES DU PANAI AVEC LES SEQUENCES RECODEES POUR LE CLONAGE PAR GOLDENBRAID .....	268
ANNEXE 7 : SEQUENCES DES 22 CONTIGS ISSUS DU TRI DE LA BANQUE RNASEQ (454) AMPLIFIES PAR PCR A PARTIR D'ADN GENOMIQUE DE PANAI .....	269
ANNEXE 8: LISTE DES AMORCES UTILISES EN PCR .....	276
ANNEXE 9: SEQUENCES CODANTES DES 5 CANDIDATS ISSUES DE L'ANALYSE DE LA BANQUE RNASEQ ILLUMINA .....	276
ANNEXE 10: SEQUENCE CODANTE DU GENE <i>CYP71AT</i> .....	279
ANNEXE 11 : LISTE DES SUBSTRATS UTILISES POUR LE CRIBLAGE METABOLIQUE DE CYP71AT, DE C10 ET DE CYP71AZ4.....	280



## Liste des abréviations

4CL:4-coumarate-CoA ligase

ACP: Analyse en composante principale

ADN: Acide désoxyribonucléique

ADNc: ADN complémentaire

ANOVA: Analysis of Variance

ARN : Acide Ribonucléique

BCIP : 5-bromo-4-chloro-3-indole-phosphate

BET: Bromure d'éthidium

BLAST: Basic local alignment search tool

C2'H: *p*-coumaroyl-CoA 2'-hydroxylase

C4H: Cinnamate 4-hydroxylase

cM: centiMorgan

CNRGV : Centre National de Ressources Génomiques Végétales

CoA: Coenzyme A

CYP: Cytochrome P450

DAG: DiacylGlycerol

DIBOA: 2,4-dihydroxy-1,4-benzoxazin-3-one

DIMBOA: 2,4-Dihydroxy-7-methoxy-1,4-benzoxazin-3-one

DMS: Demethylsuberosine

DMSO: Diméthylsulfoxyde

DO: Densité optique

DTT: Dithiothréitol

EDTA : Acide éthylène diamine tétra-acétique

ESI: Electrospray ionization

FMN: Flavine mononucléotide

GB : GoldenBraid

GW : Gateway

HRMS : Spectromètre de masse à haute résolution

INRA : Institut National de la Recherche Agronomique

IP3: Inositol trisphosphate

kDa: KiloDalton

LAE : Laboratoire Agronomie et Environnement

LIST: Luxembourg Institute of Science and Technology

MAPK: Mitogen-activated protein kinase

MS: Mass spectrometry

NADH: Nicotinamide adénine dinucléotide

NADPH: Nicotinamide adénine dinucléotide phosphate

NaPi : Phosphate de sodium

NBT : 4-Nitroblue tétrazolium chloride

Nt : Nucléotide

PAL : Phénylalanine ammonia lyase

Pb : Paires de bases

pb: paires de bases

PCR : Polymerase chain reaction

PPB : Partir par billions

PPO : Pipéronyl butoxyde

PVDF : Polyfluorure de vinylidène

qPCR: PCR quantitative

RNS: Reactive Nitrogen Species

ROS: Reactive Oxygen Species

SDS : Dodécylsulfate de sodium

SMDA: Signal métabolique différenciellement accumulé

SRS: Substrate Recognition Site

T-DNA: ADN de transfert

TAE : Tris acétate EDTA

T<sub>m</sub> : Température de fusion

UPLC-MS: Ultra Performance Liquid Chromatography- Mass Spectrometry

UV: Ultraviolet

# **Introduction Générale**



Le changement climatique global est un phénomène clairement établi depuis maintenant un demi-siècle. De même que les causes de ce phénomène (gaz à effets de serre, déforestation, émission de particules, ...), les conséquences environnementales engendrées sont relativement bien identifiées. Parmi elles, les plus importantes sont l'augmentation globale de la température, de la concentration dans l'air de CO<sub>2</sub> et d'ozone ou encore une intensification des rayonnements UV (Holopainen *et al.*, 2018). Face à ces contraintes, certains organismes ont modifié leurs comportements voire changé d'environnement pour survivre. En effet, plusieurs études ont mis en relation des modifications phénologiques de certains animaux ou insectes en lien avec ces bouleversements climatiques. Par exemple, en Pologne (Tryjanowski *et al.*, 2003) ou aux Etats-Unis (Green, 2017), il a été mis en évidence que la précocité de la période d'émergence de certains amphibiens pouvait être reliée à l'augmentation des températures. Une autre étude réalisée par (Pureswaran *et al.*, 2018) décrit la migration de lépidoptères vers des latitudes moins chaudes et moins humides comme une réponse à ces conditions environnementales nouvelles. Chez les plantes, qui sont des organismes sessiles, l'adaptation aux modifications climatiques en changeant de niche écologique est moins évidente, même si des exemples de migrations de plantes ont été décrits (Fitzpatrick *et al.*, 2008; Pitelka and Group, 1997; Renton *et al.*, 2013). Les modifications climatiques étant rapides à l'échelle de l'évolution, les plantes y répondent principalement en utilisant des armes dont elles sont déjà pourvues. Le métabolisme spécialisé fait partie de ces outils.

Le métabolisme spécialisé, ou secondaire, est le métabolisme de l'interaction des plantes avec l'environnement, et est spécifique à chaque plante. Sa plasticité permet de réagir aux contraintes extérieures (insectes pollinisateurs, allélopathie, ...) et de se prémunir contre diverses conditions de stress (biotiques et abiotiques). Cette adaptation, opportuniste dans un premier temps (détoxification d'herbicides par des enzymes impliquées dans la synthèse de certaines molécules, par exemple) est souvent devenue spécifique au cours de l'évolution (production de molécules toxiques contre certains agresseurs, production de molécules pour se protéger contre les rayonnements UV...). Sa complexité le rend difficile à appréhender lorsqu'on veut le décrypter au niveau moléculaire.

Ce travail a cependant été initié par de nombreuses équipes de chercheurs et est nécessaire pour répondre aux défis environnementaux et sociétaux auxquels nous sommes confrontés. L'acquisition de telles connaissances permettra, par exemple, d'identifier des leviers à mettre en œuvre pour permettre aux plantes de faire face à des stress comme la sécheresse ou l'attaque d'herbivores émergents. Les plantes de nouvelle génération ainsi sélectionnées permettraient de préserver l'environnement en limitant l'usage d'intrants (irrigation, pesticides). Certaines pistes ont déjà été évoquées dans la littérature comme la production d'osmoprotectants (Bartwal *et al.*, 2013; Chan *et al.*, 2013) contre la sécheresse, ou encore la synthèse de composés insecticides, antimicrobiens ou antifongiques (Anzlovar et Koce, 2014; Ashiq *et al.*, 2017; Mendoza et Silva, 2018; Tintino *et al.*, 2015).

Dans ce contexte, mon travail de thèse a pour cadre principal l'étude d'une voie de biosynthèse particulière du métabolisme spécialisé du panais. Pour réaliser ce projet j'ai fait appel à des approches de bioinformatique, de génomique, de transcriptomique, de métabolomique, de physiologie moléculaire et de transgénése que je détaillerai dans ce manuscrit. Après une analyse bibliographique permettant de faire un état des connaissances actuelles, je présenterai et discuterai les différents résultats obtenus. Je conclurai ce manuscrit par une discussion générale qui permettra de projeter l'impact de ces résultats dans un contexte plus général d'acquisition de connaissances fondamentales et de perspectives plus appliquées.

# **I. Synthèse bibliographique et objectifs de thèse**





# **1. Le métabolisme spécialisé**

## **1.1. Introduction**

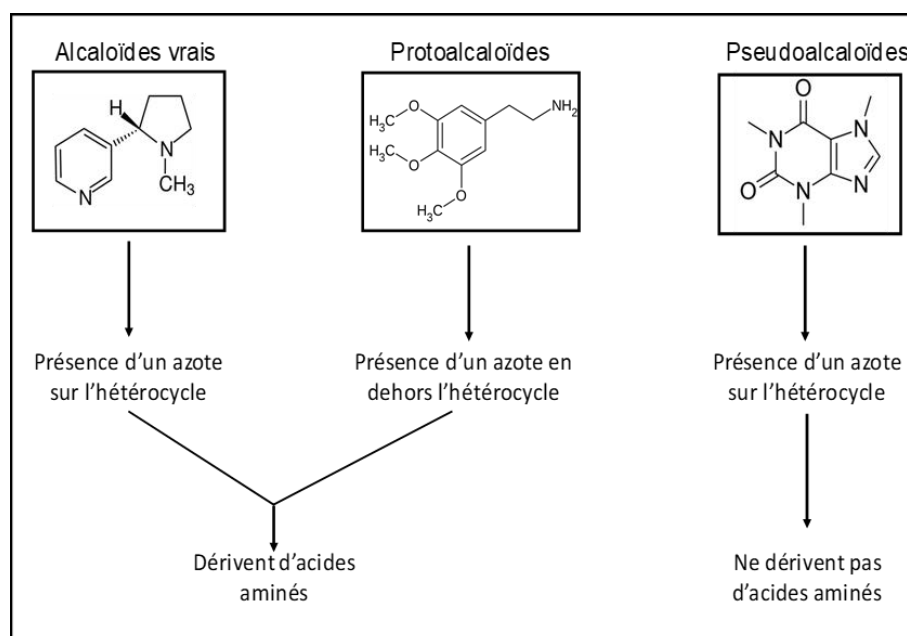
Les plantes sont des organismes sessiles et soumis à de nombreuses contraintes environnementales. Pour s'y adapter, celles-ci ont développé diverses stratégies dont la production d'une grande variété de composés actifs connus sous les termes de métabolites secondaires ou métabolites spécialisés. Depuis la fin du XIX<sup>ème</sup> siècle, ce métabolisme spécialisé a été opposé au métabolisme primaire (notamment par le biologiste allemand Albrecht Kossel (1891)). Le métabolisme primaire est décrit comme impliqué dans la survie et le développement de la plante au travers, par exemple, de la synthèse de sucres, d'acides gras ou de nucléotides (Verpoorte and Alfermann, 2000). De plus, le métabolisme primaire ainsi que les molécules associées, sont très conservés d'un point de vue phylogénétique, en opposition avec les métabolites spécialisés, plus ou moins spécifiques d'une espèce donnée, voire d'une variété donnée (Beauvoit *et al.*, 2018). Les métabolites spécialisés sont très minoritaires au regard du contenu métabolique total de la plante (Bourgau *et al.*, 2001), mais se caractérisent par une remarquable diversité de structures moléculaires et représentent par conséquent un élément essentiel de l'identité chimique de chaque espèce végétale (Al Khayri and Naik, 2016; Murthy *et al.*, 2014).

## **1.2. Description des métabolites spécialisés**

Les métabolites spécialisés se répartissent en trois principales familles : les terpènes, les composés azotés (majoritairement représentés par les alcaloïdes), et les composés phénoliques. Les terpènes constituent une famille de métabolites très diversifiée d'un point de vue structural (Aharoni *et al.*, 2005). Tous les terpènes dérivent d'une structure commune, l'iso-pentenyl-pyrophosphate, et sont classés en fonction de la longueur de leurs chaînes carbonées.

Ces composés disposent par exemple de propriétés antiprolifératives (taxol, Huang *et al.*, 2017), antimalariques (artémisinine, Kumar *et al.*, 2015) ou antibactériennes (Patel, 2015) et sont de fait très largement utilisés en médecine. Leurs propriétés physico-chimiques, et plus particulièrement leurs propriétés olfactives, ont également été à l'origine d'une utilisation massive dans l'industrie agroalimentaire et cosmétique (Pichersky *et al.*, 1994).

Les composés azotés, majoritairement représentés par les alcaloïdes, se caractérisent par la présence d'au moins un atome d'azote dans leur structure. Ces composés sont catégorisés en trois sous-familles en fonction de la nature de leur hétérocycle et de leurs précurseurs : les alcaloïdes vrais, les pseudoalcaloïdes et les protoalcaloïdes (Figure 1).

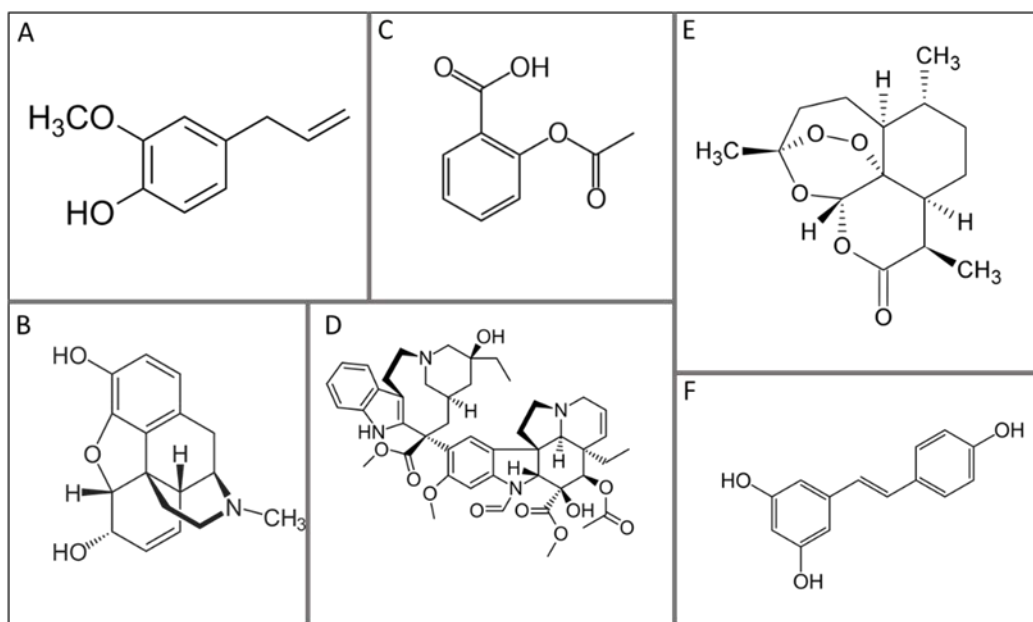


**Figure 1 : Caractéristiques des trois sous-familles d'alcaloïdes**

A l'instar des terpènes, les alcaloïdes présentent des intérêts médicaux, notamment des propriétés analgésiques (ex : morphine) ou anticancéreuses (ex : vincristine, vinblastine) (Bajwa and Haldar, 2015; Geriyol *et al.*, 2015).

Enfin, les composés phénoliques représentent une famille de métabolites dont la structure de base est composée d'au moins un cycle aromatique hydroxylé appelé groupement phénol, dérivant principalement de deux voies de biosynthèse, celle du shikimate et des phénylpropanoïdes (Lattanzio, 2013). A partir de deux groupements phénols, les composés phénoliques sont appelés polyphénols.

Les composés phénoliques sont catégorisés en classe dépendant du nombre de carbones composant le squelette de la molécule, du degré de modification de ce squelette (hydroxylation, méthylation, ...), et de leurs capacités à se lier à d'autres composés (sucres, lipides, protéines, ...). Parmi les composés phénoliques les plus notoires, on retrouve l'acide salicylique, précurseur de l'aspirine (acide acétylsalicylique) (Kratz *et al.*, 2008), le resvératrol (Dytrtova *et al.*, 2018) ou les anthocyanes (Ioannou *et al.*, 2012) (Figure 2).



**Figure 2 : Structure chimique de quelques métabolites spécialisés accumulés chez les plantes**

A : Eugénol, B : Morphine, C : Acide acétylsalicylique, D: Vincristine, E : Artémisinine F : Resvératrol

En 2017, plus de 300 000 métabolites spécialisés étaient identifiés au sein de la biodiversité végétale (Hubert *et al.*, 2017). Grâce aux outils d'analyses de plus en plus performants, le catalogue de molécules identifiées ne cesse d'augmenter. Cependant, le rôle physiologique ou écologique de toutes ces molécules est encore loin d'être élucidé. De manière générale, le métabolisme spécialisé est largement impliqué dans l'interaction des plantes avec leurs environnements. C'est le cas de l'attraction des insectes pollinisateurs pour lequel les anthocyanes et les terpènes jouent un rôle important en conférant aux fleurs un pouvoir attractif visuel ou odorant (Benvenuti *et al.*, 2016, Figure 3).

Un grand nombre de métabolites spécialisés joue également un rôle dans la protection de la plante, contre des contraintes environnementales ou biotiques. On peut ainsi citer les propriétés d'absorption des UV par les flavonoïdes et les acides hydroxycinnamiques qui permettent de protéger l'ADN de la plante de mutations engendrées par ces rayonnements (Mazza *et al.*, 2000). Enfin, de nombreux travaux ont démontré le rôle de ces composés dans la défense de la plante contre des bioagresseurs (insectes, microorganismes, herbivores) soit en tant que répulsifs, avec l'exemple des glucosinolates (Lee *et al.*, 2018) soit par des activités insecticides (ex : nicotine, Yang *et al.*, 2012), antibactériennes (ex :  $\alpha$ -pinène, Mastelic *et al.*, 2005) ou antifongiques (ex : humulène, Abraham, 2001).



**Figure 3 : Exemples de couleurs de pétales induites par la présence d'anthocyanes**  
 (1) *Ageratum houstonianum*, (2) *Antirrhinum majus*, (3) *Begonia semperflorens*, (4) *Borago officinalis*, (5) *Calendula officinalis*, (6) *Dianthus*  $\times$  *barbatus*, (7) *Fuchsia* *hybrid*, (8) *Pelargonium peltatum*, (9) *Petunia*  $\times$  *hybrid*, (10) *Tagetes erecta*, (11) *Tropaeolum majus*, (12) *Viola*  $\times$  *wittrockiana*. D'après Benvenuti *et al.*, 2016.

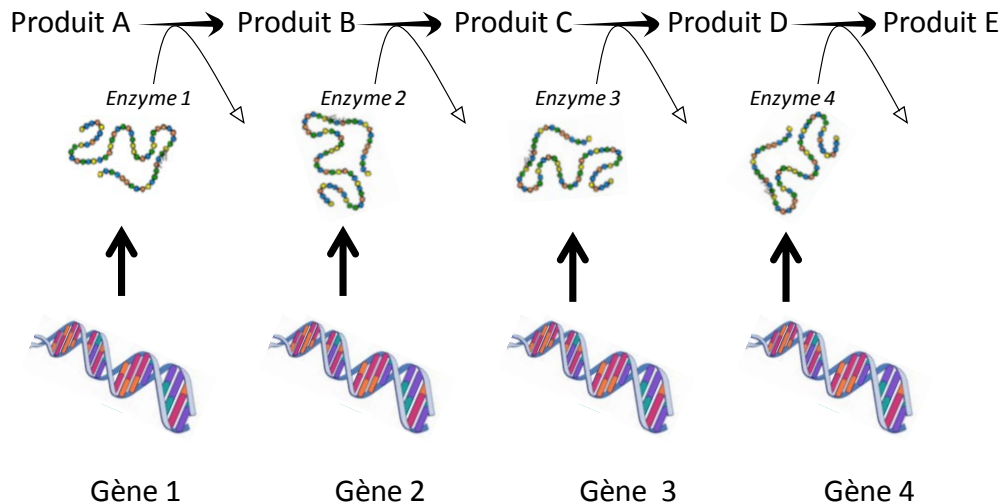
Bien que le métabolisme spécialisé confère à la plante une formidable capacité d'adaptation à son environnement, il met la plante face à un dilemme. En effet la plante prélève les ressources disponibles dans son environnement (azote, carbone, minéraux, ...) et les investit principalement dans sa croissance, son développement, sa reproduction (Karasov *et al.*, 2017).

Lorsque la plante subit un stress, tel qu'un stress biotique, une partie de ces ressources est réallouée à la synthèse de molécules de défense. En effet, les mécanismes moléculaires mis en place par la plante pour répondre à un stress et ceux impliquant la croissance de la plante utilisent le même pool de ressources, voire les mêmes précurseurs (Herms and Mattson, 1992; Margna, 1977). Ainsi, l'utilisation des ressources de la plante, par la plante, est soumise à un principe dichotomique: se développer suffisamment vite pour rivaliser avec les autres plantes et pouvoir se défendre rapidement et efficacement pour survivre aux stress, biotiques ou abiotiques (Herms and Mattson, 1992). Cette dichotomie métabolique est appelée « Balance Croissance/Défense » et gouverne le comportement de la plante par la modulation du métabolisme primaire et spécialisé (Huot *et al.*, 2014).

Le profil métabolique de chaque espèce végétale est unique et dépend notamment du contexte écologique et environnemental. Cette spécificité témoigne d'une formidable capacité d'adaptation des plantes, et s'est mise en place, au cours de l'évolution, par de nombreux processus génétiques, moléculaires et physiologiques.

## **2. Evolution du métabolisme secondaire**

La synthèse des molécules du métabolisme spécialisé implique de nombreuses étapes enzymatiques dont l'enchaînement aboutit à un ensemble de voies de biosynthèse plus ou moins complexes (Figure 4). Ces voies de biosynthèse ont pour origine des molécules du métabolisme primaire (acide organiques, acides aminés, sucres) présentes dans toutes les plantes. Les enzymes catalysant les différentes étapes de ces voies sont apparues au fur et à mesure de l'évolution par le biais de mutations des gènes codant ces enzymes spécifiques.



**Figure 4 : Représentation schématique de la synthèse de métabolites spécialisés organisée en voie métabolique**

L'apparition de mutations dans un génome est un phénomène aléatoire pouvant conduire à plusieurs cas de figure : (i) La mutation n'a aucun effet sur l'expression ou le produit du gène et peut alors être maintenue, on parle alors de mutation silencieuse, (ii) la mutation affecte l'expression ou la séquence d'un gène essentiel à la vie de la plante (mutation faux-sens ou non-sens), elle est alors éliminée au cours de l'évolution, (iii) la mutation sur un gène conduit à l'apparition de nouvelles propriétés ou de nouvelles fonctions contribuant ainsi à l'accroissement de la diversité du matériel génétique (Gorlov *et al.*, 2006).

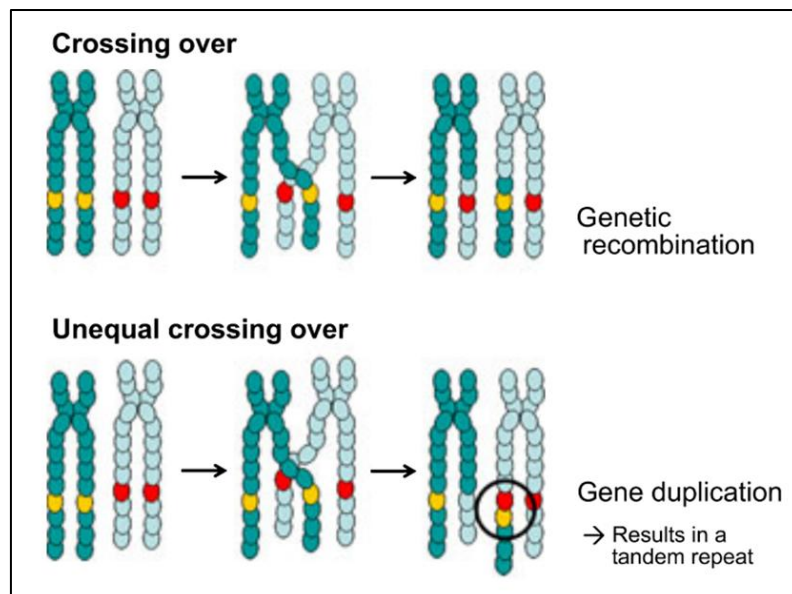
## 2.1. Diversification du matériel génétique

La duplication génique est un des phénomènes majeurs ayant contribué à l'évolution génétique chez les plantes (Ober, 2005). Elle permet l'émergence de nouveaux mécanismes moléculaires au même titre (et dans certain cas en complément) que les mutations génétiques. Ce phénomène est récurrent à l'échelle de l'évolution (Lynch and Conery, 2000), même si l'apparition d'une nouvelle fonction d'un gène après duplication reste de l'ordre de l'anecdotique à l'échelle d'un génome (Lynch and Conery, 2000).

La duplication peut concerner un gène, un chromosome voire un génome entier (Ober, 2010) et résulte de deux mécanismes principaux chez les plantes : la duplication tandem et la rétroduplication.

### 2.1.1. Duplication tandem

La duplication tandem, ou enjambement inégal, est un processus intervenant pendant la méiose. Classiquement, pendant cette étape, des fragments de chromosomes homologues sont interchangeés pour former des chromosomes recombinés, engendrant un brassage intra ou interchromosomique. Cependant, lorsque les fragments recombinés sont à des positions différentes sur chacun des deux chromosomes recombinés, la recombinaison aboutie à la formation d'un chromosome dont une partie a été déléetée et d'un chromosome avec une région génomique dupliquée (Figure 5)(Ober, 2010).



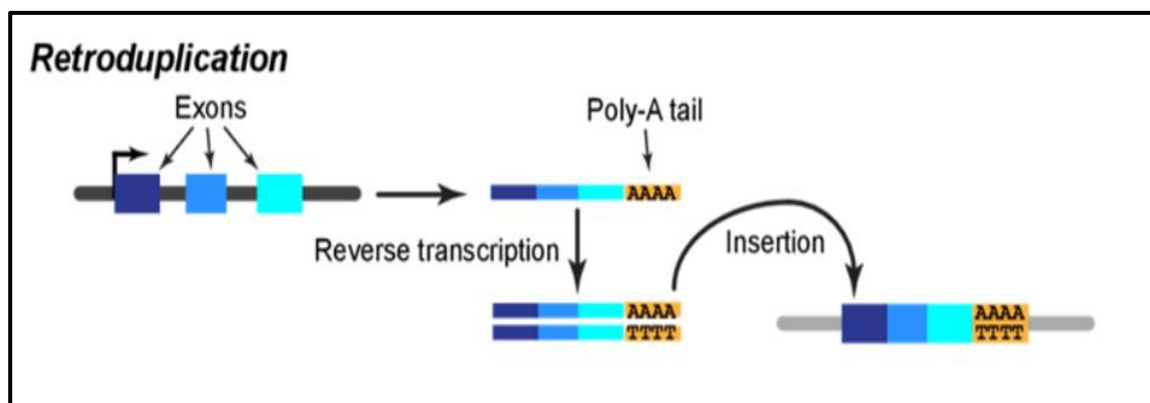
**Figure 5: Représentation simplifiée d'un enjambement et d'un enjambement inégal entre deux bras de chromosomes.** Les chromosomes sont représentés en prophase I de la méiose, chaque couleur correspond à un fragment différent des chromosomes. L'enjambement entraîne l'échange réciproque de matériel génétique entre les chromatides de deux chromosomes homologues. L'enjambement inégal apparaît lorsque deux chromosomes homologues ne sont pas parfaitement appariés, et entraîne l'échange de fragments de chromosomes de taille inégale.  
D'après Ober, 2010.



La duplication en tandem peut être illustrée par deux exemples. Premièrement, chez l'eucalyptus (*Eucalyptus grandis*), le séquençage du génome a permis de révéler qu'au total, 34% du génome était issu de duplication en tandem, soit 12 570 gènes, parmi lesquels sont retrouvés des gènes de réponse au stress, des facteurs de transcription ou encore des gènes impliqués dans la synthèse de phénylpropanoïdes (Myburg *et al.*, 2014). De la même façon, chez le peuplier (*Populus trichocarpa*), 15.6% du génome provient de ce phénomène (Tuskan *et al.*, 2006).

### 2.1.2. Rétroduplication

Dans le cas de duplications géniques, la rétroduplication est un phénomène beaucoup moins représenté que l'enjambement inéga (Panchy *et al.*, 2016). Au cours de la transcription, l'ARN messager formé peut être rétro-transcrit en ADNc double brin, susceptible alors de s'intégrer aléatoirement dans le génome (Figure 6). Le plus souvent, l'ADNc nouvellement intégré dans le génome, ou rétrogène, ne s'exprime pas et est considéré comme un pseudogène (Benovoy and Drouin, 2006; Zhang, 2003). Cependant, le rétrogène peut utiliser la machinerie transcriptionnelle d'un gène actif à proximité pour s'exprimer, générant des fonctions génétiques redondantes (Drouin and Dover, 1990; Zhang *et al.*, 2005). Si cette insertion se fait en amont d'une séquence promotrice active, l'expression de ce gène peut également se faire selon un schéma spatio-temporel totalement différent du gène initial.



**Figure 6: Représentation schématique du mécanisme de rétroduplication.** La rétroduplication est issue de la duplication d'un gène via la rétrotranscription d'un ARNm épissé  
D'après Panchy *et al.*, 2016.

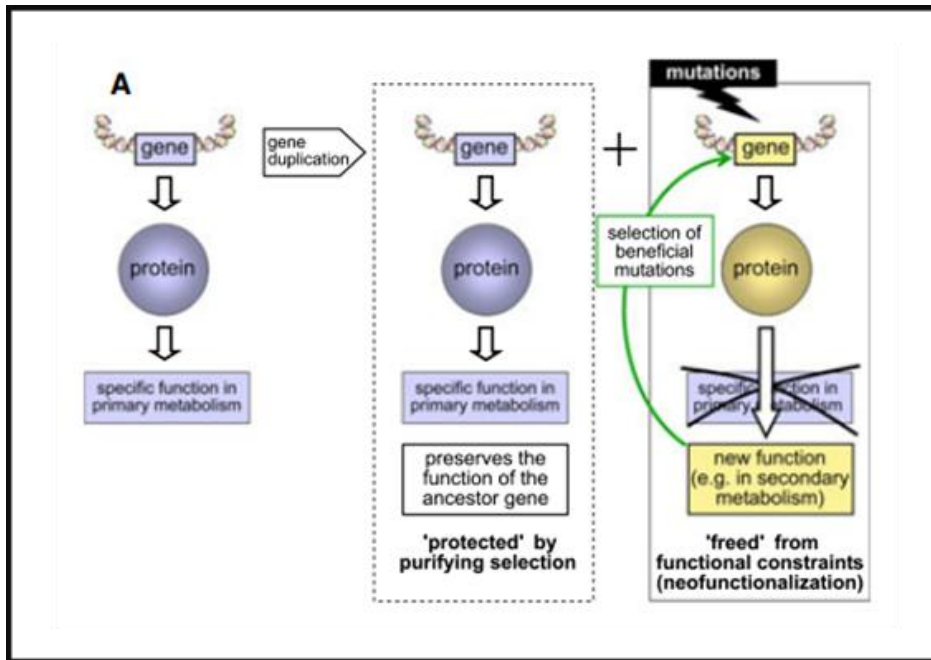
Enfin dans d'autres cas, ces gènes peuvent acquérir de nouvelles fonctions à la suite de mutations complémentaires. C'est le cas par exemple d'un gène codant initialement une enzyme impliquée dans la synthèse de la lignine chez *Arabidopsis* (*CYP98A3*) qui, après rétroposition a évolué vers un rôle dans le développement du pollen (*CYP98A8* et *CYP98A9*) (Matsuno, *et al.*, 2009). Généralement, ces nouveaux gènes issus de rétroposition proviennent d'ARN messagers épissés et sont de fait dénués d'introns.

### **2.1.3. Acquisitions de nouvelles fonctions**

#### **2.1.3.1. Néofonctionnalisation**

La présence de deux copies du même gène n'est *a priori* pas nécessaire pour un organisme vivant mais se révèle être un élément important pour permettre son adaptation à un environnement particulier.

Ainsi, la néofonctionnalisation permet à une des deux copies, par le biais de mutations, de générer des protéines avec de nouvelles fonctions (Conant *et al.*, 2014) (Figure 7). Cette étape, pendant laquelle le deuxième gène n'est pas exprimé (pseudogène) est toujours précédée par une période où la pression de sélection est nulle (Ohno, 1973). Du fait de l'absence de pression de sélection, l'existence du gène supplémentaire peut être considérée comme transitoire dans le contexte évolutif.

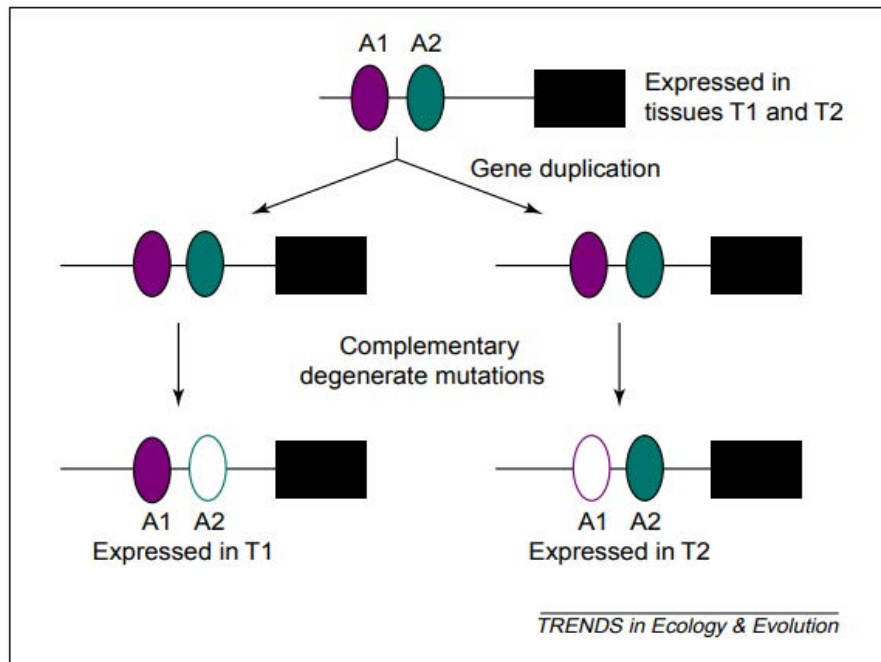


**Figure 7: Représentation schématique du mécanisme de néo-fonctionnalisation.**  
*L'hypothèse de ce modèle d'évolution des fonctions enzymatiques a été suggérée  
 premièrement par Ohno en 1970.  
 D'après Ober 2010.*

Cependant, les mutations conduisant à une néofonctionnalisation sont extrêmement difficile à distinguer des autres mutations occurrents dans le génome, et les mécanismes qui les régissent sont à ce jour mal connus (Cheng *et al.*, 2018).

### 2.1.3.2. Sub-fonctionnalisation

L'apparition de deux copies d'un même gène ne conduit pas systématiquement à l'émergence de nouvelles fonctions, comme en font état les travaux de Hugues (1994, 2002) et de Lynch (2000). La théorie de sub-fonctionnalisation issue de ces travaux postule qu'à l'issue d'une duplication génique, les deux copies formées partagent les fonctions du gène ancestral. Ces deux copies subissent alors une pression de sélection active et permanente et accumulent des mutations dégénératives. De ce fait, la fonction initiale du gène est décomposée en deux fonctions complémentaires, chacune attribuée à une copie du gène (Figure 8). Ainsi, les deux fonctions restantes, portées par chacune des copies, peuvent être exprimées séparément dans l'organisme.



**Figure 8 : Représentation schématique du mécanisme de la sub-fonctionnalisation.** Les carrés noirs représentent les gènes, les ovales pleins représentent les éléments d'actions en cis régulant la transcription des gènes et les ovales vides, les éléments cis désactivés.  
D'après Zhang, 2003.

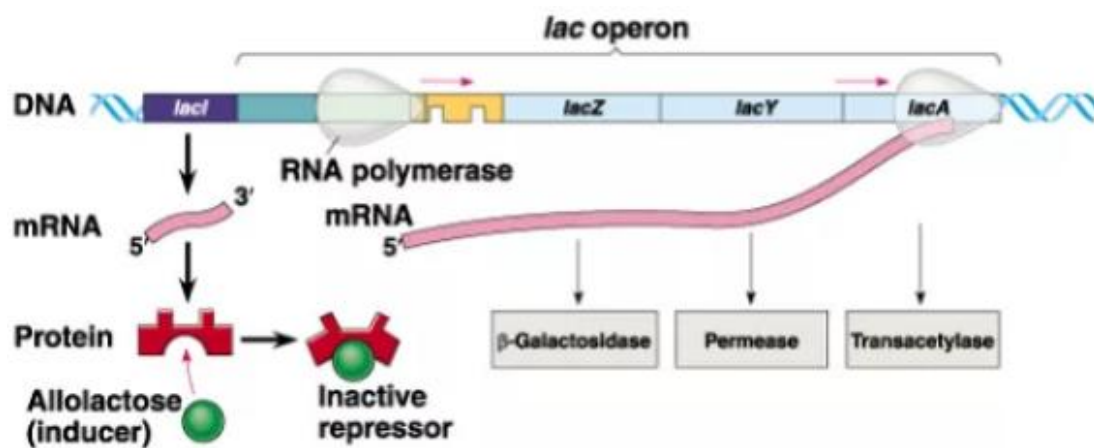
Ce phénomène est introduit par le modèle de « Duplication-Dégénération-Complémentarité » (Force *et al.*, 1999), décrivant la possible division d'un gène ancestral en deux gènes distincts. Ainsi, la sub-fonctionnalisation ne concerne que les gènes impliqués dans au moins deux fonctions biologiques. Après duplication, chaque fonction peut alors être réattribuée, voire optimisée. Elle peut apporter un avantage évolutif dans la sélection des copies du gène ancestral, notamment lorsque la division du gène ancestral, et donc des fonctions associées, favorise la spécialisation de ces fonctions. Cette spécialisation entraîne alors une activité protéique accrue due à une diminution des effets pleïotropes subit par le gène codant cette protéine (Ober, 2010). On retrouve ce phénomène chez *Solanum commersonii*, pour lequel le gène ancestral *R2R3 MYB* a donné naissance, après duplication et sub-fonctionnalisation, aux gènes *AN1* et *AN2*, respectivement impliqués dans la régulation de la synthèse d'anthocyanine et dans la résistance au froid (D'Amelia *et al.*, 2018).

## 2.2. Organisation en Cluster

### 2.2.1. Description des clusters

Il est admis de nos jours que la position de nombreux gènes sur le génome bactérien n'est pas le fruit du hasard mais répond plutôt à une organisation en groupe de gènes appelés clusters. Ces gènes sont transcriptionnellement corégulés et ont généralement des fonctions complémentaires dans la vie de la bactérie (voies de biosynthèse de molécules d'intérêt, protéines impliquées dans des mécanismes cellulaires comme le transfert du T-DNA chez les agrobactéries, ...).

Cette structure génomique est souvent mentionnée comme correspondant à des opérons (Yi *et al.*, 2007). Ce terme a été utilisé pour la première fois en 1960 par Jacques Monod et son équipe (Jacob *et al.*, 1960) pour décrire l'organisation de trois gènes, *lacZ*, *lacY* et *lacA*, impliqués dans le métabolisme du lactose chez *Escherichia coli* (Jacob and Monod, 1961) (Chhabra, 2013, Figure 9).



\* Lactose present, repressor inactive, operon on

Figure 9: Organisation génomique de l'opéron lactose chez *Escherichia coli*.  
D'après Chhabra, 2013.

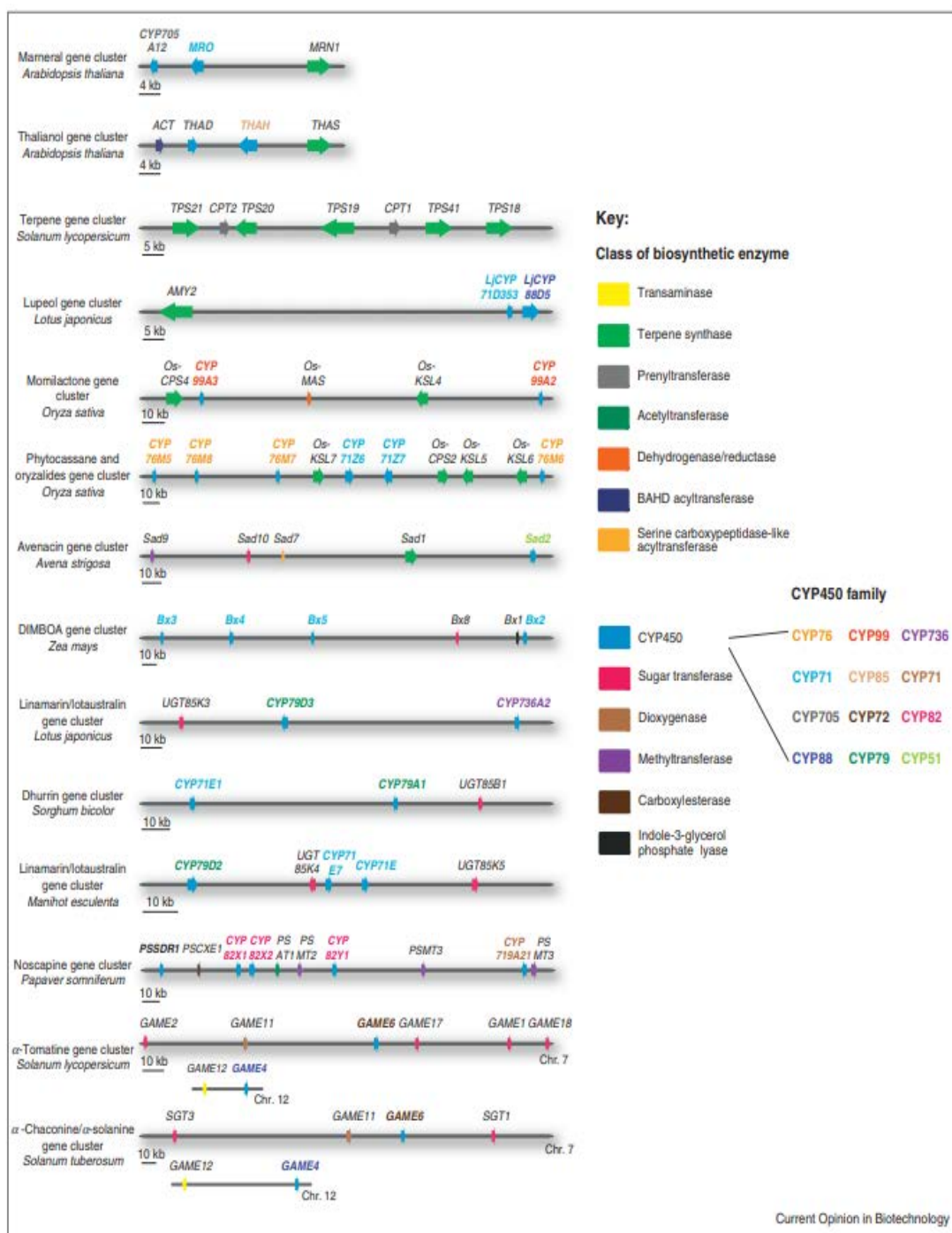
Depuis, de nombreuses études se sont focalisées sur la recherche de clusters de gènes chez les bactéries et les champignons.

L'apparition récente des nouvelles méthodes de séquençage (454, Illumina, PacBio...) ainsi que l'émergence de techniques de biologie moléculaire plus performantes ont permis de compléter les bases de données et de progresser dans la compréhension du mode de fonctionnement de ces structures génétiques. Une vingtaine de clusters ont ainsi été mis en évidence chez *Streptomyces coelicolor*, dont certains sont impliqués dans la synthèse de métabolites spécialisés, notamment à activités antibiotiques (Okamoto *et al.*, 2009; Osbourn, 2010). L'exploitation de ces données a permis l'essor de la production à grande échelle d'un grand nombre de molécules thérapeutiques telles que la céphalosporine, un anticholestérol (Vijayalekshmy *et al.*, 1992), ou les ergopeptides, impliquées dans l'inhibition de la croissance tumorale (Schiff, 2006).

Chez les plantes, la distribution des gènes sur le génome a longtemps été considérée comme aléatoire. Ce point de vue a peu à peu évolué à partir de 1997 lorsque Monica Frey a initié des travaux sur la synthèse de benzoxazinoïdes, métabolites spécialisés ayant des propriétés de pesticides chez les céréales (Frey, 1997). Les différents gènes impliqués dans la synthèse du 2,4-dihydroxy-7-méthoxy-1,4-benzoxazin-3-one (DIMBOA) appartiennent majoritairement à une famille d'enzymes très proches et résultent probablement d'événements de duplication (Dutartre *et al.*, 2012). Par ailleurs, ses travaux démontrent que tous les gènes sont localisés sur le chromosome 4 chez le maïs. Ces travaux précurseurs ont ouvert la voie à d'autres études et à la description de cas de clusters chez différentes autres plantes (Figure 10) (Oliveira *et al.*, 2018; Takos and Rook, 2014; Itkin *et al.*, 2011, 2013; Gonzales-Vigil *et al.*, 2012; Winzer *et al.*, 2012; Field *et al.*, 2011; Field and Osbourn, 2008; Sawai *et al.*, 2006; Thorsoe *et al.*, 2005). Par exemple, chez l'avoine, 4 gènes (*sad1*, *sad2*, *sad3* et *sad7*) intervenant dans la synthèse d'avenacine, une molécule avec des propriétés antimicrobiennes, sont co-ségrégés et colocalisés sur une distance de 3.6 cM (Qi *et al.*, 2004). De la même façon, le génome du riz présente un cluster de gènes impliqués dans la synthèse d'une phytoalexine terpénique, la momilactone (Wilderman *et al.*, 2004).

En règle générale, chez les plantes, un cluster de gènes impliqués dans la synthèse de métabolites spécialisés est composé du premier gène de la voie, et d'au moins un autre gène intervenant dans des réactions enzymatiques en aval de la voie de biosynthèse.

Ces clusters de gènes végétaux sont cependant très différents des opérons bactériens. Si les gènes procaryotes n'ont pas d'intron, les génomes de plantes passent par des étapes d'édition des ARN après transcription (Splicing). Par ailleurs, chez les bactéries, et contrairement aux plantes, l'expression de l'ensemble des gènes d'un opéron est régulée par un unique promoteur. Enfin, chez les végétaux, la notion de colocalisation est toute relative dans la mesure où les gènes, bien que distribués sur un même bras de chromosome peuvent être distants de plusieurs centaines, voire milliers de nucléotides. Un cluster végétal peut, de fait, être réparti sur plusieurs dizaines de milliers de nucléotides (Osbourn, 2010).



**Figure 10: Organisation génomique de certains clusters de cytochrome P450 impliqués dans la synthèse de métabolites secondaires.** Les gènes sont représentés par les flèches. Les gènes codant les enzymes intervenant dans la première étape des voies de biosynthèses décrites sont indiqués en rouge. La voie DIMBOA du maïs comprend trois gènes qui ne sont pas représentés dans la figure: Bx7 séparé du noyau du cluster par une région de 15kb, la glucoside transférase Bx9, localisé sur un chromosome différent, et Bx6 dont la localisation génomique n'est pas connue. Abréviations : OSC, oxidosqualene cyclase; DTS, diterpene synthase (classe I et classe II illustrées); IGPL, phosphate d'indole 3-glycérol lyase; DKS, dicétone synthase; AT (BAHD), BAHD-acyltransferase; AT (SCPL), SCPL-acyltransferase; MT, méthyltransferase; UGT, sucre dépendant de l'UDPtransférase, DHO, déshydrogénase/réductase; L/CT, lipase/carboxyltransférase; CES, carboxylestérase; DOX, dioxygénase; TA, transaminase; CYP, cytochrome P450. D'après Nutzmann et al., 2016



## **2.2.2. Intérêt de l'agencement en clusters**

Les clusters de gènes découverts chez les plantes sont majoritairement impliqués dans la synthèse de molécules de défense (Schlöpfer *et al.*, 2017). Plusieurs études se sont penchées sur l'intérêt d'une telle organisation dans la mise en place de systèmes de défense de la plante lors de la perception par celle-ci de stress biotiques, d'après lesquelles plusieurs théories ont émergé.

### **2.2.2.1. Hérité**

L'organisation de gènes sous forme de clusters confère à la plante un avantage sélectif, notamment vis-à-vis de la résistance aux pathogènes ou aux herbivores. En effet, cet agencement permet de diminuer voire d'éviter les risques d'interruption de la cascade de réactions de biosynthèse *via* l'apparition de mutations, pouvant conduire par exemple à une forte diminution de la bioactivité des métabolites synthétisés (Sue *et al.*, 2011), ou encore à une accumulation d'intermédiaires potentiellement toxiques pour la plante (Kristensen *et al.*, 2005; Xu *et al.*, 2012). Si la mise en place de clusters semble être un atout pour la survie de la plante, il apparaît primordial que cette combinaison favorable de gènes soit transmise à la descendance. La transmission d'un ensemble de gènes physiquement liés sur le génome, et impliqués dans un même processus de défense, permettrait alors de réduire les risques d'apparition de voies métaboliques incomplètes chez la descendance, augmentant la compétitivité de celle-ci (Qi *et al.*, 2004).

### **2.2.2.2. Co-régulation**

La régulation de l'expression de gènes fait intervenir de nombreux éléments moléculaires (ex : promoteurs), métaboliques (ex : facteur de transcription) et cellulaires (ex : chromatine) (Gatehouse, 1997; Geiduschek, 2010), chacun recrutant des éléments de régulation différents, liés à ses conditions d'expression. Dans le cas d'une voie métabolique, le nombre plus ou moins important de gènes impliqués nécessite une régulation coordonnée et fine de chacun de ces gènes pour éviter une expression incomplète de la voie.

L'agencement en clusters des gènes d'une même voie de biosynthèse permet alors d'optimiser la régulation de ces différents gènes, et d'ainsi éviter tout dérèglement de l'expression. En effet, certaines études ont révélé l'existence de mécanismes de co-régulation de gènes physiquement proches sur le génome (Cohen *et al.*, 2000; Kosak et Groudine, 2004). Ces mécanismes de co-régulation peuvent être illustrés, par exemple, par une structure bien connue chez les levures, mais encore peu caractérisée chez les plantes : les promoteurs bidirectionnels. Cet élément est décrit comme une séquence promotrice encadrée par deux gènes et capable d'initier la transcription dans deux sens différents (Hurst *et al.*, 2004). De telles structures permettent alors d'activer ou de réprimer l'expression de deux gènes adjacents, et de limiter le nombre de promoteurs unidirectionnels impliqués dans une même voie de biosynthèse (Bonnelye et Laudet, 1994). Cette co-régulation permet également de ne faire intervenir qu'un seul facteur de transcription pour plusieurs gènes, comme il a été démontré dans le cas de la synthèse du thalianol et du marnéral (Field *et al.*, 2011; Hurst *et al.*, 2004). A ce jour, plusieurs promoteurs bidirectionnels ont été décrits chez différentes plantes telles que *Arabidopsis thaliana* (Wang *et al.*, 2009b), le riz (Wang *et al.*, 2016) ou le melon (Wang *et al.*, 2008).

La co-régulation de gènes agencés en clusters peut également être favorisée par des modifications locales au niveau de la chromatine (Montavon and Duboule, 2013). Pour rappel, la chromatine est une structure composée d'ADN et de protéines, principalement représentées par les histones (Niu et Zheng., 2018). Cette structure permet le compactage de l'ADN en vue de son intégration dans le volume nucléaire de la cellule (Cooper, 2000). Dans le cas de clusters, la chromatine peut subir des modifications localisées permettant notamment d'isoler le cluster des autres gènes adjacents. Ce remodelage permet alors d'augmenter l'efficacité de transcription et de faciliter l'activation ou l'extinction de l'expression d'un ensemble de gènes impliqués dans le même processus métabolique (Nützmann et Osbourn, 2014). Chez les levures, ce mécanisme a été très étudié, notamment au travers du cluster *DAL*, composé de 6 gènes permettant à *Saccharomyces cerevisiae* d'utiliser l'allantoïne comme source d'azote (Wong et Wolfe, 2005). Chez les plantes, la co-régulation de gènes d'un même cluster a par exemple été caractérisée chez *Arabidopsis thaliana* (Tran *et al.*, 2005), chez l'avoine (Wegel *et al.*, 2009), l'*aegilops* speltoïdes (Pollak *et al.*, 2018) ou encore le poivron (Mohan *et al.*, 2018).

### 2.2.2.3. Métabolons et Channeling

Il a été démontré que la formation de complexes enzymatiques (métabolons) ainsi que l'interaction physique de plusieurs composés issus de la même voie de biosynthèse (channelling) est un mécanisme permettant d'accroître l'efficacité locale de molécules de défense (Jørgensen *et al.*, 2005). Pour exemple, ce phénomène a été observé chez le soja dans le cas de la synthèse d'isoflavones (Mameda *et al.*, 2018), connu pour leurs activités antifongiques (Meragelman *et al.*, 2005) ou encore chez le sorgho avec la synthèse de dhurrine, un glucoside cyanogénique (Laursen *et al.*, 2016). Une organisation en clusters permettrait alors de favoriser l'expression concomitante de gènes impliqués dans la synthèse de métabolons et d'améliorer la réponse de la plante en cas d'attaques (Nutzmann *et al.*, 2016). Cependant, plusieurs contre-exemples ont été décrits. Par exemple, chez le maïs ou l'avoine, les gènes respectivement impliqués dans la synthèse de DIBOA/DIMBOA ou d'avenacine ne sont pas tous colocalisés et sont réparti à travers différents compartiments subcellulaires (Frey, 1997; Mugford *et al.*, 2009, 2013). Ainsi, à ce jour, l'implication des clusters dans la formation de métabolons n'est pas encore totalement décryptée.

Au cours de l'évolution, la mise en place de mécanismes moléculaires chez les plantes a permis le développement et la diversification du métabolisme spécialisé. En raison de leur diversité et de leur complexité, la compréhension globale de ces mécanismes est en l'état actuel extrêmement laborieuse. Il apparaît donc nécessaire d'étudier ces mécanismes au travers d'un modèle. Un des exemples les plus marquants quant à son implication dans l'adaptation de la plante à son environnement est représenté par les cytochromes P450.

### 3. Les cytochromes P450

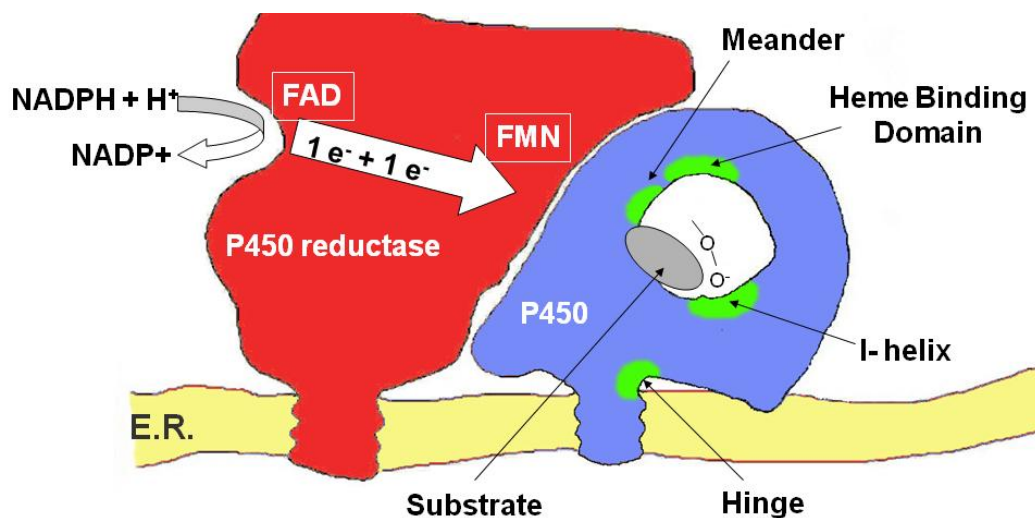
Les cytochromes P450 représentent une superfamille d'hémoprotéines enzymatiques, mise en évidence chez quasiment tous les êtres vivants (Werck-Reichhart et Feyereisen, 2000). La grande majorité des cytochromes P450 présents chez les eucaryotes sont liés à la membrane plastidiale ou à celle du réticulum endoplasmique (Schuler, 1996), mais peuvent être également ancrés à la membrane mitochondriale interne chez certains mammifères (Williams *et al.*, 2000a). Dans le cas des procaryotes, ces enzymes sont solubles (Werck-Reichhart et Feyereisen, 2000). Les cytochromes P450 ont un poids moléculaire compris entre 45 et 60 kDa, pour une séquence codante associée d'environ 1500 pb. La terminologie « P450 » est issue de l'observation par Klingerberg (1958) de l'apparition d'un pigment jaune dans des cellules de foie de rat et de cochon en présence de monoxyde de carbone. Quelques années plus tard, cette coloration jaune a été attribuée à la capacité d'absorption que possèdent ces enzymes à 450 nm lorsque le fer, à l'état ferreux, ( $\text{Fe}^{2+}$ ) se lie au monoxyde de carbone (Omura et Sato, 1964). Ces enzymes sont des monooxygénases catalysant des réactions d'oxydoréduction par l'ajout d'un atome d'oxygène et sont impliquées dans des réactions très diverses telles que l'hydroxylation, l'époxydation ou la désamination (Bernhardt, 2006). D'après Wickramasinghe et Vilee (1975), l'origine de l'activité oxydoréductrice des cytochromes P450 daterait de plusieurs milliards d'années, et serait liée à la détoxification de l'oxygène, molécule rare à cette époque. Par la suite, les cytochromes P450 se sont multipliés et leurs fonctions se sont diversifiées pour intégrer des processus de synthèse ou de dégradation de composés. Pour appuyer cette théorie, Nelson a, par des analyses phylogéniques, daté l'ancêtre commun des cytochromes P450, qui serait apparu il y a 3.5 milliards d'années (Nelson, 1999; Nelson *et al.*, 1993). Ces travaux sur l'origine des P450 et les relations phylogéniques entre les différents représentants ont permis de définir un système de classification des cytochromes P450.

### 3.1. Nomenclature des cytochromes P450

Ce système de classification est basé sur le degré d'homologie entre les séquences primaires protéiques. Ainsi, chaque cytochrome P450 débute par les trois lettres « CYP » (**C**ytochromes **P**450). A ces trois lettres vient s'ajouter un chiffre qui définit l'appartenance à une famille, régie par un degré d'homologie de séquence protéique de 40% minimum. Un code à une ou deux lettres permet ensuite de déterminer la sous-famille à laquelle l'enzyme appartient. Deux enzymes appartiennent à une même sous-famille quand elles partagent au moins 55% d'homologie de séquences. Le dernier code détermine l'ordre dans lequel un cytochrome P450, au sein d'une même sous famille, a été découvert (Nelson *et al.*, 2004). En plus de cette nomenclature, un autre système permet de discriminer les cytochromes P450, en fonction de leurs mécanismes d'activités.

### 3.2. Classification en fonction du pouvoir réducteur

Les cytochromes P450 sont organisés en 4 classes (I, II, III et IV), en fonction de la source de leur pouvoir réducteur. Dans le cas des classes I et II (Figure 11), les différentes activités catalytiques sont réalisées en présence d'un cofacteur, respectivement NADH (Nicotinamide adénine dinucléotide), et NADPH (Nicotinamide adénine dinucléotide phosphate) porté par une cytochrome P450 réductase NADPH-dépendante (CPR) et permettant le transfert des électrons (Bak *et al.*, 2011; Sono *et al.*, 1996) vers le site catalytique, constitué d'un noyau protoporphyrine IX renfermant un atome de fer.



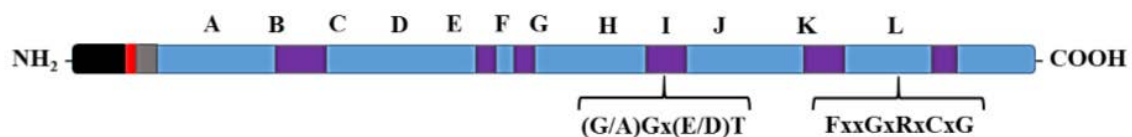
**Figure 11: Représentation schématique de la structure d'un cytochrome P450 de classe II.**  
 Les cytochromes P450 de plantes et leurs réductases sont ancrés à la membrane du réticulum endoplasmique via leurs extrémité N-terminale. Les domaines très conservés chez les P450 sont représentés en vert. Certains résidus interagissant avec les réductases sont également conservés au cours de l'évolution. La plupart des autres régions sont très variables, notamment celles impliquées dans la reconnaissance du substrat.  
 D'après Werck-Reichhart et al., 2000.

Les représentants de la classe III sont exclusivement eucaryotes et ne nécessitent pas d'apport exogènes d'électrons et d'oxygène (Werck-Reichhart et Feyereisen, 2000). La classe IV quant à elle ne comprend qu'un seul individu, P450nor, découvert chez *Fusarium oxysporum*, et capable de réduire le monoxyde d'azote en N<sub>2</sub>O (Takaya et al., 1999). Malgré ces divergences de mécanismes réducteurs et l'hétérogénéité de séquences entre les diverses familles et sous-familles, les cytochromes P450 présentent plusieurs domaines consensus, conservés au cours de l'évolution, et propres à cette superfamille d'enzymes.

### 3.3. Structures conservées

Des études cristallographiques de cytochromes P450 ont permis de déterminer la structure tertiaire de ces enzymes, très conservées entre les différents représentants (Schoch et al., 2004; Wester et al., 2004; Williams et al., 2000a, 2000b; Yano et al., 2005). Ainsi, les cytochromes P450 sont constitués d'hélices  $\alpha$  (Figure 12) et de feuillets  $\beta$ , qui s'associent en deux domaines globulaires. Ces structures sont associées à des séquences évolutivement très conservées.

Une première séquence, appelée « Cys-pocket » (FxxGxRxCxG), et située du côté C-terminal de la protéine contribue au maintien de l'hème dans la protéine (Figure 12). Ce maintien est assuré par des liaisons thioesters établies entre l'atome de soufre présent sur la cystéine des « Cys-pocket » et l'atome de fer de l'hème (Wester *et al.*, 2004). Une autre séquence, (G/A)Gx(E/D)T, en amont de la « Cys-pocket », assure l'activation de l'oxygène lors du mécanisme réactionnel. Enfin, la séquence ExxR, flanquée des deux séquences précédentes et située au niveau de l'hélice K (Figure 12), contribue à la stabilisation de l'hème de la protéine (Yano *et al.*, 2005).



**Figure 12: Représentation schématique de la structure primaire d'un P450 membranaire eucaryote.**

*En noir : ancre membranaire ; en rouge : acides aminés basiques ; en gris : coude proline ; en violet : séquences SRSs. Les hélices  $\alpha$  sont annotées de A à L.  
D'après Gotoh, 1992.*

Un motif de 4 acides aminés, le motif «PERF » (Proline-Acide glutamique-Arginine-Phénylalanine), est également très conservé au sein des P450. L'arginine du domaine « PERF », avec l'arginine et l'acide glutamique de la séquence ExxR, forme la triade E-R-R, intervenant également dans la stabilisation de l'hème (Hasemann *et al.*, 1995). Les études cristallographiques ont permis de mettre en évidence des régions conservées responsables de la fixation du substrat au site actif du cytochrome P450, appelés SRS pour « Substrat Recognition Site ». Ces régions sont mises en évidence pour la première fois en 1992 (Gotoh) à partir de l'alignement de 51 séquences de CYP2 bactériens. Les SRS, numérotés de 1 à 6, participent activement à la fonctionnalité de l'enzyme au niveau de son site actif. Plus que leurs séquences, ce sont les positions de ces différents SRS qui sont très conservées entre les cytochromes P450 d'une même sous-famille, voire entre tous les cytochromes P450. En effet, les SRS sont impliqués dans la reconnaissance et la fixation du substrat d'une enzyme donnée. Il apparaît donc logique que les séquences de ces motifs divergent en fonction de l'affinité de substrat des enzymes étudiées.

Pour exemple, une étude s'est penchée sur l'évolution de la séquence des SRS de plusieurs représentants de la sous-famille des *CYP71AJ* d'Apiacées (Dueholm *et al.*, 2015). Malgré une homologie de séquences globales de ces protéines supérieures à 75%, l'homologie entre les différents SRS est beaucoup moins conservée (Figure 13).

	SRS1	SRS2	SRS3	SRS4	SRS5	SRS6	Group
<i>CYP71AJ6</i>	ANRPNSSIWDKIFYPNGKDVVFAPY	QNIQAQLI	VEKAANEADLF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMTSETGMTLHKKSP	II-6
<i>CYP71AJ5</i>	ANRPNSSIWDKIFYPNGKDVVFAPY	QNIQAQLI	VEKAANEADLF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMTSETGMTLHKKSP	
<i>CYP71AJ15</i>	ANRPNSSIWDKIFYPNGKDVVFAPY	QNIQAQLI	VEKAANEADLF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMTSETGMTLHKKSP	
<i>CYP71AJ7</i>	ANRPNSSIWDKIFYPNGKDVVFAPY	QNIQAQLI	VEKAANEADLF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDLTSETGMTLHKKSP	
<i>CYP71AJ9</i>	ANRPNSSIWDKIFYPNGKDVVFAPY	QNIQAQLI	VEKASAEVDVF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMDAETGMTLHKKSP	
<i>CYP71AJ8</i>	ANRPNSSIWDKIFYPNGKDVVFAPY	QNIQAQLI	VEKASAEVDVF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMDAETGMTLHKKSP	
<i>CYP71AJ11</i>	ANRPNSSIWDKIFYPNGKDVVFAPY	QNIQAQLI	VEKASAEVDVF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMDAETGMTLHKKSP	II-5
<i>CYP71AJ14</i>	ANRPVSSIWSRIFYPNGKDVAFAPY	QNIQAQLI	VEKAANEIDAF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMASETGMTVHKKSP	
<i>CYP71AJ12</i>	ANRPDSSIWSRIFYPNGKDVAFAPY	QNIQAQLI	VEKAANEIDAF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMASETGMTVHKKSP	
<i>CYP71AJ13</i>	ANRPDSSIWSRIFYPNGKDVAFAPY	QNIQAQLI	VEKAANEIDAF	KDSIRGVILNMFFDGTDTTSAVLEWT	IYTPVPLLARE	LDMDSETGMTVHKKSP	I-4
<i>CYP71AJ3</i>	ADRPYSSVANRIFYPNGKDMVFARY	LQIMDML	VEKAANGVDAF	KECIKSLIWDMLGAGTETIATALEWT	LYFTAPLLVPRE	MDMTASTGITFHKKSP	
<i>CYP71AJ27</i>	ADRPYSSVANRIFYPNGKDMVFARY	LQIMDML	VEKAANGVDAF	KECIKSLIWDMLGAGTETIATALEWT	LYFTAPLLVPRE	MDMTASTGITFHKKSP	
<i>CYP71AJ2</i>	ADRPYSSVANKVFPYNGKDMVFARY	LEIMDML	VAKAAGVDDF	NECIKSLIWDMLGAGTETIATALEWT	LYFTAPLLVPRE	LDMTAVSGITLKKSP	
<i>CYP71AJ1</i>	ADRPYSSVANKVFPYNGKDMVFARY	LEIMDML	VAKAAGVDDF	NECIKSLIWDMLGAGTETIATALEWT	LYFTAPLLVPRE	LDMTAVSGITLKKSP	I-3
<i>CYP71AJ4</i>	SQRPTSIAIDKLCYKGRDVAFSRY	AEVTEML	VEKTAKEVDFAF	EECIKAVIWDMLGGTETTSTTLEWI	LYITA-FLLPRE	LDMSGVPGISLYRKYKSP	
<i>CYP71AJ21</i>	ADRPFLSVANRIFYPNGKDMVFARY	GEIMELI	AEKAKEVDTF	KDCIKAVIWDMLFVAGTDTTSSSTLEWA	LYMTAPLLVPRE	LDMTGVTGLTVRRKSP	I-2
<i>CYP71AJ25</i>	SDRPITSSVANRIFYPNGKDMVFARY	GEIMELI	VEKAANEVDTF	KDCIKAVIWDMLFVAGTDTTSSSTLEWT	LYMTAPLLVPRE	LDMTGVTGLTVRRKSP	
<i>CYP71AJ40</i>	SDRPSSSVANRIFYPNGKDMVFARY	GDIMNML	VEKLAKEMDTF	KECIKAVIWDMLFVAGTDTTSSSTLEWT	LYFTAPLLVPRQ	LDMTAVTGLTLRRSP	I-1
<i>CYP71AJ32</i>	SDRPSSLVVEYKLFYPNGKDVGFARY	GKVMELL	VEKVAKEVDFAF	KDCIKAVIWDMLFVAGTDTTSSSTLEWT	LNFTAPFLLPRE	LDMTAVTGLTLRRSP	
Consensus	:::* * : *:*:* *:	: : : :	. : : *	::*:*:* : *	: .. : : : *	::* :*:* : :	

**Figure 13: Alignement de la séquence protéiques de 19 cytochromes P450 appartenant à la sous-famille *CYP71AJ*. SRS signifie « Substrate Recognition Site ». Les résidues d'acides aminés qui constituent les sites actifs sont indiqués par les triangles. Les triangles noirs sont ceux avec les positions les plus conservés par rapport au modèle, à la différence des triangles gris**  
D'après Dueholm *et al.*, 2015.

Ainsi, les SRS participent à la spécificité d'action et à la diversification de fonction des cytochromes P450. Le nombre grandissant de cytochromes P450 végétaux associés à des fonctions très variées font que ces enzymes sont impliquées dans de nombreuses réactions biologiques.



## 3.4. Rôles biologiques

### 3.4.1. Synthèse de métabolites spécialisés

Chez les plantes, les cytochromes P450 sont largement impliqués dans de nombreuses voies de biosynthèse du métabolisme spécialisé. Par exemple, pour les terpènes, une trentaine de ces enzymes ont été caractérisées. On retrouve CYP76B10, intervenant dans la synthèse de géraniol chez *Swertia mussottii* (Wang *et al.*, 2010), ou encore CYP71D13 et CYP71D15, responsable de la synthèse de limonène chez la menthe (*Mentha x piperita*, Karp *et al.*, 1990).

De la même façon, de nombreux cytochromes P450 sont impliqués dans la synthèse d'alcaloïdes, comme pour la camalexine, une phytoalexine, chez *Arabidopsis thaliana* (Glawischnig *et al.*, 2004), l'isoquinoline chez le pavot de Californie Ikezawa *et al.*, 2009) ou encore l'ibogaïne, utilisée dans les traitements anti-addiction, et accumulée par l'iboga (*Tabernanthe iboga*, Farrow *et al.*, 2018).

Dans le cas des composés phénoliques, ces enzymes sont retrouvées dans des processus métaboliques plus larges que le métabolisme spécialisé. Les cytochromes P450 participent par exemple au maintien du port dressé de la plante par la synthèse de lignine, notamment *via* l'action de CYP98A3, impliqué dans l'hydroxylation d'esters d'acide *p*-coumarique pour former des monomères de lignine (Schoch *et al.*, 2001) ou encore CYP84A, qui hydroxyle la fonction aldéhyde de l'alcool coniférylique pour former le S-monolignol (Maruta *et al.*, 2014; Mayer *et al.*, 1999). A l'instar des terpènes et des alcaloïdes, certains P450 dédiés aux composés phénoliques sont impliqués dans l'attraction de pollinisateurs comme les flavonoïdes (Seitz *et al.*, 2007) ou les anthocyanes (Tanaka, 2006). De nombreux cytochromes P450 participent également à la synthèse de molécules de défense chez les polyphénols, comme les furocoumarines (Larbat *et al.*, 2007, 2009) ou les stilbènes (Jian *et al.*, 2016).

### 3.4.2. Détoxification des xénobiotiques

Un certain nombre de travaux décrit le rôle des P450 dans la dégradation de composés exogènes chez les plantes, notamment des herbicides. Par exemple, des réactions d'alkylation permettent à l'artichaut de se prémunir de l'action d'un herbicide, l'Afalon, par l'action de CYP76B1 (Didierjean *et al.*, 2002). Une activité similaire a été décrite chez *Amaranthus palmeri* en présence de Tembotrione (Küpper *et al.*, 2018). De la même façon, les travaux de Höfer *et al.* (2014) ont démontré l'implication de tous les CYP76 décrits chez *Arabidopsis thaliana* dans le catabolisme des herbicides phénylurées. Chez le blé, CYP71C6v1 a été démontré comme étant impliqué dans la métabolisation de deux herbicides sulfonyles, le Chlorsulfuron et le Triasulfuron (Wen-Sheng *et al.*, 2006), bien que cette activité ait été démontrée *in vitro* uniquement.

La capacité de certains cytochromes P450 à dégrader des xénobiotiques met en évidence le caractère opportuniste de ces enzymes qui se traduit par une spécificité de substrat parfois peu élevée. Cet opportunisme renforce l'hypothèse de l'implication des cytochromes P450 dans l'adaptation de la plante à son environnement.

### 3.4.3. Adaptation des P450 et coévolution

A l'instar des plantes, certains insectes sont capables de se protéger contre des molécules toxiques *via* l'intervention de cytochromes P450. Par exemple, *Helicoverpa zea*, un insecte généraliste, est capable de métaboliser plusieurs insecticides (cyperméthrine, diazinon, aldrine), grâce à la présence de CYP321A1 et CYP6B8 (Niu *et al.*, 2008; Rupasinghe *et al.*, 2007; Sasabe *et al.*, 2004). Ces propriétés sont connues depuis de nombreuses années et ont notamment été contournés par les industriels en incluant dans la formulation finale de leurs insecticides des inhibiteurs de P450 comme le pyperonylbutoxide ou PPO (Sun *et al.*, 2019).

Si les P450s sont liés à l'adaptation des plantes à leurs environnements, il a également été mis en évidence que ces enzymes jouaient un rôle essentiel pour l'adaptation des insectes. Par exemple, May Berenbaum et Mary Schuler ont montré qu'au moins 8 insectes étaient en mesure de cataboliser les furocoumarines, molécules de défense produites par certaines plantes. Cette métabolisation est réalisée par une sous-famille de cytochromes P450 commune à tous ces insectes, CYP6B (Berenbaum, 1990; Schuler, 1996). Cette capacité à dégrader des molécules de défenses est intimement liée à des mécanismes d'adaptation génomique (Calla *et al.*, 2017). En effet, l'étude du génome de *Depressaria pastinacella*, qui se nourrit principalement de panais, a révélé que 6 sous-familles de cytochromes P450, dont CYP6B, présentaient une forte homologie de séquences.

Ces sous-familles sont probablement apparues par duplication de gènes suivies de néofonctionnalisation (Calla *et al.*, 2017). L'apparition au cours de l'évolution de nouvelles molécules de défenses chez les plantes et concomitamment de systèmes de détoxification chez les insectes fait inévitablement penser à une notion de course à l'armement et de fait, à une coévolution dans laquelle les P450s semblent jouer un rôle clé. Ce phénomène a été décrit pour la première fois en 1964 (Ehrlich et Raven) comme un trait évolutif d'une espèce en réponse à un trait d'une autre espèce et destiné à évoluer. Cette définition peut être illustrée par les travaux de Berenbaum et Feeny (Berenbaum et Feeny, 1981). En effet, d'après cette étude, l'émergence d'un isoforme angulaire à partir des furocoumarines linéaires serait liée à l'existence de cette compétition entre les plantes productrices de furocoumarines et les insectes se nourrissant de ces plantes. Par leurs structures différentes, les furocoumarines angulaires permettraient alors aux plantes productrices de contourner les mécanismes de résistance mis en place par les insectes vis à vis des furocoumarines linéaires. Cette hypothèse est renforcée par le fait qu'à ce jour, aucun individu ne synthétisant uniquement des furocoumarines angulaires n'a été caractérisé.

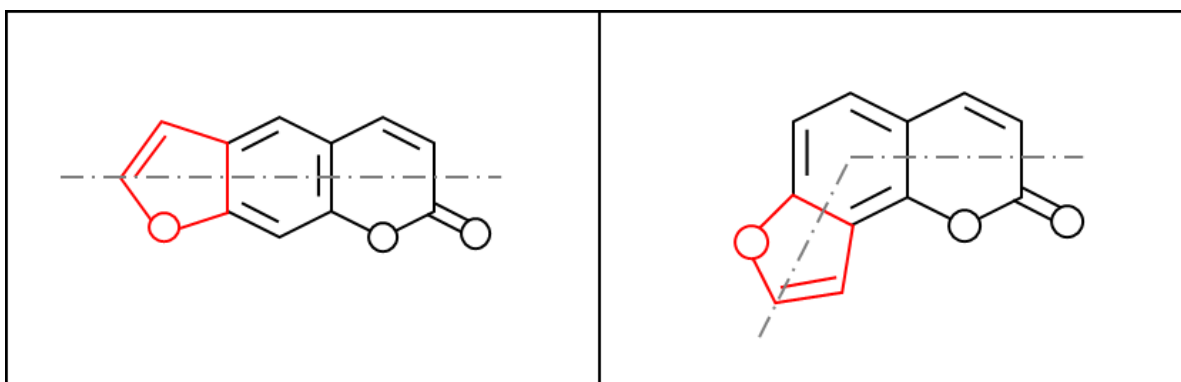
D'autres études décrivant ce phénomène ont également été publiées. Par exemple, il a été mis en évidence que la production de terpènes serait issue de l'interaction entre un coléoptère, *Blepharida*, et le genre botanique *Bursera*, duquel il est prédateur (Becerra, 2007). Cette étude démontre une forte corrélation entre l'apparition de nouvelles molécules et l'émergence de nouvelles espèces de *Blepharida*.

Les cytochromes P450 sont des enzymes clés du métabolisme spécialisé, notamment par leurs diversités de rôles biologiques, mais également par leur place prépondérante dans le phénomène de co-évolution, véritable « course à l'armement » entre plantes et insectes. Comme décrit précédemment, les plantes doivent développer de nouvelles stratégies pour contourner la résistance des insectes à leurs arsenaux chimiques. Les furocoumarines représentent un des exemples emblématiques de cette adaptation.

## 4. Les Furocoumarines

### 4.1. Structures chimiques

Les furocoumarines sont des molécules du métabolisme spécialisé et appartenant à la famille des phénylpropanoïdes. Elles résultent de la condensation d'un cycle furane sur un noyau coumarinique. La position du cycle furane permet de distinguer structuellement deux isomères : les furocoumarines linéaires et les furocoumarines angulaires (Figure 14) (Eisenbrand, 2007). La présence de ces deux hétérocycles confère aux furocoumarines une absorbance maximale aux UV dont la longueur d'onde est comprise entre 300 et 320nm.



**Figure 14: Représentation schématique de la structure de base des furocoumarines linéaires (gauche : psoralène) et angulaires ( droite : angélicine)**

## **4.2. Répartition botanique et localisation des furocoumarines dans la plante**

Même si la littérature scientifique indique que ces molécules ont été identifiées dans 74 familles de plantes différentes (Article en préparation, Munakata *et al.*), les furocoumarines sont majoritairement retrouvées dans quatre familles botaniques, phylogénétiquement éloignées, à savoir les Moracées (ex : le figuier), les Rutacées (ex : la bergamote), les Fabacées (ex : la coronille) et les Apiacées (ex : le panais) (Bourgaud *et al.*, 2006).

La répartition des deux isomères de furocoumarines n'est pas homogène au sein de ces différentes familles. En effet, on retrouve des furocoumarines linéaires et angulaires chez des Apiacées (Cook *et al.*, 2017; Melough *et al.*, 2018), chez certaines Fabacées (Tava *et al.*, 2007) ainsi que chez quelques Moracées (Heinke *et al.*, 2012), alors que les Rutacées n'accumulent que des furocoumarines linéaires (Dugrand *et al.*, 2013). Les furocoumarines ont été mises en évidence dans la majorité des organes de la plante. Cependant, elles sont principalement présentes avec des concentrations très variables dans les feuilles, les fruits et les graines ainsi que, dans une moindre mesure, dans les tiges, les racines et les organes floraux (ovules) (Reinold and Hahlbrock, 1997; Roselli *et al.*, 2017; Zobel and Brown, 1990, 1991a, 1991b).

D'un point de vue histologique, les furocoumarines ne sont pas présentes dans l'organe en entier mais dans des tissus spécifiques.

Par exemple chez le persil (*Petroselinum crispum*, apiacées), les furocoumarines sont accumulées dans la graine et le fruit respectivement dans l'albumen et le péricarpe. Dans le cas des feuilles, les furocoumarines sont localisées au niveau des conduits oléifères, des cellules épithéliales et parenchymateuses, mais également à la surface des feuilles et dans les trichomes (Reinold et Hahlbrock, 1997). De manière plus générale, la localisation des furocoumarines au niveau des organes les plus soumis à la prédation renvoie à une fonction biologique supposée : la réponse aux stress biotiques.

#### **4.2.1. Les Apiacées**

Les Apiacées représentent une famille botanique très diversifiée d'angiospermes et compte près de 4000 espèces et 450 genres (Figure 15). Les représentants de cette famille botanique sont caractérisés par des folioles alternes insérées sur une tige creuse, de petites fleurs organisées en inflorescence, sous forme d'ombelle, et des fruits indéhiscents oléifères (Ibrahim *et al.*, 2017; Sayed-Ahmad *et al.*, 2017). Les Apiacées sont principalement originaires de la région méditerranéenne et de l'Est asiatique mais ont largement été cultivés par la suite dans le reste du monde de par leur intérêt commercial, notamment en industrie pharmaceutique (Cecchini *et al.*, 2010; Christensen et Brandt, 2006; Jana et Shekhawat, 2010) et cosmétique (Hunault *et al.*, 1989).

Les Apiacées représentent l'une des plus anciennes familles botaniques cultivées et sont utilisées depuis l'antiquité en médecine traditionnelle, notamment dans le cas de troubles spasmodiques, diarrhéiques ou neurologiques (Hajlaoui *et al.*, 2010; Koppula and Choi, 2011).



**Figure 15: Exemples d'apiacées. Gauche : Fleur de carotte (*Daucus carota*), milieu : Fleur de panais (*Pastinaca sativa* sp.), Droite : Racine de panais.**

En effet, les graines de la coriandre ou du cumin sont très riches en acides gras variés, comme l'acide pétrosélinique ou l'acide linoléique, deux composés majeurs chez ces plantes (Nguyen *et al.*, 2015). D'autres acides gras tels que l'acide palmitique ou myristique sont également décrits en concentration plus faible. Ces différents acides gras sont très utilisés sous forme d'huile dans l'industrie chimique et entrent dans la composition de lubrifiants industriels ou encore de détergents (Sayed-Ahmad *et al.*, 2017).

Les apiacées sont également très riches en métabolites spécialisés, et notamment en composés phénoliques pouvant être utilisés dans la formulation de produits anti-inflammatoires (ex : quercétine, Hussain *et al.*, 2016) analgésiques (ex : flavonoïdes, Arct et Pytkowska, 2008) antioxydants (ex: resvératrol, Sadowska-Bartosz et Bartosz, 2014) ou d'anticancéreux (ex: acide chlorogénique, Li *et al.*, 2014).

#### **4.2.2. *Pastinaca sativa***

Le panais est une plante herbacée bisannuelle monocarpique. Il dispose d'un port dressé mesurant de 50 cm à 2 m avec des feuilles composées de folioles alternes tridentées insérées sur une tige creuse. Les différentes tiges sont insérées sur une racine pivotante tubérisée de couleur blanche. La floraison du panais s'étale de Juin à Octobre. Les fleurs, de couleur jaune, sont organisées en ombelle de 10 à 20 cm de diamètre composées de 6 à 25 rayons. La reproduction est hermaphrodite et aboutit à la formation d'akènes indéhiscents (Cain *et al.*, 2010).

Le genre *Pastinaca sativa* est principalement représenté par trois sous-espèces : *Pastinaca sativa* L. subsp. Sylvestris, ou panais sauvage, *Pastinaca sativa* L. subsp. Urens, ou panais brûlant et *Pastinaca sativa* L. subsp. Sativa., représentant l'espèce de panais cultivé.

### 4.3. Rôles biologiques des furocoumarines

De par les nombreuses activités antimicrobiennes décrites (McCloud *et al.*, 1992; Ojala *et al.*, 2000; Oliva *et al.*, 2003), les furocoumarines sont considérées comme des phytoalexines, capable de protéger la plante contre des stress biotiques. Des activités insecticides ont également été démontrées en appliquant des solutions de furocoumarines plus ou moins pures (extraits de plantes, huiles essentielles, furocoumarines uniques ou en mélanges, ...) sur des larves d'insectes qui ne sont pas des spécialistes des plantes productrices de ces molécules (Berenbaum et Feeny, 1981; Berenbaum *et al.*, 1991; Guo *et al.*, 2012). La toxicité des furocoumarines envers les bioagresseurs se traduit par différents mécanismes. Le premier est la capacité des furocoumarines à inhiber l'activité de cytochromes P450 (Neal and Wu, 1994). Par exemple, une étude a démontré que la xanthotoxine inhibait fortement l'activité de CYP6D1, un P450 impliqué dans le catabolisme de xénobiotiques tels que les insecticides chez la mouche domestique (Scott, 1996). D'autres travaux ont mis en évidence que chez les plantes, l'activité de la C4H de *Ruta graveolens* (CYP73A32) est partiellement inhibée par le psoralène et la xanthotoxine (Gravot *et al.*, 2004). Chez l'homme, les furocoumarines, contenues par exemple dans le jus de pomelo, peuvent inhiber l'activité de CYP3A4, une enzyme intestinale impliquée dans la métabolisation de certains médicaments (Bailey *et al.*, 2013). Les furocoumarines ont également des propriétés phototoxiques. Sous l'action d'un rayonnement UV, ces molécules sont capables d'établir des liaisons covalentes avec une grande variété de matrice (lipides, protéines, acides nucléiques).

Dans le cas de l'ADN, les furocoumarines vont créer des liaisons covalentes entre deux bases complémentaires engendrant un blocage de la réplication, de la transcription et, *in fine*, la mort cellulaire (Dardalhon *et al.*, 1998).



L'ensemble des propriétés de ces molécules confèrent à la plante une protection contre des agressions biotiques. Cependant, cette toxicité avérée peut également être néfaste pour la plante productrice elle-même. Ainsi, des mécanismes de contournement de cette toxicité ont été mis en place par la plante. Par exemple, chez les Rutacées, il a été démontré que les furocoumarines pouvaient être stockées dans la vacuole sous leur forme glycosylée (Taguchi *et al.*, 2000). Dans ce cas, lors de l'attaque d'un herbivore, le tonoplaste est rompu et les molécules sont remobilisées rapidement après déglycosylation par des  $\beta$ -glucosidases présentes dans le cytosol (Wittstock et Gershenzon, 2002). Chez la berce (*Heracleum sosnowskyi*), une apiacée, il a été démontré que les trichomes étaient capable d'excréter les furocoumarines à leur surface, sous forme de mucilage (Weryszko-Chmielewska et Chwil, 2017). Au contact de l'air, ces furocoumarines s'agencent en structure cristalline. En plus d'être une stratégie de défense en surface, ces phénomènes d'excrétion et de cristallisation des furocoumarines peuvent être reliés au caractère toxique de ces molécules et constitueraient un moyen pour la plante de se prémunir contre la toxicité de ces molécules (Sasaki *et al.*, 1988; Serrano-Pérez *et al.*, 2008).

#### **4.4. Voie de biosynthèse des furocoumarines**

La voie de biosynthèse des furocoumarines (Figure 16), est relativement bien caractérisée d'un point de vue biochimique en raison de nombreux travaux menés entre les années 1960 et 1990 (Bourgaud *et al.*, 2006; Hamerski et Matern, 1988; Wendorff et Matern, 1986; Floss et Mothes, 1966). L'utilisation de précurseurs radioactifs et d'inhibiteurs enzymatiques sur des cultures cellulaires d'*Ammi majus*, Hamerski et Matern (1988), a permis de démontrer que plusieurs étapes de cette voie de biosynthèse étaient catalysées par des enzymes de la famille des cytochromes P450. Il aura fallu attendre les années 2000 et l'amélioration des techniques de biologie moléculaire pour confirmer leur présence dans cette voie de biosynthèse et pour identifier d'autres familles enzymatiques comme celle des dioxygénases, des prényltransférases et des méthyltransférases. La voie de biosynthèse des furocoumarines débute par la conversion du *p*-coumaroyl-CoA en umbelliférone.

La première enzyme catalysant cette réaction, décrite chez *Ruta graveolens*, est une dioxygénase alpha kéto-glutarate dépendante appelée C2'H, capable de convertir à la fois le *p*-coumaroyl-CoA en umbelliférone et le feruloyl-CoA en scopolétine (Vialart *et al.*, 2012). Une enzyme similaire a été décrite chez l'ipomée (*Ipomea batatas*), capable également de synthétiser de la scopolétine à partir de feruloylCoA. L'umbelliférone obtenue est prénylée au niveau des carbones 6 ou 8, conduisant à la formation de la deméthylsubérosine (DMS) ou d'osthénol, respectivement précurseurs de la voie des furocoumarines linéaires ou angulaires, *via* l'intervention d'une prényltransférase. Une première prényltransférase a été décrite chez le persil (*Petroselinum crispum*) et catalyse exclusivement la prénylation en C6 de l'umbelliférone pour donner de la DMS (Karamat *et al.*, 2014). D'autres travaux ont montré que chez le panais, deux prényltransférases, appelées PsPT1 et PsPT2, étaient responsables de la prénylation en C6 et en C8 de l'umbelliférone. Dans ce cas, l'efficacité des enzymes est différente, faisant de PsPT1 une enzyme préférentiellement impliquée dans la synthèse de DMS et PsPT2 dans la synthèse d'osthénol. Chez le citronnier (*Citrus limon*, rutacées), une prényltransférase ayant comme substrat le bergaptol et l'umbelliférone a également été caractérisée (Munakata *et al.*, 2012, 2014).

Par la suite, la DMS et l'osthénol sont convertis respectivement en marmesine et en columbianétine au cours d'une étape de cyclisation du groupement prényl sur le groupement hydroxyle en C7. Cette étape est catalysée par un cytochrome P450 chez *Ammi majus*, mais à ce jour ces résultats n'ont pu être confirmés au niveau moléculaire (Hamerski et Matern, 1988). L'étape suivante de la voie est la conversion de la marmesine en psoralène pour les furocoumarines linéaires et la conversion de la columbianétine en angélicine pour les furocoumarines angulaires. Ces réactions sont catalysées par des cytochromes P450s et libèrent une molécule d'acétone (Hamerski et Matern, 1988 ; Steck et Brown, 1970). D'un point de vue moléculaire, le premier gène impliqué dans la synthèse de psoralène a été isolé et caractérisé à partir de cellules en culture d'*Ammi majus* (CYP71AJ1, Larbat *et al.*, 2007).

La recherche de séquences homologues chez d'autres Apiacées a permis d'identifier les gènes CYP71AJ2 et CYP71AJ3 capables de synthétiser du psoralène chez le céleri et le panais respectivement (Larbat *et al.*, 2009).

Chez le panais, une autre séquence homologue, CYP71AJ4, a été identifiée et catalyse la synthèse d'angélicine, première furocoumarine angulaire de la voie (Larbat *et al.*, 2009).

Enfin, CYP71AJ5 a été caractérisé chez *Thapsia garganica*, une apiacée qui n'accumule pas de furocoumarines, et pour lequel aucune fonction n'a été décrite (Dueholm *et al.*, 2015). Cette observation indique que les CYP71AJ ne sont pas tous impliqués dans la synthèse de furocoumarines reflétant par-là les processus évolutifs des cytochromes P450. Pour les furocoumarines linéaires, le psoralène est ensuite converti en xanthotoxol ou en bergaptol par hydroxylation, respectivement en position 5 ou 8. La conversion du psoralène en xanthotoxol est catalysée par un cytochrome P450, CYP71AZ4 chez le panais (Krieger *et al.*, 2018). Les groupements hydroxyle du bergaptol et du xanthotoxol sont ensuite méthylés pour former le bergaptène (Hehmann *et al.*, 2004) et la xanthotoxine. Chez le panais, la xanthotoxine est, par l'intervention de CYP71AZ6, convertie en 5- hydroxy-xanthotoxine (Krieger *et al.*, 2018).

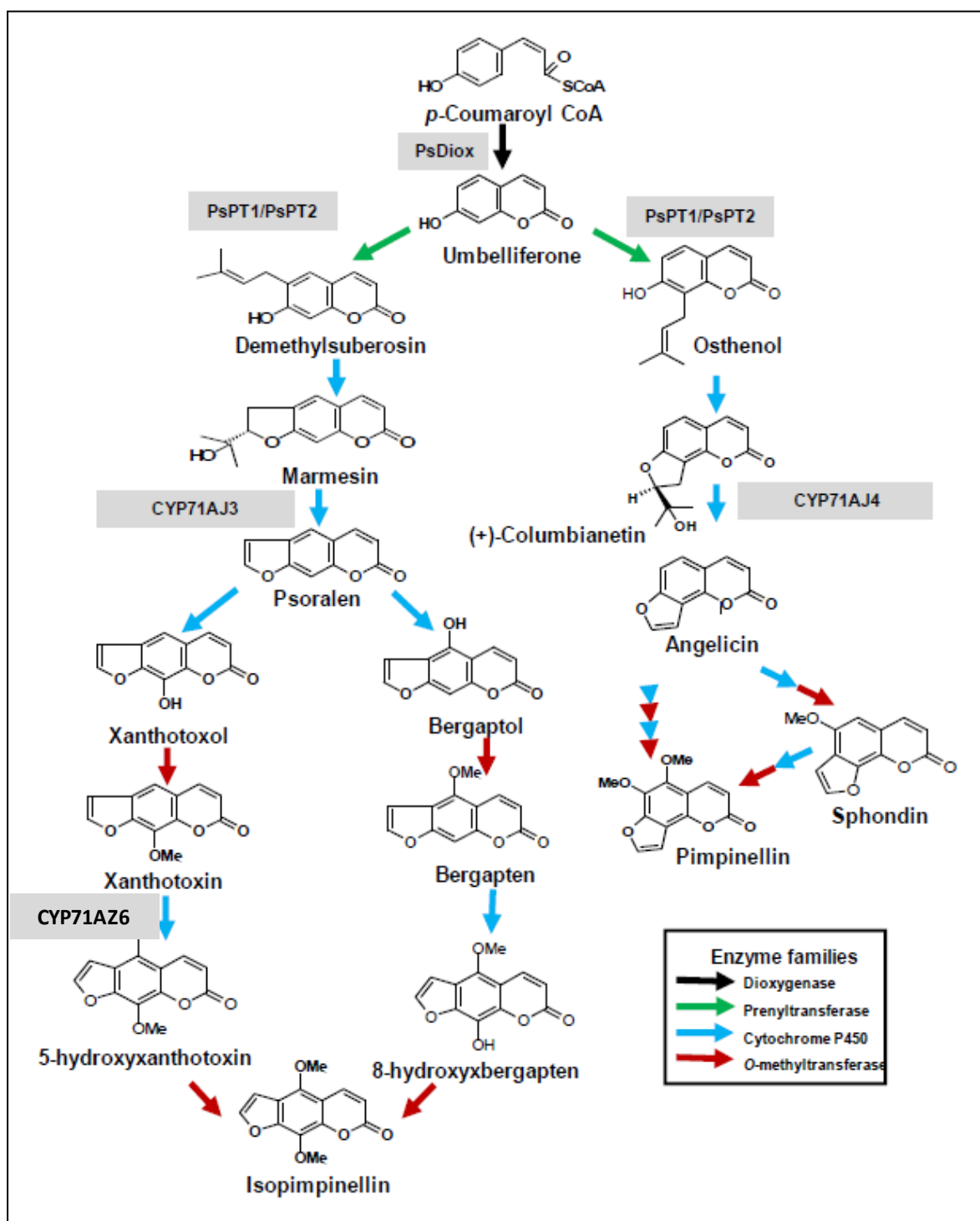


Figure 16 : Représentation de la voie de biosynthèse des furocoumarines chez le panais. Les enzymes caractérisées sont encadrées en gris.

Chez le pomélo, producteur de furocoumarines, et chez *Arabidopsis thaliana*, non producteur, une enzyme catalysant la même réaction a été caractérisée (Audray Dugrand, thèse, Kruse *et al.*, 2008). Cependant, les gènes codants ces deux enzymes appartiennent à la sous-famille CYP82. La 5- hydroxy-xanthotoxine et le bergaptène aboutissent finalement à la formation de l'isopimpinelline. Chez *Ruta graveolens*, il a été avancé que le bergaptène pouvait être converti en 8-hydroxy-bergaptène sous l'action d'une O-méthyltransférase, isolée par chromatographie d'affinité (Sharma et Brown, 1979). Pour les furocoumarines angulaires, l'angélicine est convertie en sphondine et/ou en pimpinelline.

## 5. Objectifs de thèse

Le métabolisme spécialisé est un métabolisme d'adaptation en interconnexion avec le métabolisme primaire. Souvent inductible, il intervient dans l'interaction de la plante avec son environnement, par la synthèse de composés plus ou moins spécifiques de la plante. Par la grande diversité des métabolites spécialisés, ainsi que leurs multiples rôles dans la vie de la plante, leur production implique pour la plante un coup énergétique non négligeable, pouvant sérieusement impacter son métabolisme. Ainsi, ma thèse est positionnée autour de la compréhension de ce métabolisme à travers l'étude d'une famille de molécules intervenant dans la défense de la plante contre des stress biotiques, les furocoumarines, chez une plante modèle, le panais. Mon travail de thèse s'articule autour de trois objectifs distincts.

- Le premier objectif consiste à évaluer les modifications métaboliques engendrées par la perception de deux stress environnementaux par la plante, l'ozone et la blessure mécanique. Pour ce faire, j'ai réalisé une analyse métabolomique, d'abord globale puis ciblée sur les furocoumarines, à partir de différents organes du panais soumis à ces stress.
- Dans un deuxième objectif, j'ai élaboré une approche basée sur le clonage multigénique permettant d'introduire une voie de biosynthèse, complète ou partielle, dans le génome d'une plante hôte. Cette approche a été mise en place dans le but d'étudier l'impact de la synthèse de molécules issues du métabolisme spécialisé, telles que les furocoumarines, sur le statut global de la plante, ici le tabac, naturellement dépourvu de ces composés.
- La voie de biosynthèse des furocoumarines chez les apiacées reste à ce jour lacunaire. De manière à approfondir les connaissances actuelles, je me suis attaché, dans la troisième partie de ma thèse, à rechercher, isoler et caractériser de nouveaux gènes impliqués dans cette voie de biosynthèse par deux approches complémentaires, génomique et transcriptomique.

## **II. Résultats et Discussion**

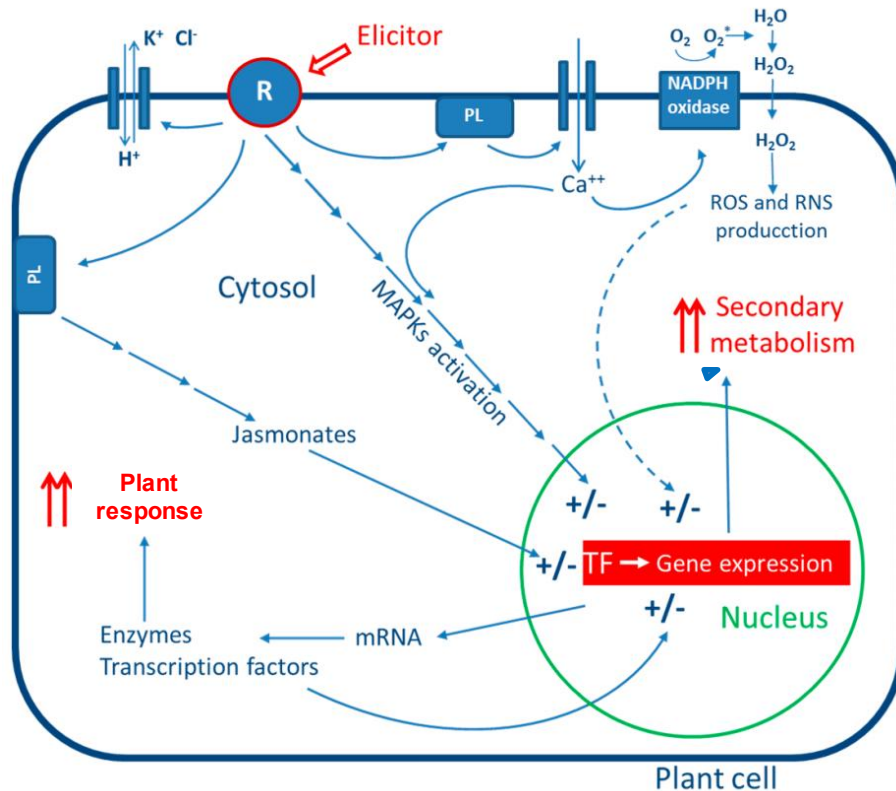
## **A. Etude des modifications métaboliques en condition de stress abiotique chez le panais**

### **1. Introduction**

Le métabolisme spécialisé est très souvent inductible par différents types d'éliciteurs. Le terme "élicitation" caractérise l'action réalisée par un agent physique ou chimique qui est capable d'induire ou de stimuler la défense des plantes (Ramirez-Estrada *et al.*, 2016; Thakur et Sohal, 2013). L'élicitation est notamment utilisée pour des applications industrielles, pour surproduire des métabolites spécialisés à forte valeur ajoutée. Par exemple, l'artémisinine, utilisée pour le traitement du paludisme, peut être suraccumulée dans des chevelus racinaires d'*Artemisia annua* cultivés en présence de cérébrosides fongiques et de monoxyde de carbone (Wang *et al.*, 2009a). De la même façon, l'élicitation de culture cellulaire d'*Ocimum* (Lamiacées) par l'ajout de méthyljasmonate au milieu de culture permet d'accroître la production d'acide bétulinique, un triterpène à activité antitumorale décrite pour la première fois chez le bouleau (*Betula*, betulacées) (Pandey *et al.*, 2015).

La réponse de la plante élicitée est initiée par la perception du ou des éliciteurs par des récepteurs spécifiques présents au niveau de la membrane plasmique. La transduction du signal par les récepteurs associés entraîne l'induction de messagers chimiques permettant d'amplifier ce signal d'élicitation (Figure 18) (Ramirez-Estrada *et al.*, 2016).





**Figure 18: Mécanisme global simplifié de la réponse de la plante à un stimulus extérieur.** R : récepteur, PL : phospholipase, MAPKs : mitogen activated protein kinases, ROS : reactive oxygen species, RNS : reactive nitrogen species, TF : facteur de transcription. D'après Ramirez-Estrada et al., 2016.

Le mécanisme met en œuvre la fixation des éliciteurs sur les récepteurs spécifiques qui conduit à la dépolarisation de la membrane plasmique. Cette dépolarisation engendre l'activation de la chaîne d'antiports  $K^+/H^+$  localisée sur la membrane plasmique. Ces antiports sont impliqués dans le transport de potassium vers le compartiment extracellulaire et de protons vers le cytoplasme, conduisant à une acidification du contenu cytoplasmique et à un relargage d'ions chlorure ( $Cl^-$ ) (Zhao et al., 2005). En parallèle, cette fixation induit l'activation de protéines-G, couplées aux récepteurs, ainsi que de la cascade de MAPKs (mitogen activated protein kinase), dont les réactions de phosphorylation se déroulent séquentiellement (Zhang et Klessig, 2001). Ces réactions conduisent à un relargage d'ions calcium ( $Ca^{2+}$ ) qui, lors de leur fixation sur des senseurs tels que la calmoduline, vont stimuler l'activité de certaines protéines (Protéines kinases Calcium Dépendantes, protéines membranaires, phosphatases,...) impliquées dans la synthèse de métabolites

spécialisés, et notamment de molécules de défense (Ludwig *et al.*, 2004; Vandelle *et al.*, 2018).

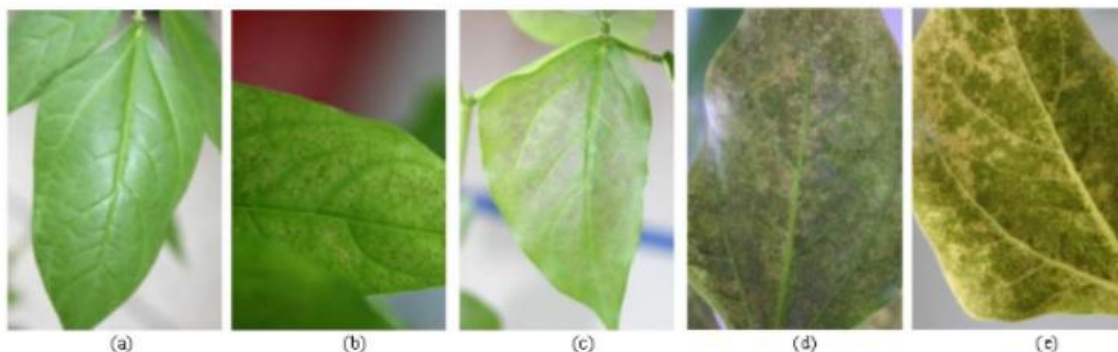
Les récepteurs aux protéines G sont connus pour propager des signaux déclenchant la synthèse de molécules de défenses, comme le diacylglycérol (DAG) ou l'inositol-1,4,5 triphosphate (IP3) (Meijer and Munnik, 2003) intervenant dans l'activation de protéines kinase C et de pompes à calcium ( $\text{Ca}^{2+}$ ). Le calcium transporté par ces pompes active différentes protéines associées à la membrane comme des NADPH oxydase, impliquées dans la formation d'espèces réactives de l'oxygène (ROS) à partir de dioxygène ( $\text{O}_2$ ) moléculaire, mais également d'espèces réactives de l'azote (RNS) à partir d'oxyde d'azote ( $\bullet\text{NO}$ ) et d'ion superoxyde ( $\text{O}_2^{\bullet-}$ ) (Tuteja and Mahajan, 2007). Les ROS et le RNS vont finalement avoir un rôle d'accélération de l'activité des facteurs de transcriptions, conduisant à une surexpression des gènes impliqués dans la synthèse de molécules de défense du métabolisme spécialisé, contrôlés par ces facteurs de transcription (Ferrari, 2010). Pour résumer, la réponse de la plante après perception d'un ou plusieurs stress peut engendrer des modifications conséquentes de son métabolisme global, tant au niveau du métabolisme primaire que spécialisé.

Très peu de données expérimentales concernant la plante modèle de ce projet de thèse, le panais, sont disponibles dans la littérature. Ainsi, pour tenter d'étoffer ces connaissances, j'ai évalué l'impact de deux stress environnementaux, l'ozone et la blessure mécanique, sur le métabolisme global du panais.

## **2. Etude de la réponse métabolique du panais à deux stress environnementaux**

Bien qu'il ne soit pas considéré comme une plante de grande culture, le panais représente un intérêt agronomique certain. A l'instar de nombreuses plantes cultivées, le panais subit les conséquences du changement climatique global. Un facteur impactant pour l'agriculture est l'augmentation de polluants atmosphériques, principalement représenté par l'ozone.

Les effets de l'augmentation de la concentration de l'ozone sur les cultures, et sur les végétaux en générale, sont notoires, et regroupent principalement des effets phénotypiques, comme des chloroses (Figure 19), voire des nécroses des parties aériennes des plantes, impactant sérieusement leurs rendements (Umponstira *et al.*, 2006).



**Figure 19 : Apparition de chloroses (figure (b) à (e)) après application d'ozone sur des feuilles de Cornille (*Vigna Unguiculata*, fabacées).**

*La figure (a) représente la condition contrôle. D'après Umponstira et al., 2006.*

D'autres conséquences du changement climatique global concernent l'accroissement des précipitations, la force des vents ou encore, de manière plus indirecte, de l'herbivorie. Elles influent sur la qualité agronomique des cultures, en étant à l'origine notamment de blessures ou, de manière plus générale, de dégradations physiques de la plante.

Dans ce contexte, nous avons choisi de focaliser notre étude sur l'impact de l'ozone et de la blessure mécanique sur le métabolisme du panais, avec l'hypothèse que les résultats seraient extrapolables à un grand nombre de conditions de stress.

## **2.1. Analyses préliminaires**

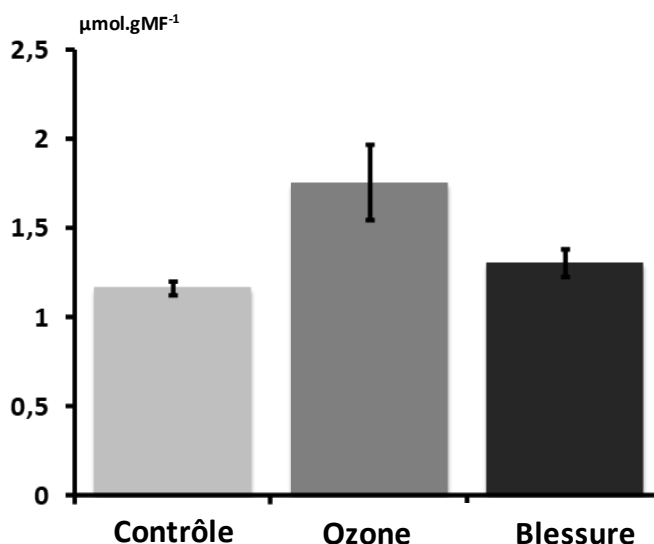
### **2.1.1. Détermination de la dose d'ozone sub-létale chez le panais**

Etant donné les effets délétères que peut avoir l'ozone atmosphérique sur les plantes cultivées, une première étape permettant de déterminer la dose d'ozone maximale à partir de laquelle des symptômes d'oxydation apparaissent chez le panais a été réalisée. Pour ce faire, des panais « Demi-long de Guernesey » âgés de 8 semaines ont été placés dans des chambres climatiques, dispensant différentes concentrations en ozone (0, 80, 120, 150, 200 et 250 particules par milliards (ppb), pendant 72h. A titre de référence, la concentration moyenne en ozone atmosphérique mesurée à Nancy sur l'année 2017-2018 est de 11.2 ppb (ATMO Grand Est Open Data, <https://data-atmograndest.opendata.arcgis.com/>). Chacune des doses d'ozone est appliquée pendant 6h toutes les 24h sur 4 plantes. Les résultats obtenus indiquent que même après 72 h à la plus forte concentration, aucun symptôme d'oxydation (chlorose, nécrose, plasmolyse...) n'est visible sur les plantes, comparativement aux témoins (0 ppb). Ainsi, pour la suite des expérimentations, seule la concentration la plus élevée (250 ppb) sera utilisée.

### **2.1.2. Evaluation du statut oxydatif**

Dans le cadre de cette étude je me suis attaché à analyser l'impact d'un stress ozone et de blessures mécaniques sur la plante. Dans un premier temps nous avons évalué le statut oxydatif par la mesure de la concentration d'une espèce réactive de l'oxygène, le peroxyde d'hydrogène ( $H_2O_2$ ) dans les différentes plantes suite aux différents traitements. La présence de ROS a été décrite comme étant un des marqueurs de stress chez la plante (cf paragraphe 1.) et l'évaluation du statut oxydatif permet d'obtenir des résultats préliminaires quant à l'efficacité des différents traitements. D'après les résultats obtenus (Figure 20), la quantité de peroxyde d'hydrogène dans les folioles de panais traitées à l'ozone est 1.5 fois supérieure à celle présente dans la condition contrôle, après trois jours de traitement.

Une tendance se dégage, bien que ces résultats ne soient pas statistiquement significatifs ( $p\text{-value} > 0.05$ ). En revanche, cette concentration ne varie pas dans les plantes blessées.



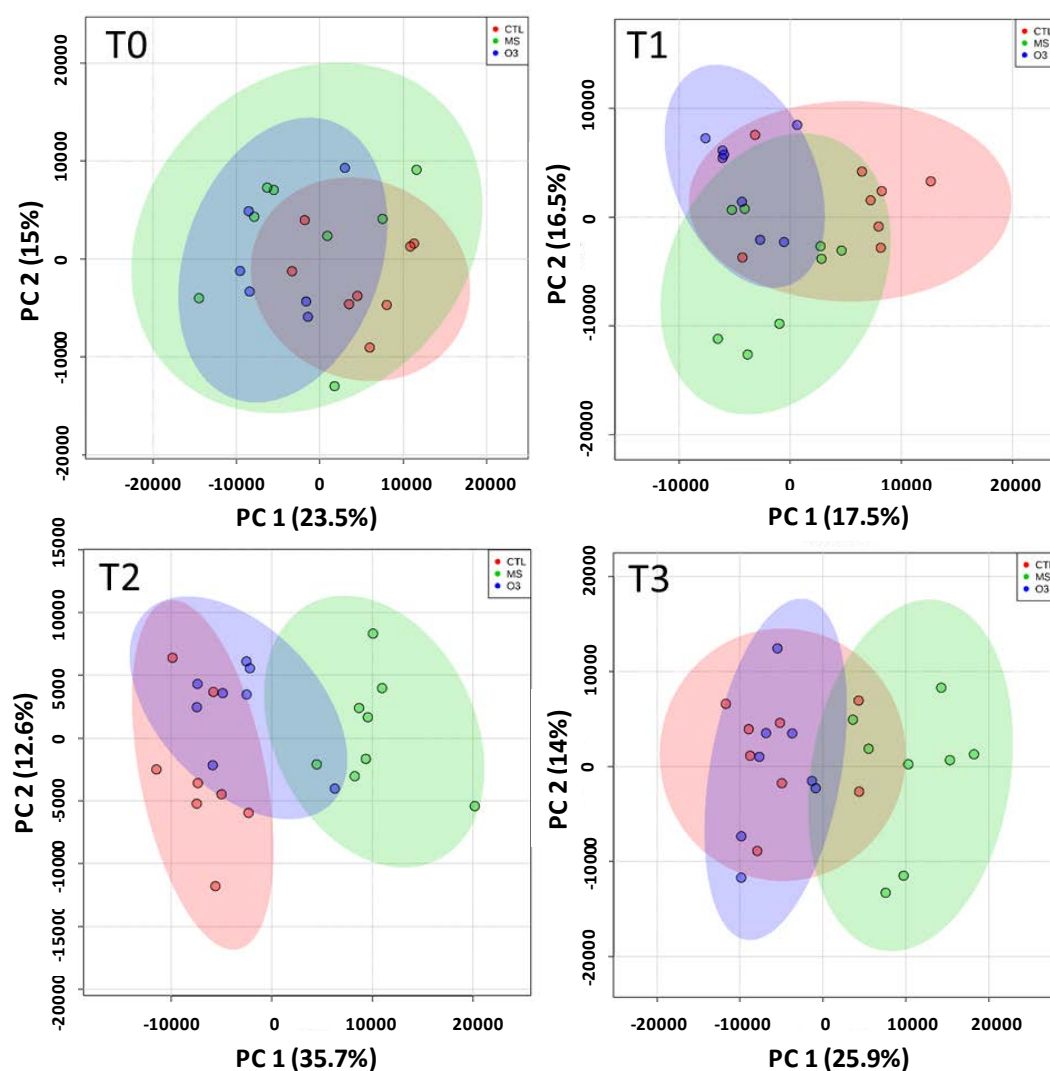
**Figure 20: Mesure de la concentration en  $\text{H}_2\text{O}_2$  dans les feuilles de panais.**  
*L'analyse a été réalisée par le Dr. Anthony Gandin 72h après l'application de la première dose d'ozone («ozone») ou des blessures mécaniques («blessures»). Des plantes contrôles ont été traitées dans les mêmes conditions («contrôle»). Les barres verticales représentent l'erreur standard. Un test de Kruskal Wallis n'ayant abouti à aucun résultat significatif ( $p\text{-value} > 0.05$  pour chaque condition) a été réalisé.*

Dans le but d'évaluer la corrélation entre la sensibilité à l'élicitation et le statut oxydatif, nous avons réalisé une analyse non ciblée du contenu métabolique de folioles de panais traités avec les deux méthodes d'élicitation citées précédemment.

## 2.2. Evolution du profil métabolique global

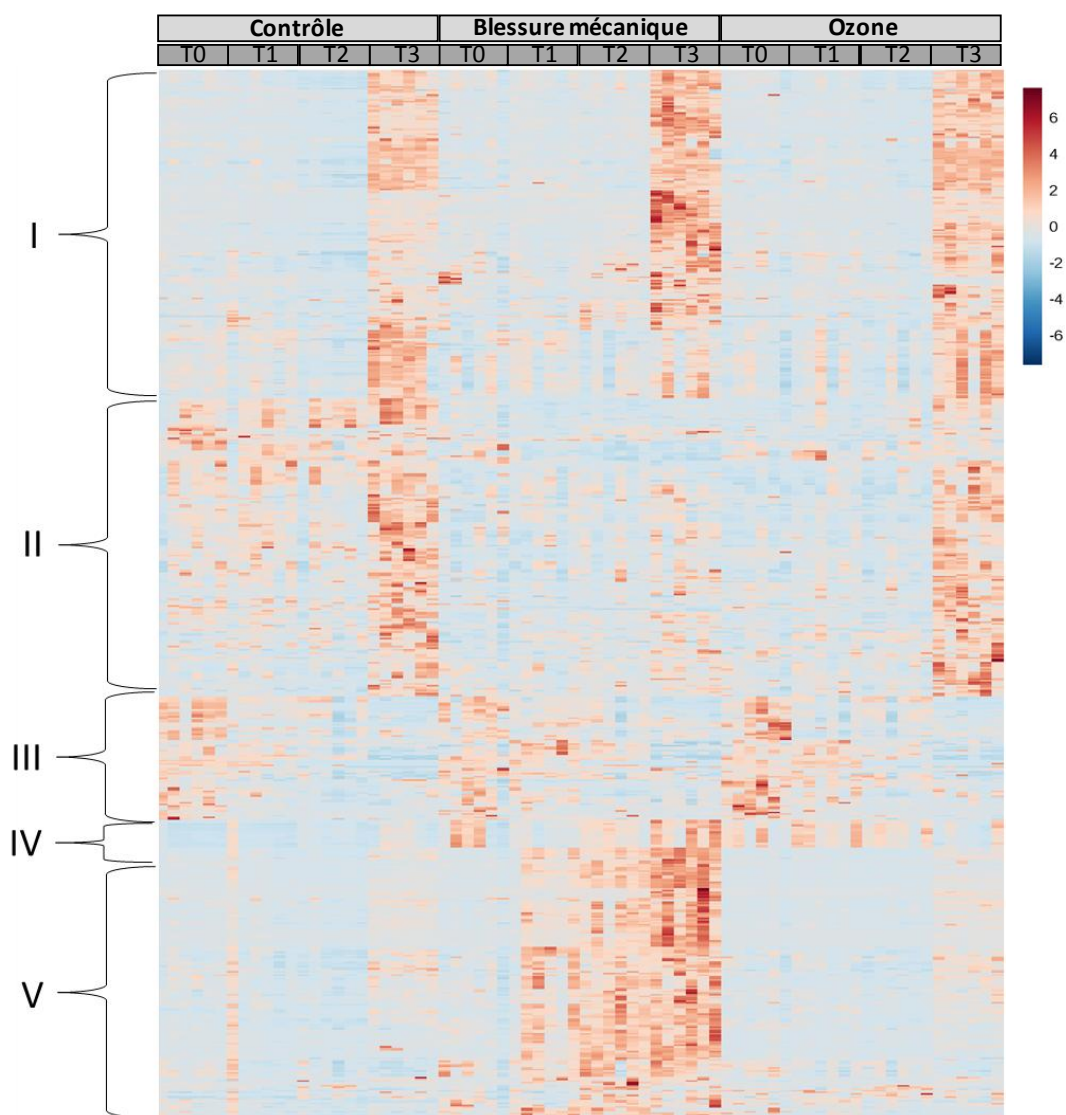
Pour déterminer les modifications globales du métabolisme du panais en réponse aux stress « Ozone » et « Blessures Mécaniques », nous avons réalisé une analyse métabolique non ciblée. Pour ce faire, nous avons utilisé des données obtenues par UHPLC couplée à de la spectrométrie de masse issues de l'analyse des échantillons hydro-méthanoliques de folioles de plantes contrôles, ou soumises aux deux éliciteurs cités précédemment après 1,2 ou 3 jours de traitement (le protocole expérimental est détaillé dans la partie Matériel et Méthodes). Les données UHPLC-MS ont été traitées par le logiciel MetaboAnalyst.

A partir de ces données, et dans un premier temps, nous avons réalisé une analyse en composante principale (ACP) permettant de visualiser la répartition des traitements les uns en fonctions des autres (Figure 21). Les résultats obtenus indiquent que les 3 traitements, qui se superposent à T0, commencent à diverger à partir de 24h post traitement. Après 48h de traitement (T2), les échantillons traités par blessures mécaniques se distinguent clairement de ceux de la condition contrôle, alors que dans le cas de l'ozone, la différence avec la condition contrôle est beaucoup moins marquée.



**Figure 21: Analyse en composante principale (ACP) à partir de données métaboliques non ciblées issues du traitement de folioles de panais dans différentes conditions.** Les analyses ont été réalisées à T0, T1 (24h), T2 (48h) et T3 (72h) après l'application des stress. Les couleurs rouge, vert et bleu correspondent respectivement à la condition contrôle, blessure mécanique et ozone (250 ppb). Pour chaque ACP, 24 échantillons ont été utilisés (8 réplicats par traitement). Le nombre de variables est de 956, 641, 655 et 755 pour T0, T1, T2 et T3 respectivement. Le nombre de composantes principales est de 8 pour chaque ACP. Le pourcentage de la variance cumulative expliquée est de 76.6%, 77.9%, 83.1% et 76.8% pour les ACP à T0, T1, T2 et T3 respectivement.

A T3, la démarcation de la condition « Blessures Mécaniques » par rapport à la condition contrôle persiste, à l'inverse de la condition « Ozone » pour laquelle très peu de différences sont encore observables. Ces premiers résultats suggèrent une réponse métabolique différente et transitoire de la plante aux différents traitements, avec un maximum atteint à T2. Une analyse plus fine a ensuite été réalisée à partir de tous les signaux métaboliques identifiés à chacun des temps de prélèvements.



**Figure 22 : Représentation en HeatMap de 646 SMDAs déterminés par une ANOVA à deux facteurs en fonction du temps et du traitement ( $p < 0.001$ ) à T0, T1 (24h), T2 (48h) et T3 (72h) après l'application des stress (ozone ou blessures mécaniques).**

La couleur rouge indique une intensité élevée, la couleur bleue une intensité faible. Chaque ligne correspond à un SMDA. Chaque colonne correspond à un échantillon. Les différents groupes de signaux métaboliques sont identifiés par les symboles sur la gauche. Les couleurs rouge, vert et bleu correspondent respectivement à la condition contrôle, blessure mécanique et ozone.



La représentation en Heatmap qui en découle permet de visualiser plus précisément le nombre de SMDAs (Signaux Métaboliques Différentiellement Accumulés) dans chacune des conditions, ainsi que leurs concentrations relatives (Figure 22). Cette analyse est restreinte aux signaux significatifs obtenus à la suite d'une ANOVA à deux facteurs ( $p < 0.001$ ). D'après les résultats de cette analyse, plusieurs groupes de signaux métaboliques ont pu être identifiés, en fonction de leurs profils d'accumulation temporels. Tout d'abord, le groupe I (Figure 22) présente des signaux métaboliques qui s'accumulent de plus en plus en fonction du temps (de T0 à T3), indépendamment de la condition considérée. Le groupe III présente la même insensibilité aux différents traitements, mais l'accumulation des signaux métaboliques est dans ce cas corrélée négativement au temps. Pour les groupes II, IV et V, des profils d'accumulations différents sont observables. En effet, les signaux métaboliques du groupe II sont sous-représentés dans les conditions « Blessure mécanique » et « Ozone », en comparaison de la condition « Contrôle ». Le groupe IV représente un petit nombre de signaux métaboliques s'accumulant préférentiellement en condition « Ozone », alors que le groupe V comporte des signaux s'accumulant uniquement dans la condition « Blessure mécanique » et progressivement avec le temps.

Pour avoir une vision encore plus fine, le nombre de SMDAs entre chacun des stress et la condition « Contrôle » a été déterminé pour les différents temps de prélèvements par le biais d'une analyse ANOVA simple ( $p < 0.05$ ). Globalement, les résultats de cette analyse (Tableau 1) indiquent que la blessure mécanique induit de plus grandes modifications métaboliques dans les feuilles de panais que l'application d'ozone. En effet, tous temps confondus, la condition « Blessure mécanique » induit la suraccumulation de 600 signaux métaboliques (« Up »), et présente 141 signaux sous-accumulés (« Down »), soit respectivement, 3 fois et 9 fois plus que la condition « Ozone ». Le nombre de SMDAs augmente dès le temps T1 dans les deux conditions, pour atteindre un maximum à T3 pour la condition « Blessures mécaniques ». Pour la condition « Ozone », la réponse de la plante semble plus furtive, avec une diminution du nombre de SMDAs dès T2.



**Tableau 1 : Nombre de SMDAs en condition « Ozone » et « Blessure mécanique » par rapport au contrôle.** Les SMDAs sont catégorisés en fonction du prélèvement (T0 à T3), du traitement, et du profil d'accumulation. Les SMDAs ont été identifiés par une ANOVA à un facteur ( $p < 0.05$ , test post-hoc de Fisher d'égalité des deux variances)

	Ozone/Contrôle		Blessures mécaniques/Contrôle	
	Down	Up	Down	Up
T0	1	1	1	0
T1	19	42	17	111
T2	19	6	41	304
T3	10	16	82	185

Les résultats obtenus dans cette étude indiquent que le métabolisme global du panais réagit différemment en fonction des stress générés et de la temporalité considérée. La blessure mécanique semble être le stress le plus impactant, notamment par la présence de nombreux signaux métaboliques différentiellement accumulés. Certaines variations métaboliques sont, dans une moindre mesure, observables dans les échantillons de plantes soumis à un stress oxydatif par l'ozone. Pour évaluer plus précisément la réponse métabolique de la plante face à ces deux stress, nous avons tenté d'identifier *in silico* plusieurs des SMDAs détectés.

## 2.3. Identification *in silico* des SMDAs

L'identification des métabolites a été réalisée à partir des SMDAs présents dans les échantillons à T1, T2 et T3 pour les trois conditions, représentant un ensemble de données de 544 signaux métaboliques non-redondants. Pour ce faire, les échantillons hydro-méthanoliques utilisés précédemment par UHPLC-MS ont été utilisés pour générer des données de fragmentation moléculaires *via* des analyses par LTQ (voir Matériel et Méthodes). A partir de ces résultats, j'ai fait un tri en éliminant les signaux métaboliques ayant une intensité de pics trop faible pour être correctement identifiés ou ayant un spectre de fragmentation non exploitable. Ce travail d'épuration a mené à 175 SMDAs exploitables.

**Tableau 2 : Résumé des métabolites putativement annotés différemment accumulés en fonction des traitements et du temps.** Les métabolites listés sont répartis en trois groupes (I, II et III). Le groupe I décrit les métabolites sous-accumulés en condition O<sub>3</sub> et Blessures Mécaniques; le groupe II comprend des métabolites sur-accumulés en condition O<sub>3</sub> et Blessures Mécaniques; le groupe III décrit les métabolites uniquement sur-accumulés en condition Blessures Mécaniques. Pour chaque métabolite identifié, le ratio entre traitement et contrôle est uniquement indiqué lorsqu'il est significatif (p<0.001).

	DAMFs	Ozone			Blessure mécanique			Poids moléculaire	Formule chimique	Nom assigné	Classe
		T1	T2	T3	T1	T2	T3				
I	M293T11	—	—	—	—	—	0.41	270.28	C16H14O4	Impératorine	Coumarines et dérivés
	M595T8	—	—	—	—	—	0.67	594.52	C27H30O15	Kaempferol 3-rhamnoside-7-glucoside	Flavonoïdes
	M611T6	—	—	—	—	—	0.48	610.52	C27H30O16	Rutine	Flavonoïdes
	M627T5	—	—	—	—	—	0.62	626.52	C27H30O17	Quercétine 3, 4-diglucoside	Flavonoïdes
	M625T8	0.55	0.61	—	—	—	0.53	624.54	C28H32O16	Myricétine 7-méthyl éther 3,4'-di-O-alpha-L-rhamnopyranoside	Flavonoïdes
	M757T6	—	0.48	—	—	0.74	0.51	756.66	C33H40O20	Quercétine-3-O-alpha-L-rhamnopyranosyl(1-2)-beta-D-glucopyranoside-7-O-alpha-L-rhamnopyranoside	Flavonoïdes
	M751T13	—	—	—	—	0.38	0.28	728.99	C43H68O9	CID 11104554	Acyls gras
	M585T16	—	0.69	—	—	0.25	0.37	584.87	C37H60O5	DG(20:5(5Z,8Z,11Z,14Z,17Z)/14:1(9Z)/0:0)	Glycerolipides
	M583T15	—	—	—	—	—	0.55	560.84	C35H60O5	DG(14:1(9Z)/18:3(9Z,12Z,15Z)/0:0)	Glycerolipides
	M769T16	—	—	—	—	0.28	0.44	747.01	C43H70O10	18:3/16:3-MGD	Glycerolipides
II	M783T15	—	—	—	—	0.48	0.37	760.99	C43H68O11	Acide oxyphytydienoïque/16:3-MGD (2S,2(1)R)-2(1),2(2)-dicarboxy-8-ethenyl-2,7,12,18-tetraméthyl-2,2(1)-dihydrobenzo[b]porphyrin-13,17-Acide dipropanoïque	Glycerolipides
	M677T8	3.04	2.20	0.44	—	2.10	1.63	676.72	C38H36N4O8		Tétrapyrroles et dérivés
	M661T9	3.95	2.64	—	—	2.38	—	660.76	C42H36N4O4	Catabolites de chlorophylle	Tétrapyrroles et dérivés
	M707T9	3.23	2.96	0.45	—	2.68	2.20	706.74	C36H42N4O11	Catabolites de chlorophylle	Tétrapyrroles et dérivés
	M693T9	3.03	2.44	—	—	2.13	—	692.71	C35H40N4O11	Catabolites de chlorophylle	Tétrapyrroles et dérivés
	M691T10	2.96	2.48	0.61	—	3.14	2.45	690.70	C35H38N4O11	Catabolites de chlorophylle	Tétrapyrroles et dérivés
	M135T11	—	—	—	3.24	6.87	4.19	134.22	C10H14	p-cymene	Lipides prénoles
	M151T11	—	—	—	2.57	3.94	2.01	150.22	C10H14O	Thymol	Lipides prénoles
	M387T6	—	—	—	1.71	2.19	1.94	364.30	C17H16O9	Xanthotoxol glucoside	Coumarines et dérivés
	M205T4	—	—	—	—	1.83	—	204.23	C11H12N2O2	Tryptophane	Indoles et dérivés
III	M197T7	—	—	—	—	—	1.81	196.24	C11H16O3	4-(3-hydroxybutyl)-2-methoxyphenol	Phenols
	M237T7	—	—	—	2.52	4.61	3.97	236.22	C12H12O5	Triméthoxycoumarine	Coumarines et dérivés
	M267T11	—	—	—	—	3.16	3.07	244.28	C15H16O3	Osthol	Coumarines et dérivés
	M335T10	—	—	—	—	8.36	6.16	334.36	C21H18O4	Anhydronotoptol (dérivés) 1 Anhydronotoptol (dérivés) 2	Coumarines et dérivés
	M335T12	—	—	—	2.25	10.19	2.49	334.36	C21H18O4		Coumarines et dérivés
	M337T11	—	—	—	2.88	6.26	3.01	336.38	C21H20O4	Anhydronotoptol	Coumarines et dérivés
	M353T10	—	—	—	3.45	17.39	11.04	352.38	C21H20O5	Lansiumarine A isomère 1	Coumarines et dérivés
	M353T11	—	—	—	—	7.87	2.30	352.38	C21H20O5	Lansiumarine A isomère 2	Coumarines et dérivés
	M339T13	—	—	—	—	1.68	2.86	338.39	C21H22O4	8-geranyloxy psoralène	Coumarines et dérivés
	M613T15	—	—	—	—	2.74	6.22	612.92	C39H64O5	DG(18:3(9Z,12Z,15Z)/18:3(9Z,12Z,15Z)/0:0)	Glycérolipides
	M277T12	—	—	—	2.53	4.27	3.35	276.41	C18H28O2	Acide stéaridonique	Acyls gras
	M279T12	—	—	—	—	6.04	5.66	278.43	C18H30O2	Acide linoléinique	Acyls gras
	M279T13	—	—	—	—	5.44	—	278.43	C18H30O2	Acide linoléinique	Acyls gras
	M283T14	—	—	—	3.61	14.30	15.02	282.46	C18H34O2	Acide oléique	Acyls gras
	M295T12	—	—	—	3.01	17.50	21.44	294.42	C18H30O3	13-HOTE	Acyls gras
	M496T14	—	—	—	2.99	4.81	2.71	495.63	C24H50NO7P	LysoPC(16:0)	Glycérophospholipides
	M518T12	—	—	—	—	—	4.90	517.63	C26H48NO7P	LysoPC(18:3)	Glycérophospholipides
	M534T12	—	—	—	3.33	7.21	5.25	533.67	C27H52NO7P	LysoPC(18:2/0:0)	Glycérophospholipides
	M536T12	—	—	—	4.03	11.08	6.48	535.65	C26H50NO8P	PC(16:1(9Z)/2:0)	Glycérophospholipides
	M791T16	—	—	—	2.01	3.82	2.93	790.06	C42H80NO10P	PS(18:0/18:1(9Z))	Glycérophospholipides

Les spectres de fragmentation des signaux métaboliques retenus ont été utilisés pour interroger différentes bases de données publiques (UNPD, LipidMAPS, HMDB, MassBank, METLIN et ReSpect), aboutissant finalement à l'identification de 39 métabolites (Tableau 2). A noter que la majorité d'entre eux sont spécifiques du stress mécanique.

Ces résultats indiquent que les différents métabolites identifiés appartiennent à 9 classes majeures de molécules : les flavonoïdes, les acyls gras, les coumarines (et dérivés), les phénols, les indols (et dérivés), les tétrapyrroles (et dérivés), les lipides prénoles, les glycérolipides et les glycérophospholipides. En fonction de leurs conditions d'accumulation (Ozone ou Blessures mécaniques), les métabolites identifiés peuvent être regroupés en 3 classes (I, II et III, Tableau 2).

Le premier groupe (I) réunit les métabolites dont l'accumulation est fortement réduite dans les deux conditions de stress, par rapport à la condition « Contrôle ». Parmi ces métabolites sont retrouvés les flavonoïdes ainsi que certains acyles gras et glycérolipides. Dans le cas des blessures mécaniques, ces métabolites sont fortement réduits à T2 et T3. Dans le cas de l'ozone, seule l'accumulation d'un petit nombre de métabolites est réduite à partir de T1 jusqu'à T2.

Le deuxième groupe (II) inclut des métabolites fortement accumulés sous stress ozone à T1 et T2 et dans le cas des blessures mécaniques à T2 et T3. Parmi ces métabolites, un SMDA est identifié comme un dérivé de tétrapyrrole, basé sur sa formule chimique et son profil de fragmentation (SMDA M677T8, tableau 2). Cette famille de métabolites, également nommée phyllobiline ou encore « catabolites de chlorophylle non-fluorescent » (NCC), a été découverte dans les feuilles sénescentes ou les fruits en maturation (Kräutler, 2008; Kräutler et Hörtensteiner, 2006).

Enfin, le groupe III contient des métabolites dont l'accumulation est induite exclusivement par la blessure mécanique. Ce groupe inclut deux métabolites identifiés comme des monoterpènes ainsi que plusieurs coumarines et furocoumarines. De plus, certains composés du métabolisme primaire, comme le tryptophane, des acides gras et des glycérophospholipides, sont également accumulés dans ces conditions. Globalement, l'accumulation de ces métabolites est induite dès T1 pour atteindre un maximum à T2, suivi d'une légère diminution jusqu'à T3.

En résumé, les blessures mécaniques et l’ozone induisent de nombreuses modifications métaboliques chez le panais, tant au niveau du métabolisme primaire que secondaire. Cependant, l’application de blessures mécaniques semble induire la réponse la plus forte et la plus rapide dans les folioles de panais. Les résultats de cette étude confirment que les blessures mécaniques permettent d’induire la synthèse de molécules de défense, telles que les coumarines et les furocoumarines. Le travail réalisé dans le cadre de ce premier chapitre a fait l’objet d’une publication dans la revue *Metabolites* (cf annexe 1)

### 3. Discussion et Conclusion

Les résultats des travaux présentés dans ce chapitre ont permis de déterminer les modifications métaboliques globales induites, chez le panais, par deux stress : l’ozone et la blessure mécanique.

**Détermination de la dose d’ozone sub-létale** - Dans un premier temps, j’ai tenté d’évaluer la dose d’ozone à partir de laquelle des symptômes d’oxydation (chlorose, nécrose) pouvaient apparaître sur des feuilles de panais. Les résultats obtenus montrent que même à une dose 25 fois supérieure à la concentration moyenne atmosphérique (250ppb) aucun symptôme n’apparaît. Des études similaires ont été menées chez *Melissa officinalis* (Lamiacées) et *Betula pendula* (Bétulacées) avec des concentrations en ozone inférieures à 250 ppb (respectivement 200 ppb et 150 ppb). D’après les résultats de ces travaux, de nombreux signes d’oxydation du tissu végétale, notamment des chloroses et des nécroses, apparaissent après 48h d’application d’ozone (Pellegrini *et al.*, 2011; Tuomainen *et al.*, 1996). Bien que ces deux plantes n’appartiennent pas à la même famille botanique que le panais, ces résultats tendent à conférer au panais une certaine résistance à l’oxydation.

**Evaluation du statut oxydatif** – L’analyse de la concentration en peroxyde d’hydrogène est généralement un bon indicateur pour mesurer l’induction de la réponse à un stress par la plante.

Bien qu'ils ne soient pas significatifs, les résultats que j'ai obtenus indiquent que les folioles traitées à l'ozone présentent une plus forte concentration en  $H_2O_2$  que les plantes blessées. Ils indiquent que si la plante ne présente pas de signe extérieur d'oxydation des tissus, celle-ci répond tout de même au stress de manière significative. Ces résultats peuvent trouver plusieurs explications. Une fois assimilé par la plante, l'ozone peut rapidement être dégradé d'abord en dioxygène et en radical hydroperoxyde ( $HO_2\cdot$ ) puis, par des réactions en chaîne, en peroxyde d'hydrogène (Cremasco and Mochi, 2013). L'accumulation de peroxyde d'hydrogène n'est donc pas une réelle surprise. Pour ce qui concerne les réponses aux blessures, l'explication pourrait être plus complexe. Le panais est une plante riche en composés antioxydants, tels que les flavonoïdes, les coumarines ou l'acide chlorogénique (Reyes *et al.*, 2007; Nikolic *et al.*, 2014). Dans le cas de la condition « blessure », l'application du stress a été réalisée uniquement le premier jour. Il est de fait envisageable que la quantité de peroxyde d'hydrogène ait diminué au cours des 3 jours d'incubation, via des mécanismes de réponse au stress oxydatif de la plante.

**Evolution du profil métabolique global et identification des SMDAs** – L'analyse métabolique globale nous a permis d'identifier 39 métabolites répartis en 3 classes (Tableau 2). Dans le cas des flavonoïdes (Classe I), la diminution de concentration en condition de stress ozone par rapport au contrôle peut être reliée au caractère antioxydant de ces molécules. Ainsi, nous pouvons émettre l'hypothèse que les flavonoïdes permettent de réduire l'oxydation causée par les ROS, produits par l'ozone, ce qui entraîne une diminution de leur concentration. Au-delà de la production de ROS, l'ozone induit également une augmentation de NCC (Nonfluorescent Chlorophyll Catabolites, tétrapyrrole, classe II). Cette augmentation peut être interprétée comme une conséquence de la dégradation de la chlorophylle en réponse à l'ozone, bien qu'aucun symptôme de chloroses n'ait été observé (cf. paragraphe 2.1.1.). Le rôle des NCC chez les plantes n'est à ce jour pas clairement identifié, même s'il semblerait que certains d'entre eux possèdent une forte activité antioxydante (Müller *et al.*, 2007). L'accumulation de NCC en condition « Ozone » pourrait alors participer chez le panais à la résistance à l'ozone. Dans le cas du stress mécanique, la concentration des flavonoïdes est également statistiquement inférieure à la condition contrôle.

Comme cela a déjà été décrit dans la partie bibliographique, les furocoumarines et les flavonoïdes sont des composés phénoliques issus de la voie des phénylpropanoïdes, initiée par la conversion de la phénylalanine en acide trans-cinnamique par une phénylalanine ammonia-lyase (PAL). Ainsi, ces deux familles possèdent des précurseurs communs, notamment le *p*-coumaroylCoA. Les furocoumarines sont décrites comme inductibles à la blessure mécanique (Krieger *et al.*, 2018; Roselli *et al.*, 2017; Zangerl, 1990; Zangerl and Berenbaum, 1998, 2005). Nous pouvons émettre l'hypothèse d'une compétition entre les flavonoïdes et les furocoumarines pour l'utilisation du *p*-coumaroylCoA. Un tel exemple de compétition métabolique a déjà été décrit chez *Arabidopsis thaliana* génétiquement modifié pour exprimer un gène bactérien bifonctionnel codant une chorismate mutase/phrephenate deshydratase. Ce gène intervient dans un mécanisme du métabolisme primaire et entre en compétition avec la voie du shikimate et des phénylpropanoïdes pour la disponibilité en phénylalanine (Tzin *et al.*, 2009). La suraccumulation de furocoumarines (osthol, 8-geranyloxypsoralène, xanthotoxol glucoside), et de coumarines (anhydronotoptol, lansiumarin A) en condition de blessures mécaniques tend à valider cette hypothèse. La blessure mécanique engendre également l'accumulation statistiquement significative de deux monoterpènes (thymol et *p*-cymène, Classe III). Cette accumulation s'opère 24 à 48h après le stress. La surproduction de monoterpènes après blessures mécaniques a déjà été décrite chez *Myrica cerifera* et *Artemisia annua* comme impliquée dans la réponse, directe ou indirecte, de la plante face à des contraintes mécaniques (Ruan *et al.*, 2016; Wang et Lincoln, 2004). Certaines familles de lipides sont également suraccumulées après blessures mécaniques, notamment les acyl gras et les glycérophospholipides. D'autres, comme les glycérolipides sont plutôt sous-représentées dans ces conditions (Classe I). Ces modifications du profil lipidique sont concordantes avec des mécanismes de réponse à la blessure mécanique décrits dans la littérature (Hou *et al.*, 2016; Lee *et al.*, 1997). Ainsi, l'accumulation de lysophospholipides (LysoPC (16 :0), lysoPC (18 :3), lysoPC(18 :2) tableau 1) et d'acides gras (Acide stéaridonique, acide linoléique, acide oléique, tableau 1) peut être la conséquence de l'activation de phospholipases, intervenant dans la formation de phospholipides à partir du catabolisme d'acides gras (Ryu and Wang, 1996).

De plus, la présence d'acides gras libres globalement insaturés pourrait être reliée à l'activation d'acides gras désaturases, comme décrit chez la tomate (Zhang *et al.*, 2018). L'accumulation de 13-HOTE, produit *via* l'intervention de lipoxygénase au travers de l'oxydation d'acide linoléique, peut être reliée à l'induction de la voie de biosynthèse des oxylipines. Cette voie conduit notamment à la synthèse d'acide jasmonique, une phytohormone intervenant dans la réponse de la plante aux stress biotiques et abiotiques (Marchica *et al.*, 2019; Savatin *et al.*, 2014). Enfin, la diminution de glycérolipides, et notamment de diacylglycérols et de monogalactosyldiacylglycerols, peut être également liée à l'activation de phospholipases, comme décrit précédemment, ou à la dégradation de structures plastidiques, riches en glycérolipides. Une telle modification du métabolisme des lipides a déjà été décrit chez dans le cas d'*Arabidopsis thaliana* soumise à des blessures mécaniques (Vu *et al.*, 2014, 2015).

**Dans cette étude, nous avons mis en évidence la flexibilité métabolique de la réponse de la plante en fonction du stress appliqué. Bien que les mécanismes impliqués dans cette réponse soient extrêmement complexes, l'analyse métabolomique globale a permis de mieux les comprendre et de mettre en évidence une réponse originale du panais aux stress.**

## **B. Analyse fine du profil métabolique : cas des furocoumarines**

Pour aller plus loin dans la description de la réponse de la plante au stress, de nouvelles analyses ont été conduites en appliquant des blessures mécaniques sur des folioles, les tiges et les racines. Pour éviter d'avoir un effet systémique qui résulterait du transport de molécules entre les différents organes, ceux-ci ont préalablement été détachés de la plante avant les blessures, puis placés en atmosphère humide, à température ambiante, pendant 24h.

### **1. Comparaison inter-organes en condition contrôle**

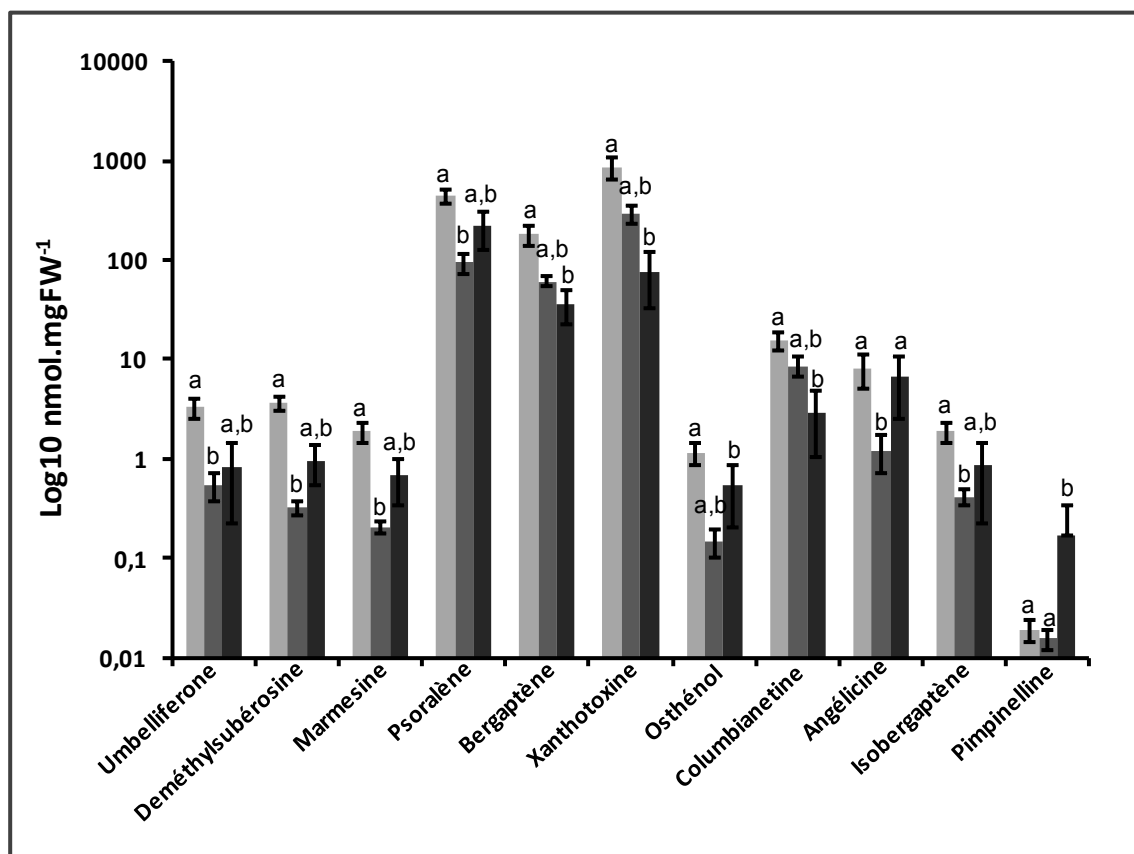
#### **1.1. Analyse phytochimique des folioles, des tiges et des racines.**

Dans un premier temps, la concentration en furocoumarines a été analysée par UHPLC-MS à partir d'extraits réalisés sur des folioles, des tiges et des racines de panais en condition contrôle permettant ainsi de déterminer les différences de concentration basale de ces composés dans la plante.

Les analyses statistiques (Figure 23) mettent en évidence que le profil d'accumulation des furocoumarines est très variable en fonction du composé considéré, mais également en fonction de l'organe. Globalement, et à l'exception de la pimpinelline, la concentration en furocoumarines est plus élevée dans les feuilles que dans les autres organes. Parmi les 11 composés analysés, 64 % sont moins abondants dans les tiges, que dans les folioles et dans les racines (umbelliférone, DMS, marmesine, psoralène, osthénol, angélicine, isobergaptène). Les racines, accumulent moins de bergaptène, de xanthotoxine et de columbianétine que les autres organes.



En revanche, la pimpinelline est beaucoup plus accumulée dans les racines que dans les autres organes. Tout organe confondu, le psoralène, le bergaptène et la xanthotoxine sont les trois furocoumarines les plus abondantes.



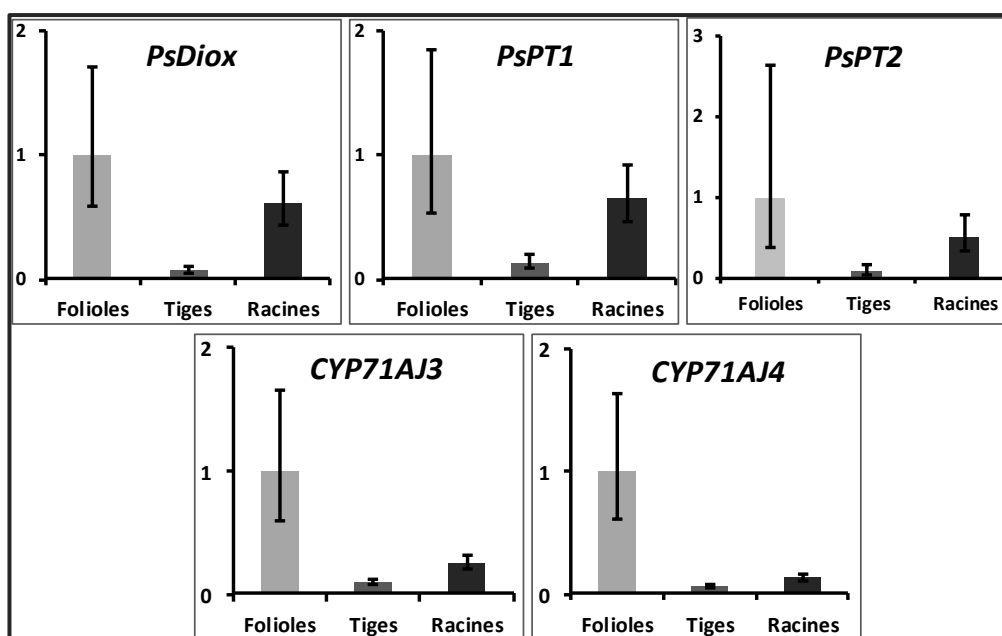
**Figure 23: Mesure de la concentration en furocoumarines dans les folioles (gris clair), tiges (gris) et racines (gris foncé) de panais en condition contrôle.**  
 Les barres verticales représentent l'erreur standard. Les valeurs représentées par deux lettres identiques ne sont pas statistiquement significatives (test Kruskal-Wallis,  $p$ value < 0.05). L'échelle en ordonnée est représentée en log10.

## 1.2. Analyse du niveau d'expression de gènes impliqués dans la voie de biosynthèse des furocoumarines en conditions contrôle.

Pour établir un lien entre la concentration des différentes furocoumarines et l'expression de certains gènes, nous avons analysé par PCR quantitative (Figure 24) le niveau d'expression de *PsDiox*, *PsPT1*, *PsPT2*, *CYP71AJ3* et *CYP71AJ4*.

Pour rappel, ces 5 gènes catalysent respectivement la formation d'umbelliférone, de demethylsuberosine, d'osthénol, de psoralène et d'angélicine. Cette analyse a été réalisée à partir d'ARNs extraits de folioles, de tiges et de racines .

Les résultats obtenus confirment une relation étroite entre l'expression des gènes et l'accumulation des produits associés Dans le cas des 5 gènes cibles, la plus faible expression est retrouvée dans la tige et coïncide avec des résultats décrits précédemment (Figure 23). Le niveau d'expression le plus élevé est constaté dans les folioles, bien que d'après l'erreur standard, l'expression dans les folioles présente une grande variabilité. A l'exception de *CYP71AJ3* et *CYP71AJ4*, la différence d'expression entre folioles et racines n'est que très peu marquée.



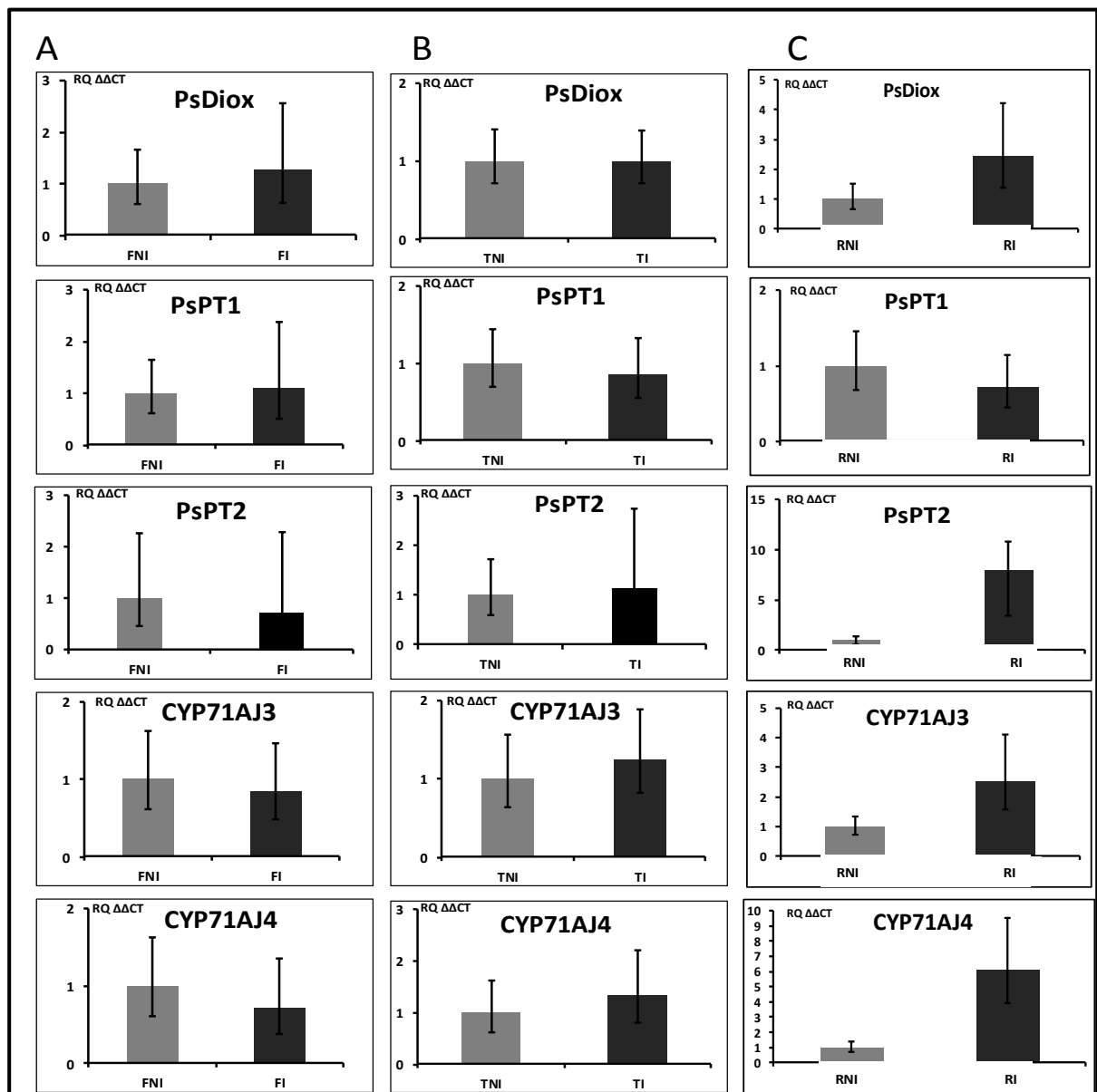
**Figure 24: Analyse du niveau d'expression des gènes *PsDiox*, *PsPT1*, *PsPT2*, *CYP71AJ3* et *CYP71AJ4* dans les folioles, tiges et racines de panais en condition contrôle par PCR quantitative. Le gène codant l'ARN 5.8s de panais a été utilisé comme gène de ménage. Les barres verticales représentent l'erreur standard.**

## **2. Etude de l'induction de la voie de biosynthèse en fonction des organes**

Pour analyser l'impact de la blessure sur la production de furocoumarines chez le panais à l'échelle de la plante, j'ai analysé en parallèle l'expression de 5 gènes connus de la voie des furocoumarines et la production de 11 de ces molécules dans 3 tissus différents : les racines, les feuilles et les tiges.

### **2.1. Expression différentielle des gènes spécifiques de la voie de biosynthèse dans différents tissus en réponse à des blessures**

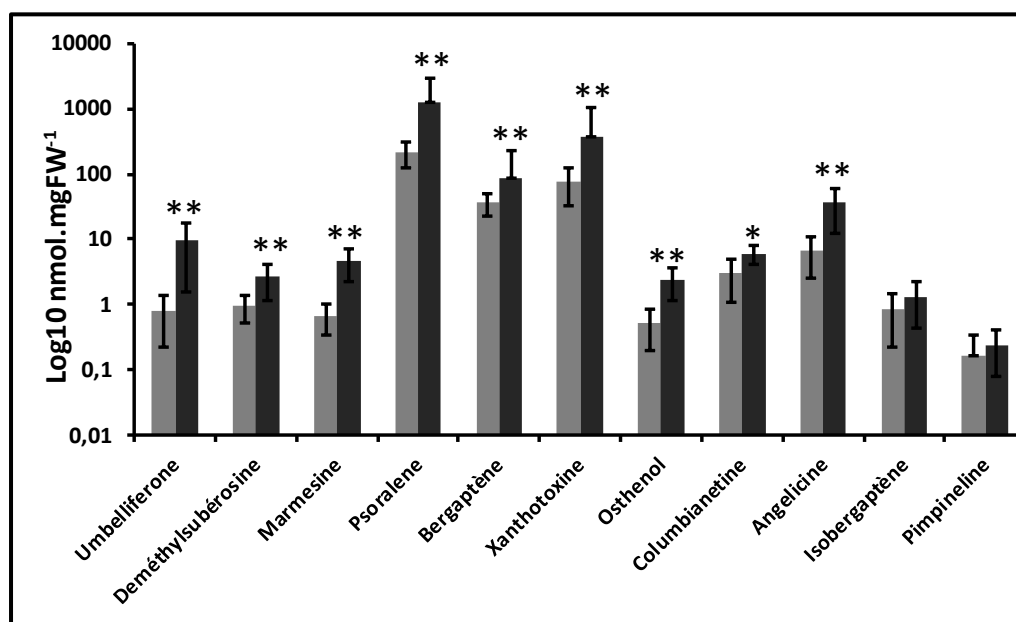
Les résultats obtenus démontrent que les 5 gènes cibles étudiés (*PsDiox*, *CYP71AJ3*, *CYP71AJ4*, *PsPT1*, *PsPT2*) ne sont pas surexprimés dans les folioles et les tiges en comparaison avec les plantes contrôles (Figure 25). En revanche, dans les racines de panais, le niveau d'expression de *PsDiox* et de *CYP71AJ3* est multiplié par 2.5 (Figure 25) et pour *CYP71AJ4* et *PsPT2*, l'expression est multipliée par 6 et 7, respectivement.



**Figure 25: Analyse en PCR quantitative de l'expression de des gènes *PsDiox*, *PsPT1*, *PsPT2*, *CYP71AJ3* et *CYP71AJ4* dans les folioles, tiges et racines de panais en condition contrôle (respectivement FNI, TNI et RNI, gris) et après l'application de blessure mécanique (respectivement FI, TI et RI, noir). Le gène codant l'ARN 5.8s de panais a été utilisé comme gène de ménage. Les barres verticales représentent l'erreur standard.**

## 2.2. Analyse différentielle de la production de furocoumarines dans les racines

La plus forte expression des gènes étudiés est retrouvée dans les racines (Figure 25). J'ai donc choisi de me focaliser uniquement sur cet organe pour la suite. Le stress mécanique induit une augmentation générale des concentrations en furocoumarines dans les racines après 24h. Cette induction est significative pour toutes les molécules à l'exception de l'isobergaptène et de la pimpinelline (Figure 26). Dans le cas de l'umbelliférone, le psoralène, l'angélicine et l'osthéol, la suraccumulation en condition d'éllicitation est corrélée avec la surexpression des gènes impliqués dans leurs synthèses (Figure 26), respectivement *PsDiox*, *CYP71AJ3*, *CYP71AJ4* et *PsPT2*. En revanche, on constate une accumulation de la DMS qui est déconnectée de la transcription de *PsPT1*.



**Figure 26: Mesures de la concentration en furocoumarines dans des racines de panais en condition contrôle (gris) et après l'application de blessures mécaniques (noir).** Les barres verticales représentent l'erreur standard. Les étoiles représentent la significativité des différences entre les conditions « contrôle » et « blessure mécanique » selon le test de Wilcoxon (\*renvoie à une p-value < 0.05, \*\* à une p-value < 0.01)

Pour résumer, les blessures mécaniques induisent la synthèse de chaque furocoumarines avec une efficacité variable. Elles semblent être le plus efficace au niveau de la racine. Par ailleurs, je confirme que cette suraccumulation après des blessures mécaniques des racines est globalement corrélée avec l'expression des gènes connus de la voie métabolique de ces molécules. Les résultats obtenus dans ce chapitre ont été valorisés dans 2 articles, Roselli *et al* et Krieger *et al*. (cf annexes 2 et 3)

### 3. Discussion et conclusion

Pour se protéger contre l'agression par un certain nombre de prédateurs, les plantes ont développé un arsenal de molécules chimiques dont font partis les furocoumarines. Si ce métabolisme de défense est souvent inductible, pour des raisons de coût métabolique, dans certains cas, les molécules toxiques peuvent également être produites de manière basale pour assurer une défense immédiate. L'analyse réalisée chez le panais montre que c'est le cas des furocoumarines. Ainsi l'analyse réalisée en conditions contrôle a montré la présence de différentes molécules de cette famille dans tous les tissus. Néanmoins, cette analyse montre aussi que le psoralène, le bergaptène et la xanthotoxine sont présents en quantité plus importante. Cette différence de concentration peut s'expliquer par le fait que ces trois composés sont les furocoumarines de la voie les plus actifs biologiquement, par leur phototoxicité accrue et leur capacité à générer des mono-adduits d'ADN. Une concentration plus élevée de ces composés peut alors permettre à la plante de se défendre plus efficacement contre des bioagresseurs.

L'analyse différentielle réalisée sur les 3 organes cibles indique que, à l'exception de la pimpinelline, la concentration en furocoumarines est la plus forte dans les folioles. Ce résultat est cohérent avec le fait que, lors d'une attaque d'herbivores, les folioles du panais sont préférentiellement sujettes à la prédation, contrairement à la tige, dans laquelle la concentration en furocoumarines est globalement la plus faible. La racine, en revanche, peut être la cible de microorganismes telluriques.

Ce risque peut alors expliquer la forte concentration en pimpinelline, en comparaison avec les autres organes. En effet, plusieurs études ont démontré la forte activité antimicrobienne et antifongique de la pimpinelline, notamment envers certains bacilles, mycobactéries et levures (O'Neill *et al.*, 2013; Walasek *et al.*, 2015). De telles variations d'accumulation de furocoumarines inter-organes ont également été démontrées chez *Ruta graveolens* (Rutacées), également accumulatrice de furocoumarines (Milesi *et al.*, 2001)

Les analyses du niveau d'expression des gènes impliqués dans la synthèse des furocoumarines est cohérente avec la production des molécules. Ainsi, la plus faible expression est retrouvée dans les tiges et la plus importante dans les folioles. On constate cependant une très grande variabilité de cette expression qui peut être reliée à la différence d'âge entre différentes folioles d'une même tige. En effet, l'analyse transcriptionnelle a été réalisée à partir d'un pool de folioles élicitées appartenant à une même tige. Chez le panais, les folioles les plus apicales d'une même tige sont plus âgées que les folioles proximales. Ce résultat va dans le sens de certaines études démontrant que l'activité transcriptionnelle d'une foliole est dépendante de son âge. Ce phénomène a notamment été constaté chez le soja, dans le cas de gènes intervenant dans la synthèse de glyceolline, impliquée dans la défense contre *Phytophthora megasperma* (Bhattacharyya and Ward, 1986). De la même façon, le niveau d'expression de plusieurs gènes impliqués dans le transport du fer chez *Arachis hypogaea* varie fortement en fonction de l'âge du foliole (Akhtar *et al.*, 2015).

Pour finir, la DMS est suraccumulée alors que *PsPT1*, principalement impliquée dans la synthèse de ce composé, n'est pas significativement surexprimée dans les mêmes conditions (cf Figure 26). Chez le panais, les prényltransférases *PsPT1* et *PsPT2* initient préférentiellement la synthèse de furocoumarines linéaires et angulaires, respectivement, à partir de l'umbelliférone (Munakata *et al.*, 2016), bien que chacune soit capable de catalyser les deux réactions. Ainsi, nous pouvons émettre l'hypothèse que *PsPT2*, surexprimée après un stress mécanique, synthétise de l'osthénol mais également de la DMS dans ces conditions d'élicitation. Ces différentes études indiquent clairement que les furocoumarines sont produites de manière constitutive par les panais. Néanmoins, leur synthèse peut également être induite.

Pour déterminer quels sont les tissus les plus inductibles, j'ai analysé le niveau d'expression des gènes spécifiques à la voie de biosynthèse dans différents organes.

Les résultats obtenus montrent clairement une forte surexpression dans les racines. Ce type de résultat a déjà été décrit dans la littérature. En effet, les travaux de Zangerl indiquent que le niveau d'accumulation des furocoumarines, notamment la xanthotoxine, dans les racines de panais est bien supérieure à celui des folioles ou des tiges après une blessure mécanique.

Ces travaux indiquent également que, dans notre cas, la sensibilité à l'éllicitation des organes étudiés est inversement proportionnelle à la teneur basale en furocoumarines. En effet, la plus grande concentration en furocoumarines est retrouvée dans les folioles de panais qui sont pourtant les moins inductibles. Par ailleurs, les racines sont les organes les moins concentrés en furocoumarines mais possèdent le plus fort niveau d'induction. Ce phénomène peut être expliqué par la théorie de la défense optimale (ODT). L'ODT (Kaplan *et al.*, 2008) a été formulée pour expliquer les implications évolutives et environnementales conduisant à des différences de réponse à un stress biotique, en l'occurrence l'herbivorie, au sein d'une même plante. Cette théorie est basée sur un principe définissant les défenses moléculaires comme des processus très coûteux pour la plante. Ainsi, les molécules de défense sont préférentiellement synthétisées dans des organes de la plante ayant le plus de risque d'être attaqués par des herbivores, comme les feuilles, et les plus importants à la pérennisation de son matériel génétique comme les graines. Par exemple, les jeunes feuilles sont plus appétentes et donc plus riches en métabolites secondaires que les feuilles plus âgées (Ballhorn *et al.*, 2009). A l'inverse, les tiges, très peu concentrées en furocoumarines, ne sont pas des cibles privilégiées des herbivores. Ainsi, une faible inductibilité de la synthèse de molécules de défense telles que les furocoumarines après un stress mécanique est naturellement corrélée avec une forte concentration. En effet, étant donné que la synthèse de molécules de défense utilise les mêmes ressources que le développement de la plante, une forte induction de la synthèse de furocoumarines dans les feuilles, déjà très concentrées, risque de fortement impacter le métabolisme de la plante (Barto et Cipollini, 2005). De plus, la toxicité avérée des furocoumarines pourrait également être néfaste pour la plante.



Dans le cas des analyses phytochimiques différentielles dans la racine, cette théorie peut également expliquer les profils d'accumulation de la pimpinelline et de l'isobergaptène très peu sensibles à la blessure mécanique mais présentant une des plus fortes concentrations dans la racine.

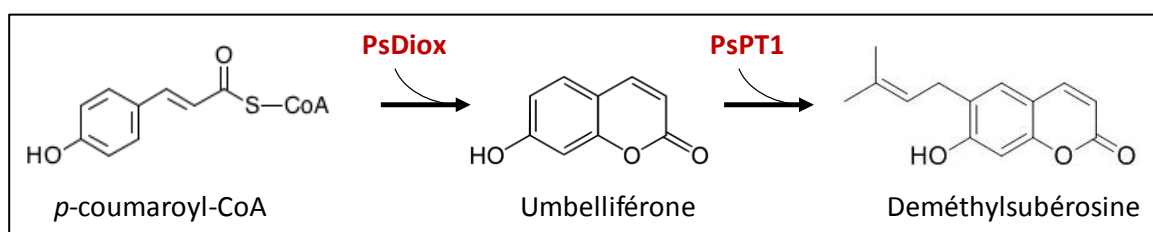
Les résultats des travaux d'élicitation de la synthèse de furocoumarines par blessures mécaniques ont été utilisés pour mettre en place un outil d'expression différentielle permettant la sélection de gènes candidats (cf partie D).

## C. Ingénierie métabolique : transposition partielle de la voie de biosynthèse des furocoumarines chez le tabac

### 1. Introduction

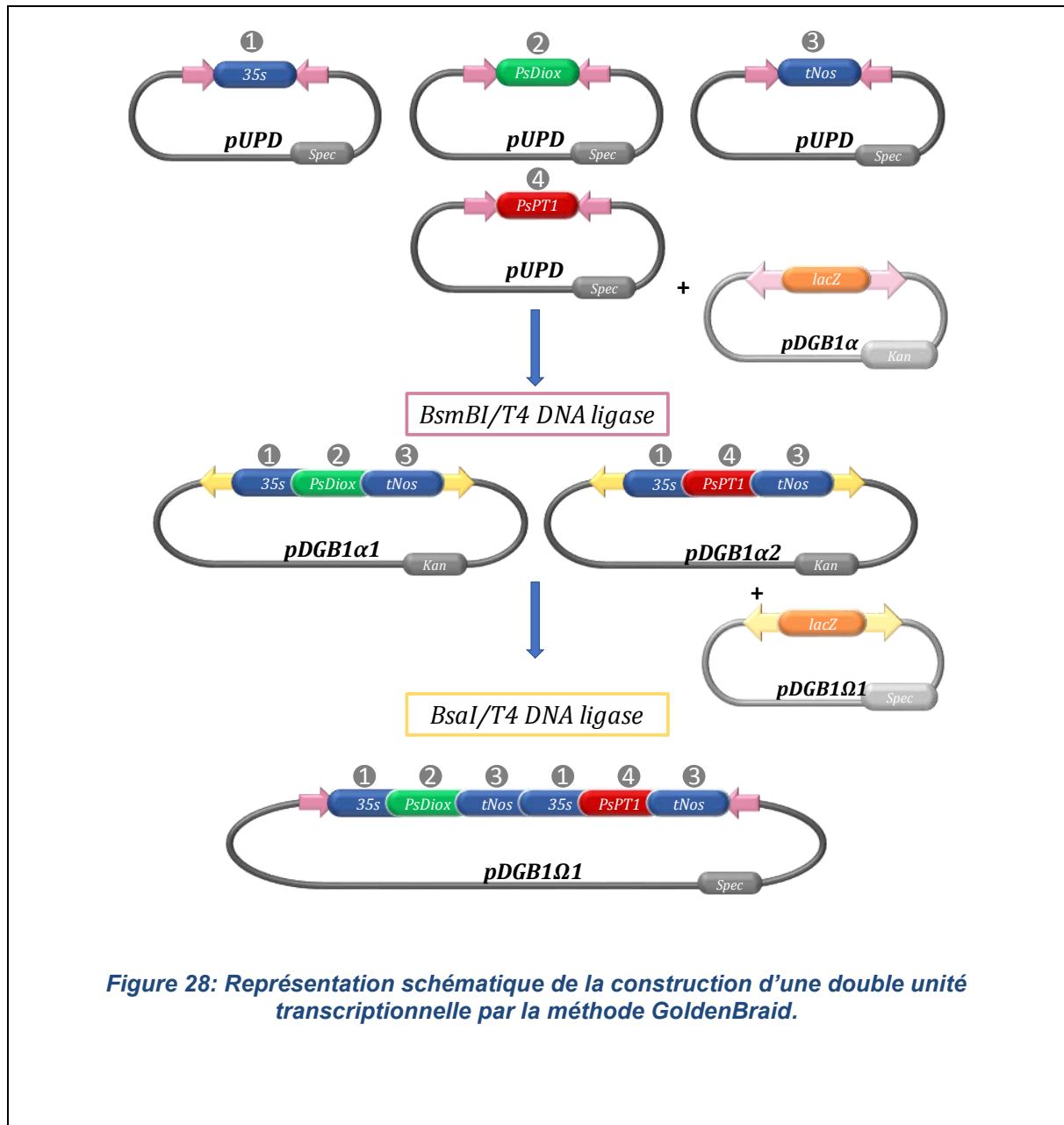
La théorie de la « Balance Croissance/Défense » repose sur le fait que la synthèse de molécules de défense par la plante requiert les mêmes ressources carbonées et énergétiques que le métabolisme primaire (croissance, développement, reproduction) et influe donc inéluctablement sur la physiologie de la plante. Pour caractériser les impacts métaboliques, physiologiques voire phénotypiques induites par la mise en place d'une telle voie, nous avons choisi de générer un système permettant la synthèse de molécules de défense, ici les furocoumarines, par un organisme qui en est naturellement dépourvu, le tabac.

La synthèse des furocoumarines est organisée en voie métabolique, pour laquelle plusieurs étapes ont été caractérisées au niveau moléculaire. A partir de ces connaissances, nous avons choisi, dans un premier temps, de nous focaliser sur les deux premiers gènes intervenant dans cette voie chez le panais, *PsDiox* et *PsPT1*, respectivement impliqués dans la synthèse d'umbelliférone et de DMS (Figure 27), afin de les intégrer dans le génome du tabac (*Nicotiana tabacum*).



**Figure 27: Production de deméthylsubérosine à partir de p-coumaroyl-CoA par l'intervention de PsDiox et PsPT1.**

Pour permettre l'expression de ces deux gènes et une synthèse séquentielle des protéines correspondantes, nous avons opté pour l'utilisation du système de clonage GoldenBraid (Figure 28) permettant, par le biais d'une seule transformation, d'insérer simultanément dans le génome d'une plante hôte plusieurs séquences codantes au sein d'un même vecteur d'expression (Sarrion-Perdigones *et al.*, 2011).



La méthode GoldenBraid est basée sur des réactions successives de digestion, de restriction et de ligation permettant d'intégrer plusieurs cassettes d'expression dans un seul vecteur. Les enzymes de restriction utilisées dans cette méthode, *Bsmbl* et *Bsal*, sont de classe IIs et disposent d'un site de reconnaissance et d'un site de clivage différents.

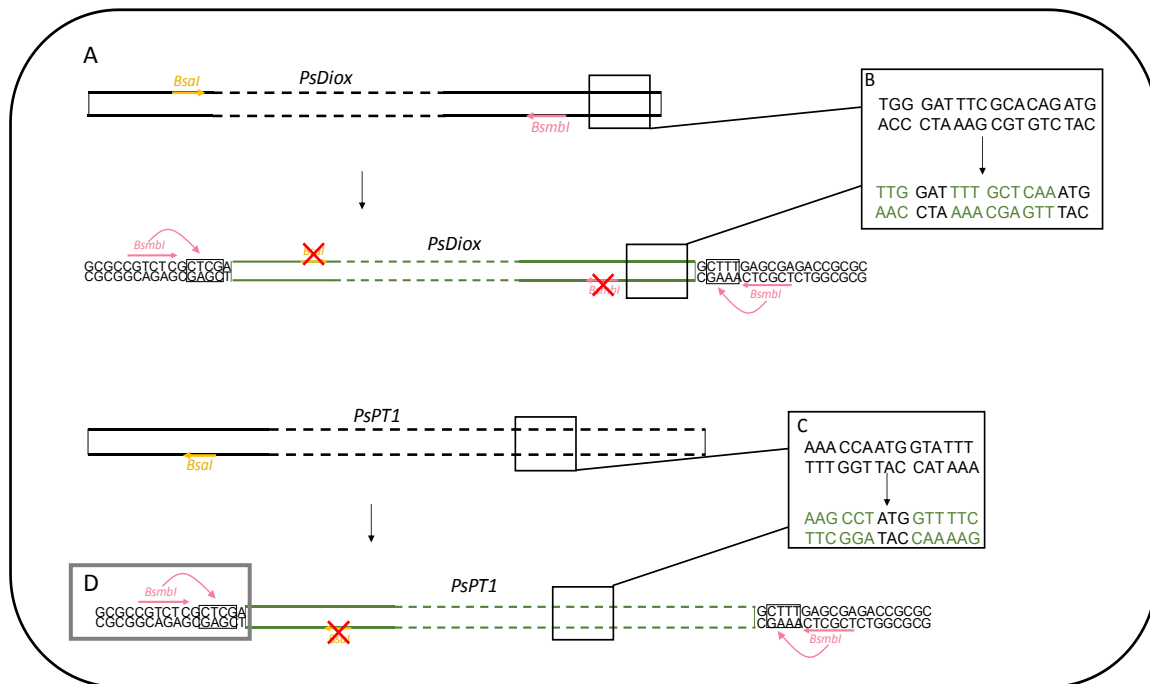
Cette dissociation permet de générer des extrémités cohésives de 4 nucléotides dont la séquence peut être décidée par l'expérimentateur lors de la conception des cassettes d'expression. Des sites de fixation et de restriction sont alors ajoutés de chaque côté de la séquence à cloner. L'ajout de ces sites est appelé la domestication.

## **2. Construction des vecteurs GoldenBraid**

### **2.1. La domestication**

La domestication consiste à générer des constructions nucléotidiques composées de la séquence à cloner flanquée des sites de fixation et de restriction de *Bsmbl*, la première enzyme utilisée dans le système GoldenBraid. Dans le cadre de ce projet de recherche, les séquences codantes de *PsDiox* et *PsPT1* ont été domestiquées par la société Eurofins. La domestication est composée de trois étapes. Dans un premier temps, les sites de restrictions *Bsal* et *Bsmbl* internes aux séquences codantes ont été retirés et remplacés par des codons synonymes, c'est à dire des codons différents qui désignent le même acide aminé (Figure 29 A). Etant donné que les gènes *PsDiox* et *PsPT1* sont destinés à être exprimés dans le tabac, une optimisation de la séquence de ces deux gènes pour la production de protéine dans *Nicotiana* a également été réalisée par Eurofins (Figure 29 B et C) (Annexe 6). Cette optimisation se traduit par l'introduction de mutations silencieuses *via* des codons synonymes des codons originels mais préférentiellement utilisés chez la plante hôte (Tableau 3).

Enfin, des adaptateurs préfixes et suffixes, respectivement de 16 et 19 nucléotides sont insérés en amont et en aval de la séquence codante (Figure 29 D). Ces adaptateurs permettront ultérieurement aux enzymes *Bsal* et *Bsmbl* de se fixer et de libérer par clivage les séquences à cloner.



**Figure 29: Domestication des séquences codantes de *PsDiox* et *PsPT1*.**

A. La domestication consiste en la suppression des sites de restrictions *Bsal* et *Bsmbl* (symbolisée par les croix rouges) internes aux séquences d'intérêt, et à l'ajout de séquences préfixes et suffixes correspondant aux sites de reconnaissance et de restriction de *Bsmbl*. Les traits verts symbolisent le recodage des séquences d'intérêt dans le but d'optimiser la production de protéines recombinantes dans le tabac. Des exemples de codons utilisés dans ce cadre sont présentés en figures 1B et 1C. La figure D représente un exemple d'adaptateurs.

**Tableau 3: Récapitulatif de la fréquence d'apparition des codons chez le tabac**

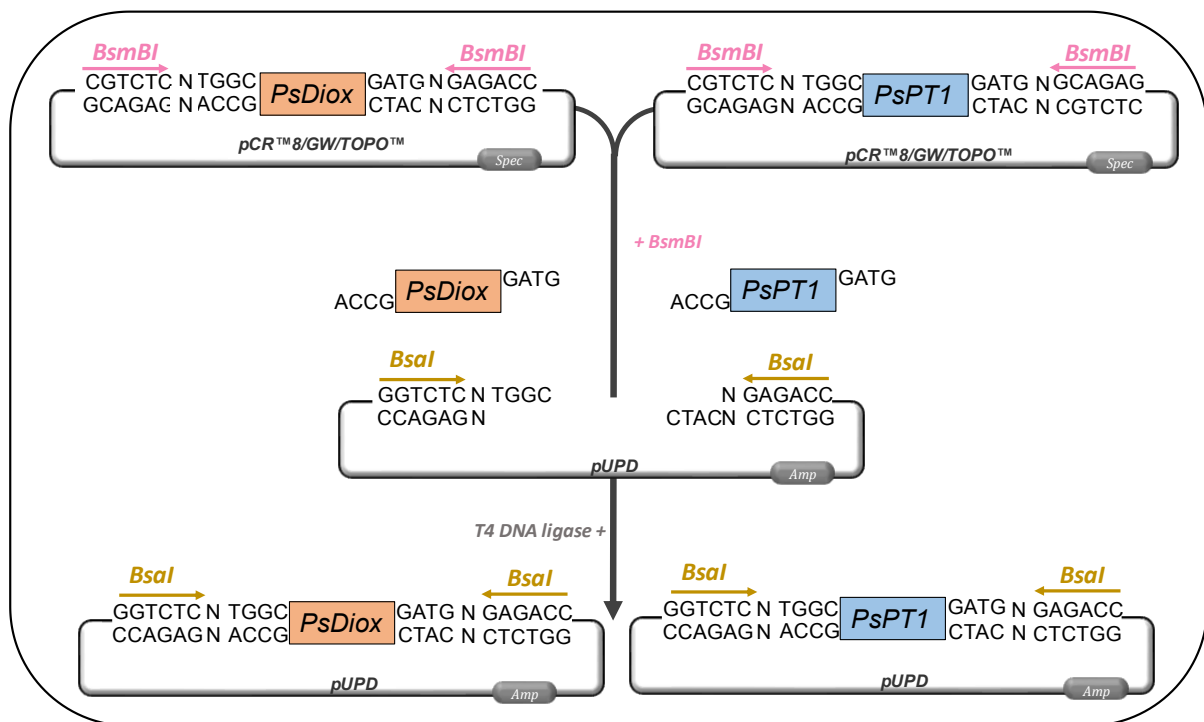
	U		%	C		%	A		%	G		%	
U	UUU	Phénylalanine	23.7	UCU	Sérine	22.3	UAU	Tyrosine	15.8	UGU	Cystéine	9.2	U
	UUC		17.6	UCC		10.4	UAC		12.8	UGC		7.2	C
	UUA	Leucine	12.8	UCA		17.2	UAA	STOP	0.7	UGA	STOP	0.9	A
	UUG		24.3	UCG		5.6	UAG		0.7	UGG		Tryptophane	12.4
C	CUU	Leucine	24.9	CCU	Proline	18.9	CAU	Histidine	12.9	CGU	Arginine	7.7	U
	CUC		12.5	CCC		6.4	CAC		8.2	CGC		4.2	C
	CUA		9.2	CCA		16.8	CAA	Glutamine	18.1	CGA		5.8	A
	CUG		11.9	CCG		6.5	CAG		17.0	CGG		5.3	G
A	AUU	Isoleucine	26.7	ACU	Thréonine	17.4	AAU	Asparagine	29.1	AGU	Sérine	14.7	U
	AUC		13.9	ACC		10.1	AAC		16.9	AGC		10.7	C
	AUA		12.1	ACA		15.2	AAA	Lysine	29.0	AGA	Arginine	15.8	A
	AUG	Méthionine	23.9	ACG		5.4	AAG		38.0	AGG		13.0	G
G	GUU	Valine	26.1	GCU	Alanine	33.2	GAU	Acide Aspartique	38.5	GGU	Glycine	24.3	U
	GUC		10.6	GCC		12.6	GAC		16.4	GGC		11.4	C
	GUA		9.9	GCA		23.5	GAA	Acide Glutamique	35.2	GGA		22.5	A
	GUG		15.6	GCG		6.5	GAG		30.9	GGG		11.1	G

## 2.2. Assemblage des constructions GoldenBraid

Une fois domestiquées, les séquences *PsDiox* et *PsPT1* ont été insérées dans le plasmide pUC57. Ces deux constructions ont été amplifiées dans *E.coli* et utilisés comme vecteur donneur pour l'insertion des gènes *PsDiox* et *PsPT1* dans le premier plasmide intervenant dans la construction GoldenBraid, à savoir le plasmide pUPD.

### 2.2.1. Clonage des séquences d'intérêt dans pUPD

Le plasmide pUPD est un plasmide receveur dans lequel est incorporé séparément chaque élément utilisé dans les constructions GoldenBraid (promoteur, séquence codante, terminateur, ...).

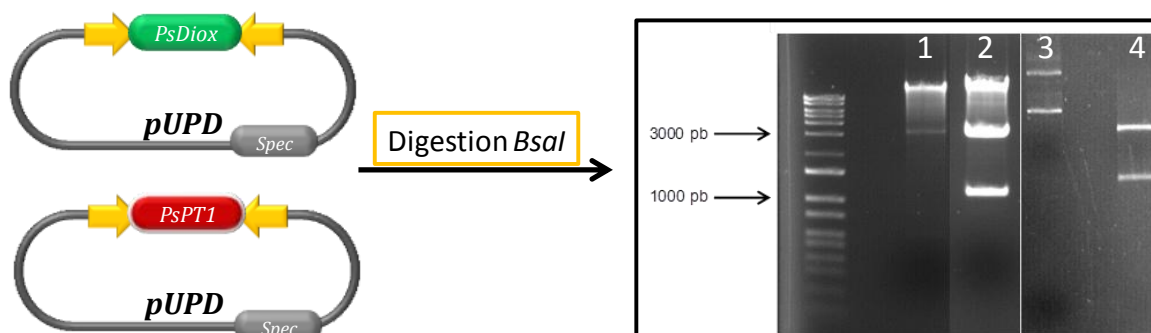


**Figure 30: Clonage de *PsDiox* et *PsPT1* dans le plasmide *pUPD*.** Le *pUPD* est le plasmide de destination dans lequel est inséré, après digestion par *BsaI* et ligation par T4 DNA ligase, les éléments qui constituent l'unité transcriptionnelles (ici les gènes *PsDiox* et *PsPT1*).

Le promoteur et le terminateur utilisés sont des produits commerciaux et sont fournis prêt à l'emploi dans le plasmide *pUPD*. Nous nous sommes donc uniquement attachés à cloner les séquences *PsDiox* et *PsPT1* dans le vecteur *pUPD*.

Pour ce faire, le plasmide *pUPD* vide, les plasmides donneurs présentant les séquences d'intérêt (*PsDiox* ou *PsPT1*), l'enzyme de restriction *Bsmbl* et la T4 DNA ligase sont mélangés dans un seul tube selon les paramètres décrits dans le paragraphe 2.1.8.3 du Matériel et Méthodes.

L'action de *Bsmbl* va générer des bords cohésifs au niveau des extrémités 5' et 3' de chacune des séquences d'intérêts, ainsi qu'au niveau des sites de restriction correspondants présents sur *pUPD*. Ainsi, par complémentarité de base et sous l'action de la T4 DNA ligase, les séquences d'intérêts sont liguées individuellement dans le plasmide *pUPD* (Figure 31). Les plasmides *pUPD* recombinants sont introduits dans *E.coli* et amplifiés. Une digestion par *BsaI* permet de vérifier la présence du transgène (Figure 31).



**Figure 31: Digestion des constructions *pUPD : PsDiox* et *pUPD : PsPT1* par l'enzyme *Bsal*.**  
Piste 1 et 3: *pUPD* non digéré. Piste 2 et 4 : *pUPD : PsDiox* et *pUPD : PsPT1* digéré par *Bsal*, respectivement. La bande avec le plus haut poids moléculaire en piste 2 correspond au plasmide *pUPD : PsDiox* non digéré, probablement dû à une concentration trop élevée et/ou à une forme superenroulée de ce plasmide.

### 2.2.2. Construction des unités transcriptionnelles

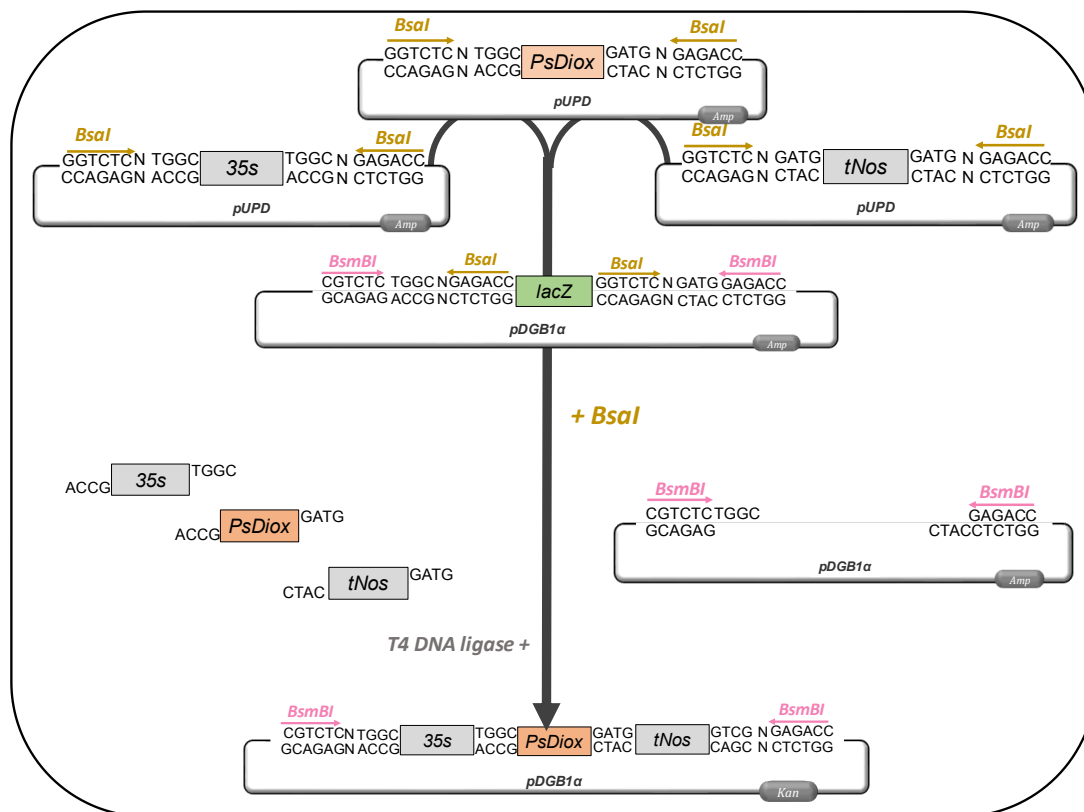
Un des intérêts de la méthode GoldenBraid est de pouvoir assembler "à façon" des unités transcriptionnelles, c'est-à-dire des cassettes d'expression composées d'un promoteur, d'une séquence codante et d'un terminateur.

Pour construire de telles unités transcriptionnelles, j'ai utilisé les 2 plasmides *pUPD* recombinants contenant les séquences codantes *PsDiox* et *PsPT1*, ainsi que des plasmides *pUPD* commerciaux contenant le promoteur constitutif 35S ou le terminateur tNos.

Pour cette étape, le plasmide *pUPD* devient le plasmide donneur, et les plasmides receveurs sont les plasmides *pDGB1α1* et *pDGB1α2* qui diffèrent par l'orientation des sites de restriction de l'enzyme *Bsal*.

De la même manière que pour l'étape précédente, les différents constituants de l'unité transcriptionnelle sont mélangés dans un tube unique, en présence du plasmide receveur (*pDGB1α1* pour *PsDiox* et *pDGB1α2* pour *PsPT1*), de la T4 DNA ligase et de l'enzyme de restriction *Bsal*. A l'instar de *Bsmbl*, *Bsal* va générer des extrémités cohésives au niveau des sites de restriction présents de chaque côté du promoteur 35S, des séquences codantes, du terminateur tNos et du plasmide *pDGB1α* (Figure 32). La ligation des différents éléments dans le plasmide receveur est assurée par la T4 DNA ligase.





**Figure 32: Assemblage du promoteur 35s, de la séquence codante *PsDiox* et du terminateur *tNos* en unité transcriptionnelle dans le plasmide pDGB1α.**

Ces ligations ne pourront se faire que selon une organisation prédéfinie. Le promoteur ne pourra se liguer qu'à l'extrémité 3' terminale du plasmide pDGB1α et l'extrémité 5' terminale du gène d'intérêt.

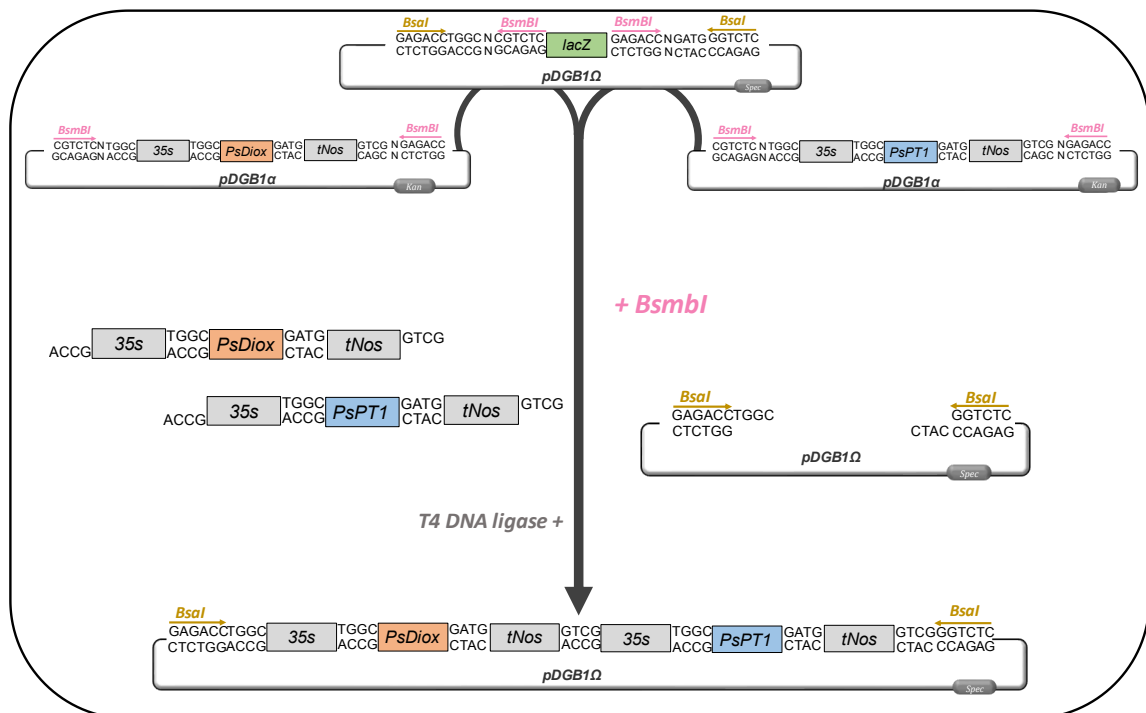
De la même manière le gène d'intérêt ne pourra se liguer qu'au promoteur à son extrémité 5' terminale et au terminateur à son extrémité 3' terminale, etc.

Les produits de ligation obtenus à l'issue des 25 cycles de digestions/ligations sont ensuite introduits et amplifiés dans des bactéries. Les plasmides recombinants sont analysés par une digestion de restriction.

### 2.2.3. Assemblage des unités transcriptionnelles

La dernière étape consiste à générer une construction plasmidique unique contenant les deux unités transcriptionnelles 35S-*PsDiox*-tnos et 35S-*PsPT1*-tnos. Pour cela, les plasmides pDGB1 $\alpha$  recombinants ainsi que le plasmide de destination pDGB1 $\Omega$  ont été digérés par l'enzyme de restriction *BsmBI* et ligués selon un protocole similaire au précédent (cf paragraphe 2.2.2.). La construction générée est appelée [35S-*PsDiox*-tnos : 35S-*PsPT1*-tnos]. Une fois les réactions d'assemblage terminées, l'insertion de la construction GoldenBraid dans le plasmide est vérifiée par digestion. Les sites de reconnaissance de l'enzyme *BsmBI* disparaissent lors de la recombinaison, étant donné qu'ils sont présents soit sur le plasmide receveur de chaque côté du gène *lacZ*, éliminé lors de la recombinaison, soit sur le plasmide donneur (Figure 33). Ainsi, la digestion a été réalisée par l'enzyme *BsaI*, dont les sites de reconnaissance sont présents sur le plasmide receveur pDGB1 $\Omega$ .

Une fois le plasmide recombiné final amplifié, l'intégration de la cassette dans le plasmide receveur est vérifiée par digestion, cette fois avec *BsaI*, dont les sites de clivage sont également présents sur le plasmide receveur pDGB1 $\Omega$ .



**Figure 33: Insertion des unités transcriptionnelles 35s-*PsDiox*-tnos et 35s-*PsPT1*-tnos dans le plasmide pDGB1 $\Omega$ .**

Par la méthode GoldenBraid, 3 cassettes d'expressions ont été générées et insérées dans des plasmides distincts. Ces trois cassettes appelées 35s-*PsDiox*-tnos, 35s-*PsPT1*-tnos et 35s-*PsDiox*-tnos : 35s-*PsPT1*-tnos ont alors été testées en expression transitoire dans des feuilles de *Nicotiana benthamiana* dans le but de déterminer leur efficacité à produire de l'umbelliférone et de la DMS.

### 3. Expression transitoire des unités transcriptionnelles

Avant de réaliser des transformations stables de tabac, j'ai souhaité vérifier que les différentes constructions étaient fonctionnelles. Pour cela j'ai mis en place une expérimentation préliminaire basée sur de l'expression transitoire. Les plasmides pDGB sont des plasmides dédiés à la transformation d'*Agrobacterium tumefaciens*.

Chacune des constructions générées, ainsi qu'un plasmide contenant une unité transcriptionnelle portant le gène codant la protéine GFP sous contrôle du promoteur 35S et du terminateur tNos, ont été introduits dans la souche EHA105 d'agrobactéries. Ces bactéries ont été cultivées selon un protocole décrit dans Karamat *et al* (Karamat *et al.*, 2014), puis infiltrées dans des feuilles de *N. benthamiana* âgées de 8 semaines (Figure 34). Trois répétitions biologiques sont réalisées pour chacune des constructions.

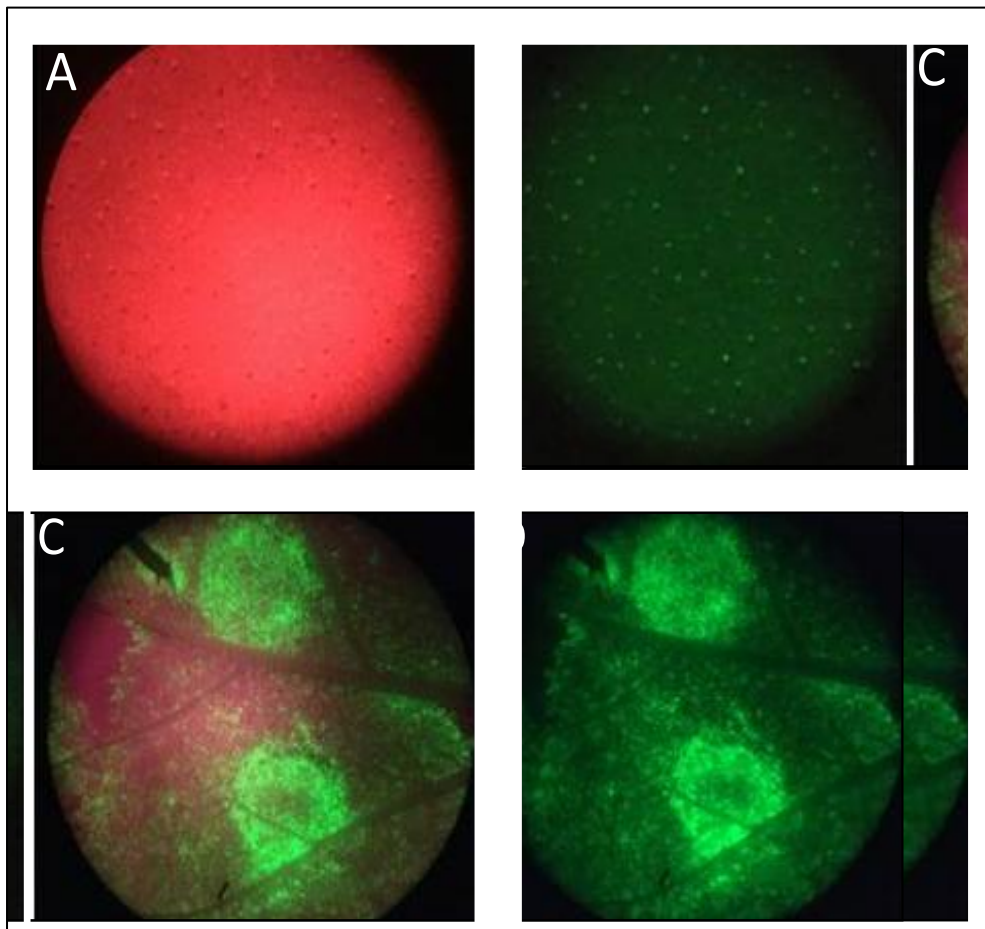


**Figure 34: Infiltration de suspensions d'agrobactéries dans des feuilles de *Nicotiana benthamiana*.** La suspension d'agrobactéries prélevée par la seringue est infiltrée sur la face abaxiale des feuilles.

### 3.1. Test préliminaire

Afin de déterminer la fenêtre temporelle permettant l'expression des unités transcriptionnelles GoldenBraid, nous avons dans un premier temps infiltré des agrobactéries contenant une construction permettant l'expression de la protéine GFP. Les résultats que nous avons obtenus indiquent une autofluorescence des feuilles (Figure 35 A et B). En parallèle de ces contrôles, nous avons suivi la fluorescence de 4 à 6 jours post agroinfiltration. Nous avons ainsi pu observer une fluorescence qui s'est révélée être maximale 4 jours après infiltration (Figure 35 C et D), et bien plus intense que dans les feuilles témoins, non transformées (A et B) démontrant que les constructions réalisées étaient fonctionnelles. Le signal de fluorescence observable sur la figure A correspond à l'autofluorescence naturelle de la plante.

Ces résultats permettent de valider l'efficacité du système GoldenBraid pour la production de protéines recombinantes dans le tabac *via* la transformation transitoire dans nos conditions de laboratoire.



**Figure 35: Observation de la fluorescence de la GFP dans des feuilles de *Nicotiana benthamiana* à la loupe binoculaire (x100).**  
*A et B : Feuille non transformée (respectivement sans et avec filtre rouge), C et D: Feuille transformée avec TU-GFP et incubée pendant 4 jours (respectivement sans et avec filtre rouge). Le filtre rouge permet de soustraire à l'image l'autofluorescence de la chlorophylle autour de 650 nm. La longueur d'émission de la GFP est de 510 nm. La GFP et la chlorophylle ont une gamme de spectre d'absorption commune autour de 490 nm.*

La fluorescence de la GFP dans les feuilles de tabac transformées se poursuit jusqu'à 6 jours après agro infiltration, puis décroît rapidement. Nous avons donc utilisé ce cadre temporel (entre 4 à 6 jours) pour la suite des expérimentations.

### 3.2. Expression transitoire de PsDiox et de PsPT1 dans des feuilles de *N. benthamiana*

Afin de vérifier que les enzymes PsDiox et PsPT1 sont fonctionnelles dans les feuilles de *N. benthamiana* infiltrées, j'ai mis en œuvre 2 approches réalisées en parallèle. La première est basée sur la mise en évidence de l'expression des gènes par RT-PCR, la seconde sur la mise en évidence de l'activité des enzymes *in vivo*.

Pour évaluer le système, j'ai, dans un premier temps, infiltré les feuilles de *N. benthamiana* avec des agrobactéries contenant i) la construction [35S-**PsDiox**-tnos], ii) la construction [35S-**PsPT1**-tnos], iii) la construction [35S-**PsDiox**-tnos : 35S-**PsPT1**-tnos]. Les feuilles ont été récoltées 96h après l'infiltration. Dans un second temps, j'ai apporté aux feuilles les différents substrats des enzymes. Pour cela, le pétiole de chaque feuille agroinfiltrée a été placé dans une solution de substrat spécifique de l'enzyme considérée pendant 24H pour que la réaction de conversion puisse s'effectuer *in vivo*. Afin de permettre une bonne diffusion des molécules substrats dans les feuilles, celles-ci sont placées dans une enceinte climatisée stimulant les processus d'évapotranspiration (cf paragraphe 2.3.2.1, Matériels et Méthodes). Des feuilles témoins, non transformées, ont été récoltées en parallèle des feuilles agroinfiltrées et incubées dans de l'eau distillée stérile avec les différents substrats utilisés.

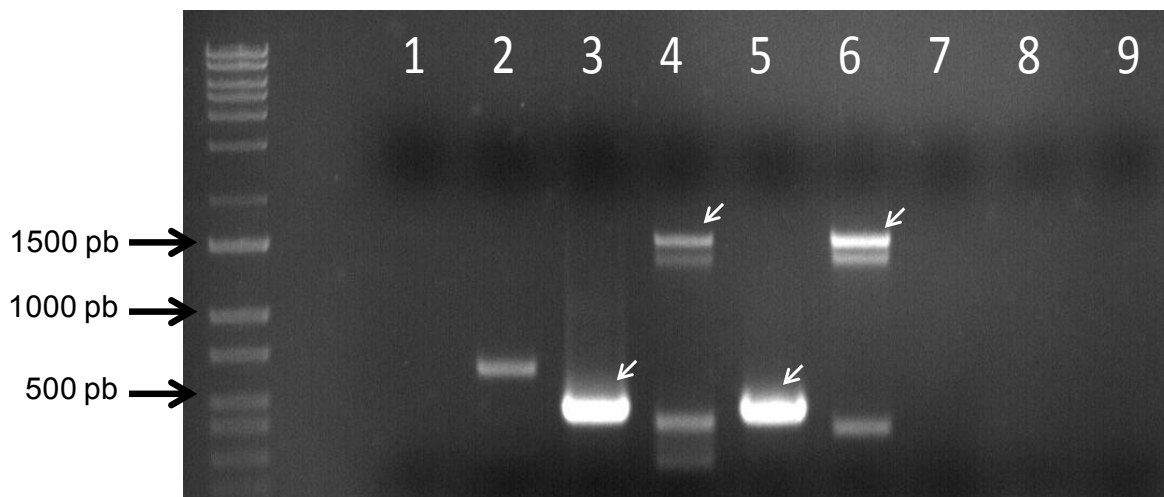
Dans le cas de PsPT1, le substrat est la DMS. Dans le cas de la PsDiox, j'ai utilisé de l'acide *p*-coumarique plutôt que du *p*-coumaroylCoA étant donné la faible stabilité de ce dernier. L'acide *p*-coumarique et le *p*-coumaroylCoA sont des composés naturellement présents dans le tabac, qui dispose donc des outils enzymatiques assurant la conversion de l'acide *p*-coumarique en *p*-coumaroylCoA. L'apport exogène d'acide *p*-coumarique permet théoriquement d'augmenter la concentration en *p*-coumaroylCoA dans la plante, évitant alors la limitation du substrat de PsDiox.

Après 24h, les feuilles ont été rincées à l'eau distillée stérile, séparées de leurs nervures centrales, puis congelées dans de l'azote liquide et broyées. Le broyat est alors utilisé pour vérifier l'expression des gènes cibles par RT-PCR mais également pour faire des analyses phytochimiques.

### 3.2.1. Analyse de l'expression des gènes

Pour mettre en évidence l'activité transcriptionnelle en lien avec les constructions GoldenBraid dans les feuilles de tabac transformées, j'ai extrait les ARN totaux à partir des feuilles prélevées. Comme indiqué dans le matériel et méthodes (cf paragraphe 2.1.2.), des amplifications par PCR sont réalisées directement à partir d'ARN pour s'assurer de l'absence d'ADN génomique résiduel .

Pour chacune des constructions GoldenBraid étudiées, y compris la construction portant le gène de la GFP, un signal d'amplification correspondant à la taille attendue a été mis en évidence (Figure 36). Dans le cas de la double construction, j'ai utilisé les deux couples d'amorces spécifiques des gènes *PsDiox* et *PsPT1* (Figure 36, Piste 5 et 6) (Annexe 8). A noter que l'utilisation des amorces spécifiques de *PsPT1* (Figure 36, Piste 4 et 6) génère, en plus de l'amplification à la taille attendue d'environ 1600 pb, des hybridations aspécifiques. Etant donné que les ARNs utilisés proviennent, dans ce cas, de deux échantillons distincts, ces hybridations aspécifiques sont très certainement liées à des conditions suboptimales d'amplification des cibles par PCR, notamment la température d'hybridation des amorces.



**Figure 36: RT-PCR réalisées sur de l'ARN extraits de feuilles de tabac après 4 jours d'incubation.**

Piste 1 : plante non transformées avec amorces spécifiques du gène codant la GFP. Piste 2 : Construction GFP, Piste 3 : Construction PsDiox, Piste 4 : Construction PsPT1, Piste 5 : Construction PsDiox-PsPT1 avec amorces spécifiques de PsDiox, Piste 6 : : Construction PsDiox-PsPT1 avec amorces spécifiques de PsPT1. Piste 7 à 9 : Amorces GFP, PsDiox et PsPT1, respectivement, sans ARN.

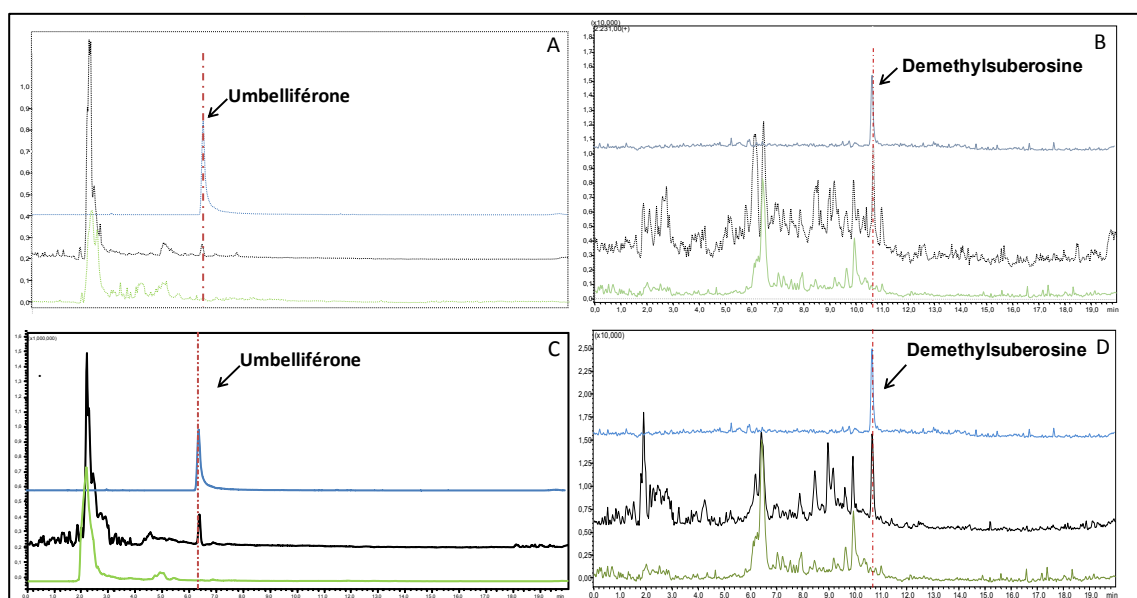
### 3.2.2. Analyses phytochimiques

En parallèle de l'extraction des ARNs, j'ai également extrait les composés phénoliques totaux à partir des feuilles infiltrées. L'extrait méthanolique obtenu a été analysé par une approche ciblée en UHPLC couplée à la spectrométrie de masse destinée à détecter le produit des deux réactions seules, ou combinées. Ainsi nous avons tenté de détecter la présence d'umbelliférone dans les échantillons de feuilles surexprimant PsDiox, et de la DMS dans les échantillons surexprimant PsPT1. Enfin nous avons suivi spécifiquement ces deux molécules dans les extraits de feuilles exprimant les deux constructions.

Premièrement, nous avons détecté de l'umbelliférone dans les échantillons de feuilles de tabac exprimant le gène *PsDiox*, seul ou associé à *PsPT1*, 4 jours après agroinfiltration et incubées en présence d'acide *p*-coumarique (Figure 37 A et C). En revanche, l'umbelliférone n'est plus détectable après 6 jours. De plus, étant donné que l'umbelliférone est un des substrats de *PsPT1*, il serait logique de détecter, dans le cas de la double construction, de la DMS dans les échantillons. Or, aucun des résultats obtenus ne va dans ce sens.



Dans le cas de PsPT1, après 4 jours, aucun pic de DMS n'est détectable dans les échantillons de feuilles incubées avec de l'acide *p*-coumarique ou de l'umbelliférone (Tableau 4). En revanche, dans les mêmes conditions et après 6 jours, nous pouvons observer un pic correspondant à de la DMS, sur la base de son ratio m/z et de son temps de rétention, avec le gène seul (Figure 37 B) et la double construction (Figure 37 D).



**Figure 37: Analyse des produits de métabolisation dans le tabac des différentes unités transcriptionnelles GoldenBraid par UPLC/MS.**

A et C correspondent à l'analyse du produit de métabolisation après 4 jours d'incubation en présence d'acide coumarique respectivement pour TU-Diox et TU-PsDiox: TU-PsPT1. B et D correspondent à l'analyse des produits de métabolisation en présence d'umbelliférone après 6 jours d'incubation respectivement de TU-PsPT1 et de TU-PsDiox: TU-PsPT1. Les courbes vertes représentent les témoins non transformés; la courbe bleue représente l'umbelliférone (A et C) et la DMS (B et D) en tant que standard. Les analyses ont été réalisées selon le mode « Selected Ion Monitoring » en ionisation positive. Les ratio m/z de l'umbelliférone et de la demethylsubérosine sont respectivement 163,14 (+1) et 230,263 (+1).

**Tableau 4: Récapitulatifs des résultats de l'incubation des feuilles de tabac transformées avec les constructions GoldenBraid en présence du substrat associés.**

La condition témoin correspond à des feuilles de tabac infiltrées avec de l'eau. JAI : Jours après infiltration de la suspension d'agrobactéries

	Construction GoldenBraid							
	Témoin		PsDiox		PsPT1		PsDiox+PspT1	
	JAI		JAI		JAI		JAI	
	4	6	4	6	4	6	4	6
Umbelliférone	-	-	+	-	-	-	+	-
Demethylsubérosine	-	-	-	-	-	+	-	+

L'activité des deux protéines a été validée *in vivo* en transformation transitoire dans le tabac. Cependant l'apparition des produits réactionnel à des temps différent suggèrent que les constructions GoldenBraid ont une fenêtre temporelle d'expression différente.

## 4. Construction de tabacs mutants, producteurs de furocoumarines : résultats préliminaires

### 4.1. Introduction

En utilisant les constructions générées précédemment, et dont la fonctionnalité a été démontrée, j'ai réalisé des transformations stables de plantules de tabac *in vitro* via *Agrobacterium tumefaciens* pour déterminer l'impact de la synthèse de furocoumarines sur le métabolisme global de la plante.

## 4.2. Transformation stable de *Nicotiana tabacum*

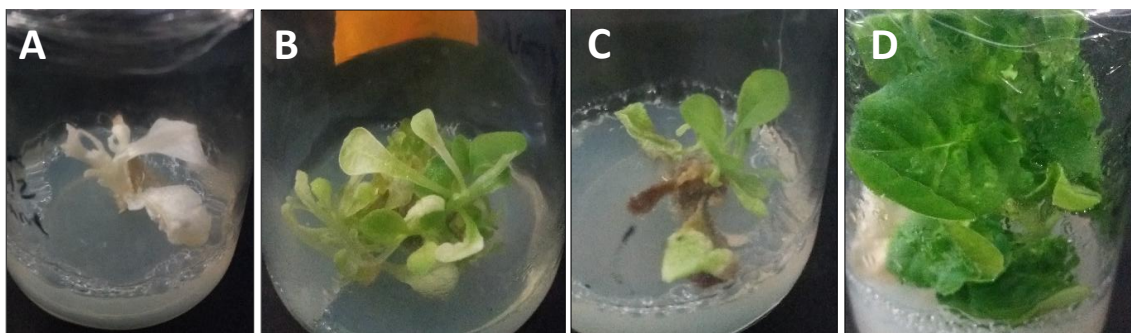
### 4.2.1. Obtention de plantes transgéniques

Des *Agrobacterium tumefaciens* EHA105 transformées avec les constructions [35s-***PsDiox***-tnos], [35s-***PsPT1***-tnos] ou [35s-***PsDiox***-tnos : 35s-***PsPT1***-tnos] ont été incubées en présence d'explants de feuilles de *Nicotiana tabacum* comme décrit dans la partie Matériel et Méthodes (cf. paragraphe 2.2.3.). Les explants potentiellement transformés ont été placés sur un milieu permettant la callogenèse et la régénération. Cette expérimentation a été réalisée en parallèle avec la construction [35s-***GFP***-tnos]. La sélection des plantes régénérantes transformées avec les constructions d'intérêt a été réalisée sur un milieu sélectif contenant d'une part de l'ampicilline (qui permet une sélection des cellules transformées) et d'autre part de la céfotaxime (qui permet de ralentir la croissance des agrobactéries) (Cassinat and Nicolas, 1994; Spicer *et al.*, 1984). Afin d'avoir un autre contrôle des explants similaire de plantes non transformées ont été placées dans conditions similaires.

Malgré les précautions d'usage prises lors des lavages et des conditions de cultures, seuls les explants transfectés avec la construction [35s-***PsDiox***-tnos : 35s-***PsPT1***-tnos] ont été préservés de la contamination et ont permis d'obtenir des cals. Les résultats présentés dans la suite de ce chapitre, ne sont donc que des résultats préliminaires qu'il conviendra de renforcer et de confirmer.

### 4.2.2. Analyse phénotypique

Les cals obtenus ont permis la régénération de plantules résultants d'événements de transformations indépendants. Les différents individus obtenus présentent des phénotypes différents par rapport aux plantes témoins non transformées. Ces plantules ont des phénotypes plus ou moins albinos (Figure 38 A) et présentent des morphologies foliaires particulières avec des limbes plus allongés et plus étroits (Figure 38 B). Trois d'entre elles ont été transférées sur des milieux d'enracinement afin de générer des plantes plus grandes permettant de faire des analyses plus fines.



**Figure 38: Plantules de *Nicotiana tabacum* régénérées à partir d'explants transformés par la construction GoldenBraid [35s-PsDiox-tnos : 35s-PsPT1-tnos].**

*A,B et C : Plantules transformées, D : Plante sauvage , non transformée.*

Ces résultats préliminaires ont permis de générer des plantes génétiquement transformées ayant potentiellement intégrées les gènes cibles. Ces plantes présentent des phénotypes particuliers d'éclaircissement et de déformation des feuilles. Ces résultats nécessitent à présent d'être reproduits avec les témoins nécessaires et d'être approfondis.

## 5. Discussion et conclusion

Dans le premier chapitre j'ai analysé l'impact de stress sur la synthèse de métabolites secondaires et j'ai fait un focus sur la production de furocoumarines dans ce contexte. Dans ce second chapitre j'aborde l'étude de cette voie de biosynthèse selon un autre angle qui consiste à apporter dans une plante le matériel génétique nécessaire pour produire de nouvelles molécules.

Afin de tester leurs fonctionnalités, j'ai réalisé des transformations transitoires de feuilles de *Nicotiana benthamiana* avec des agrobactéries présentant les différentes constructions GoldenBraid. Les analyses phytochimiques et transcriptomiques faites sur les différents tissus ont démontré que, dans ces conditions, les différentes protéines étaient produites et fonctionnelles. Cependant, les résultats ont mis en évidence un décalage temporel entre l'activité de PsDiox et de PsPT1. Ce décalage peut provenir de différents éléments.

- Il pourrait être lié à la nature différente de ces deux enzymes. En effet, PsDiox est une dioxygénase  $\alpha$ -cétoglutarate dépendante soluble (Vialart *et al.*, 2012), alors que PsPT1 est une prényltransférase membranaire (Munakata *et al.*, 2016).
- Les séquences codantes de *PsDiox* et *PsPT1* ont été au préalable recodées pour optimiser la production de protéines recombinantes dans le tabac. Ainsi, présents sur les séquences d'intérêts ont été remplacés par des codons préférentiellement utilisés par la machinerie cellulaire dans *Nicotiana sp* , représentant 21% de la séquence originale dans le cas de *PsDiox* et 23% pour *PsPT1*.

Si l'optimisation du biais de codon est une approche intéressante, certains auteurs indiquent aussi que l'utilisation exclusive de codons optimaux peut engendrer une diminution de l'activité enzymatique des protéines correspondantes, liée notamment à une traduction trop rapide par les ribosomes (Webster *et al.*, 2017). En effet, il a été démontré, notamment chez *E.coli* , que l'utilisation par l'organisme de codons prépondérants ou rares n'est pas le fruit du hasard mais est liée à des éléments structuraux de la protéine, tels que les hélices  $\alpha$  ou les feuillets  $\beta$  (Thanaraj et Argos, 1996). De la même façon, Tuller *et al.* (2010) ont démontré que la vitesse de traduction n'est pas linéaire au cours de la production de la protéine mais augmente au fur et à mesure de la progression des ribosomes. Ainsi, une utilisation plus modérée de codons optimisés dans le recodage des séquences *PsDiox* et *PsPT1* permettrait peut-être d'améliorer l'expression de ces unités transcriptionnelles et de la production des protéines associées.

- Les unités transcriptionnelles ont été construites en utilisant un promoteur commun, celui du virus de la mosaïque du chou-fleur, ou 35s. Etant donné qu'il s'agit d'un promoteur constitutif, doté d'une expression ubiquitaire, un grand nombre d'ARNm sont produits lors de la transcription. Il a été démontré que l'utilisation de ce promoteur peut induire la formation de dimères d'ARNm, pouvant être dégradés par la plante et donc induire une extinction du signal de transcription (Mlotshwa *et al.*, 2010). L'utilisation de promoteurs différents, tels que MHP, issu de *Medicago truncatula* (Xiao *et al.*, 2005), permettrait de réduire ce phénomène d'extinction.

L'analyse par expression transitoire *in vivo* livre également d'autres résultats inattendus. Par exemple, en plus des produits de métabolisation attendus (DMS et umbelliférone) on peut observer l'apparition de pics (figure 37 D, courbe noire) qui sont absents dans l'échantillon contrôle (figure 37 D, courbe verte). Ces produits inconnus, de même masse que la DMS, pourraient être des isomères. Leur présence soulève la question de la spécificité *in vivo* des enzymes surexprimées et de fait de l'applicabilité de notre approche. Est-ce qu'il est réellement possible de transposer une voie de biosynthèse dans un contexte cellulaire différent du contexte originel sans risquer la production de produits secondaires ?

La purification et l'analyse par résonnance magnétique nucléaire (RMN) permettra de les identifier et de déterminer si PsDiox et PsPT1 sont capable de métaboliser d'autres composés que leurs substrats préférentiels.

Au vu des résultats précédents, j'ai abordé par la suite la question de la transformation stable. Pour cela, la construction [35s-**PsDiox**-tnos : 35s-**PsPT1**-tnos] a été insérée dans le génome d'une autre variété de Nicotiana, *N. tabacum*. Les plantules ont des phénotypes très marqués, pour la plupart albinos. Ici, l'albinisme est défini par l'incapacité de la plante à produire des chloroplastes et donc d'enclencher des mécanismes photosynthétiques (Makowska et Oleszczuk, 2014). Ce phénotype est souvent retrouvé en culture *in vitro*, notamment dans le cas d'hybridation, d'androgenèse ou de culture de double haploïdes (Kumari *et al.*, 2009). Dans notre cas, plusieurs hypothèses peuvent être envisagées.

- Il peut tout simplement être lié à l'expression des gènes intégrés dans le génome du tabac. La synthèse de furocoumarines potentiellement induites par l'expression des gènes cibles peut en effet entraîner une modification du métabolisme de la plante, notamment du métabolisme primaire, qui pourrait entraîner l'apparition de chloroses, voire d'un phénotype albinos. Les furocoumarines, bien que tolérées par les plantes productrices, sont des molécules toxiques, aussi bien pour certains insectes que pour les plantes elles-mêmes. La plante hôte utilisée ne produit pas naturellement de furocoumarines est donc pas adaptée à ces molécules.

- Pour tout de même utiliser cette voie de biosynthèse modèle pour notre étude, il serait judicieux d'utiliser des promoteurs inductibles, permettant alors de contrôler plus finement l'expression des différents gènes intégrés dans le tabac.
- Plusieurs études ont démontré l'implication de la composition du milieu de culture dans l'apparition de plantes albinos. En effet, le ratio entre chlorophylle *a* et chlorophylle *b* varie fortement en fonction du milieu utilisé, comme cela a été démontré chez l'avoine de mer (*Uniola paniculata*, (Nishiyama and Motoyoshi, 1962). Cette variation peut être causée, premièrement, par la nature et la concentration de la ressource carbonée du milieu. Pour exemple, une étude menée chez *Triticum turgidum* (Poacées) a démontré que des plantes partiellement albinos étaient capables de retrouver un statut photosynthétique complet en variant la nature des sucres apportés au milieu de culture (Saidi *et al.*, 1997). De la même façon, (Cistué *et al.*, 1994) ont démontré une corrélation positive entre la concentration en mannitol et le recouvrement de la coloration verte des feuilles chez l'orge. Cependant, les plantes non transformées restent vertes tout au long de leurs croissances. D'autres études se sont penchées sur l'implication des facteurs de régulation de la croissance des plantes dans l'apparition de plantes albinos. Parmi ces facteurs, les hormones telles que les auxines et les cytokinines peuvent fortement impacter la division cellulaire et le développement de la plante, même à faible concentration (Kumari *et al.*, 2009). De plus, (Asif *et al.*, 2013) ont mis en exergue l'impact de la combinaison d'un antibiotique, la céfotaxime, et d'une phytohormone, l'acide abscissique, sur l'apparition du phénotype albinos. Ils ont également mis en relation l'accroissement de la concentration en céfotaxime avec l'augmentation de la fréquence d'apparition de plantes albinos.
- On ne peut pas exclure non plus l'implication un impact du site d'insertion aléatoire du T-DNA. Cette localisation de l'insert varie en fonction de l'événement de transformation. La synthèse de la chlorophylle est médiée par de nombreuses étapes enzymatique, organisées en voie de biosynthèse, et faisant donc appel à de nombreux gènes (Kim *et al.*, 2013).

Etant donné que la chlorophylle est le produit final de cette voie de biosynthèse, il est envisageable que l'insertion d'un fragment d'ADN, dans notre cas la cassette d'expression [35s-*PsDiox*-tnos : 35s-*PsPT1*-tnos], à l'intérieur d'un des gènes de cette voie peut aboutir à une faible production, voire à une disparition de la chlorophylle, causant ainsi un phénotype albinos.

Cette théorie a été démontrée chez *Arabidopsis thaliana* par (Qin *et al.*, 2007). En effet, les auteurs ont établi que l'interruption d'un gène de la voie de biosynthèse de la chlorophylle, en l'occurrence le gène codant la phytoène désaturase, induit un phénotype albinos, mais également une diminution de la synthèse de carotènes et de gibbérélines. Pour rappel, les gibbérélines sont des phytohormones impliquées dans le développement de la plante (Pimenta Lange and Lange, 2006). Une diminution de la synthèse de ces hormones pourrait alors potentiellement expliquer les différences de morphologies entre les feuilles des plantules transformées et témoin (Figure 38).

Dans notre cas, cette hypothèse semble néanmoins peu probable dans la mesure où les différentes plantes obtenues sont issues d'événements de transformation indépendants (à partir d'explants différents).

- Du fait de la présence d'un hème de fer au cœur de la chlorophylle, l'activité photosynthétique est fortement liée à la présence cet élément dans la plante. Ainsi, une diminution de la disponibilité en fer dans la plante conduit inéluctablement à une diminution de la synthèse de chlorophylle (Pushnik *et al.*, 1984). Dans ce cadre, nous nous sommes questionnés sur des possibles modifications du métabolisme du fer dans les plantes régénérées. Dans cette optique, nous avons envisagé l'hypothèse d'un possible « gene silencing », *via* l'interaction de deux séquences d'ARNm avec un certain degré d'identité. Ainsi, à l'aide de l'algorithme BLAST, j'ai recherché des séquences connues de *Nicotiana tabacum*, plus ou moins identique aux séquences utilisées dans les constructions GoldenBraid, *PsDiox* et *PsPT1*. D'après les résultats obtenus, le gène *PsDiox* est identique à 69.20% à un gène de *Nicotiana tabacum* codant une feruloylCoA *ortho*-hydroxylase (Accession XM 016606995.1). Cette enzyme est impliquée, chez *Arabidopsis thaliana*, dans la synthèse de scopolétine à partir de feruloyl\_CoA (Kai *et al.*, 2008).



La scopolétine a récemment été caractérisée pour son intervention dans le transport du fer (Siwinska *et al.*, 2018).

De façon plus large, cette homéostasie peut être assurée par la familles des coumarines en général (Clemens et Weber, 2015). Bien que l'identité entre *PsDiox* et la feruloylCoA *ortho*-hydroxylase soit relativement faible, nous pouvons émettre l'hypothèse d'une hybridation entre les ARNm de ces deux gènes, conduisant à des mécanismes de « gene silencing ».

Ces mécanismes peuvent alors entraîner une diminution de la synthèse de scopolétine, impactant alors négativement le transport du fer dans la plante et entraînant finalement une diminution de la synthèse de chlorophylle.

Pour vérifier cette hypothèse, nous pouvons envisager de mesurer le niveau d'expression du gène de tabac codant la feruloylCoA *ortho*-hydroxylase dans les plantes transformées et les plantes contrôle. Une diminution de l'expression de ce gène dans les plantes transformées témoignerait d'un phénomène de silencing et tendrait à valider cette hypothèse.

Ces différentes théories devront bien sûr être validées ou invalidées expérimentalement. Pour cela plusieurs expérimentations devront être réalisées.

- Tout d'abord, différents milieux, avec des sources de carbones variées et des modifications de la nature et de la concentration en hormones et en antibiotiques, peuvent être testés pour évaluer leurs implications dans la diminution ou la perte d'activité photosynthétique.
- Ensuite, l'interruption d'un gène impliqué dans la synthèse de chlorophylle par l'insertion d'un fragment d'ADN peut être étudiée par la réalisation de PCR quantitative ciblé sur différents gènes de cette voie. De la même façon, l'expression du gène codant la feruloyl CoA *ortho* hydroxylase peut être suivie par cette méthode dans les plantes transformées. Cependant, actuellement, nous ne disposons que de très peu de matériel végétal, ce qui rend la réalisation de ces différentes expérimentations problématiques. De nouvelles séries de transformation de plante dans les mêmes conditions que citées précédemment, ainsi que la multiplication *in vitro* des plantes disponibles sont en cours de réalisation.



## **D. Identification de nouveaux gènes et caractérisation fonctionnelle**

### **1. Introduction**

Nous avons démontré dans les chapitres précédents que la blessure mécanique était une méthode efficace pour induire la synthèse des furocoumarines chez le panais, ainsi que pour la surexpression des gènes connus associés à cette voie (*PsDiox*, *PsPT1*, *PsPT2*, *CYP71AJ3* et *CYP71AJ4*). Ce dispositif m'a paru pertinent pour entamer, par une approche différentielle, la recherche d'autres gènes impliqués dans la synthèse de ces molécules. Dans le cadre de ce travail, je me suis intéressé plus particulièrement à la marmesine synthase car la caractérisation moléculaire de cette étape permettrait de faire le lien entre la synthèse de DMS et de psoralène, première furocoumarine linéaire chez le panais et ainsi poursuivre le travail de reconstruction de voie métabolique chez le tabac. Pour ce faire, j'ai employé deux approches complémentaires, l'une transcriptomique, l'autre génomique.

### **2. Recherche de gènes candidats**

#### **2.1. Approche transcriptomique**

Dans le cadre de ce projet de recherche, et notamment grâce aux travaux de Hamerski et Matern (1988b, 1988a), j'ai focalisé l'identification de gènes candidats sur les cytochromes P450 et plus particulièrement sur le clan CYP71. La raison de ce choix repose sur le fait que chez les poacées, et en particulier chez le maïs, tous les P450 impliqués dans la synthèse de benzoxazinoid-2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) et de 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) appartiennent à une seule famille de P450, les CYP71C (Dutartre *et al.*, 2012).

Il est envisageable que pour la synthèse des furocoumarines, ce schéma soit identique, dans la mesure où, à ce jour, tous les cytochromes P450 impliqués dans la synthèse de ces molécules chez le panais sont des représentants de cette famille (Krieger *et al.*, 2018; Larbat *et al.*, 2007, 2009).

En plus de la famille CYP71, j'ai également ciblé la famille CYP81, connue pour intervenir dans des mécanismes de réponses de la plante face à un bio agresseur, notamment *via* la synthèse de répulsifs contre les insectes tels que les glucosinolates chez *Arabidopsis thaliana* (Bak *et al.*, 2011). Enfin, j'ai recherché des P450 de la famille CYP82, majoritairement impliqués dans la réponse de la plante à l'herbivorie (Bak *et al.*, 2011).

Pour identifier des gènes candidats, j'ai dans un premier temps, recherché les séquences appartenant à ces trois familles de cytochromes P450 dans une banque RNAseq de panais (Méthode 454).

Dans un second temps, j'ai évalué le profil d'expression de chacun des gènes identifiés par RTqPCR à partir d'ARNs issus de racines de panais élicitées par blessures mécaniques en comparaison avec des tissus témoins.

### **2.1.1. Analyse de banque RNAseq (séquençage 454) de panais**

Les données bio-informatiques chez le panais sont extrêmement restreintes en raison du faible nombre d'études portant sur ce modèle. Célia Krieger (2014) a, dans le cadre de sa thèse, contribué à étoffer ces données au travers de la construction d'une banque RNAseq normalisée (Méthode de séquençage 454) de panais. Cette banque a été construite à partir d'ARNs de panais issus de jeunes plantules traitées au méthyljasmonate, un éliciteur chimique, notamment du métabolisme spécialisé. Cette banque RNAseq a servi de point de départ à l'identification de gènes candidats.

Si la normalisation permet de limiter les risques de passer à côté de gènes exprimés à des niveaux très faibles, elle empêche aussi l'exploitation de la banque pour des approches quantitatives. Cette banque a été constituée par pyroséquençage 454.

Cette technologie permet de générer des séquences d'une taille moyenne de 400 pb qu'il a d'abord fallu assembler pour générer des séquences plus grandes ("Large contigs", Figure 39), réduisant ainsi le nombre de séquences de 99.5%.

Dans un premier temps, ces « large contigs » ont été classés en fonction de leur appartenance à la superfamille des cytochromes P450 (Tableau 5). Ce premier crible a permis de mettre en évidence **336** séquences présentant des homologues importantes avec les P450s. Parmi ces séquences j'ai restreint les gènes candidats aux enzymes de la famille CYP71, CYP81 et CYP82.

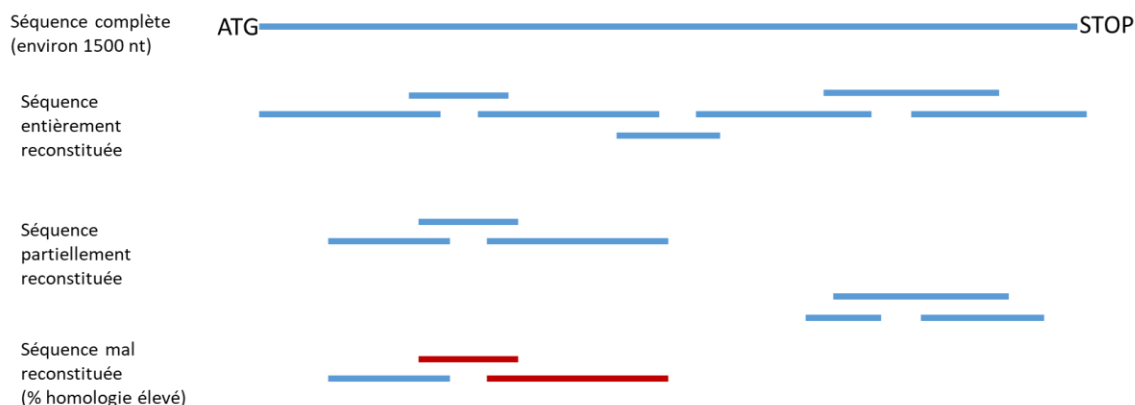
Le tri des 336 séquences initiales a permis de réduire à **93** le nombre de séquences cibles. J'ai ensuite utilisé un troisième filtre en me basant sur les fonctions biologiques décrites pour chacune des 93 séquences. Seules les séquences proches des protéines impliquées dans la synthèse de composés phénoliques, de molécules de défenses ou dans la réponse de la plante à un stress ont été retenues. Des séquences non annotées mais homologues aux trois sous-familles de cytochromes P450 précédemment citées ont également été retenues. Finalement, **79 séquences** sont conservées (Tableau 5).

Bien que déjà assemblées en "large contigs", les séquences issues de la banque RNAseq de panais sont très souvent des séquences partielles. Pour pallier cette limite, j'ai réalisé un assemblage complémentaire qui m'a permis finalement d'obtenir **45 séquences d'une longueur moyenne de 800 nt (Tableau 5)**.

**Tableau 5: Récapitulatif des différentes étapes de tri de la banque RNAseq de panais générée avec la méthode de séquençage 454**

	Nombre de séquences
Séquences totales de la banque (Large contigs)	3047
Séquences cytochrome p450	336
Séquences homologues au CYP71/CYP81/CYP82	93
Séquences annotées	79
Séquences assemblées en contigs	45
Séquences validées par PCR	22

Cette méthode présente deux limites. La première est que la superposition des séquences partielles peut aboutir à la formation de séquences chimériques, c'est-à-dire un assemblage de plusieurs fragments nucléotidiques issus de gènes différents (Figure 39). La seconde est en lien direct avec la méthode de séquençage des ADNc qui peut générer des erreurs aboutissant à une séquence plus ou moins différente de la séquence originelle.



**Figure 39: Schéma représentant les différents cas de figure dans la démarche d'assemblage de séquences RNAseq.**

De fait, pour vérifier que ces contigs existent réellement, j'ai réalisé des PCR sur chacun d'entre eux. Les résultats obtenus confirment la présence de **22** d'entre eux (Tableau 6). La séquence de chacun de ces 22 contigs est détaillée en annexe 7.

**Tableau 6 : Récapitulatif des résultats d'amplification par PCR des différents contigs candidats à partir d'ADN génomique de panais.** La croix rouge indique une absence de signal d'amplification après révélation du gel d'agarose.

Nom du Contig	Amplification	Nom du Contig	Amplification	Nom du Contig	Amplification
C1	✓	C16	✓	C31	✗
C2	✗	C17	✗	C32	✓
C3	✗	C18	✓	C33	✓
C4	✗	C19	✗	C34	✓
C5	✗	C20	✓	C35	✓
C6	✗	C21	✗	C36	✗
C7	✗	C22	✗	C37	✗
C8	✓	C23	✗	C38	✓
C9	✗	C24	✗	C39	✓
C10	✓	C25	✓	C40	✗
C11	✗	C26	✓	C41	✓
C12	✓	C27	✗	C42	✓
C13	✓	C28	✗	C43	✗
C14	✓	C29	✓	C44	✗
C15	✓	C30	✗	C45	✗

## 2.1.2. Mesures quantitatives de l'expression des séquences candidates

Afin d'étudier l'implication potentielle de ces gènes dans la voie de biosynthèse des furocoumarines, j'ai ensuite évalué leur niveau d'expression dans des tissus élicités mécaniquement ou non, et j'ai comparé leur niveau d'induction à ceux de gènes impliqués dans la voie de biosynthèse des furocoumarines.

### 2.1.2.1. Mise au point méthodologique : évaluation de l'efficacité des amorces de PCR quantitative

Dans un premier temps, l'efficacité des couples d'amorces spécifiques de chaque séquence a été évaluée à partir d'ADN génomique de panais (variété « Demi-long de Guernesey ») suivant une série de dilution. Les séquences des amorces sont détaillées en annexe 5. A partir des résultats, le pourcentage d'efficacité et le coefficient de détermination  $R^2$  ont été calculés pour chaque couple d'amorces.

Le coefficient de détermination, compris entre 0 et 1, permet de juger de la qualité d'une régression linéaire simple.

Ce coefficient détermine l'adéquation entre les données expérimentales et théoriques, pour lesquelles  $R^2=1$ . Un coefficient de détermination proche de 1 correspond à une distribution des données expérimentales resserrées autour de la régression linéaire, traduisant d'une corrélation avec les données théoriques.

Les résultats (Tableau 7) indiquent que, pour les amorces C18, C32, C38 et C39, aucune amplification n'a été détectée avec les conditions d'amplification utilisées. Pour les autres couples d'amorces, l'efficacité est supérieure à 85%, avec une moyenne de 95%, et un coefficient de détermination compris entre 0.985 et 0.998. Ces valeurs indiquent que, au minimum, 98,5% des valeurs mesurées correspondent aux données théoriques, ce qui traduit d'une très bonne spécificité des différents couples d'amorces.

**Tableau 7: Résultats de l'analyse de l'efficacité des couples d'amorces spécifiques des séquences candidates par PCR quantitative à partir de 100ng d'ARN de panais.**  
Les couleurs verte et rouge indiquent respectivement la présence et l'absence d'amplification.

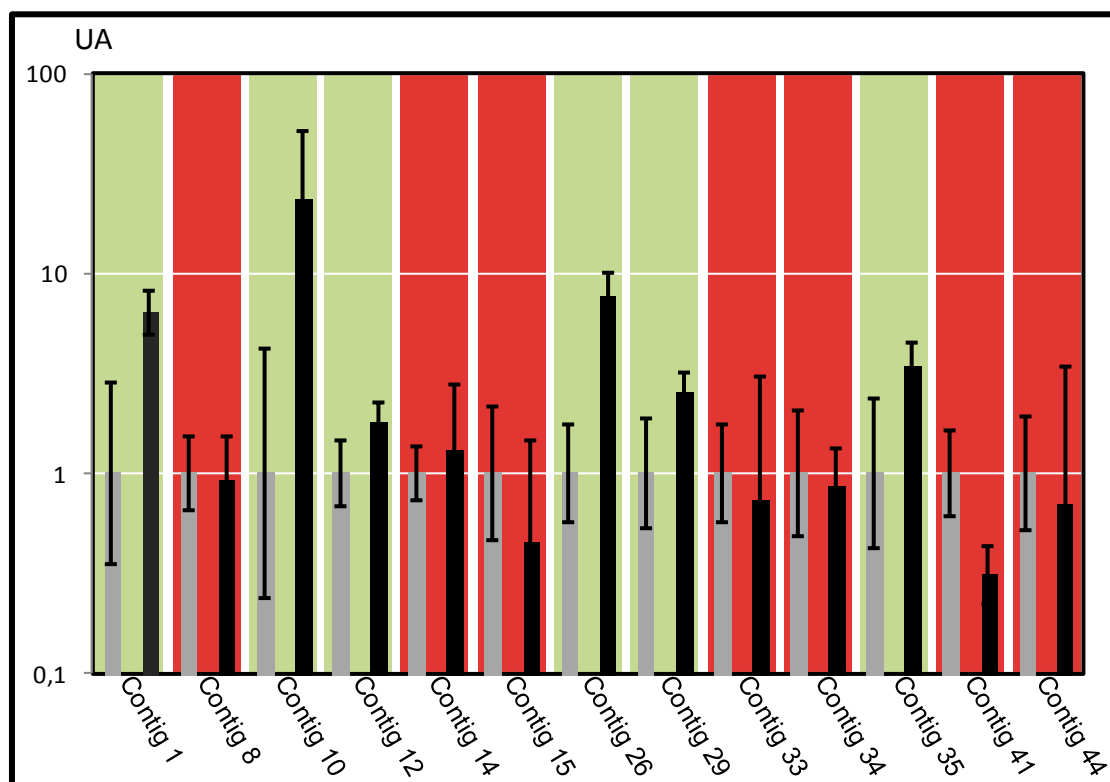
Amorces	% efficacité	$R^2$		10 <sup>ème</sup>	100 <sup>ème</sup>	1000 <sup>ème</sup>	10000 <sup>ème</sup>	100000 <sup>ème</sup>
C1	97,612	0,99						
C8	95,825	0,994						
C10	101,269	0,997						
C12	93,625	0,997						
C13	94,271	0,993						
C14	96,261	0,993						
C15	100,231	0,985						
C16	122,935	0,988						
C18								
C20	103,269	0,989						
C25								
C26	102,34	0,992						
C29	95,165	0,996						
C32								
C33	94,619	0,996						
C34	100,804	0,998						
C35	100,962	0,998						
C38								
C39								
C41	84,017	0,996						
C42	99,511	0,996						
C44	98,764	0,998						



Pour la majorité de ces couples d'amorces, l'absence d'amplification est à noter pour une quantité d'ADN génomique matriciel de 1 pg, ce qui semble être la quantité limite pour l'hybridation de ces amorces. En compilant ces différents résultats, 17 couples d'amorces ont été retenus sur les 22 initiaux, soit **17 séquences candidates**.

#### 2.1.2.2. Analyse du profil d'expression des séquences candidates

Le profil d'expression des séquences candidates a été évalué par RT-qPCR en utilisant le système d'expression différentielle validé précédemment pour la surexpression de gènes impliqués dans la synthèse de furocoumarines (cf partie A paragraphe 2.3.4.). Les ARNs totaux extraits de racines de panais stressés ou non ont été rétrotranscrits en ADN complémentaires (ADNc).



**Figure 40: Quantification relative des différentes séquences candidates dans les racines de panais élicitées mécaniquement (gris foncé) ou non élicitées (gris clair).**

L'axe des abscisses est représenté en log10. La couleur verte indique un résultat statistiquement significatif, la couleur rouge un résultat non statistiquement significatif. L'ARN 5.8s de panais a été utilisé comme gène de référence. L'erreur standard est représentée par les barres noires.

Les résultats de la PCR quantitative (Figure 40) n'ont pas permis d'observer une amplification pour les contigs C13, C16, C20 et C42. Etant donné que l'efficacité des amorces a été démontrée précédemment à partir d'ADN génomique, l'absence de signal pour ces 4 contigs peut s'expliquer soit par une expression trop faible pour être détectée par PCR quantitative soit par une absence d'expression des gènes candidats associés aux contigs.

Concernant les 13 autres séquences, les résultats indiquent une surexpression plus ou moins marquée pour les contigs 1, 10, 12, 26, 29 et 35 qui est cohérente avec celle des gènes déjà caractérisés de la voie de biosynthèse des furocoumarines (*PsDiox*, *PsPT1*, *PsPT2*, *CYP71AJ3*, *CYP71AJ4*) (Figure 25). Également, l'intensité de la réponse à l'élicitation est très variable en fonction des séquences étudiées. Par exemple, le niveau d'expression du contig C1 et C26 est 7 fois plus élevé que la condition contrôle, alors que celui de C10 est multiplié par 23.

**L'ensemble de ce travail m'a finalement amené à me focaliser sur ces seules 6 séquences candidates : C1, C10, C12, C26, C29 et C35.**

#### **2.1.2.3. Affiliation des candidats à des familles de P450 : vers une prédiction fonctionnelle ?**

Afin de déterminer leur appartenance à une famille de P450 en particulier, j'ai réalisé une analyse par l'algorithme BLAST à partir de différentes bases de données (Tableau 8).

**Tableau 8: Résultats de l'analyse des séquences nucléotidiques des contigs candidats par l'algorithme BLAST.** Le pourcentage d'identité est basé sur la comparaison de séquences nucléiques.

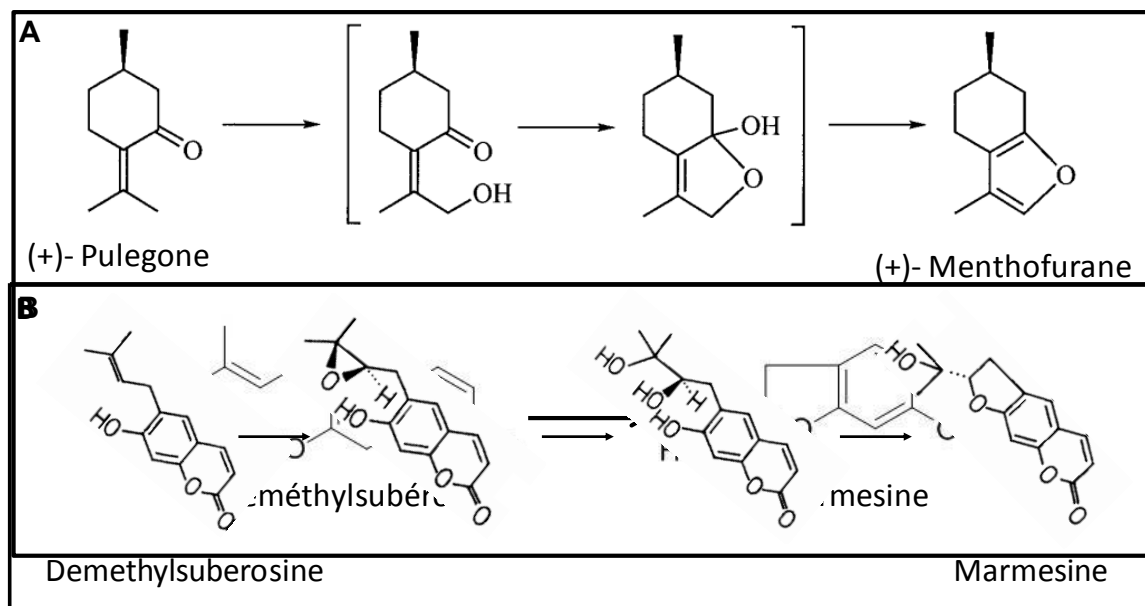
Séquences candidates	Résultats BLAST	Identité
Contig C1	CYP71A9-like ( <i>Daucus carota</i> )	87%
Contig C10	Menthofuransynthase-like ( <i>Daucus carota</i> )	86%
Contig C12	CYP71A8-like ( <i>Daucus carota</i> )	87%
Contig C26	CYP71D98 ( <i>Ammi majus</i> )	91%
Contig C29	CYP81D1-like ( <i>Daucus carota</i> )	90%
Contig C35	CYP82H1 ( <i>Ammi majus</i> )	88%

Les résultats indiquent que C1 et C12 sont respectivement homologues à un gène identifié comme un *CYP71A9*-like et un *CYP71A8*-like de la carotte (*Daucus carota*). Comme prévu initialement, ces deux gènes appartiennent au Clan 71 des cytochromes P450 et plus précisément à la famille CYP71A. Le gène *CYP71A9* a été caractérisé pour la première fois chez le soja (Fabacées) (Schopfer and Ebel, 1998). Ce gène est décrit comme étant impliqué dans la synthèse de glyceollines, une famille de composés terpéniques considérés comme des phytoalexines. La transcription de *CYP71A9* chez le soja est élicitable par blessure mécanique. *CYP71A8* a lui été isolé du pois Mascate (*Mucuna pruriens*, fabacées) mais n'a pas été étudié à ce jour.

Les contig C26 et C35 sont homologues aux gènes *CYP71D98* et *CYP82H1*. Ces gènes, identifiés chez *Ammi majus*, ont été cités par Sandra Kellner et Ulrich Matern dans le cadre d'approches de DD-RT-PCR (Differential Display Reverse Transcription PCR) (Travaux de Sandra Kellner et Ulrich Matern, 2006, Thèse). Ces gènes ont déjà été exprimés au Laboratoire Agronomie et Environnement par Romain Larbat et Guilhem Vialart dans le cadre de leurs thèses respectives.

CYP71D97 est en mesure de métaboliser la 5-hydroxyflavone en une molécule hydroxylée qui n'a pas été identifiée. CYP82H1 est en revanche encore orphelin de fonction. Le contig C29 est homologue à 90% au gène *CYP81D1*-like de la carotte. Ce dernier a été cloné chez *Arabidopsis thaliana* et caractérisé comme étant impliqué dans la synthèse de glucosinolates (Wiesner *et al.*, 2014), intervenant dans la protection de la plante contre des bioagresseurs.

Enfin, le contig **C10** est homologue à 86% à la menthofurane synthase-like (MFS) de la carotte. La menthofurane synthase, mise en évidence chez une lamiacée (*Mentha x piperita*, ou menthe poivrée) est impliquée dans la synthèse de menthofurane, un monoterpène à haute toxicité utilisé notamment comme insecticide (Bernays, 2017), à partir de pulégone (Bertea *et al.*, 2001). L'intérêt de l'homologie entre le contig C10 et la menthofurane synthase réside dans les mécanismes réactionnels de cette dernière, particulièrement similaire à ceux de la marmesine synthase (Figure 41) et de la columbianetine synthase. Ces mécanismes réactionnels sont définis par la cyclisation d'un groupement prényle sur un groupement hydroxyle. Cette similarité de mécanismes réactionnels fait du contig C10 un sérieux candidat.



**Figure 41: Comparaison des mécanismes réactionnels de la menthofurane synthase (A) et de la marmesine synthase (B).**

### **2.1.3. Recherche de séquences codantes complètes**

Bien que résultant de l'assemblage de différents contigs, la longueur moyenne des gènes cibles ainsi obtenus est de 800 nt environ, et ne correspondent donc pas à un gène complet (environ 1500 nt). Pour cloner les séquences codantes complètes, et étant donné que le génome du panais n'est à ce jour pas séquencé, une nouvelle banque transcriptomique de panais a été générée, en collaboration avec le Dr. Sylvain Legay du Luxembourg Institute of Science and Technologies (LIST, Belvaux, Luxembourg).

#### **2.1.3.1. Construction de la banque, séquençage et assemblage de la banque**

Cette banque transcriptomique a été établie à partir d'ARN totaux extraits de racines de panais élicitées mécaniquement. Bien qu'il ne s'agisse pas ici d'une banque d'expression différentielle, cet outil peut permettre d'avoir un premier aperçu des gènes exprimés dans ces conditions d'élicitation, démontrées précédemment comme favorables à la suraccumulation de furocoumarines. Préalablement au séquençage, la qualité et la concentration de la banque a été évaluée. La concentration en ARNm a été évaluée à 37nM.

Les segments d'ADNc composant cette banque ont une taille comprise entre 177 et 1000 pb, avec une moyenne de 379 pb. La banque d'ADNc a été séquençée par la méthode Illumina. Ce séquençage a généré 630 millions de séquences d'une taille moyenne de 150 pb. Dans le but d'obtenir des segments d'ADNc plus long, les séquences brutes issues du séquençage ont été assemblées en contigs selon les paramètres décrits au paragraphe 2.1.14 du Matériel et Méthodes. L'assemblage a permis de générer 71 579 contigs, de taille comprise entre 257 et 15398 nucléotides, pour une valeur moyenne de 867 pb. La distribution des données est définie par un premier quartile égal à 2094 pb, une valeur médiane de 1232 pb et un deuxième quartile de 582 pb. Finalement, les différents contigs ont été analysés et annotés par l'algorithme Blast *via* la suite logiciel Blast2Go.

### 2.1.3.2. Analyse de la banque RNAseq de panais (Illumina)

#### 2.1.3.2.1. Analyse globale

Dans un premier temps, j'ai entrepris la recherche de plusieurs gènes connus de la voie de biosynthèse des furocoumarines pour évaluer la validité de cette banque. Pour cela, les séquences des gènes *CYP71AJ3*, *PsDiox* et *PsPT1* ont été analysées par l'algorithme BLAST à l'aide du logiciel BioEdit. Les résultats indiquent que, pour chacun des gènes, une séquence homologue à 100% avec une e-value de  $0^E0$  est présente dans la banque transcriptomique (*CYP71AJ3* : contig 1214, *PsDiox* : contig 48, *PsPT1* : contig 1098). La e-value est définie par la probabilité que deux séquences soient homologues par hasard. Ainsi, plus la e-value est proche de 0, plus le résultat du BLAST tend vers une unique possibilité.

Etant donné que les séquences retrouvées dans la banque RNAseq sont parfaitement identiques aux séquences cibles connues, nous pouvons considérer que l'assemblage de la banque est robuste. La banque peut donc être utilisée pour la recherche de nouveaux gènes. Dans un second temps, j'ai répertorié les différentes familles de gènes retrouvés dans la banque. Pour commencer, parmi les 71 579 séquences disponibles, seules 44 564 ont été annotées par BLAST via le logiciel Blast2Go.

Les familles de gènes putatives représentées dans la banque transcriptomique de panais sont extrêmement diversifiées, couvrant de nombreux processus du métabolisme primaire, tels que la modification de la chromatine (38 gènes), la synthèse protéique (ex: facteurs d'élongation, 73 gènes), ou la synthèse de cellulose (55 gènes). On y retrouve également des gènes impliqués dans le métabolisme spécialisé comme la germacrène synthase (12 gènes), composé à activité antimicrobienne et insecticides (Morteza-Semnani and Ghanbarimasir, 2019), ou de la synthèse du shikimate (17 gènes), précurseurs de plusieurs voies métaboliques du métabolisme spécialisé.

Dans le cadre de mes travaux de thèses, je me suis focalisé sur les gènes apparentés aux cytochromes P450. Sur les 312 séquences de cytochromes P450 disponibles dans la banque, 46 correspondent à des séquences de CYP71, 7 de CYP81 et 19 de CYP82.

#### **2.1.3.2.2. Détermination de la séquence complète des contigs candidats**

Cette collection a permis de cibler les séquences les plus homologues aux séquences candidates. L'obtention de ces séquences issues de la banque RNAseq Illumina a été déterminante pour compléter les séquences partielles des contigs candidats. La recherche d'homologie entre les séquences candidates et les séquences issues de la banque RNAseq Illumina a été réalisé par l'algorithme BLAST *via* le logiciel BioEdit. De cette façon, la séquence de chacun des contigs candidats a été complétée, exceptée pour CYP71A-C12. Chacune des séquences extraites de la banque Illumina a été analysée par le logiciel « Expasy translate tool » permettant de prédire le cadre de lecture le plus probable. Ces prédictions ont été confirmées par l'alignement des séquences génomiques d'intérêt, incomplètes, avec les gènes desquels les contigs candidats sont le plus homologues (Figure 42). Finalement, cette approche transcriptomique a permis de sélectionner **5 séquences candidates** (Annexe 9). Les différentes étapes de cribles sont résumées dans le Tableau 9.

1	130
GG10	
Sequence	ATGCAGTTAT CAGCATGGCT GCAACCTCTC ATTATCTTTC TCATACCCCT TTCTTAACT ATCTTCGTAT TTAACCATA TTTTGCTACT AAACCTGGTA AAAACCTCCC ACCCTCTCCG CCAAGCTTC
Consensus	.....
131	260
GG10	
Sequence	CTATCATTGG CCATTTGCAC AAACCTGGAA GCTTACCACA CCGTCCGCTC CTCAGCTTGT CTAGAAAATA TGGACCACTC ATGCTACTTC ACCTTGGAA GATTCCGACT CTAGTTGTTT CATCTGCTGA
Consensus	.....
261	390
GG10	
Sequence	GGCAGCTCAA GAGATCATGA AAACCTACGA TGTCAAGTTT GCGGACAGGG TTCGAACATA TGCTAATGAG AAGCTTCTTT ATTGTTATAA GGATGTGTCT ATGGCACCTT ATGGCGACTA CTGGAGAAAA
Consensus	.....
391	520
GG10	
Sequence	ATGAAAAGCT TTTGTGTTGT CCAGCTTTTG AGTAACAAAA GGGTTGAAGA TATGGGAACG ATAAGAGAGG AAGAACTGC TTTTCTACTG AAAATATTC TAGACTCTAA TTCTTCGGTT GTGAACCTTGA
Consensus	.....
521	650
GG10	
Sequence	GTGAACTGAT AATCACACTC ACATTCAATC TTGTTCTAC AGCAGCTTTT GGGAAAAAG TTGATGAATG GGGGAAAGGC AGAAAGTTAT GGATGCTAAT AAAGAGACG GAGGAATTAC TTGGCCGATT
Consensus	.....
651	780
GG10	
Sequence	TGATGTTGGA ACCTTTACTC CATGGCTTGC ATGGATCAAT ACTGTTTCAG GTTCAATGC CAAAGTAAAA AACATTGCAA AAGAGAACGA TGAATTTTC GAACGGATTG TGGAGAGCA CATAAATGCT
Consensus	.....
781	910
GG10	
Sequence	TCAACAAGAA AGAAGAAAGG CGGTGAGAAA ATTAAGAAAA CAGAGGATT TCTCGATATC TTGCTTGGAG CACAAGAGGA TGCCATCGGT GGCTTTCATA TTGATAGAGA TGGCATCAAA GCTCTGATCC
Consensus	.....
911	1040
GG10	
Sequence	TGGATATGAT TGCTGGTGA ACTGACACTA CCTACACAGC TCTAGAATGG ACAATGACGG AACTTTTAAG GCACCCGAAA GCCATGAAGG AAGTGCAAAA TGAGATAAGG AATATTGCA GAGGAAAAATC
Consensus	.....
1041	1170
GG10	
Sequence	TAACATATCT TATCCAGATG ACTTGGAAAA CATGCACTAC TTGAAGCAG TGATTAAGAA GGCATACGG TTACATCCAC CACTTCCGTT ATTAGTCCCA CGAACTGCAA GACAGGATGT TAAAGTTATG
Consensus	.....ACGG TTACATCCAC CACTTCCGTT ATTAGTCCCA CGAACTGCAA GACAGGATGT TAAAGTTATG
1171	1300
GG10	
Sequence	GGCTACGATA TTGCAGCTGG AACTATGGTC ATCATCAACG CTTGGGGAAT CGGAAGGGAC CCTAGTTTAT GGGATGAACC GGATGAGTTT AGGCCAGAAA GGTTCTTGAA TTCGACAATA GATTACAGAG
Consensus	GGCTACGATA TTGCAGCTGG AACTATGGTC ATCATCAACG CTTGGGGAAT CGGAAGGGAC CCTAGTTTAT GGGATGAACC GGATGAGTTT AGGCCAGAAA GGTTCTTGAA TTCGACAATA GATTACAGAG
1301	1430
GG10	
Sequence	GTCATGACTT CCAGTTGATC CCAATTTGGTG CTGGGAGGAG GGGTTGTCCC GGAATAAGTT TTGCCATGTC AAATTACGAG TATGTGTTGG CAAATCTTCT GCATAAGTTT GATTGGAAAC TTCCTAATGG
Consensus	GTCATGACTT CCAGTTGATC CCAATTTGGTG CTGGGAGGAG GGGTTGTCCC GGAATAAGTT TTGCCATGTC AAATTACGAG TATGTGTTGG CAAATCTTCT GCATAAGTTT GATTGGAAAC TTCCTAATGG
1431	1560
GG10	
Sequence	AATGAAGAGT GATGAGCTTA ATATGGATGA GGGCACTGGT CTTACTTGTC ATAGAAAAGT CCCTCTACTT GTAGTTGCAA CTCCAATTTT TTGTTAGCA GACTGTCTTT TATTCTTCTA TCATGTGATG
Consensus	AATGAAGAGT GATGAGCTTA ATATGGATGA GGGCACTGGT CTTACTTGTC ATAGAAAAGT CCCTCTACTT GTAGTTGCAA CTCCAATTTT TTGTTAGCA GACTGTCTTT TATTCTTCTA TCATGTGATG
1561	1669
GG10	
Sequence	CAGTTCGGGT CCTCTCTGTT TCAGGTCCAT TTAATGATT TACAAAATAA AGGTCAAGAA TACAAAGATA ATTAATAGN AAAAAAAAAA AAAAAACAAA AAAAAAAA
Consensus	CAGTTCGGGT CCTCTCTGTT TCAGGTCCAT TTAATGA.....

**Figure 42 : Alignement de la séquence du contig C10 avec la séquence correspondante issue de la banque transcriptomique de panais générée avec la méthode de séquençage Illumina. La couleur rouge indique une parfaite homologie entre les séquences alignées. La couleur bleue indique les différences de nucléotides entre les séquences. Les points noirs indiquent des zones non homologues.**



**Tableau 9 : Synthèse des étapes de sélection des séquences candidates à partir des séquences validées par PCR**

	Nombre de séquences
Séquences validées par PCR	22
Séquences conservées après vérification de la validité des amorces	17
Séquences amplifiées par RT qPCR	13
Séquences surexprimées en condition d'éllicitation par blessures mécaniques	6
Séquences codantes déterminées grâce à la banque RNAseq de panais Illumina	5

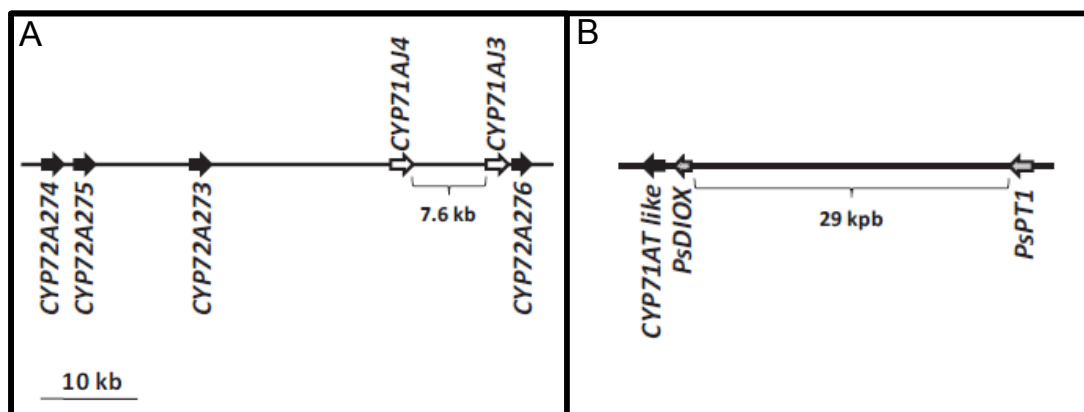
En résumé, le travail réalisé dans cette partie a permis de mettre en évidence 22 gènes codants pour des P450s sur exprimés dans des tissus élicités. L'exploitation des bases de données disponibles ou construites dans le cadre de ma thèse m'a amené à identifier **5** séquences candidates complètes, à savoir *CYP71A-C1* (homologue à *CYP71A9*), *CYP71-C10* (homologue à une MFS), *CYP71D-C26* (homologue à *CYP71D98*), *CYP81D-C29* (homologue à *CYP81D1*) et *CYP82H-C35* (homologue à *CYP82H1*).

## 2.2. Approche génomique

### 2.2.1. Contexte et identification de *CYP71AT*

L'objectifs des travaux décrits dans Roselli *et al* ( 2017) était de voir si les gènes impliqués dans la synthèse des furocoumarines étaient organisés sous la forme de clusters. Pour atteindre cet objectif, une banque génomique (BAC) a été construite en collaboration avec l'équipe d'Hélène Bergès (CNRGV, Toulouse). L'exploration de cette banque a mis en évidence la proximité des gènes *CYP71AJ3* et *CYP71AJ4* sur le même chromosome et qui ne sont distants que de 7.6 kb (Figure 43 A). L'analyse a également montré que *PsPT1* était très proche *PsDiox*, distant de 29 kpb.

Cette analyse nous a amené à formuler l'hypothèse de 2 clusters de gènes pour la voie de biosynthèse des furocoumarines: un premier cluster pour les gènes en aval, communs à de nombreuses voies de biosynthèse et un second cluster spécialisé pour la voie de biosynthèse de furocoumarines.

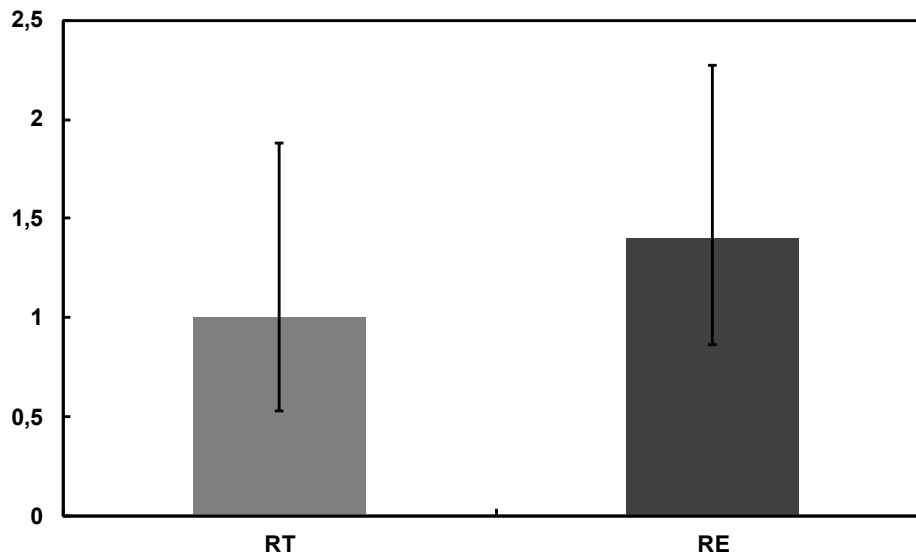


**Figure 43: Localisation des gènes *CYP71AJ3*, *CYP71AJ4*, *PsDiox*, *PsPT1* et *CYP71AT* sur le génome du panais. Les figures A et B correspondent à deux BACs différents.**

L'analyse de la séquence du BAC portant *PsPT1* et *PsDIOX* a permis de mettre en évidence la présence d'une séquence codant un cytochrome P450. Ce gène est localisé à 800 pb du gène *PsDiox* et a été affilié à la sous-famille CYP71AT (D. Nelson, communication personnelle) (Annexe 10). Dans une logique d'organisation en cluster dédié à la synthèse de furocoumarines, ce gène nous a semblé particulièrement intéressant.

### 2.2.2. Analyse du profil d'expression

Une fois la présence de *CYP71AT* établie sur le génome du panais, j'ai analysé son profil d'expression par RT-qPCR à partir d'ARNs extraits de racines de panais élicitées ou de racines non traitées. Les résultats (Figure 44) n'indiquent pas de différence significative entre les 2 modalités testées.



**Figure 44: Profil d'expression de *CYP71AT* en PCR quantitative à partir d'ARN extraits de racine de panais contrôle (RT) et élicitées mécaniquement (RE).**  
*L'ARN 5.8s de panais a été utilisé comme gène de référence. L'erreur standard est représentée par les barres noires.*

En complément de l'approche transcriptomique, l'approche génomique m'a permis de cibler un gène candidat, *CYP71AT*, mis en évidence grâce à l'analyse d'une banque génomique de panais. Ce gène est distant de seulement 800 pb de *PsDiox*, un autre gène impliqué dans la synthèse de furocoumarines chez le panais. Bien que la surexpression de ce gène en condition d'élicitation ne soit pas avérée, sa proximité avec *PsDiox* et son appartenance à la famille des CYP71 en fait un sérieux candidat.

### 3. Clonage et caractérisation fonctionnelle des gènes candidats

L'étape suivant l'identification de gènes candidats et l'obtention des séquences codantes complètes consiste à les cloner et exprimer les protéines correspondantes avant de réaliser un criblage fonctionnel.

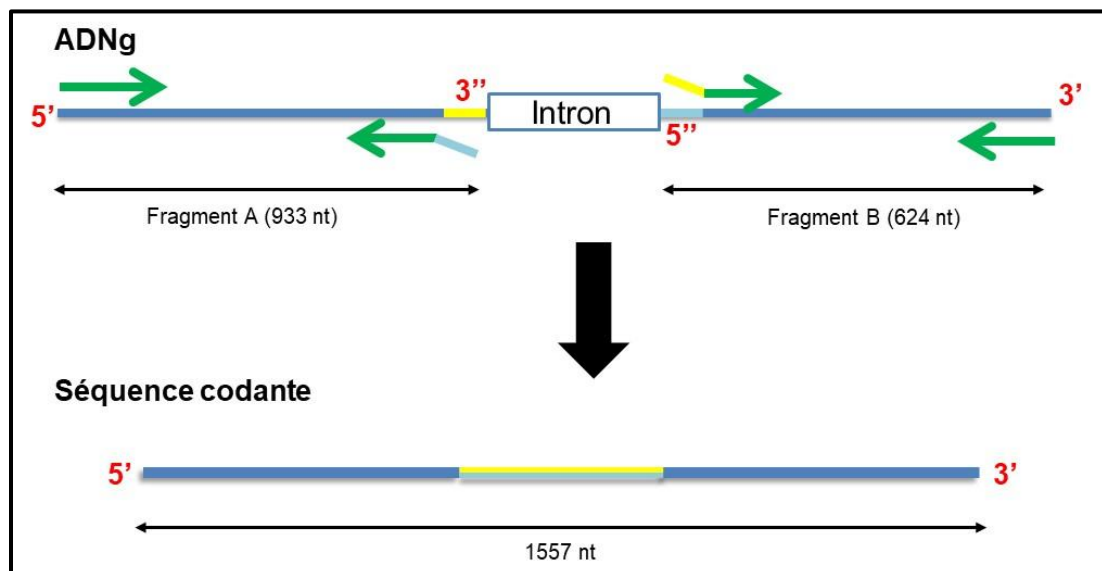
### 3.1. Clonage dans le plasmide pYeDP60\_GW<sup>®</sup>

#### 3.1.1. Clonage de CYP71-C10.

Dans un premier temps, j'ai entrepris, à partir d'ADNc issus de racines de panais élicité, d'amplifier la séquence codante des différentes séquences candidates. Seule la séquence du contig CYP71-C10 a pu être amplifiée dans ces conditions. Une fois cloné dans le plasmide pCR<sup>™</sup>8/GW/TOPO<sup>™</sup> et séquencé, la séquence codante a été introduite dans le plasmide pYeDP60\_GW<sup>®</sup> par recombinaison puis introduit dans des levures WAT11.

#### 3.1.2. Clonage de *CYP71AT*

Bien que l'expression de *CYP71AT* ait pu être mise en évidence par RT-qPCR, le clonage de la séquence codante complète a été impossible. Pour contourner ce problème, j'ai reconstruit la séquence codante de ce gène à partir d'ADN génomique. Pour cela, j'ai synthétisé deux couples d'amorces à partir des séquences présentes de chaque côté de l'intron du gène *CYP71AT*. La position de l'intron a été déterminée par la plateforme en ligne Gene Ontology.



**Figure 45 : Reconstruction de la séquence du gène *CYP71AT* à partir d'ADN génomique de panais.** Les flèches vertes désignent les amorces utilisées pour l'amplification par PCR de chacun des fragments. Le fragment jaune correspond à la séquence reconstruite.

Pour le couple d'amorces du fragment A (Figure 45), une séquence de 20 nt, complémentaire aux 20 premiers nucléotides du fragment B, a été ajoutée à la séquence de l'amorce antisens.

De la même façon, une séquence complémentaire aux 20 derniers nucléotides du fragment A a été ajoutée à la séquence de l'amorce sens du fragment B. Une première réaction PCR a permis d'amplifier chacun des deux fragments séparément. Les fragments ont ensuite été purifiés à partir du gel d'agarose et incorporés dans une nouvelle réaction PCR.

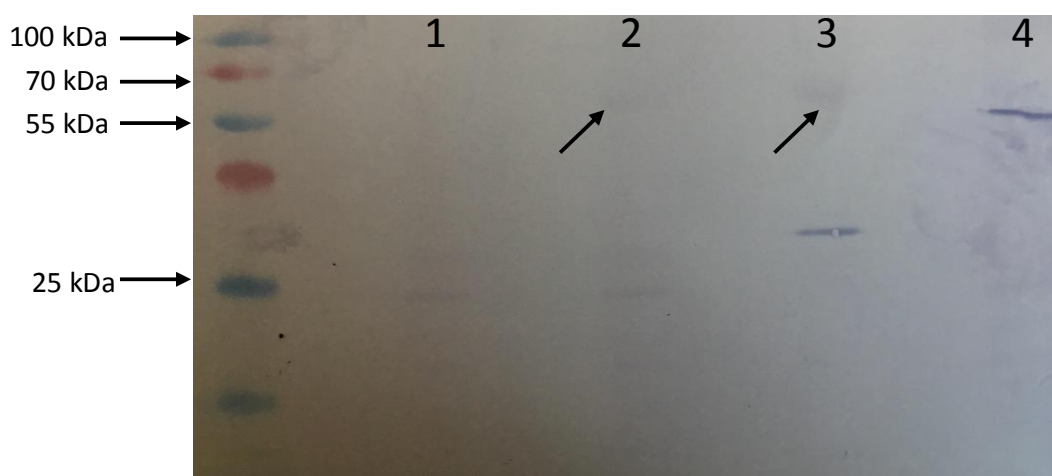
Cette dernière permet, grâce aux extrémités débordantes additionnelles des amorces antisens du fragment A et sens du fragment B, d'accoler les deux fragments pour reconstruire la séquence codante du gène *CYP71AT*. La seconde PCR est fractionnée en deux étapes. La première étape va, grâce à une température d'hybridation spécifique, permettre l'appariement des amorces portant une extrémité débordante avec les régions des fragments A et B desquelles elles sont complémentaires. L'action de la polymérase va alors permettre l'accolement des deux fragments. La deuxième étape fait intervenir les amorces dépourvues d'extrémités débordantes qui, grâce à une température d'hybridation différente de celle de la première étape, vont permettre d'amplifier les deux fragments accolés à partir de l'extrémité 5' du fragment A et 3' du fragment B, et ainsi reconstruire la séquence codante complète du gène *CYP71AT*. Le produit de PCR a été cloné dans le plasmide pCR<sup>TM</sup>8/GW, séquencé puis transféré dans le vecteur pYeDP60\_GW.

## **3.2. Expression des protéines dans la levure**

### **3.2.1. Analyse de l'expression**

Les 2 gènes candidats *CYP71AT* et *CYP71-C10* ont été exprimés dans WAT11. La culture des levures a été réalisée suivant le protocole décrit dans le matériel et méthode. J'ai confirmé l'expression des protéines par immunodétection en utilisant un sérum dirigé contre une série de 6 histidines introduite dans la séquence codante en amont du codon stop. Les résultats obtenus confirment la production d'un cytochrome P450 contrôle (*CYP71AZ4*).

Son poids moléculaire apparent est de 55 kDa et correspond à la taille prédite de la protéine. Dans le cas de CYP71AT et CYP71-C10 (piste 2 et 3), on peut observer un signal très faible à 55 kDa (flèches noires) et qui est absent du témoin négatif. Ce résultat très mitigé ne permet pas de conclure quant à la présence ou à l'absence des protéines cibles dans les échantillons analysés. Dans le cas de CYP71-C10, on observe aussi un signal beaucoup plus intense de 30 kDa qui pourrait correspondre à un produit de dégradation de la protéine.



**Figure 46: Immunodétection des protéines CYP71AT et MFS (C10).**

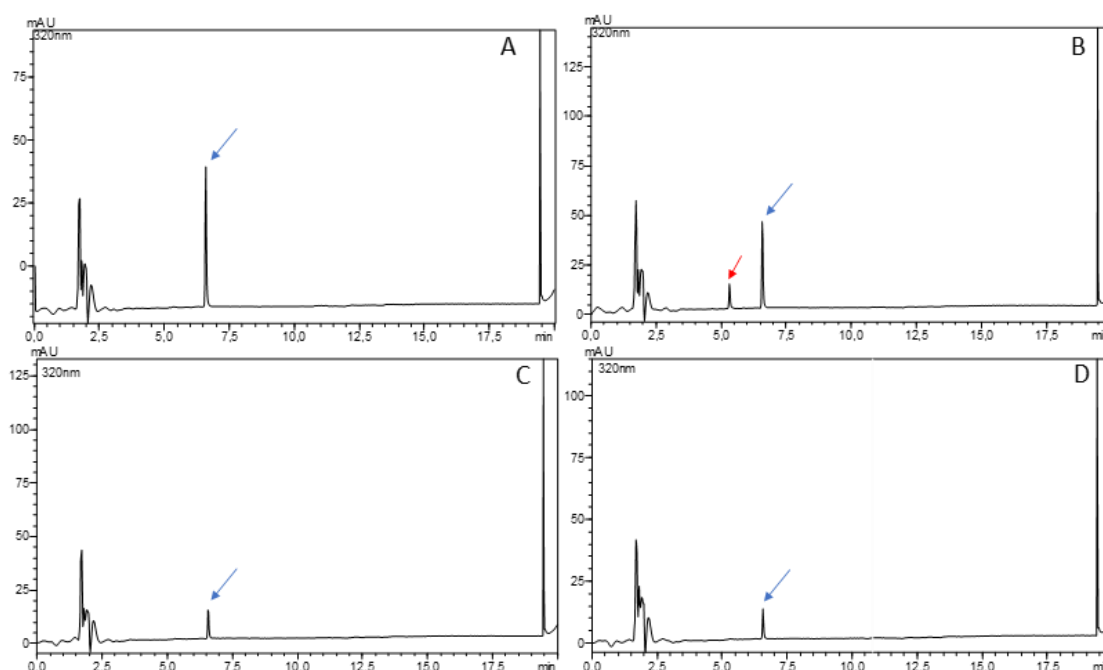
*Piste 1 : PyeDP60\_GW vide, piste 2 : C10, piste 3 : CYP71AT, piste 4 : témoin positif CYP71AZ4.*

### **3.2.2. Criblage métabolique : Etude de l'activité métabolique des gènes candidats**

Malgré plusieurs tentatives de production et la faible expression des protéines, nous avons tout de même choisi de réaliser un criblage fonctionnel de CYP71AT et CYP71-C10. Pour cela, les microsomes préparés à partir de levure « produisant » ces P450s ont été incubés par une large gamme de molécules incluant notamment les intermédiaires de la voie de biosynthèse des furocoumarines (DMS, psoralène, angélicine, fraxétine, osthénol, xanthotoxol, xanthotoxine, bergaptol et bergaptène).

Afin de confirmer le système, nous avons systématiquement réalisé des incubations témoins avec le couple CYP71AZ4/psoralène permettant de suivre la synthèse de xanthotoxol (Figure 47 A et B) en présence de NADPH.

Les criblages réalisés dans les mêmes conditions avec CYP71AT et CYP71-C10 n'ont à ce jour pas permis de mettre en évidence de métabolisation. Un exemple des résultats obtenus pour CYP71AT, incubé en présence de psoralène, sans et avec NADPH est présenté en figure 47(C et D respectivement). L'ensemble des résultats de métabolisation des différents substrats, ainsi que leurs formules chimiques et leurs masses sont résumés en annexe 11.



**Figure 47 :Analyse des produits de métabolisation de CYP71AT par UPLC (320nm).**  
A : CYP71AZ4 sans NADPH, B ; CYP71AZ4 avec NADPH, C : CYP71AT sans NADPH, D : CYP71AT avec NADPH. Les flèches bleues correspondent au substrat utilisé (psoralène). La flèche rouge correspond au produit de métabolisation, ici le xanthotoxol.

## 4. Discussion et conclusion

L'objectif de ce travail était d'identifier d'autres gènes impliqués dans la voie de biosynthèse des furocoumarines afin d'en permettre la reconstruction complète dans une plante qui n'en produit pas. Identifier des gènes candidats qui pourraient jouer un rôle dans la synthèse de ces molécules est un défi. J'ai très rapidement été confronté à 2 freins qu'il a fallu débloquent. Le premier concerne l'absence d'information sur cette voie de biosynthèse. Ne pouvant réaliser un travail exhaustif sur l'ensemble des P450s, j'ai émis l'hypothèse que les P450s impliqués dans leur synthèse appartenaient à un clan en particulier. Le second frein concerne le peu de ressources nucléotidiques disponibles concernant le panais. L'analyse d'une banque RNAseq construite par Célia Krieger a permis d'identifier 22 séquences candidates.

L'inductibilité de la synthèse de furocoumarine mise en évidence dans le premier chapitre de ce manuscrit m'a permis de restreindre à 6 le nombre de gènes ayant un profil d'expression corrélé avec celui de gènes déjà connus de la voie de biosynthèse des furocoumarines chez le panais. Malheureusement, la qualité de la banque RNAseq 454 n'a pas permis d'accéder aux séquences complètes. Pour pallier ce problème, nous avons construit en collaboration avec le LIST une nouvelle banque reposant sur la méthode de séquençage Illumina. En dehors du fait de fournir les séquences codantes complètes des gènes candidats identifiés cette banque permet maintenant de disposer de nouvelles ressources jusqu'ici inaccessibles. Cette banque permettra notamment d'approfondir et d'élargir les connaissances sur la voie de biosynthèse des furocoumarines chez les panais, sur les éléments qui la régule (facteurs de transcription, ...) mais également sur d'autres processus du métabolisme primaire et secondaire et ainsi de compléter les connaissances sur le métabolisme de cette plante obtenue dans le cadre du premier chapitre de mémoire. Par exemple, dans cette banque sont retrouvés des gènes codant la 4-CoumaroylCoA ligase (4CL, 23 gènes) ou la Cinnamate-4-hydroxylase (C4H, 5 gènes), deux enzymes pas encore caractérisées chez le panais, et qui sont des protéines clés de plusieurs voies de biosynthèses telles que les flavonoïdes (Sircar et Mitra, 2009), la lignine (Hu *et al.*, 2017) mais également les furocoumarines.



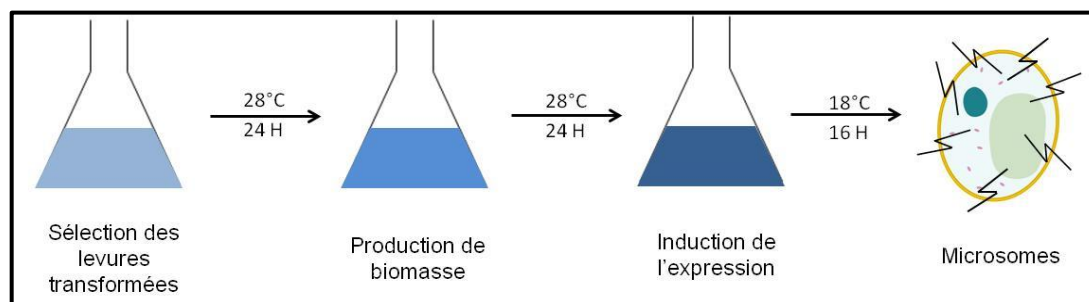
Cette description de la banque transcriptomique Illumina de panais est loin d'être exhaustive, et une analyse plus profonde permettrait d'identifier de nombreux leviers quant à la compréhension moléculaire de la voie de biosynthèse des furocoumarines chez le panais. Egalement, cette banque pourra être combinée avec les résultats de l'approche métabolomique présentée en partie A dans le but de construire un réseau associant métabolomique et transcriptomique. Potentiellement, ce réseau nous permettrait d'identifier de nouveaux gènes de la voie des furocoumarines mais également d'autres voies de biosynthèses du métabolisme primaire et spécialisé.

Cette approche a notamment été utilisée chez le tabac pour identifier les différentes voies métaboliques impliquées dans la réponse au froid (Jin *et al.*, 2017).

La production des protéines CYP71AT et CYP71-C10 a été évaluée par immunodétection. Malgré de nombreux essais et à la différence du témoin positif CYP71AZ4, nous n'avons jamais pu conclure fermement quant à la production des enzymes d'intérêt même si un signal correspondant au poids moléculaire attendu est observable. L'absence de protéine détectable peut ne pas être limitant pour mettre en évidence une métabolisation de substrat si l'enzyme est très efficace. Malheureusement, dans notre cas, nous n'avons pu détecter aucune activité enzymatique avec les différentes molécules testées. Si on part de l'idée que la protéine est tout de même produite en quantité très faible, l'absence de métabolisation de molécules peut être expliquée de différentes manières.

- L'explication la plus évidente est que les substrats utilisés ne sont pas les substrats des enzymes étudiées. Le métabolisme secondaire étant extrêmement riche il est envisageable que des intermédiaires non connus interviennent dans cette voie de biosynthèse. Ainsi, un élargissement de la gamme de substrats testée, principalement d'autres coumarines et furocoumarines dans un premier, pourrait permettre de caractériser l'activité de ces enzymes. Le cas échéant, la gamme pourra être étendue à d'autres voies biosynthèses partageant des précurseurs communs avec celle des furocoumarines. Une analyse métabolique différentielle, à partir de panais élicité ou non par blessures mécaniques, permettra de cibler les composés phénoliques suraccumulés dans ces conditions et ainsi de préciser la gamme de substrats à tester.

- Etant donné que l'amplification du gène *CYP71AT* à partir d'ADNc s'est révélée négative, la séquence codante du gène a été reconstruite à partir d'ADN génomique et sur la base de données *in silico*. Ainsi, la position de l'intron est issue d'une prédiction de la plateforme Gene Ontology. Dans le cas où la prédiction de la position de l'intron est erronée, la reconstruction de la séquence codante de *CYP71AT* peut entraîner un décalage du cadre de lecture, conduisant à la formation d'une protéine tronquée et donc non fonctionnelle. Il est de fait important de confirmer cette séquence en trouvant un moyen d'amplifier le cDNA natif à partir de différents types de tissus ou de différentes variété de panais.
- Le niveau d'expression des protéines peut être limitant. Pour tenter de l'améliorer, les séquences peuvent être optimisées de différentes manières. Par exemple l'optimisation du biais de codon est une approche largement utilisée et décrite dans la littérature. Du fait de la dégénérescence du code génétique, un acide aminé peut être codé par plusieurs codons, pouvant aller jusqu'à 6 codons alternatifs. De plus, l'utilisation plus ou moins affine de chacun des différents codons codant un acide aminé varie selon l'organisme étudié (Kumar and Sharma, 1995; Song *et al.*, 2018). Ainsi l'utilisation d'un codon préférentiel d'un organisme donné permet d'influer sur la vitesse de traduction. Néanmoins, cette optimisation doit être mise en œuvre de manière réfléchie car elle peut également devenir un frein comme je l'ai indiqué dans la discussion de la partie C de ce mémoire.
- Une autre manière d'améliorer la production des protéines consiste à adapter les conditions de cultures des levures. Le système d'expression a initialement été mis au point pour CYP73A1 (La cinnamate 4-hydroxylase de Topinambour) mais de nombreuses publications rapportent les difficultés liées à la mauvaise expression de P450 dans ce système (Figure 48).



**Figure 48: Représentation schématique de la production protéines recombinantes dans la levure**

L'étape de sélection des levures transformées, réalisée pendant 24 h à 28°C est suivie d'une deuxième étape pendant laquelle la production de biomasse est favorisée. L'augmentation de biomasse se traduit par une accélération de la reproduction des levures, asexuée, par bourgeonnement. Durant la multiplication du nombre de cellules de levures, le plasmide portant le gène d'intérêt peut être perdu par celle-ci au grès des divisions cellulaires. Le gène d'intérêt n'étant pas intégré de manière stable dans le génome de la levure, la perte du plasmide recombiné induit une perte de l'expression de ce gène. Ainsi, une diminution du temps d'incubation pour la production de biomasse est à envisager pour limiter la perte du plasmide recombiné (Kawahata *et al.*, 1999; Zhang *et al.*, 1996). De plus, les cytochromes P450 sont des enzymes fragiles, dotées d'une stabilité peu élevée et d'une sensibilité élevée à la température (Kumar, 2010; Rekka *et al.*, 2002). Un ajustement de la température et du temps d'incubation pour la production de protéines permettrait d'adapter cette production.

- Dans le but d'améliorer la production des protéines d'intérêts dans la levure WAT11, (Batard *et al.*, 2000) ont démontré que le remplacement de l'extrémité N-terminale, qui porte l'ancre membranaire, d'une C4H de blé (CYP73A17) par celle de topinambour (CYP73A1) permettait d'obtenir une meilleure production de la protéine. En effet, l'ancre membranaire est destinée à fixer la protéine dans la membrane du réticulum endoplasmique. Le remplacement de cette ancre membranaire pourrait permettre de modifier le positionnement de la partie globulaire de la protéine par rapport à la membrane du RE, entraînant de fait une modification du canal d'accès au site actif de l'enzyme.

Cette méthode a également été utilisée par (Larbat *et al.*, 2007) dans le but d'obtenir un niveau d'expression détectable de CYP71AJ1.

- Enfin, d'autres systèmes d'expression hétérologues peuvent être envisagés. L'utilisation du système basé sur les Baculovirus (Stolt-Bergner *et al.*, 2018) est décrites pour certains P450s. L'intérêt de ce système repose sur l'existence d'un gène codant la polyhedrine, protéine de structure de la capsid du virus, dont le promoteur induit une très forte expression lors du cycle d'infection (Caron *et al.*, 1990). Dans le cas où le gène de la polyhedrine est remplacé par la séquence codante d'un gène d'intérêt, l'infection de cellule d'insectes par le baculovirus recombiné permet dans certains cas une forte expression de ce gène d'intérêt. Cette méthode a déjà été éprouvée pour l'expression de cytochromes P450 de plantes (Mizutani and Ohta, 1998). Une autre approche consiste à réaliser l'expression hétérologue dans les feuilles de *Nicotiana benthamiana* à l'image de ce que j'ai décrit dans le second chapitre de ce manuscrit pour la PsDIOX et la PsPT. Cette méthode a déjà été appliquée au laboratoire pour caractériser CYP98A22, impliqué dans la synthèse d'acide chlorogénique chez les Rutacées (Karamat *et al.*, 2012).

## Conclusion générale et perspectives

Le métabolisme spécialisé est un réseau matriciel extrêmement complexe et différent d'une plante à une autre. Il repose sur un équilibre mis en place chez les plantes au cours de leur évolution. En perturbant cet équilibre, les impacts peuvent être divers et variés. Pour comprendre ces interactions dans la plante, une des approches possibles consiste à étudier les interactions de la plante dans son environnement. Bien que de nombreux outils « Omique » aient fait leurs apparitions au cours de ces dernières années, une étude globale et exhaustive reste encore difficile à mettre en œuvre. Dans le cadre de mon travail de thèse, j'ai choisi de l'aborder en travaillant sur une voie de biosynthèse, celle des furocoumarines, en espérant que de tirer sur ce fil permettra de débobiner à plus long terme les mécanismes mis en jeu.

Les furocoumarines sont des molécules issues du métabolisme spécialisé et considérées comme des phytoalexines. Comme de nombreuses molécules de ce métabolisme, la synthèse des furocoumarines est organisée en voie métabolique comprenant de nombreuses étapes. Des travaux antérieurs réalisés notamment au LAE ont contribué à la caractérisation de certaines d'entre elles. Pour compléter les connaissances sur les mécanismes moléculaires mis en jeu, j'ai entrepris la recherche de nouveaux gènes candidats potentiellement impliqués dans cette voie. Pour orienter mes recherches, je me suis inspiré des travaux de Monika Frey qui montrent que l'ensemble des gènes impliqués dans la synthèse du DIMBOA chez le blé faisait partie d'une même famille de P450s. J'ai choisi de focaliser mes investigations à des gènes proches des gènes déjà décrits pour les furocoumarines, les CYP71. J'ai ainsi pu identifier plus d'une vingtaine de séquences candidates 1) en explorant une base de données existante (Banque RNAseq 454 construite par Célia Krieger), 2) en exploitant une base de données génomique construite par Sandro Roselli dans le cadre du projet SaarLorBiotech et 3) en construisant ma propre banque RNAseq en collaboration avec Sylvain Legay du LIST (Luxembourg). Ce nombre de séquences candidates a finalement été réduit à 6 en me basant notamment sur des critères d'expressions et de corrélation avec la présence de furocoumarines dans les plantes.

Ces travaux ont été inclus dans deux articles déjà publiés (Krieger *et al.*, 2018; Roselli *et al.*, 2017) et dans un troisième article (Munakata *et al.*, actuellement en cours d'évaluation).

La caractérisation fonctionnelle d'enzyme est une tâche relativement compliquée à mettre en œuvre notamment à cause de la richesse du métabolisme spécialisé et de la spécificité de la plupart des enzymes. Un exemple bien décrit concerne l'ortho-hydroxylation du p-coumarate dans la synthèse de lignine qui passe par une étape d'estérification avec du quinate ou du shikimate. Cette complexité est telle que plus de 70% des P450s d'*Arabidopsis* restent encore orphelin à ce jour (Mizutani et Ohta, 2010). La recherche de la fonction des 6 gènes candidats n'a pas fait exception à cette règle.

Il est nécessaire de poursuivre ce travail et la caractérisation de *CYP71AT* et *CYP71A-C10* (MFS) me semble être prioritaire à court terme. Les arguments en faveur de cette orientation sont dans le cas de *CYP71AT* la proximité avec d'autres gènes identifiés de la voie des furocoumarines. L'option de cluster de gènes est cohérente avec de nombreux autres exemples chez les plantes. Dans le cas de *CYP71A-C10* l'argument repose sur la proximité du mécanisme enzymatique de la Menthofurane synthase avec celle de la marmesine synthase. Les 4 autres gènes ne doivent pas non plus être négligés et une recherche plus large doit être engagée. Pour la faciliter, il serait envisageable de construire une banque d'expression différentielle en se reposant sur les résultats décrits dans la partie B du premier chapitre de cette thèse. L'analyse métabolique ciblée sur les furocoumarines à partir d'échantillons de panais traités ainsi que l'analyse de l'expression de différents gènes impliqués dans la synthèse de furocoumarines mettent en évidence que les racines blessées de panais pourraient être un bon support pour construire une nouvelle banque RNAseq.

De nombreux autres aspects de la compréhension de ce formidable réseau métabolique peuvent être envisagés. Il a été démontré de longue date que la synthèse des molécules du métabolisme spécialisé est un processus dynamique se réalisant dans de nombreux tissus voire compartiment cellulaire. Le cas de la synthèse des alcaloïdes chez *Catharantus* en est l'exemple le plus emblématique (El-Sayed et Verpoorte, 2007). Il est vraisemblable que, au vu de la toxicité des furocoumarines, des processus analogues soient mis en jeu.

Dans le but d'obtenir des résultats plus fins, il serait envisageable de réaliser une analyse par une approche de MALDI-imaging (Matrix-Assisted Laser Desorption Ionization). Cette méthode, couplée à un spectromètre de masse, permet de localiser précisément des composés donnés sur une coupe de tissus. Cette méthode a déjà été utilisée chez le pois ou chez *Arabidopsis thaliana* pour localiser des flavonoïdes dans la graine (Seyer *et al.*, 2010). Une fois ciblés, les tissus accumulateurs de furocoumarines pourront être microdisséqués et utilisés pour générer une banque d'expression différentielle plus précise et plus ciblée sur les furocoumarines, et pourraient permettre d'identifier et de caractériser de nouveaux gènes impliqués dans cette voie.

Toujours dans cette notion de métabolisme dynamique il serait également intéressant d'étudier la synthèse des furocoumarines dans le contexte cellulaire. En effet, les différentes étapes de la synthèse de ces molécules se fait par différents types d'enzymes. Les dioxygénases sont des enzymes cytosoliques, les prényltransférases sont des enzymes transmembranaires plastidiales tandis que les P450 sont des enzymes membranaires du réticulum endoplasmique. Il paraîtrait logique que ces molécules ne transitent pas au hasard dans la cellule au moment de leur synthèse mais qu'elles soient prises en charge dans une structure de type métabolon. Pour explorer cette possibilité, une méthode basée sur le FLIM-FRET (Fluorescence Lifetime Imaging Microscopy – Förster Resonance Energy Transfer) permettrait d'étudier les interactions compétitives entre les enzymes (channeling) ainsi que les potentielles associations en complexes moléculaires (métabolons) (Bücherl *et al.*, 2014). La première démonstration de l'existence de ces interactions par cette méthode a été réalisée chez *Arabidopsis thaliana* dans le cas de flavonoïdes (Crosby *et al.*, 2011).

Mais le métabolisme spécialisé n'est pas restreint aux seuls furocoumarines. Les ressources générées dans le cadre de mes travaux de thèse vont permettre d'aller au-delà. Une exploration plus poussée de la banque RNAseq de panais construite en collaboration avec le LIST permettrait d'identifier d'autres gènes potentiellement intéressants. Par exemple, l'identification et la caractérisation de gènes codant des molécules et protéines impliquées dans la réponse au stress, comme décrit dans la partie A (Protéines kinases Calcium Dépendantes, MAPKs, calmoduline, ...)

permettrait d'éclaircir la réponse de la plante à différents stress et de définir des marqueurs moléculaires de stress.

En croisant les données de cette banque avec celles de la banque d'expression différentielle, nous pourrions étudier le comportement d'autres voies de biosynthèses du métabolisme spécialisé. Ces études permettraient de fait de mieux caractériser le réseau métabolique de cette plante, voire d'en modéliser le fonctionnement. En effet, l'analyse métabolique globale que nous avons réalisé à partir de panais soumis à deux stress environnementaux a amené à l'identification de plusieurs familles de molécules impactées par ces différents stress, comme les flavonoïdes, les coumarines ou les monoterpènes. Dans le cadre de cette étude, nous avons déterminé comment le métabolisme du panais réagissait en présence de deux stress abiotiques : l'ozone et la blessure mécanique. Dans un premier temps, nous avons réalisé une analyse métabolique globale des folioles de panais, soumises à ces deux stress, suivant une cinétique (T0, 24h, 48 et 72h). Pour ce faire nous avons mis au point une méthode basée sur des analyses en spectrométrie de masse d'échantillons hydro-méthanoliques couplées à des plateformes de métabolomiques en ligne (XCMS, MetaboAnalyst, MsDial, MsFinder). Cette méthode nous a permis d'établir un premier état des lieux des différents métabolites différentiellement accumulés en condition de stress, par rapport à la condition contrôle et entre les stress, en fonction du temps. Les résultats nous ont également permis d'identifier un composé inédit chez le panais, le 8-géranyloxypsoralène. Ces travaux ont été réalisés en collaboration avec l'UMR SILVA (Facultés des Sciences et Technologies de Nancy) et ont été valorisés dans un article en cours d'évaluation. Ces travaux globaux ont permis de déterminer les modifications, à la fois du métabolisme primaire et spécialisé, engendrées par l'impact d'un stress chez la plante. Ils posent ainsi les premières briques qui pourront mener à terme à la compréhension de l'équilibre Croissance/Défense chez le panais.

Pour aller plus loin dans cette compréhension, j'ai engagé des travaux visant à transposer la voie de biosynthèse des furocoumarines du panais dans le génome d'une plante non productrice, *Nicotiana tabacum*. Les furocoumarines sont des molécules toxiques et Gravot *et al* (2004) ont mis en évidence que l'équipement enzymatique de plantes non productrice de ces molécules y était plus sensible que les plantes spécialisées.



Dans un premier temps, j'ai souhaité faire la preuve de concept de dispositif en ne travaillant que sur 2 gènes et dans un contexte d'expression « basique » en utilisant un promoteur constitutif fort (Promoteur 35S du CaMV).

Afin de réaliser la transformation en une étape et de limiter le temps d'expérimentation, j'ai choisi d'intégrer les gènes codant pour les *PsDiox* et *PsPT1* dans un plasmide unique. Pour cela, j'ai mis en œuvre une approche basée sur le GoldenBraid (Sarrion-Perdigones *et al.*, 2011). Avant de réaliser une transformation stable relativement longue à mettre en œuvre, j'ai fait des expressions transitoires dans des feuilles de *Nicotiana benthamiana*. Ces travaux ont permis de valider la fonctionnalité des cassettes d'expression mais ont aussi mis en exergue un questionnement sur le temps d'expression des protéines et soulèvent notamment le problème de l'utilisation des promoteurs constitutifs pour les 2 protéines.

Cette étude d'expression transitoire a, dans un second temps, été complétée par la construction de plantes transformées de manière stable de *Nicotiana tabacum*. Les plantules régénérées présentent des phénotypes atypiques : albinos, blanches et vertes, ou complètement vertes. Les plantes étant trop petites dans le temps impartis, il faudra le moment venu faire des analyses complémentaires permettant de comparer les profils métaboliques des plantes transgéniques et des plantes sauvages.

S'il s'avérait que ces plantes étaient en mesure de produire des furocoumarines il sera alors possible de les comparer à des plantes sauvages en considérant différents critères comme la croissance ou la protection de ces plantes vis-à-vis d'agresseurs. Ces expérimentations apporteraient ainsi des informations importantes sur la notion de balance croissance/défense et permettraient d'imaginer de nouvelles pratiques culturales agroécologiques permettant d'optimiser les rendements et la protection des plantes. Par ailleurs, le fait de disposer d'un modèle de plante produisant des molécules nouvelles permettrait d'obtenir des informations sur le vrai rôle de ces molécules dans les plantes.



# **III. Matériel et Méthodes**

# 1. Matériel

## 1.1. Matériel végétal

### 1.1.1. *Pastinaca sativa*

Les plants de panais utilisés sont issus de graines commerciales (labonnegraine.com) de *Pastinaca sativa* L. subsp. *sativa* variété "Demi-Long de Guernesey"

### 1.1.2. *Nicotiana benthamiana*

Les graines de *Nicotiana benthamiana* proviennent de l'IBMP à Strasbourg.

### 1.1.3. *Nicotiana tabacum* var. **Samsung**

Les plants de *Nicotiana tabacum* sont issus de graines fournies par Plant Advanced Technologies (Vandœuvre-lès-Nancy, France).

## 1.2. Souches de bactéries

### 1.2.1. *Escherichia coli* MC1022

La souche MC1022 de *E.coli* (génotype : *araD139*,  $\Delta(\textit{ara, leu})7697$ ,  $\Delta(\textit{lacZ})\text{M15}$ , *galU*, *galK*, *strA*) est un dérivé de la souche M182 (génotype:  $\Delta(\textit{lacI}POZY)$  X74, *galK*, *galU*, *strA*) et a été utilisée pour l'amplification de plasmides.

### 1.2.2. *Agrobacterium tumefaciens* EHA105

Nous avons utilisé la souche d'*Agrobacterium tumefaciens* EHA105 (Hood *et al.*, 1993) pour l'expression transitoire dans *Nicotiana benthamiana* et pour la transformation stable de *Nicotiana tabacum*. L'agrobactérie porte sur son génome le gène *Rif<sup>R</sup>* qui confère la résistance à la rifampicine.

### **1.3. Souches de levures *Saccharomyces cerevisiae* WAT11**

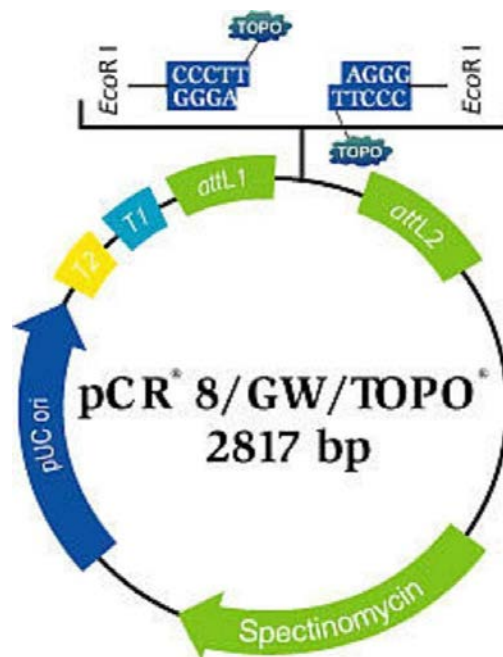
La souche *Saccharomyces cerevisiae* WAT11 a été utilisée pour l'expression des cytochromes P450. Elle dérive de la souche W303 qui possède une mutation du gène *ADE2* et *URA3* empêchant la croissance en absence d'adénine et d'uracile (Pompon *et al.*, 1996).

Le gène natif *CPR* codant une NADPH P450 réductase chez la souche W303 a été remplacé par le gène *AthR1* d'*Arabidopsis thaliana* par une approche de recombinaison homologue. Le remplacement du gène *CPR* réalisé à l'aide d'un plasmide intégratif (pYeDP110) a mené à l'introduction concomitante d'un gène *URA3* fonctionnel, conférant à la souche WAT11 une autotrophie à l'uracile. Le gène *AthR1* est placé sous le contrôle du promoteur *GAL10-CYC1*.

### **1.4. Vecteurs et plasmides**

#### **1.4.1. pCR<sup>TM</sup>8/GW/TOPO<sup>TM</sup>**

Le kit pCR<sup>®</sup>8/GW/TOPO<sup>®</sup> TA cloning (Invitrogen, Carlsbad, CA, Etats-Unis (Figure 49) est utilisé pour le clonage de produits PCR. Dans ce kit, le plasmide pCR<sup>®</sup>8/GW/TOPO<sup>TM</sup> est fourni prêt à l'emploi sous une forme linéarisée et possédant deux extrémités 3'-thymidine (T).



**Figure 49: Représentation schématique du plasmide pCR<sup>TM</sup>8/GW/TOPO<sup>TM</sup>**

La présence de ces T débordants empêche sa circularisation et permet la fixation d'un fragment d'ADN issu d'une PCR disposant de A excédentaires à leur extrémité 5' terminale. Ce fragment d'ADN est ligué au plasmide grâce à une topoisomérase (Topo) fixée sur les extrémités du plasmide.

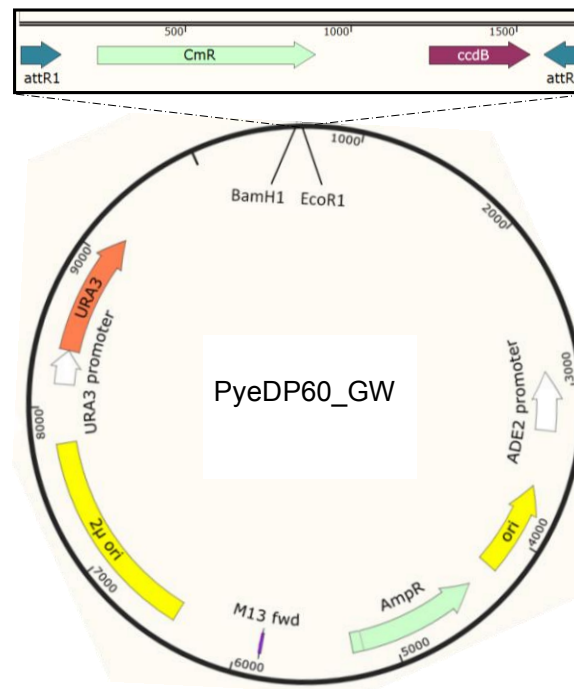
L'origine de réplication du plasmide pUC permet d'obtenir un grand nombre de copies lorsque le plasmide pCR<sup>®</sup>8 recombinant est introduit dans *E. coli*.

La pression de sélection est assurée par la présence du gène conférant la résistance à la spectinomycine. Ce kit de clonage permet également, par l'utilisation de la technologie Gateway<sup>®</sup>, le transfert de la séquence d'intérêt vers un plasmide d'expression ou tout autre vecteur de destination possédant les sites Gateway<sup>®</sup>.

#### **1.4.2. pYeDP60\_GW<sup>®</sup>**

Le plasmide pYeDP60\_GW<sup>®</sup> (Célia Krieger, travaux de thèse, figure 50) est un vecteur de destination dérivé du plasmide pYeDP60 (Urban *et al.*, 1990) conçu pour l'expression de cytochrome P450 dans plusieurs souches de *S. cerevisiae*, comprenant la souche WAT11. Ce plasmide navette est un vecteur d'expression pouvant se répliquer dans *E. coli* grâce à la présence de l'origine de réplication ColE1 mais également dans la levure grâce à l'origine de réplication 2 $\mu$ .

La cassette de recombinaison Gateway<sup>®</sup> comportant les sites attR1 et attR2 est encadrée par le promoteur *GAL10-CYC1* et par le terminateur *PGK*. Sur cette cassette de recombinaison sont également présents le gène de résistance au chloramphénicol et le gène *ccdB*. Il dispose du gène de résistance à l'ampicilline et des gènes conférant l'autotrophie pour l'adénine et l'uracile.

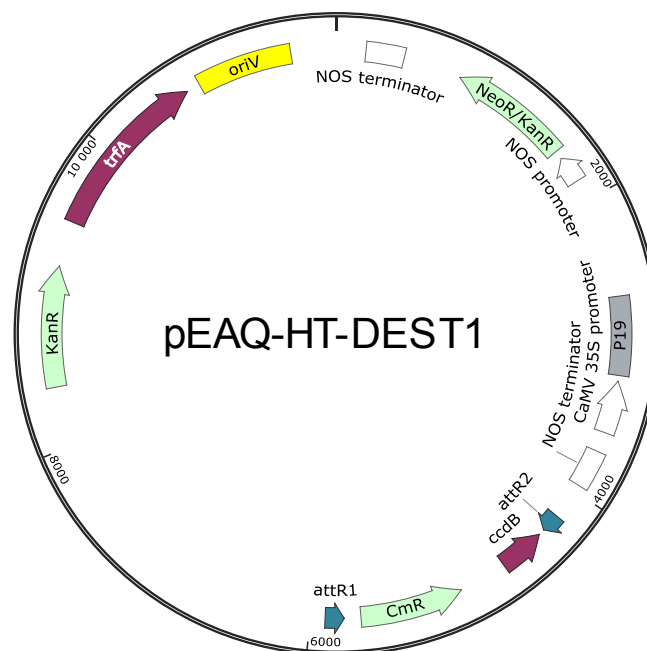


**Figure 50: Représentation schématique du plasmide PyeDP60\_GW.**

### 1.4.3. pEAQ-HT-DEST1

Le vecteur pEAQ-HT-DEST1 (Figure 51) est un vecteur d'expression optimisé pour la production de protéines recombinantes par expression transitoire dans la plante (Peyret and Lomonossoff, 2013). La présence de l'origine *OriV* et du gène *TrfA* codant la protéine d'initiation de la réplication permet au plasmide pEAQ-HT-DEST1 de se répliquer dans *E. coli*. La présence de site de recombinaison attR1 et attR2 et d'une cassette de clonage permet le transfert optimisé de séquences d'intérêt dans le plasmide en utilisant la technologie Gateway<sup>®</sup> (cf paragraphe 1.4.2.). Ce plasmide contient le gène codant la protéine p19 sous contrôle du promoteur constitutif de la mosaïque du chou-fleur "CaMV 35s" et du terminateur de la nopaline synthase Nos.

La protéine p19, découverte chez le TBSV ("Tomato Bushy Stunt Virus"), est utilisée pour améliorer l'expression transitoire de protéines chez les plantes.



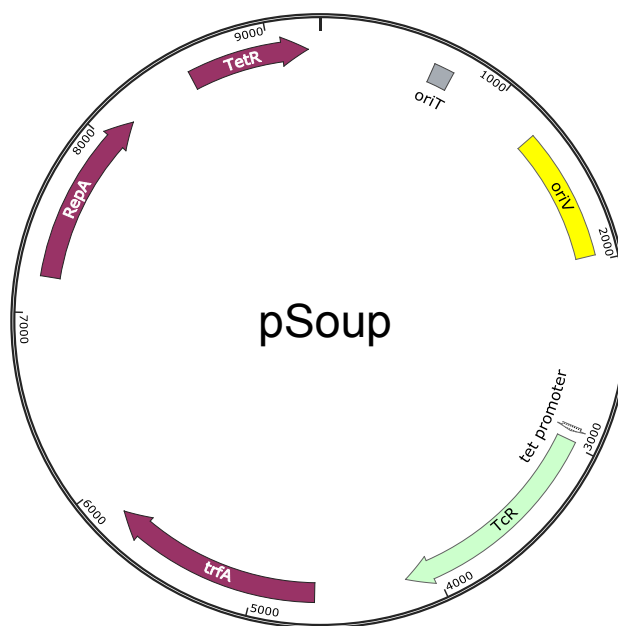
**Figure 51: Représentation schématique du plasmide pEAQ-HT-DEST1.**

Cette protéine joue un rôle de suppression de gène silencing, empêchant ainsi la dégradation d'ARN messagers surnuméraires et permettant une plus forte et plus longue expression du gène hétérologue d'intérêt (Lakatos *et al.*, 2004; Pumplin and Voinnet, 2013). Enfin, le plasmide pEAQ-HT-DEST1 porte un gène conférant la résistance à la kanamycine lorsqu'il est introduit dans des bactéries.

#### 1.4.4. pSoup

Le plasmide pSoup (Figure 52) nous a été fourni par Mark Smedley (John Innes Center, Norwich, Royaume-Uni). Ce vecteur est dérivé du plasmide pRK2 et porte le gène *pSa* codant la réplicase A d'*E. coli*. Il dispose d'un gène conférant la résistance à la tétracycline *Tet<sup>R</sup>* permettant la sélection de bactéries transformées. Le plasmide pSoup est un plasmide dit « helper » et fonctionne en système binaire avec d'autres plasmides dont l'architecture est basée sur le système pGreen. Le système binaire pGreen a été conçu spécifiquement pour la transformation de plantes *via Agrobacterium tumefaciens* (Hellens *et al.*, 2000).



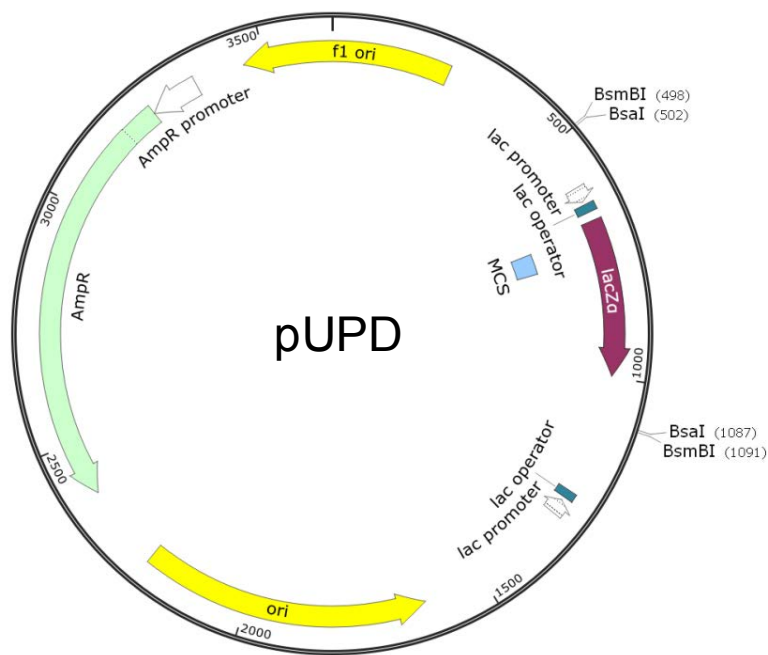


**Figure 52: Représentation schématique du plasmide pSOUP.**

#### 1.4.5. pUPD

Le plasmide pUPD (Addgene, Diego Orzaez Lab Plasmids, Cambridge, MA, Etats-Unis) (Figure 53) est un vecteur de domestication permettant de cloner chaque sous-unité d'une cassette d'expression (promoteur, séquence codante, terminateur) séparément (Sarrion-Perdigones *et al.*, 2011). Ce vecteur se réplique dans *E.coli* et porte le gène de résistance à l'ampicilline, permettant de sélectionner les bactéries transformées.

La présence du gène *lacZ*, flanqué des sites de restrictions des enzymes *Bsmbl* et *Bsal*, permet de discriminer les bactéries transformées avec le plasmide vide de celle transformées avec le plasmide portant la séquence d'intérêt. En effet, le gène *lacZ* code la  $\beta$ -galactosidase qui, en présence d'IPTG dans le milieu, va cliver le X-Gal, un substrat chromogénique menant à l'apparition de colonies bactériennes bleues. L'apparition de ces colonies bleues indique que les bactéries ont intégrées un plasmide sous forme native, vide. Si les bactéries ont intégré le plasmide portant la séquence d'intérêt, la colonie est blanche.



**Figure 53: Représentation schématique du plasmide pUPD.**

#### 1.4.5.1. pUPD-35s

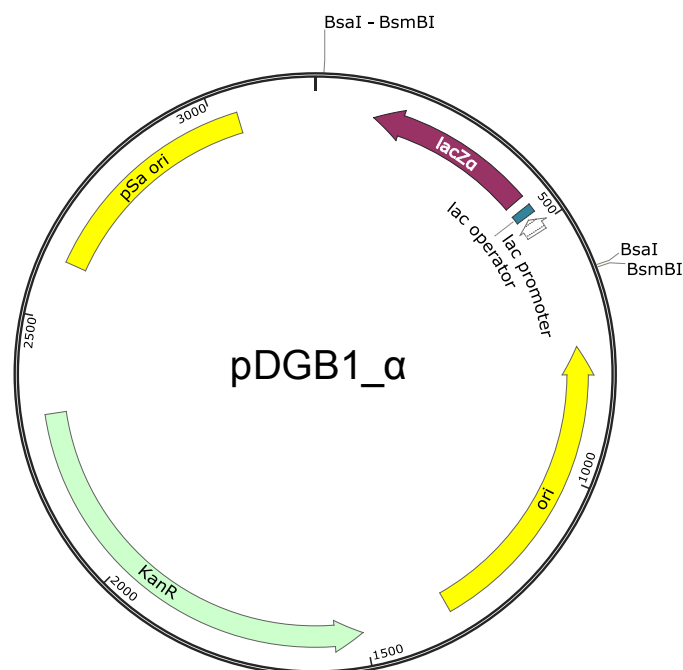
Ce plasmide dérive du plasmide pUPD et présente le promoteur 35S, (Addgene, Diego Orzaez Lab Plasmids) utilisé pour construire les unités transcriptionnelles.

#### 1.4.5.2. pUPD-tNos

Ce plasmide (Addgene, Diego Orzaez Lab Plasmids) présente le terminateur tNos et a été utilisé dans la construction des unités transcriptionnelles.

#### 1.4.6. pDBG1\_α

Les plasmides pDBG1\_α (Addgene, Diego Orzaez Lab Plasmids) (Figure 54) sont des vecteurs dans lesquels sont assemblés un promoteur, une séquence codante et un terminateur pour former une unité transcriptionnelle. Ces vecteurs fonctionnent en système binaire avec le vecteur pSoup.



**Figure 54: Représentation schématique du plasmide pDGB1\_α**

La réplicase A produite à partir du plasmide pSoup, interagit en *trans* avec l'origine de réplication pSa ori présente sur le vecteur pDGB1\_α et permet d'initier la réplication de ce vecteur. La présence du gène de résistance à la kanamycine permet de sélectionner les bactéries transformées avec ce plasmide. Il existe plusieurs versions du vecteur pDGB1\_α, dont les versions pDGB1\_α1 et pDGB1\_α2, utilisés pour ces travaux. La différence entre ces versions réside dans la séquence des sites de restrictions propre aux enzymes *Bsmbl* et *BsaI*.

#### **1.4.7. pDGB1\_α1:35s-GFP-tNos**

Ce vecteur (Addgene, Diego Orzaez Lab Plasmids) porte une construction permettant d'exprimer le gène codant la GFP. Cette construction est utilisée comme témoin positif d'expression dans le système plante (Addgene, Diego Orzaez Lab Plasmids) *via Agrobacterium tumefaciens*.

#### **1.4.8. pDGB1\_Ω**

Le vecteur pDGB1\_Ω (Addgene, Diego Orzaez Lab Plasmids) est utilisé pour l'assemblage de plusieurs unités transcriptionnelles en une seule cassette d'expression. Son mode de fonctionnement est identique à celui du vecteur pDBG1\_α. La présence du gène de résistance à la spectinomycine permet de sélectionner les bactéries transformées avec ce plasmide. Il existe plusieurs versions de ce vecteur. La version pDGB1\_Ω1 est la seule utilisée durant ces travaux. La différence avec les plasmides pDGB1\_α réside dans la séquence des extrémités débordantes libérées après restriction.

### **1.5. Conditions de culture**

#### **1.5.1. Milieux de culture de bactéries**

Tous les milieux décrits ci-dessous sont stérilisés par autoclave pendant 20 minutes, à 121 °C, 1 bar, avant utilisation, et stockés à température ambiante.

##### **1.5.1.1. Milieu LB**

Le milieu LB (Lysogeny Broth) est utilisé pour la culture d'*E. coli*. Ce milieu est préparé en dissolvant 25g de la poudre commerciale (Sigma-Aldrich, Saint-Louis, MO, Etats-Unis) dans 1L d'eau distillée. Le milieu LB est composé de tryptone (10g/L), d'extrait de levure (5g/L) et de NaCl (10 g/L). De l'agar à 10 g/L (Sigma-Aldrich) est ajouté pour la préparation de milieu solide.

##### **1.5.1.2. Milieu YEB**

Le milieu YEB (Yeast Extract Broth, pH7) est utilisé pour la culture d'*Agrobacterium tumefaciens*. Ce milieu, dont la composition est détaillée dans le tableau 10, est obtenu en mélangeant plusieurs composés commerciaux.

**Tableau 10: Composition du milieu YEB.**

<b>Composition</b>	<b>Concentration (g/L) pour 1x</b>	<b>Fournisseur</b>	<b>Référence</b>
Extrait de boeuf	1	Becton Dickinson	212303
Extrait de levure	0.5	Conda pronadisa	1702
Peptone	1	Becton Dickinson	211677
Saccharose	1.5	VWR	0335-1KG
MgSO <sub>4</sub>	0.05	VWR	AAAA14491-0b
Agar (pour milieu solide)	1	Sigma-Aldrich	A7921

### 1.5.1.3. Antibiotiques utilisés

L'addition d'antibiotiques aux milieux de cultures bactériens permet de sélectionner les bactéries recombinantes. La liste des différents antibiotiques utilisés est présentée en tableau 11.

**Tableau 11: Liste des antibiotiques utilisés pour la culture de bactéries**

<b>Antibiotique</b>	<b>Fournisseur</b>	<b>Solvant</b>	<b>Référence</b>	<b>Concentration initiale (g/L)</b>	<b>Concentration finale (mg/L)</b>
Ampicilline	Sigma-Aldrich	Eau distillée	A9518-25G	100	100
Céfotaxime	Duchefa	Eau distillée	C0111	100	100
Kanamycine	Sigma-Aldrich	Eau distillée	K1377	50	50
Rifampicine	Duchefa	DMSO	R0146	20	20
Spectinomycine	Sigma-Aldrich	Eau distillée	S4014	100	100
Tétracycline	Duchefa	Ethanol absolu	T0150	50	5

## **1.5.2. Milieux de culture de levures**

### **1.5.2.1. Milieu YPGA**

Le milieu YPGA est un milieu de culture complet optimisé pour la croissance des souches de levures *S. cerevisiae* WAT11 sauvages présentant une mutation du gène *ADE2*.

### **1.5.2.2. Milieu SGI**

Le milieu SGI est un milieu sélectif dépourvu d'adénine et permettant de sélectionner les levures transformées par les plasmides pYeDP60 recombinants.

La présence du gène *ADE2* sur le plasmide pYeDP60\_GW® permet à la levure de devenir autotrophe pour l'adénine et donc de se développer sur du milieu SGI. Ce dernier est utilisé sous forme liquide ou solide.

### **1.5.2.3. Milieu YPGE**

Le milieu YPGE est un milieu de culture complet utilisé pour la croissance des levures transformées par les plasmides pYeDP60 recombinants. Ce milieu contient du glucose qui est la source de carbone utilisée préférentiellement par les levures. Une fois le glucose entièrement consommé, l'éthanol également présent dans le milieu YPGE est utilisé comme source secondaire. Le glucose est un répresseur du promoteur *GAL10-CYC1*, inhibant alors l'expression d'un gène d'intérêt placé sous son contrôle et *a fortiori* la production de la protéine correspondante. Ce milieu est utilisé sous forme liquide exclusivement.

### **1.5.2.4. Milieu YPL**

Le milieu YPL est un milieu de culture contenant du galactose utilisé pour induire l'expression des gènes d'intérêt et la production des protéines correspondantes grâce à la présence de galactose, permettant de réguler positivement le promoteur *GAL10-CYC1*, et donc d'activer l'expression du gène d'intérêt et la production de la protéine correspondante. Ce milieu est utilisé sous forme liquide exclusivement.

**Tableau 12: Composition de 1L des différents milieux utilisés pour la culture de levure**

Composants	Fournisseur	Référence	YPGA	YPL	YPGE	SGI
Extrait de levure	Conda pronadisa	1702	10 g	10 g	10 g	X
Bactopeptone	Becton Dickinson	211677	10 g	10g	10 g	X
Acides bacto-casamino	Becton Dickinson	223050	X	X	X	1 g
Base azotée pour levure	Roth	HP26	X	X	X	7 g
Glucose	Fisher	10335850	20 g	X	5 g	20 g
Galactose	Roth	4987	X	20 g	X	X
Adénine	Sigma-Aldrich	A2786-5G	30 mg	X	X	X
Tryptophane	Alfa Aesar	A10230	X	X	X	20 mg
Ethanol 96%	Carlo Erba	3086612	X	X	3%	X
Agar (pour milieu solide)	Sigma-Aldrich	A7921	x	x	X	7 g

### 1.5.3. Culture de plantes en terre

Les plantes sont cultivées en terre dans une chambre de culture en conditions contrôlées, à une température de 20°C, une photopériode alternant 16h de lumière et 8h d'obscurité ainsi qu'un arrosage avec de l'eau trois fois par semaine (Lundi, Mercredi et Vendredi).

Dans un premier temps, les graines sont mises à germer dans un seul pot. Un mois après la germination, les plantes sont transférées dans des pots individuels pour un mois supplémentaire. Le terreau utilisé est composé de tourbe de sphaigne, de fibre de coco et de perlite (Proportions non disponibles).

#### **1.5.4. Culture de plante *in vitro***

##### **1.5.4.1. Stérilisation des graines de tabac**

Les graines sont placées dans une boîte de Pétri de 9 cm de diamètre contenant de l'éthanol 96% (30 mL). Après 30 secondes d'incubation, l'éthanol est retiré et les graines sont rincées avec un bain d'eau distillée stérile (30 mL) pendant 5 minutes.

Le bain d'eau est remplacé par un bain de chlore actif à 2% (30 mL) pendant 5 minutes. La solution de chlore est ensuite éliminée et les graines sont rincées avec un bain d'eau distillée stérile (30 mL) pendant 5 minutes. Deux autres bains d'eau distillée stérile successifs permettent d'éliminer les traces de chlore actif restant à la surface des graines. Chaque étape de bain est associée à une agitation manuelle vigoureuse.

##### **1.5.4.2. Semis des graines et culture**

Les graines décontaminées sont disposées sur un milieu MSx1 solide associé à une solution de vitamines, à du saccharose et de l'agar (Tableau 13), coulé en boîte de Pétri de 9 cm de diamètre. La germination est réalisée en chambre de culture selon les conditions décrites précédemment. Les plantules sont ensuite repiquées individuellement sur le milieu MS complémenté dans des pots en verre deux mois après germination.

##### **1.5.4.3. Conditions de culture des plantes *in vitro***

Les plantes *in vitro* sont cultivées en chambre de culture à 26°C, disposant d'une photopériode alternant 14 h de lumière et 10 h d'obscurité, avec une intensité lumineuse de 800 lux et 80 % d'humidité.



**Tableau 13: Composition du milieu de culture de *Nicotiana tabacum in vitro***

Composition	Concentration	Fournisseur	Référence
Milieu MS	X1	Duchefa	M0221
Vitamines (après autoclave)	X1	Duchefa	M0409
Saccharose	3%	VWR	470302-810
Agar HP696 (pour milieu solide)	7g/L	Kalys-Biotech	A7921

## **1.6. Induction de la synthèse des furocoumarines chez le panais**

Chacun des plants, tissus ou plantules étudié est congelé à l'azote liquide puis broyé à l'aide d'un mortier et d'un pilon. Le contenu métabolique est extrait comme décrit ci-après et analysé en UPLC couplé à un spectromètre de masse.

### **1.6.1. Elicitation mécanique**

#### **1.6.1.1. À partir de folioles, de tiges et de racines séparées**

Les feuilles, les tiges et les racines de 6 panais « Demi-long de Guernesey » sont séparées les unes des autres à partir de panais de 2 mois cultivés en terre, par une incision au scalpel au niveau du collet et du pétiole de chaque foliole.

A l'aide d'une brosse disposant de poils métalliques équidistants les uns des autres, de multiples blessures sont appliquées sur toute la surface des folioles, des tiges et des racines. Chacun des organes est alors placé pendant 24h à température ambiante dans une boîte de Pétri de 9cm de diamètre, fermée, et contenant un morceau de coton imbibé d'eau distillée pour garder une atmosphère humide dans la boîte de Pétri et éviter l'assèchement des échantillons végétaux. Le morceau de coton humide est placé de façon à ne pas être en contact avec les échantillons.

Ces expérimentations sont réalisées sur 6 plantes indépendantes. Des conditions expérimentales identiques sont appliquées aux plantes contrôles, non blessées.

#### **1.6.1.2. À partir de folioles de la plante entière**

Les blessures sont réalisées sur 8 panais « Demi-long de Guernesey » uniquement sur les folioles, qui ne sont pas détachées des tiges de la plante. L'incubation est réalisée à une température de 21°C dans des chambres de culture, pour une intensité lumineuse de 800 lux. Des prélèvements de folioles sont réalisés juste avant l'application du traitement (T0), et 24h (T1), 48h (T2) et 72 h (T3) après l'application du traitement. A chaque temps de la cinétique, les prélèvements ont été opérés à partir des mêmes plantes. Ainsi, au terme de l'expérience, 4 prélèvements (T0, T1, T2 et T3) ont été réalisés sur chacune des 8 plantes. Des conditions expérimentales identiques sont appliquées aux plantes contrôles.

#### **1.6.2. Elicitation à l'ozone**

Les plantes cultivées en terre et âgées de deux mois sont placées dans des enceintes permettant de faire varier la concentration en ozone (UMR SILVA, Faculté des sciences et Technologies, Nancy) et de contrôler la température (21°C), la photopériode (14h de jour, 8h de nuit), l'intensité lumineuse (800 lux) et l'hygrométrie (90%).

##### **1.6.2.1. Détermination de la concentration d'ozone sub-létale**

Afin de déterminer la dose d'ozone sub-létale pouvant être appliquée au panais, différentes concentrations d'ozone ont été testées, à savoir 0, 80, 120, 150, 200 ou 250 parties par milliards (ppb). Pour se faire, chaque concentration a été appliquée sur 4 plantes, qui ont été incubées dans des chambres de culture (UMR SILVA) pendant 72 h. Un contrôle visuel journalier a été réalisé durant toute la durée de l'expérience pour déceler de potentiels symptômes d'oxydation (chlorose et nécroses).

### **1.6.2.2. Etude de l'évolution de la concentration en furocoumarines chez le panais en fonction du temps**

Une concentration en ozone de 250 ppb a été appliquée sur 8 panais pendant 6 heures toutes les 24h. Des prélèvements de folioles ont été effectués juste avant l'application de l'ozone (T0) et après 24h (T1) 48h (T2) et 72h (T3) suivant le protocole représenté en figure 6. Pour chaque échantillon, le contenu métabolique est extrait et analysé en UPLC-MS (cf paragraphe 3.4.2.1.1.). Les ARNs totaux sont également extraits dans le but de quantifier l'expression de gènes cibles. Des conditions expérimentales identiques sont appliquées aux plantes contrôles.

## **1.7. Outils de bio-informatique**

Tous les logiciels cités ci-dessous ont été utilisés principalement suivant les paramètres par défaut. Dans le cas contraire, les paramètres utilisés seront précisés dans la partie "Résultats".

### **1.7.1. Logiciels en ligne**

#### **1.7.1.1. MCBF Oligo Calculator**

Le logiciel MCBF Oligo Calculator (<http://mbcf149.dfci.harvard.edu/docs/oligocalc.html>) est utilisé pour calculer la température d'hybridation, la longueur et le pourcentage de GC d'amorces nucléotidiques pour la PCR.

#### **1.7.1.2. Neb Cutter V2.0**

Le logiciel NebCutter V2.0 (<http://nc2.neb.com/NEBcutter2/>) de New England Biolabs (Ipswich, MA, États-Unis) est utilisé pour déterminer la position des sites de restriction sur une séquence d'acides nucléiques, linéaires ou plasmidiques, et pour visualiser le profil virtuel de digestion lié à l'action des enzymes sélectionnées.

#### **1.7.1.3. Multalin**

Le logiciel Multalin (<http://multalin.toulouse.inra.fr/multalin/>), crée par Florence Corpet de l'INRA de Toulouse (Corpet, 1988), est utilisé pour réaliser des alignements globaux à partir de plusieurs séquences en même temps.

#### **1.7.1.4. XCMS**

Le logiciel XCMS (<https://xcmsonline.scripps.edu/>) est utilisé pour réaliser des analyses globales de données issues d'analyses en spectrométrie de masse. Il a été utilisé pour les analyses métabolomiques.

#### **1.7.1.5. MetaboAnalyst**

MetaboAnalyst, (<https://www.metaboanalyst.ca/MetaboAnalyst/faces/home.xhtml>) crée par un groupe de recherche de l'université d'Alberta (Xia *et al.*, 2009), est une suite d'outils permettant l'analyse et l'interprétation de données issues de la spectrométrie de masse. Ce logiciel permet également la normalisation, le traitement statistique, la représentation graphique et la cartographie de voie métaboliques de données métabolomiques

### **1.7.2. Logiciels à télécharger**

#### **1.7.2.1. BioEdit**

Le logiciel BioEdit (Tom Hall, Ibis Therapeutics Carlsbad, CA, Etats-unis) est utilisé pour réaliser des recherches de séquences à partir de banques locales grâce à l'algorithme BLAST, générer des cartes de restriction de plasmides et des alignements de séquence (Clus talW)

#### **1.7.2.2. Gene Studio Professional Edition**

Le logiciel Gene Studio a permis de réaliser le contigage des séquences issues de banque 454 RNAseq de panais, réalisée durant la thèse de Célia Krieger, ainsi que l'assemblage de la banque RNAseq de panais réalisé par Sylvain Legay (LIST).

### **1.7.2.3. PrimerExpress**

Le logiciel PrimerExpress (ThermoFisher Scientific, Waltham, MA, Etats-Unis) est utilisé pour calculer la température d'hybridation, la longueur et le pourcentage de GC d'amorces nucléotidiques pour la PCR quantitative. Ce logiciel permet, à partir d'une séquence nucléique, de définir plusieurs couples d'amorces de PCR quantitative compatible avec l'utilisation de différents agents intercalent de l'ADN.

### **1.7.2.4. R**

Le logiciel R, développé Ross Ihaka et Robert Gentleman, est un logiciel libre de droits dédié aux analyses statistiques de données scientifiques. Ce logiciel a été utilisé pour réaliser des tests de Wilcoxon et de Kruskal-Wallis.

### **1.7.2.5. StepOne plus v2.3**

Le logiciel StepOne plus v2.3 (Thermofisher Scientific) a été utilisé pour acquérir et traiter des données issues d'analyses de PCR quantitatives.

### **1.7.2.6. LabSolution**

Le logiciel LabSolution (Shimadzu, Kyoto, Japon) est utilisé pour acquérir et traiter les données issues d'analyses UPLC couplé à de la spectrométrie de masse.

### **1.7.2.7. Xcalibur**

Le logiciel Xcalibur (Thermofisher, Waltham, MA, Etats-unis)) est utilisé pour acquérir et traiter les données issues d'analyses LTQ couplé à l'Orbitrap.

### **1.7.2.8. MS-Dial**

MS-Dial (Tsugawa *et al.*, 2015) est une plateforme logicielle permettant la détection et l'annotation de signaux métaboliques issues d'analyses de spectromètre de masse. Ce logiciel a été utilisé dans le cadre de l'analyse métabolique non ciblée.

#### **1.7.2.9. MsFinder**

Ce logiciel permet de prédire la structure chimique de composés identifiés par MS-Dial sur la base du profil de fragmentation MS-MS de ces composés (Tsugawa *et al.*, 2016). Ce logiciel a été utilisé dans le cadre de l'analyse métabolique non ciblée.

#### **1.7.2.10. CLC Genomics Workbench v6.5**

Cette suite logicielle (Qiagen Bioinformatics, Qiagen, Hilden, Allemagne), permet l'analyse et la visualisation de données issues de séquençage à haut débit, a été utilisée pour construire la banque transcriptomique de racines de panais.

#### **1.7.2.11. Blast2Go Pro**

Blast2Go (BioBam Bioinformatics, Valence, Espagne) est une plateforme fournissant des outils de prédiction fonctionnelle protéique et d'analyse fonctionnelle de données génomiques. Cette plateforme a été utilisée pour l'annotation de la banque transcriptomique de racine de panais

## **2. Méthodes utilisées**

### **2.1. Outils de biologie moléculaire**

#### **2.1.1. Extraction d'ADN génomique de plante**

Le matériel végétal (feuilles, tiges ou racines) est instantanément congelé à l'azote liquide puis broyé à l'aide d'un mortier et d'un pilon préalablement nettoyés avec une solution à 2% de chlore actif et une solution d'éthanol 70%. La poudre obtenue est transférée dans un microtube de 1.5mL. L'ADN génomique est extrait à l'aide du kit "E.Z.N.A ® SP Plant DNA kit" (Omega Bio-tek, Norcross, GA, États-Unis) selon les recommandations du fournisseur.

L'élution est réalisée deux fois dans 50µL d'eau ultrapure. L'ADN génomique extrait est dosé par spectrophotométrie et stocké à -20°C.

### **2.1.2. Extraction d'ARN totaux de plante**

Le plan de travail ainsi que tout le matériel utilisé sont préalablement nettoyés avec une solution à 2% de chlore actif et avec de l'éthanol 70%. Le matériel végétal (feuilles, tige ou racines) est instantanément congelé dans de l'azote liquide puis broyé à l'aide d'un mortier et d'un pilon propre. La poudre obtenue est transférée dans un microtube de 1.5mL.

Les ARN totaux sont extraits à l'aide du kit "E.Z.N.A ® SP Plant RNA kit" (Omega Bio-tek) selon les recommandations du fournisseur.

L'élution est réalisée deux fois dans 50µL d'eau RNase free fournie avec le kit. Les ARNs totaux extraits sont séparés sur gel d'agarose (VWR, 97064-256, Radnor, PA, Etats-Unis) par électrophorèse pour vérifier leur intégrité.

L'ADN génomique résiduel des échantillons d'ARN est dégradé en utilisant l'enzyme "Amplification Grade DNase" (Sigma-Aldrich) en suivant les recommandations du fournisseur.

L'absence d'ADN génomique dans les échantillons d'ARN est validée en réalisant une PCR à partir de ces échantillons en utilisant les amorces spécifiques du gène *CYP71AJ3*. Les ARNs totaux extraits sont dosés par spectrophotométrie et stockés à -80°C.

### **2.1.3. Synthèse d'ADN complémentaire**

La synthèse d'ADN complémentaire est réalisée avec 100ng d'ARN totaux exempts d'ADN génomique en utilisant le kit " High Capacity RNA-to-cDNA Kit" (Applied biosystems, Foster City, CA, États-Unis) en suivant les indications du fournisseur. Le mélange réactionnel composé de la matrice nucléique (100ng), du tampon 1x de réaction et de la rétrotranscriptase est incubé 1h à 37°C puis à 95°C pendant 5 minutes. Les ADN complémentaires générés sont dilués au 1/5<sup>ème</sup> et stockés à -20°C.

### **2.1.4. Amplification d'un fragment d'ADN**

#### **2.1.4.1. Par PCR en point final**

Les amplifications d'un fragment d'ADN par PCR ont été réalisées avec différentes polymérases dont les caractéristiques sont détaillées ci-dessous.

Ces réactions d'amplification sont réalisées dans un thermocycleur C-1000 Touch™ (Bio-rad, Hercules, CA, Etats-Unis).

La séquence des amorces utilisées ainsi que la température d'hybridation sont détaillées dans l'annexe 8 . Pour chaque enzyme, le milieu réactionnel est composé du prémix à une concentration finale de 1X , de l'amorce sens (0.2 µM), de l'amorce antisens (0.2 µM), d'ADN matriciel (50-100 ng) et d'eau distillée pour un volume final de 50 µL par réaction.

#### **2.1.4.1.1. EmeraldAmp® PCR Master Mix 2X**

L' EmeraldAmp® PCR Master Mix 2X (Takara Bio inc., Kusatsu, Japon) est un prémix, composé d'une polymérase, d'oligonucléotides et de tampon de charge mais dont la composition exacte ainsi que la concentration des différents réactifs ne sont pas publiées par le fournisseur.

Cette enzyme est une polymérase sans activité correctrice et est uniquement utilisée pour amplifier des fragments d'ADN qui ne sont pas destinés à être clonés.

Le programme utilisé pour la réaction est composé d'une première étape de dénaturation initiale à 95°C pendant 5 minutes, suivie, pendant 35 cycles, d'une étape de dénaturation à 95°C pendant 10 secondes, d'une étape d'hybridation pendant 30 secondes entre 55 et 60°C, d'une étape d'élongation à 72°C (1minute/kb), puis d'une étape d'élongation terminale de 10 minutes à 72°C.

#### **2.1.4.1.2. SapphireAmp® Fast PCR Master Mix 2X**

A l'instar de la polymérase précédente, la SapphireAmp® Fast PCR Master Mix (Takara Bio Inc.) contient une polymérase sans activité corrective et avec une vitesse de réaction plus élevée.

Le programme utilisé pour la réaction est composé d'une première étape de dénaturation initiale à 98°C pendant 5 minutes, suivie, pendant 30 cycles, d'une étape de dénaturation à 98°C pendant 5 secondes, d'une étape d'hybridation pendant 5 secondes entre 55 et 60°C puis d'une étape d'élongation à 72°C (10 secondes/kb). Cette enzyme est utilisée pour l'amplification rapide de fragment d'ADN.



#### **2.1.4.1.3. PrimeSTAR® Max DNA Polymerase 2X**

La PrimeSTAR® Max DNA Polymerase (Takara Bio Inc.) est un prémix contenant une polymérase « haute-fidélité » qui dispose d'un taux d'erreur beaucoup plus faible que les polymérases « classiques ». Cette polymérase a été utilisée pour amplifier des fragments d'ADN destinés à être clonés ou séquencés.

Le programme utilisé pour la réaction est composé d'une première étape de dénaturation initiale à 98°C pendant 5 minutes, suivie, pendant 30-35 cycles, d'une étape de dénaturation à 98°C pendant 10 secondes, d'une étape d'hybridation pendant 5 secondes entre 55 et 60°C puis d'une étape d'élongation à 72°C (5 secondes/kb).

#### **2.1.4.2. Par PCR quantitative**

La PCR quantitative est une méthode permettant d'analyser le niveau d'expression d'un gène en temps réel. Cette méthode est basée sur l'utilisation d'une molécule fluorescente, le SYBR<sup>TM</sup> Green, qui a la capacité de s'intercaler dans le petit sillon de l'ADN. Ainsi, l'intensité de la fluorescence du SYBR<sup>TM</sup> Green est proportionnelle à la quantité de produits PCR formés, permettant alors de déterminer le niveau d'expression d'un gène donné. Le mélange réactionnel est composé de SYBR® *Premix Ex Taq*<sup>TM</sup> (Tli RNaseH Plus) 1x (Takara), de 0.2 µM d'amorces sens et antisens, de "ROX Reference Dye" (Takara) 1x, d'une solution d'ADNc à 5ng/µL et d'eau ultrapure pour un volume final de 20µL pour chaque échantillon. Le Rox est utilisé comme standard de fluorescence. Les 20 µL finaux sont déposés dans un des puits d'une plaque " MicroAmp Fast Optical 96-well Reaction Plate with Barcode (0.1 mL) (Applied Biosystem). Pour s'assurer que chaque réaction démarre au même moment, la solution d'ADNc et le reste du milieu réactionnel ne sont pas mélangés directement. Dans un premier temps, le mélange réactionnel sans la matrice nucléique est déposé sur un côté du puits, la solution d'ADNc est ensuite déposée sur le côté opposé. Une centrifugation à 80 g pendant 1 minute permet d'incorporer la solution d'ADNc dans le milieu réactionnel.

La force centrifuge exercée permet également d'éliminer les bulles d'airs potentiellement présentes dans le mélange réactionnel. Un film "iCycler iQ Optical tape" (Bio-Rad) est finalement appliqué sur la plaque, permettant de protéger les échantillons d'une potentielle contamination et d'éviter l'évaporation des solutions. La plaque contenant les échantillons est placée dans l'appareil "StepOne Plus System" (ThermoFischer Scientific) puis incubée selon le programme suivant : Une étape de dénaturation initiale à 95°C pendant 30 secondes, puis, pendant 40 cycles, 95 °C pendant 5 secondes, 60°C pendant 30 secondes, suivi d'une étape de courbe de fusion. Le gène de panais codant l'ARN 5.8s a été utilisé comme gène de référence d'expression, ou "gène de ménage". Les différentes amorces utilisées sont présentées en annexe 5.

### **2.1.6. Electrophorèse sur gel d'agarose**

La migration par électrophorèse sur gel d'agarose permet de séparer des fragments d'ADN et d'ARN en fonction de leurs poids moléculaire et de leur charge électrique négative. Le gel d'agarose est préparé à 1% d'agarose dans du tampon TAE composé de Tris 40 mM acétate (VWR, 97061-174), d'EDTA 5M pH8 (ThermoFisher scientific, 15549660) et de 1.25% d'acide acétique.

La concentration en agarose peut être modulée en fonction de la taille des fragments à séparer. Une concentration élevée favorisera la séparation des fragments de faible poids moléculaire alors qu'une forte concentration sera utilisée pour séparer des fragments de haut poids moléculaire.

Un marqueur de taille Exact Ladder DNA premix 2log (Ozyme, St-Cyr-l'Ecole, France) est utilisé en parallèle, permettant de déterminer la taille des fragments séparés.

Après la migration, le gel est incubé pendant 10 minutes dans un bain composé d'eau distillée et de BET (Sigma-Aldrich, E7637) pour une concentration finale de 0.04%.

### **2.1.7. Extraction d'ADN à partir d'un gel d'agarose**

L'extraction d'ADN à partir d'un gel d'agarose est réalisée avec le kit "Nucleospin Gel and PCR Clean Up" (Macherey-Nagel, Hoerd, France) selon les recommandations du fournisseur. L'élution de l'ADN extrait est réalisée deux fois dans 30 µL d'eau ultrapure puis stockés à -20°C.

### **2.1.8. Méthodes de clonage**

#### **2.1.8.1. Clonage d'un fragment de PCR dans le plasmide pCR™8**

Les enzymes « haute-fidélité » possèdent une activité correctrice traduit par une réaction exonucléase 3'-5'. Elles permettent la production de produit de PCR à bout franc.

Pour réaliser des clonages par une approche TA il est nécessaire de rajouter des A à l'extrémité des fragments d'ADN. Pour ce faire, une adénylation des produits PCR est réalisée en utilisant le kit "PCR MasterMix 2X" (ThermoFisher Scientific). Le mélange réactionnel composé de tampon MasterMix 1X et de produit d'amplification (300-600ng/µL) est incubé à 72°C pendant 40 minutes.

La réaction de ligation du produit PCR adénylé dans le pCR™8/GW/TOPO™ est réalisée dans un volume final de 3µL. Le mélange réactionnel est composé de produit PCR adénylé (200-400ng/µL), de 0.5µL de solution saline fournie avec le kit de clonage, diluée au quart et du vecteur pCR™8/GW/TOPO™ linéaire (2ng/µL). Le mélange réactionnel est incubé à température ambiante (entre 20°C et 25°C) pendant 2h.

#### **2.1.8.2. Recombinaison dans le plasmide pYeDP60\_GW® et pEAQ-HT-DEST1**

La recombinaison d'un gène d'intérêt d'un plasmide donneur vers pYeDP60\_GW® ou pEAQ-HT-DEST1 est réalisé avec le kit "Gateway LR Clonase™ II Enzyme Mix" selon les recommandations du fournisseur (Thermofischer Scientific).

Dans un microtube de 1.5mL sont mélangés le plasmide donneur contenant le gène d'intérêt (50-150ng), 1µL du plasmide pYeDP60\_GW® (150ng/µL) et du tampon TE 1X pour un volume final de 8 µL.

A ce mélange réactionnel sont ajoutés 2  $\mu\text{L}$  d'enzyme LR Clonase™ II. Après homogénéisation à l'aide d'un vortex (2x2 secondes) et microcentrifugation (2000 g, 2 secondes), le mélange réactionnel est incubé 1 heure à 25 °C.

1  $\mu\text{L}$  de protéinase K est ensuite ajouté et le mélange réactionnel est incubé 10 minutes à 37°C. Finalement, 1 $\mu\text{L}$  de cette solution est utilisé pour la transformation de bactéries *E.coli*.

### 2.1.8.3. Par la méthode GoldenBraid

La technologie GoldenBraid est un système de clonage multigénique basé sur l'utilisation d'enzymes de restrictions de type IIS, qui ont la particularité d'avoir un site de fixation différent de leur site de restriction. L'action de ces enzymes sur un ADN double brin va permettre de libérer une extrémité cohésive de 4 nucléotides (Figure 55). Les enzymes de restriction utilisées ici sont les enzymes *BsaI* et *BsmBI*.

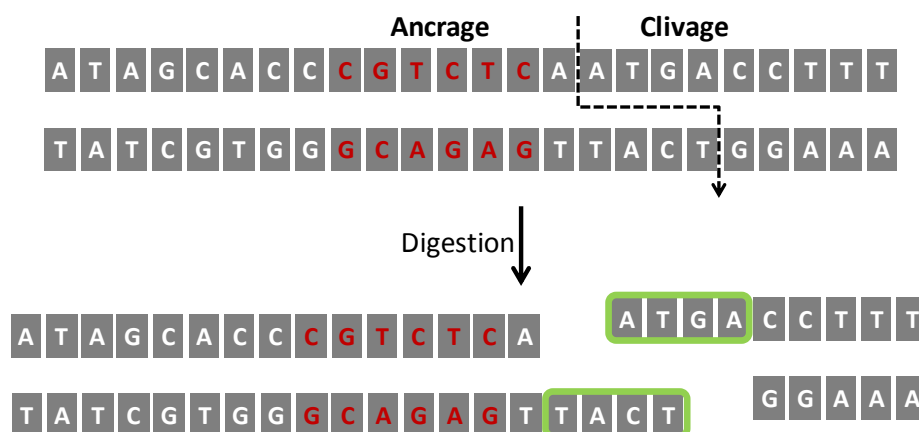


Figure 55: Représentation du mode d'action des enzymes de type IIS

#### 2.1.8.3.1. Domestication

Les clonages dans les plasmides pUPD et pDGB sont réalisés au niveau des sites de restriction *BsaI* et *BsmBI*. Pour se faire, il faut dans un premier temps générer des inserts dépourvus de ces sites de restriction. Cette étape est appelée la domestication et consiste à recoder la séquence d'intérêt en y introduisant des modifications ciblées et silencieuses. En plus de ces modifications ponctuelles, nous avons rajouté à l'extrémité des différents fragments d'ADN une séquence de 16

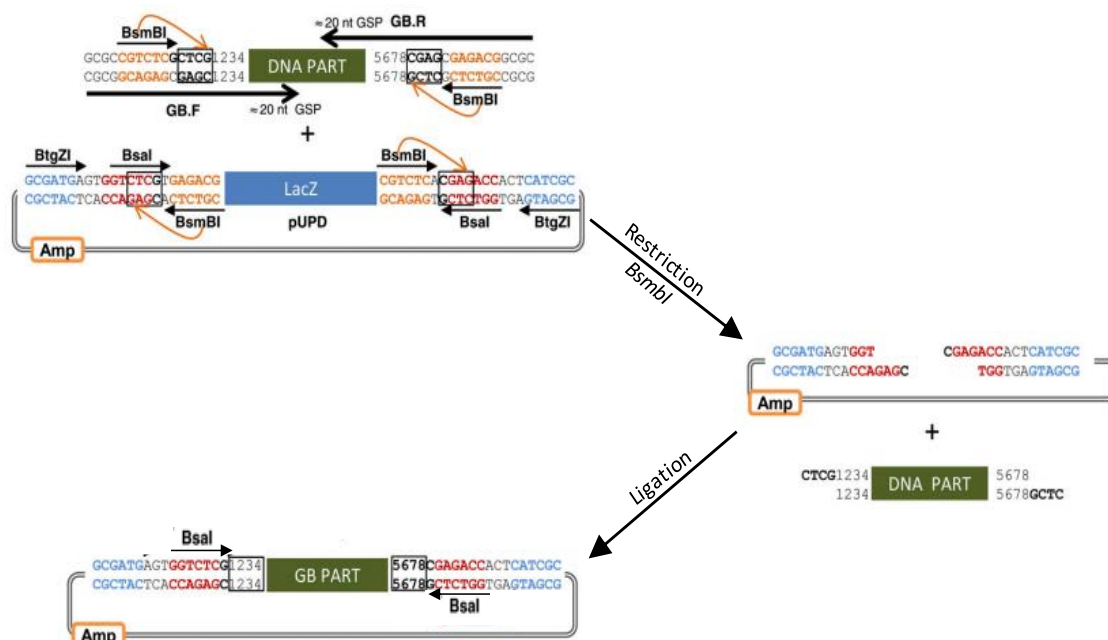
nucléotides (GCGCCGTCTCGCTCGA) et une séquence de 19 nucléotides (GCTTTGAGCGAGACCCGC) contenant les sites d'ancrage et de restriction de *Bsmbl*. La domestication a été réalisée par la société Eurofins (Luxembourg, Luxembourg).

#### **2.1.8.3.2. Ligation dans le plasmide pUPD**

La ligation du gène d'intérêt dans le plasmide de destination pUPD est assurée par des réactions successives de digestion par l'enzyme *Bsmbl* et de ligation par la T4 DNA ligase (Figure 56).

Le mélange réactionnel est composé de 1 µL de T4 DNA ligase 5U/µL (Thermoscientific), 1 µL de T4 DNA ligase buffer 10x (Thermoscientific), 1 µL *Bsmbl* 10 U/µL (Thermoscientific), du plasmide pCR8 contenant le gène d'intérêt, 5 µL du plasmide pUPD et de l'eau distillée pour un volume final de 10 µL.

Le microtube contenant le milieu réactionnel est incubé à 37°C pendant 2 minutes puis 16°C pendant 5 minutes. Ces deux étapes sont répétées 25 fois. La concentration des plasmides est comprise entre 100 et 300 ng/µL. Le volume de plasmide utilisé peut-être adapté en fonction de l'efficacité de la réaction. Les produits de ligation sont ensuite utilisés pour réaliser des transformations de bactéries *E.coli* (Cf paragraphe 3.1.10.1.).



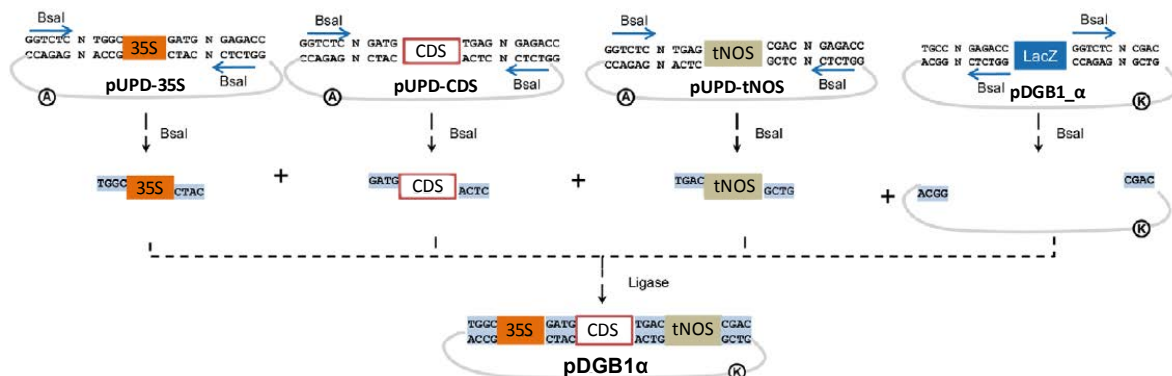
**Figure 56: Ligation d'une sous-unité de cassette d'expression dans le plasmide pUPD par la méthode GoldenBraid**

### 2.1.8.3.3. Construction des unités transcriptionnelles simples

La construction d'une unité transcriptionnelle consiste à assembler un promoteur, une séquence codante et un terminateur en une seule cassette d'expression dans un plasmide de destination (Figure 57).

Le mélange réactionnel est composé du plasmide pUPD+35S, pUPD+Tnos, pUPD contenant le gène d'intérêt, pDGB1 $\alpha$ , de 1 $\mu$ L de *Bsal* 10 U/ $\mu$ L (Thermoscientific), de 1 $\mu$ L de T4 DNA ligase 5U/ $\mu$ L, de 1 $\mu$ L de T4 DNA ligase buffer 10 x et d'eau distillée pour un volume final de 10 $\mu$ L. La concentration des plasmides est comprise entre 100 et 300 ng/ $\mu$ L. Le volume de plasmide utilisé peut être adapté en fonction de l'efficacité de la réaction.

Le microtube contenant le milieu réactionnel est incubé à 37°C pendant 2 minutes puis 16°C pendant 5 minutes pendant 25 cycles. Cette réaction conduit à la formation de deux unités transcriptionnelles, appelées 35s-***PsDiox***-Tnos et 35s-***PsPT1***-Tnos, et portées respectivement par le plasmide pDGB1 $\alpha$ 1 et pDGB1 $\alpha$ 2. Le mélange réactionnel est introduit dans des bactéries *E.coli* électrocompétentes (cf. paragraphe 3.1.10.1.).



**Figure 57: Assemblage d'une unité transcriptionnelle dans un plasmide pdgb1α par la méthode du GoldenBraid**

#### 2.1.8.3.4. Construction des unités transcriptionnelles multiples

Cette dernière étape permet d'assembler plusieurs unités transcriptionnelles dans un seul plasmide d'expression. Le mélange réactionnel est composé du plasmide pDBG1Ω1, du plasmide contenant l'unité transcriptionnelle 35s-PsDiox-Tnos et le plasmide portant l'unité transcriptionnelle 35s-PsPT1-Tnos en concentration égale. Le mélange réactionnel est additionné de 1 µL de T4 DNA ligase 5 U/µL, de 1 µL de T4 DNA ligase buffer 10 x, de 1 µL de *Bsmbl* 10 U/µL et d'eau distillée pour un volume final de 10 µL. Le microtube contenant le milieu réactionnel est incubé à 37°C pendant 2 minutes puis 16°C pendant 5 minutes pendant 25 cycles. Cette réaction aboutie à la construction d'une double unité transcriptionnelle, portée par le plasmide pDGB1\_Ω1, et appelée "35s-**PsDiox**-Tnos:35s-**PsPT1**-Tnos".

La concentration des plasmides est comprise entre 100 et 300 ng/µL et peut être ajustée en fonction de l'efficacité de la réaction. Les produits de ligation sont introduits dans *E.coli* par électroporation (cf paragraphe 3.1.10.1.).

### 2.1.9. Préparation de cellules compétentes

#### 2.1.9.1. Bactéries électrocompétentes

#### **2.1.9.1.1. *Escherichia coli* MC1022**

Lorsque les bactéries ne sont pas dans des conditions isotoniques, elles sont très fragiles et peuvent facilement être lysées. Pour éviter cela, il est nécessaire de les manipuler avec précaution et de leur éviter des chocs mécaniques (centrifugations ou resuspension trop violente) et de travailler sur glace et avec des solutions froides. Des bactéries issues d'un stock glycérol (-80°C) sont étalées sur du milieu LB solide et incubées dans une étuve à 37°C jusqu'à l'apparition de colonies isolées. Une colonie unique est utilisée pour ensemer 30 mL de milieu LB et est incubée à 37°C sous agitation à 200 rpm pendant 16 heures. Dix millilitres de cette pré-culture sont utilisés pour ensemer 200mL de milieu LB dans un erlenmeyer de 1L. Cette culture est incubée à 37°C jusqu'à atteindre une  $DO_{600nm}$  comprise entre 0.6 et 0.8. La culture bactérienne est alors placée dans la glace pendant 30 minutes puis centrifugée pendant 15 minutes à 4°C, 2200 g. Le surnageant est éliminé et le culot est suspendu délicatement avec 20 mL d'eau stérile froide. Après une nouvelle centrifugation pendant 15 minutes à 2200 g, le surnageant est éliminé. Les deux derniers points sont répétés une nouvelle fois, et le culot est suspendu dans 20 mL de glycérol 10% froid et stérile. Après une dernière centrifugation en utilisant les paramètres précédents, le culot est suspendu dans 1 mL de glycérol 10% froid et stérile. La solution est laissée au repos 30 minutes dans la glace. La suspension bactérienne est finalement aliquotée dans des tubes stériles par volume de 45 µL. Les aliquots sont congelés dans de l'azote liquide puis stockés à -80°C.

#### **2.1.9.1.2. *Agrobacterium tumefaciens***

Des agrobactéries issues d'un stock glycérol sont étalées sur un milieu YEB solide additionné de rifampicine et mises en culture dans une étuve à 28°C jusqu'à l'apparition de colonies isolées. Une colonie est utilisée pour ensemer 5 mL de milieu YEB et incubée pendant 16h à 28°C. Deux millilitres de cette pré-culture sont utilisés pour ensemer 50 mL de milieu YEB. La culture de 50 mL est incubée dans une étuve à 28°C sous agitation (200 rpm) jusqu'à atteindre une  $DO_{600nm}$  comprise entre 0.6 et 0.8. La culture bactérienne est alors placée sur de la glace pendant 30 minutes puis centrifugée 10 minutes à 4°C, 3000 g.



Le culot est suspendu dans 45 mL d'eau stérile froide. La suspension est centrifugée selon les mêmes paramètres que précédemment et le surnageant est éliminé. Les deux dernières étapes sont répétées une nouvelle fois et le culot est suspendu dans 40 mL de glycérol 10% froid et stérile. Après une dernière centrifugation en utilisant les paramètres précédents, le culot est suspendu dans 500 µL de glycérol 10% froid et stérile. La solution est finalement aliquotée par volume de 45 µL. Les aliquotes sont congelées dans de l'azote liquide puis stockés à -80°C.

#### **2.1.9.2. Levures chimiocompétentes**

Des levures *Saccharomyces cerevisiae* WAT11 issues d'un glycérol stock sont étalées stérilement sur un milieu YPGA solide et incubées à 28°C pendant 48 heures. Une colonie de levure est repiquée dans 10 mL de milieu YPGA. La solution est incubée sous agitation (180 rpm) pendant 36 heures à 28°C.

La suspension est diluée dans 50 mL de milieu YPGA pour obtenir une  $DO_{700nm}$  de 0.2. La culture est incubée sous agitation à 28°C jusqu'à une  $DO_{700nm}$  de 0.2 puis centrifugée à 3200 g pendant 5 minutes à température ambiante. Le culot est resuspendu avec 1 mL d'eau stérile puis la suspension est transférée dans un microtube de 2mL. Après une centrifugation d'une minute à 1500 g à température ambiante, le surnageant est éliminé et le culot est resuspendu dans 1 mL d'une solution composée d'acétate de lithium 100 mM et de TE 1X (Tris HCl 0.1 mM pH7.5, EDTA 10µM). Après une nouvelle centrifugation à 1500 g pendant une minute à température ambiante, le surnageant est éliminé et le culot est suspendu dans 200 µL de TE 1X additionné d'acétate de lithium 100 mM.

#### **2.1.10. Transformation de cellules compétentes**

##### **2.1.10.1. Bactéries électrocompétentes**

L'électroporation est une méthode de transformation permettant, par l'application d'un choc électrique de forte intensité et de courte durée, la déstabilisation temporaire de la structure de la membrane bactérienne, la rendant perméable. Dans une cuve d'électroporation, dont les électrodes sont séparées de 2mm, 40µl de suspension froide de bactéries électro compétentes sont additionnés de 1 à 10ng de produit de ligation.

La cuve est placée dans l'électroporateur Micropulseur<sup>TM</sup> (Biorad) sur laquelle est appliqué un courant de 2.5 kV. Immédiatement après le choc électrique, 500 µL de milieu LB froid sont ajoutés dans la cuve. La solution bactérienne est incubée 1h à 37°C puis étalée sur une boîte de Pétri contenant du milieu LB solide ainsi que les antibiotiques adéquats. La boîte de Pétri est placée à 37°C pendant une nuit. Dans le cas de la transformation d'*Agrobacterium tumefaciens*, le protocole utilisé est quasiment identique, à la différence du temps d'incubation (4h), de la température d'incubation (28°C), et du milieu utilisé (YEB).

#### **2.1.10.2. Levures chimiocompétentes**

Le mélange réactionnel composé de 10 µL d'ADN de sperme de saumon 10mg/mL préalablement dénaturé à 100°C pendant 30 minutes puis refroidi à 4°C, 5µg de plasmide portant le gène d'intérêt, 50 µL de levures chimiocompétentes et 500 µL de tampon TE 1X contenant de l'acétate de lithium 100 mM et du PEG 4000 50% est incubé 30 minutes en agitation (180 rpm) à 30°C puis 15 minutes à 42°C au bain-marie. Le mélange est placé une minute sur la glace puis centrifugé 2 minutes à 1500 g et à température ambiante. Le surnageant est éliminé et le culot est resuspendu dans 1 mL d'eau stérile. Après une nouvelle centrifugation suivant les paramètres précédents, le culot est suspendu dans 300 µL de milieu YPGA et incubé sans agitation à 30°C pendant 2 heures. La solution de levures est centrifugée pendant une minute à 1500 g, à température ambiante, et le culot est resuspendu dans 300 µL de milieu SGI liquide. Finalement les levures sont étalées sur du milieu SGI solide puis incubées à 28°C jusqu'à l'apparition de colonies.

#### **2.1.11. Extraction d'ADN plasmidiques**

L'ADN plasmidique est extrait d'*E.coli* en utilisant le kit "NucleoSpin Plasmid" (Macherey-Nagel). Une culture bactérienne de 2 ml incubée une nuit est centrifugée à 10000 g et à température ambiante. Le culot est resuspendu dans un tampon de resuspension et traité suivant les recommandations du fournisseur. L'ADN plasmidique est élué deux fois dans 50 µL d'eau stérile et conservé à -20°C. Dans le cas d'ADN plasmidique extrait d'*Agrobacterium tumefaciens*, du lysozyme en poudre est ajouté au tampon de resuspension fourni avec le kit, pour une concentration finale de 10mg/mL, permettant d'améliorer la lyse de l'agrobactérie.

### **2.1.12. Quantification par spectrophotométrie**

La détermination de la concentration en acide nucléique est réalisée avec le spectrophotomètre Biophotometer (Eppendorf, Hambourg, Allemagne) en mesurant la  $DO_{260nm}$  de 5  $\mu$ L d'échantillons dans 65  $\mu$ L d'eau distillée. Un blanc est réalisé au préalable avec 70  $\mu$ L d'eau distillée. La mesure de  $DO_{600nm}$  de cultures bactériennes est réalisée en utilisant l'Ultrospec 10 (Biochrom, Cambridge, Royaume-Uni).

### **2.1.13. Western Blot**

#### **2.1.13.1. Electrophorèse sur gel de polyacrylamide en condition dénaturante**

L'électrophorèse sur gel d'acrylamide en condition dénaturante, ou SDS-PAGE, permet de séparer les protéines d'un mélange complexe en fonction de leurs poids moléculaires. La présence de dodécylsulfate de sodium (SDS) dans le gel de polyacrylamide va permettre de dénaturer les protéines, mais également d'annuler la totalité des charges électriques naturellement présentes sur ces protéines. L'acquisition de charges négatives permet *alors* de séparer les protéines uniquement en fonction de leurs poids moléculaires. Les échantillons protéiques sont d'abord concentrés dans un gel de concentration à 5% d'acrylamide composé de Tris HCL (pH 6.8) 125mM, de SDS 1%, de tétraméthyléthylènediamine (TEMED) 0.1%, de persulfate d'ammonium (APS) 0.1% et d'eau distillée pour un volume final de 5mL.

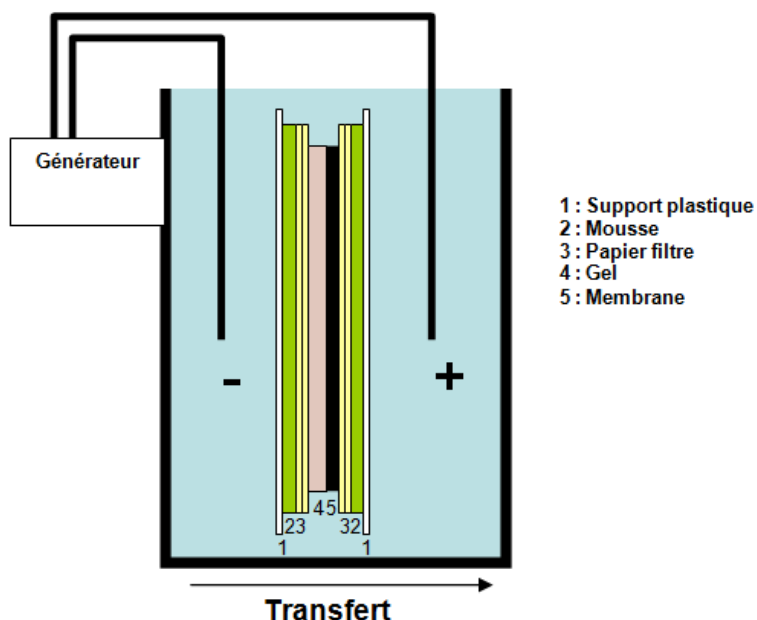
Les protéines concentrées sont ensuite séparées dans un gel de séparation à 10% d'acrylamide composé de Tris HCL (pH 8.8) 375 mM, de SDS 1%, de TEMED 0.08%, d'APS 1% et d'eau distillée pour un volume final de 10 ml. Les échantillons de protéines sont additionnés de tampon de charge 1X (Tris HCL 125 mM pH 6.8, SDS 5%, glycérol 20%,  $\beta$ -mercaptoéthanol 5%, bleu de bromophénol 0.05%, DTT 400 mM) puis chauffés à 100°C pendant 5 minutes. Un marqueur de poids moléculaire (ProSieve<sup>TM</sup> QuadColor<sup>TM</sup> protein marker, 4.6 kDa-300 kDa, Lonza Rockland Inc, Rockland, ME, Etats-Unis) est utilisé pour déterminer la taille des protéines après migration.

Le gel contenant les échantillons protéiques est placé dans une cuve de migration contenant du tampon de migration (Tris/Glycine 1X pH 8.3). Un courant de 120V est appliqué pendant 1 heure.

#### 2.1.13.2. Transfert des protéines séparées sur une membrane PVDF

Les différentes protéines séparées sont ensuite transférées sur une membrane de polyfluorure de vinylidène, ou PVDF (Amersham™ Hybond™ P 0.45 PVDF, No 10600023, GE Healthcare, Chicago, IL, Etats-Unis) ce qui va permettre d'immobiliser les protéines. Pour ce faire, le gel de séparation contenant les protéines est mis au contact de la membrane PVDF.

Le contact entre le gel et la membrane est maximisé grâce à la présence de mousse et de papier filtre, le tout inséré dans un dispositif adapté (Figure 58). L'ensemble est placé dans une cuve en présence du tampon de transfert (Tris/Glycine 1x, éthanol 20%) et soumis à un courant électrique de 100V pendant 45 minutes.



*Figure 58: Représentation schématique du dispositif expérimental de transfert des protéines du gel de séparation vers la membrane PVDF*

### 2.1.13.3. Immunodétection

La membrane de Nylon sur laquelle ont été transférées les différentes protéines d'un échantillon donné, préalablement séparées par électrophorèse, est rincée pendant 5 minutes avec un tampon PBS1x/Tween 20 issue de la dilution de la solution mère PBS10x (87.6 g de NaCl, 2g de  $\text{KH}_2\text{PO}_4$ , 14.2 g de  $\text{NaH}_2\text{PO}_4$ , 2g de KCl) auquel est ajouté 1% de Tween 20, pour un volume final de 1L, pH 7.4. Le tampon de rinçage est remplacé par un tampon PBS/Tween20 1x contenant 5% de lait en poudre. Ce dernier va permettre de saturer les sites aspécifiques. Après 1 heure d'incubation, un anticorps primaire anti-histidine (6xHistidine Epitope Tag antibody, Acris, Origene, Rockville, MD, Etats-Unis) est ajouté au tampon, pour une concentration finale de 1/2000<sup>ème</sup> pendant 2 heures. Trois rinçages successifs sont ensuite réalisés pendant 10 minutes chacun avec du tampon PBS/Tween 20 1x pour éliminer des traces résiduelles d'anticorps primaires.

La membrane est incubée avec un tampon PBS/Tween 20 1x contenant du lait en poudre 5% et additionné de l'anticorps secondaire anti-anticorps primaire ("Anti-Rabbit IC (whole molecule)-Alkaline Phosphatase produced in goat", Sigma-Aldrich) conjugué à la phosphatase alcaline pendant 90 minutes.

La membrane est ensuite incubée à deux reprises dans un tampon composé de Tris 0.1M et de NaCl 0.1M (pH 9.6) et les protéines sont révélées par l'ajout de NBT(100mg/ml)/BCIP(100mg/ml) après une incubation de 15 minutes. Le 5-bromo-4-chloro-3-indole-phosphate (BCIP) est une molécule qui va être phosphorylée par la phosphatase alcaline, couplée à l'anticorps secondaire. Le BCIP phosphorylé va se dimériser pour augmenter sa stabilité et former le 5'-5'-dibromo-4'-4'-dichloro-indigo qui va réduire l'autre molécule impliquée dans la révélation des protéines, le nitrobleu tétrazolium chloride (NBT). Le NBT réduit change de couleur pour devenir le bleu de formazan, observable à l'œil nu. Ainsi, la présence de bleu de formazan traduit la présence de la protéine d'intérêt sur la membrane de nitrocellulose.

#### **2.1.14. Construction d'une banque transcriptomique de panais par séquençage Illumina**

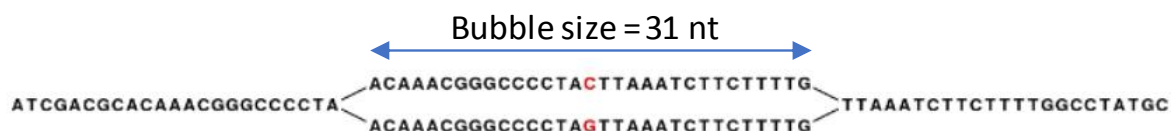
La banque transcriptomique de panais a été réalisée par le Dr. Sylvain Legay (sylvain.legay@list.lu) du Luxembourg Institute of Science and Technology ( Esch-sur-Alzette, Luxembourg).

##### **2.1.14.1. Préparation de la librairie et séquençage**

Les ARNs totaux sont extraits à partir de 100 mg de racines de panais élicitées au préalable par blessures mécaniques, selon le protocole décrit au paragraphe 4.2. L'évaluation de l'intégrité des ARNs totaux, la purification et la quantification des ARNm ont été réalisées par le Dr. Sylvain Legay. La préparation de la banque à partir d'ARNm a été réalisée à l'aide du kit « SMARTer Stranded RNA-Seq Kit (Takara Bio) selon la méthode « pair-ends ». Une fois réalisée, la qualité et la concentration de la banque ont été respectivement contrôlées *via* l'utilisation du bioanalyseur Agilent 2100 (Agilent Technologies, Santa Clara, CA, Etats-Unis) et le kit « Library Quantification-Illumina/ROX Low » (Kapa Biosystems, Boston, MA, USA) a l'aide d'un thermocycleur RT-qPCR Vii7 96-well (Life Technologies). Le séquençage a été réalisé par la méthode Illumina à partir de la librairie diluée à 4nM dans un séquenceur MiSeq Illumina (San Diego, CA, Etats-Unis).

##### **2.1.14.2. Assemblage et annotation de la banque**

L'assemblage a été réalisée avec le logiciel CLC Genomics Workbench v6.5 avec un « Word size » de 55pb et un « Bubble size » de 200 pb. Le « Word Size » est défini par le nombre de nucléotides identiques nécessaire pour considérer que deux séquences sont homologues dans le cas d'un alignement. Le « Bubble Size » détermine la longueur de la bifurcation entre deux séquences très homologues induite par la présence d'un ou plusieurs nucléotides divergents (Figure 59). Une fois le « Bubble size » atteint, le variant le plus représenté dans la banque est gardé pour construire la séquence finale.



*Figure 59: Exemple d'un « Bubble size » de 31 nucléotides lors de l'alignements de 2 séquences nucléotidiques.*

## 2.2. Production de protéines en systèmes hétérologues

### 2.2.1. Dans *Saccharomyces cerevisiae* WAT11

Une colonie de levures transformées est utilisée pour ensemer 50 mL de milieu SGI dans un erlenmeyer de 250 ml. La culture est incubée 24 heures à 28°C. La suspension de levure est utilisée pour ensemer 200 mL de milieu YPGE (Tableau 14) dans un erlenmeyer de 1 L qui est incubé 24 heures à 28°C. La culture de 200 mL est centrifugée à 2200 g, à température ambiante pendant 30 minutes, puis le culot est suspendu dans 200 mL de milieu YPL. La culture est incubée pendant 16 heures à 18°C puis centrifugée 10 minutes à 3500 g et à 4°C. Le surnageant est éliminé et le culot est suspendu dans 200 mL de tampon TEK froid (Tableau 14).

La solution est aliquotée dans des tubes de 50 mL centrifugés à 3500 g pendant 10 minutes à 4°C. Les culots sont resuspendus dans 4 mL de tampon TES (Tableau 14) froid avant que des billes de verre (Sigma-Aldrich, G9268-500G) soient ajoutées jusqu'à l'affleurement de la surface de la solution dans chaque tube.

**Tableau 14: Composition des différents tampons utilisés pour la génération de microsomes de levures**

<b>Composés</b>	<b>TEK</b>	<b>TES</b>	<b>TEG</b>
Tris HCl pH 7.5	50 mM	50 mM	50 mM
EDTA	1 mM	1 mM	1 mM
KCl	100 mM	X	X
Sorbitol	X	600 mM	X
BSA	X	1 %	X
B-mercaptoéthanol	X	20 mM	X
Glycérol	X	X	30 %

Les tubes contenant les billes de verres et la solution de levures sont agités vigoureusement à la main horizontalement pendant 5x1 minutes, avec une minute de repos sur glace entre chaque agitation. La solution est versée dans un nouveau tube de 50 mL sur glace, en évitant au maximum de transférer des billes de verres. Les billes sont lavées 3 fois avec 10 mL de tampon TES froid, qui sont à chaque étape de lavage ajoutés à la solution de levure, contenue dans le tube de 50 mL sur glace jusqu'à un volume final de 45 mL. La solution finale est centrifugée pendant 15 minutes à 8800 g et à 4°C puis filtrée, à l'aide de filtre plissé pour éliminer les billes restantes, et transférée dans un nouveau tube de 50 mL. Le filtrat est centrifugé 1 heure à 125 000 g, 4°C et le surnageant est éliminé.

Le culot est lavé deux fois avec 5 mL de tampon TES puis une fois avec 500 µL de tampon TEG (Tableau 14) froid. Le volume de tampon TEG peut être ajusté en fonction de la quantité de microsomes produits. Le culot est décroché puis transféré dans un tube de Potter, préalablement refroidi.

Par des mouvements de rotation tout en faisant monter et descendre le pilon dans le tube de Potter, le culot est délicatement suspendu dans le tampon TEG. Cette étape doit être réalisée en minimisant le plus possible l'entrée d'air dans la solution, afin de conserver l'efficacité des protéines. Une fois le culot suspendu, la solution de microsomes est placée sur glace



## **2.2.2. Dans *Nicotiana benthamiana* par expression transitoire**

### **2.2.2.1. Préparation de la suspension bactérienne à inoculer**

Une colonie d'*Agrobacterium tumefaciens* transformées avec un plasmide portant le gène d'intérêt est ensemencée dans 50 mL de milieu YEB supplémentés des antibiotiques de sélections adéquats et incubée une nuit à 28°C, sous agitation (180 rpm). La culture bactérienne est centrifugée 10 minutes à 5000 g à température ambiante. Le surnageant est éliminé et le culot est suspendu dans 2mL d'eau distillée. Un deuxième lavage est réalisé suivant les mêmes conditions que précédemment. Le culot est suspendu dans 1 mL d'eau distillée et la concentration cellulaire est mesurée au spectrophotomètre (DO à 600nm) en utilisant un aliquot de 10 µL de la suspension bactérienne diluée 100 fois.

La suspension d'agrobactéries utilisée pour infiltrer les feuilles de tabac est composée d'agrobactéries portant le gène qui exprime la protéine p19, pour une DO<sub>600nm</sub> finale de 0.2, et d'agrobactérie portant le gène d'intérêt, pour une DO<sub>600nm</sub> finale de 0.4.

### **2.2.2.2. Agro infiltration des feuilles de *Nicotiana benthamiana***

Les suspensions bactériennes sont infiltrées à l'aide d'une seringue sans aiguille de 1 mL au niveau de la face abaxiale des feuilles de tabac. Pour chaque feuille, 1 à 5 mL de suspension bactérienne sont utilisés pour l'agroinfiltration.

Des bactéries transformées avec un plasmide portant le gène codant la protéine GFP ou avec un plasmide vide sont utilisées respectivement comme contrôle positif et négatif d'agroinfiltration. Des plantes non transformées sont également utilisées comme témoin négatif. Les différentes plantes sont placées dans un caisson de confinement pendant 96 heures, à température ambiante.

### **2.2.3. Dans *Nicotiana tabacum* var. Samsung par expression stable**

#### **2.2.3.1. Préparation de la suspension bactérienne**

Une colonie d'*Agrobacterium tumefaciens* transformées avec un plasmide portant le gène d'intérêt est ensemencée dans 30 mL de milieu YEB liquide supplémenté des antibiotiques de sélections adéquats et incubée 36 heures à 28°C sous agitation (180 rpm). Après incubation, la culture bactérienne est additionnée d'acétosyringone 100 µM (D134406-5G, Sigma-Aldrich), et remise en culture pendant 4 heures. Lorsque la DO<sub>600nm</sub> atteint 0.7, la culture d'agrobactéries est centrifugée à 1700 g pendant 15 minutes à température ambiante. Le surnageant est éliminé et le culot est suspendu dans 30 mL de milieu YEB. La suspension bactérienne est centrifugée de nouveau selon les paramètres précédents et les deux dernières étapes sont répétées deux fois. A la suite de la dernière centrifugation, le surnageant est éliminé et le culot est suspendu dans 30 mL de milieu de resuspension (Tableau 15).

Des bactéries transformées avec un plasmide portant le gène codant la protéine GFP ou avec un plasmide vide sont utilisées respectivement comme contrôle positif et négatif d'agroinfiltration.

**Tableau 15 : Composition des différents milieux utilisés pour la transformation stable de tabac par *Agrobacterium tumefaciens***

	Milieu de resuspension	Milieu de co-culture	Milieu d'élimination	Milieu de sélection	Milieu d'enracinement
MS	X1	X1	X1	X1	X1
Vitamines MS	X1	X1	X1	X1	X1
Saccharose	3 %	3%	3%	3%	3%
6-BAP	<b>X</b>	0.5 mg/mL	<b>X</b>	0.5 mg/mL	<b>X</b>
Agar HP696	<b>X</b>	7.5g/L	<b>X</b>	7.5g/L	7.5g/L
Antibiotique de sélection	<b>X</b>	<b>X</b>	<b>X</b>	50 mg/L	50 mg/L
Céfotaxime	<b>X</b>	<b>X</b>	250 mg/mL	250 mg/mL	250 mg/mL
Acide naphtalène acétique	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	0.1 mg/L

### 2.2.3.2. Infection d'explants foliaires

Des rectangles de 1 cm<sup>2</sup> sont découpés dans les feuilles de différents vitroplants et immergés pendant 10 minutes dans la suspension d'agrobactérie. Les explants sont séchés sur papier filtre préalablement autoclavé et transférés sur un milieu solide de co-culture (Tableau 15) pendant 48 heures à 23°C à la lumière.

### 2.2.3.3. Sélection et régénération des plantes transformées

Les explants sont retirés du milieu de co-culture et sont placés dans 10 mL de milieu MS supplémentés de céfotaxime (Milieu d'élimination, Tableau 15), un antibiotique bactériostatique permettant d'inhiber la croissance bactérienne. Après 10 minutes d'incubation sous agitation à température ambiante, les explants sont séchés sur papier filtre stérile et disposés sur un milieu de sélection (Tableau 15) permettant, par la présence d'antibiotiques spécifiques, aux explants transformés de survivre.

Ce milieu contient également une cytokinine, le 6-benzylaminopurine (6-BAP), utilisé pour favoriser le développement de tiges feuillées à partir de cals. Une fois apparues, les plantules transformées sont détachées des cals puis repiquées sur un milieu d'enracinement (Tableau 15) contenant une auxine, l'acide naphthalène acétique (NAA) permettant de promouvoir la production racinaire.

## **2.3. Etude de l'activité des protéines recombinantes**

### **2.3.1. À partir de microsomes de levures**

#### **2.3.1.1. Criblage métabolique**

Dans un microtube de 1.5 mL sont mélangés 40  $\mu$ L de microsomes, 100  $\mu$ M de substrat(s), 2  $\mu$ L de NADPH 10mM et du tampon NaPi pH 7.5 pour un volume final de 100  $\mu$ L. Le milieu réactionnel est incubé 20 minutes à 28°C et 1 $\mu$ L d'acide trifluoroacétique (TFA) est ajouté. Le TFA va acidifier le milieu réactionnel, ce qui va permettre d'arrêter la réaction de métabolisation et de précipiter les protéines, et donc les microsomes.

#### **2.3.1.2. Extraction des composés phénoliques à l'acétate d'éthyle**

La solution de métabolisation est centrifugée 30 minutes à 22000 g, puis le surnageant est récupéré et additionné de 800  $\mu$ L d'acétate d'éthyle. Après 5 minutes de centrifugation à 100 g, la phase organique est récupérée et transférée dans un nouveau microtube de 1.5 mL. Une deuxième extraction à partir du culot est réalisée comme précédemment et la phase mobile est ajoutée à celle de la première extraction. La solution est évaporée une nuit (Concentrator plus, Eppendorf) et le culot est resuspendu dans 50  $\mu$ L de méthanol 80%.

## 2.3.2. À partir de *Nicotiana benthamiana*

### 2.3.2.1. Infiltration des substrats

Pour permettre la réaction enzymatique dans la plante, les différents substrats des protéines d'intérêt sont infiltrés dans les feuilles de tabac en utilisant le phénomène d'évapotranspiration. Ainsi, le pétiole de chaque feuille est placé dans un microtube de 1.5 mL contenant une solution de substrat unique à 5mM. Les microtubes contenant la solution de substrat et la feuille de tabac sont placés dans une enceinte confinée disposant de deux ouvertures placées l'une en face de l'autre et permettant la circulation de l'air. L'enceinte contenant les feuilles de tabac est placée à 28 °C pendant 24 heures (Figure 60).



*Figure 60: Dispositif expérimental de l'infiltration de substrats dans les feuilles de tabac transformées*

### 2.3.2.2. Traitement des feuilles transformées

Après l'incubation des feuilles transformées dans la solution de substrat, les feuilles sont dépourvues de leurs pétioles, congelées à l'azote liquide et broyées dans un mortier à l'aide d'un pilon pour permettre l'extraction du contenu métabolique (cf. paragraphe) et des ARNs, traités ensuite pour analyser respectivement les produits de métabolisation et le profil d'expression des gènes cibles. Dans le cas de la GFP, l'expression est observée à la loupe trinoculaire à fluorescence.

## **2.4. Analyses métaboliques**

### **2.4.1. Extraction des composés à analyser**

Les échantillons végétaux sont instantanément congelés à l'azote liquide puis broyés à l'aide d'un mortier et d'un pilon. A partir de la poudre obtenue, et pour chaque échantillon, 100 mg sont placés dans un microtube de 2 mL et mélangés à 800 µL de méthanol 80 %. Après avoir été vortexés pendant une minute, les microtubes sont placés dans un bain à ultrasons à 37kHz (Elmasonic S70, Elma, Wetzikon, Suisse) pendant 10 minutes. Les ultrasons vont permettre d'améliorer la lyse des cellules végétales et accroître l'efficacité d'extraction des polyphénols. Après une centrifugation de 30 minutes à 15000 g et à température ambiante, le surnageant est transféré dans un nouveau microtube de 2 mL.

Une deuxième extraction à partir du culot et selon les mêmes paramètres que précédemment est réalisée pour finalement obtenir un volume final de surnageant de 1.6 mL. Le méthanol 80% contenu dans le surnageant est mis à évaporer pendant une nuit et le culot est resuspendu dans 100 µL de ce même solvant puis analysées par UPLC-MS (3.4.2.1.1.).

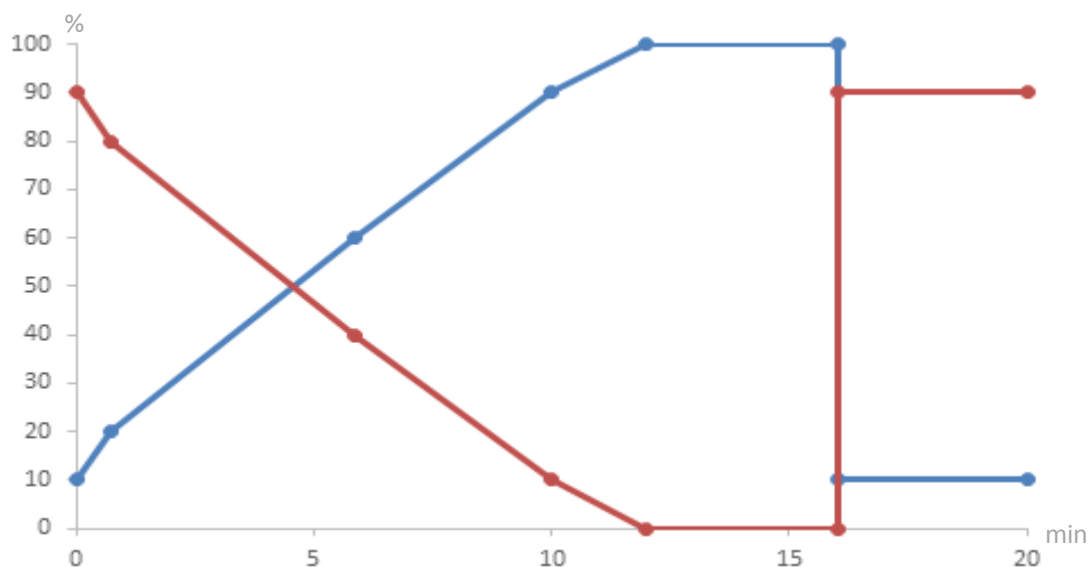
### **2.4.2. Analyse du contenu métabolique**

#### **2.4.2.1. Analyse ciblée des polyphénols**

##### **2.4.2.1.1. UPLC-MS**

Les analyses phytochimiques des différents échantillons végétaux ont été réalisées en utilisant la chromatographie liquide à haute performance (UPLC) Nexera (Shimadzu, Kyoto, préfecture de Kyoto, Japon), équipée d'un détecteur UV SPDM20A (Shimadzu) et couplée à un spectromètre de masse à quadripôle simple LCMS2020 (Shimadzu). Les composés phénoliques sont séparés dans une colonne C18 en phase inverse (ZORBAX Eclipse Plus 150 ou 100 × 2,10 mm ; taille de particule = 1,8 µm ; Agilent Technologies) surmontée d'une pré-colonne (1290 Infinity, Agilent Technologies). Ces deux parties sont placées dans une enceinte à 40°C.

Les solvants d'élution utilisés sont une solution de méthanol pur contenant 0.1% d'acide formique et de l'eau ultrapure contenant 0.1% d'acide formique. Le gradient de la phase mobile est représenté en figure 61.



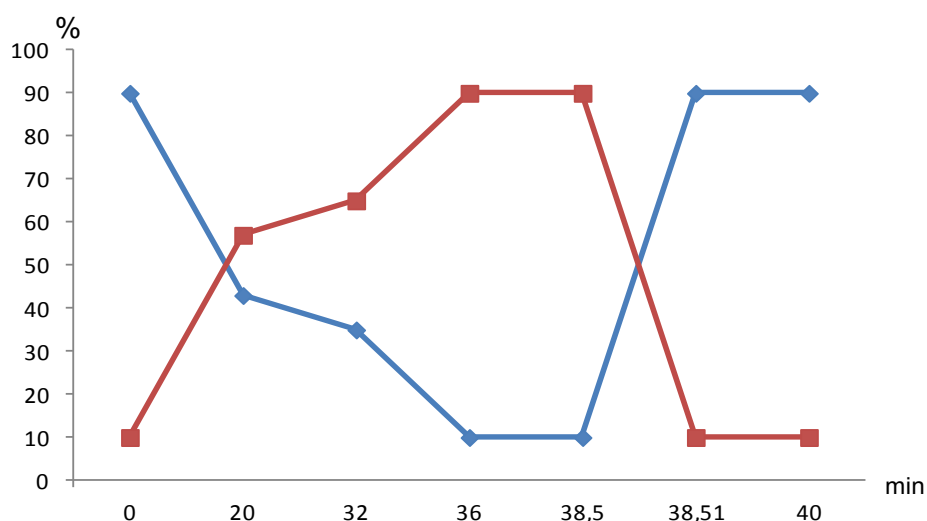
**Figure 61: Gradient de la phase mobile utilisé pour la séparation des composés phénoliques en UPLC.**

*La courbe bleue correspond au gradient de la solution de méthanol + 0.1 % d'acide formique, la courbe rouge correspond à la solution d'eau ultrapure + 0.1% d'acide formique.*

Le débit de la phase mobile est de 0.2mL/minute et le volume d'injection est de 5  $\mu$ L. La détection des composés est réalisée par le détecteur UV sur des longueurs d'onde comprises entre 200 et 600 nm. Une fois séparés, les composés sont analysés par spectrométrie de masse, en mode SIM (Selected Ion Monitoring), permettant le suivi de composés spécifiques ciblés, ce qui augmente la précision de la méthode. Le spectromètre de masse est connecté à l'UPLC par une double source d'ionisation (DUIS) réglée en mode positif, par combinaison de l'ionisation par électrospray et l'ionisation chimique à pression atmosphérique. L'ionisation par électrospray reçoit une tension de 4.5 kV, et les températures du bloc chauffant, de l'entrée et de la ligne de désolvatation du spectromètre de masse sont respectivement réglées à 400 °C, 350 °C et 250 °C. Les données sont enregistrées et traitées par le logiciel LabSolution.

#### 2.4.2.1.2. Orbitrap-LTQ

L'utilisation d'un appareil de chromatographie liquide (Dionex UltiMate® 3000, Thermo Scientific) couplé à un spectromètre de masse en tandem (LTQ XL™ Linear Ion Trap Mass Spectrometer, Thermo Scientific) a permis de déterminer le profil de fragmentation des différents molécules analysées en UHPLC-MS. Avec un volume d'injection de 40 µL, la séparation des composés est réalisée sur une colonne C18 en phase inverse (Alltima™ 150 × 2,1 mm ; taille de particule = 5 µm, Alltech® ). Les solvants d'élution utilisés sont une solution de méthanol pur contenant 0.1% d'acide formique et de l'eau ultrapure contenant 0.1% d'acide formique. Le gradient de la phase mobile est représenté en figure 63. Le spectromètre en tandem dispose d'une trappe à ions permettant de discriminer les composés en fonction de leurs ratios  $m/z$ . Ce dispositif est relié à un système de chromatographie liquide par une source d'ionisation ESI réglée en mode positif, et reçoit une tension de 4.5kV et une température de 300°C. Les débits des gaz coaxial, auxiliaire et barrière sont respectivement réglés à 97 40, 10 et 10 unités arbitraires (UA)/min. A l'entrée de la source d'ionisation, les voltages du capillaire de transfert, du split lens et du front lens sont respectivement réglés à 36 V, -44 V et -3,5 V. L'analyse balaye des ratios  $m/z$  allant de 50 à 1000, mais peut être réalisée à partir de ratios  $m/z$  ciblés. Les données d'analyses sont traitées par le logiciel Xcalibur™ (version 2.1.SP1.Build1160).



**Figure 62: Gradient de la phase mobile utilisé pour la séparation des composés phénoliques en Orbitrap.**

*La courbe bleue correspond au gradient de la solution de méthanol + 0.1 % d'acide formique, la courbe rouge correspond à la solution d'eau ultrapure + 0.1% d'acide formique.*



#### **2.4.2.1.3. Identification et quantification relative des composés phénoliques ciblés**

L'identification des composés phénoliques analysés est réalisée en utilisant une gamme de standards de composés potentiellement présents dans les échantillons.

En comparant le ratio  $m/z$  et le temps de rétention du standard avec ceux des composés analysés, il est possible d'identifier ces composés. Chacun des standards est analysé à une concentration de 10  $\mu\text{M}$ .

La quantification relative des composés analysés est rendue possible grâce à la présence d'un standard interne dans chacun des échantillons, c'est-à-dire d'un composé qui n'est pas naturellement présent dans les échantillons, ici le 4-méthylumbelliférone (10  $\mu\text{M}$ ).

La présence de ce composé permet de limiter les biais de mesure de l'équipement analytique. En parallèle des échantillons, une gamme de concentration (1 $\mu\text{M}$ , 5 $\mu\text{M}$ , 10 $\mu\text{M}$ , 20 $\mu\text{M}$ , 30 $\mu\text{M}$ ) pour chacun des standards est également analysée, et permet, par la génération d'une courbe étalon, de déterminer la concentration d'un composé pour une aire de pic donnée. Ainsi, la quantité relative d'un échantillon donné est définie en mettant en relation l'aire du pic du standard interne, divisée par l'aire du pic du composé, avec la courbe étalon correspondant à ce composé.

#### **2.4.2.2. Analyse métabolomique non ciblée**

Les données brutes issues des analyses U-HPLC-MS ont été converties en fichiers cdf *via* le logiciel LabSolution et prétraitées en utilisant la plate-forme en ligne XCMS. Le prétraitement consiste à détecter chacun des pics présents sur les spectres de masse des échantillons et corriger les temps de rétention de chacune des molécules dans le but d'aligner les différents spectres. La détection des pics est réalisée par la méthode CentWave ( $M/z = 2.5$  ppm, largeur de pics [5-50], seuil signal/bruit de fond = 20).

La correction du temps de rétention a été réalisée avec la méthode obiwarp et l'alignement des composés avec les paramètres suivants (minfrac=0.5, bw=5, mzwid=0.015). L'analyse a été réalisée en comparant les 8 répétitions de chaque traitement à un temps des récoltes donné (T0, T1, T2, T3). L'ensemble des données prétraitées est importé vers Microsoft Excel et converti en fichiers csv.

Les analyses statistiques multivariées et univariées ont été exécutées avec le logiciel en ligne Metaboanalyst 4.0. Avant les analyses statistiques, les données sont filtrées en se basant sur l'écart-type relatif (RSD/mean) dans le but d'éliminer les signaux avec une faible répétabilité. Les données filtrées ont été normalisées en utilisant la méthode Pareto, consistant à transformer chaque donnée en valeur centrée réduite et à diviser ces valeurs par l'erreur standard au carré. L'analyse en composante principale (ACP), la représentation en heatmap et la régression des moindres carrés partiels ont été conduites à partir des données filtrées.

Une analyse de variances univariées (ANOVA,  $p < 0.05$ , Différence significative minimale de Fisher (LSD), test post-hoc) a été réalisée pour identifier les signaux métaboliques différenciellement significatifs entre les différentes conditions.

L'identification des métabolites a été conduite uniquement à partir de ces données différenciellement accumulées. Les données issues de l'analyse HPLC-Orbitrap, contenant des informations concernant la masse exacte et le profil de fragmentation des signaux ont été converties en fichiers. Elles ont été analysées à l'aide des logiciels MS-Dial (v.3.20) et MS-Finder (v.3.12). Des données issues de spectrométrie de masse à haute résolution (HRMS) ont permis de déterminer la formule moléculaire des composés identifiés, tandis que les données issues du MS/MS ont été utilisées pour proposer des structures moléculaires par comparaison avec 21 bases de données métabolomiques, notamment UNPD, LipidMAPS, HMDB, MassBank et Respect. L'identité de la rutine, l'imperatorine, l'osthol, le 8-geranyloxypsoralen, l'acide oléique et l'acide linoléique a été confirmée par comparaison avec des standards.



## Références Bibliographiques

- Abraham, W.R. (2001). Bioactive sesquiterpenes produced by fungi: are they useful for humans as well? *Curr. Med. Chem.* 8, 583–606.
- Aharoni, A., Jongsma, M., and Bouwmeester, H. (2005). Volatile science? Metabolic engineering of terpenoids in plants. *Trends Plant Sci.* 10, 594–602.
- Akhtar, S., Kakei, Y., Bashir, K., Shahzad, A., Yamakawa, T., Arshad, M., Ul-Hassan, F., Nakanishi, H., and Nishizawa, N.K. (2015). Characterizing the expression of genes involved in iron transport in Pakistani peanut varieties under iron deficiency stress. *Turk. J. Bot.* 39, 996–1007.
- Al Khayri, J., and Naik, P. (2016). Impact of Abiotic Elicitors on *In vitro* Production of Plant Secondary Metabolites: A Review. *J. Adv. Res. Biotechnol.* 1, 1–7.
- Anzlovar, S., and Koce, J.D. (2014). Antibacterial and Antifungal Activity of Aqueous and Organic Extracts from Indigenous and Invasive Species of Goldenrod (*Solidago* spp.) Grown in Slovenia. *Phyton-Ann. Rei Bot.* 54, 135–147.
- Arct, J., and Pytkowska, K. (2008). Flavonoids as components of biologically active cosmeceuticals. *Clin. Dermatol.* 26, 347–357.
- Ashiq, B., Chohan, S., Perveen, R., Abid, M., and Mehmood, M.A. (2017). Chemical composition and antifungal potential of medicinal plants against seedborne mycoflora of eggplant (*Solanum melongena* L.). *Acta Bot. Croat.* 76, 72–79.
- Asif, M., Eudes, F., Randhawa, H., Amundsen, E., Yanke, J., and Spaner, D. (2013). Cefotaxime prevents microbial contamination and improves microspore embryogenesis in wheat and triticale. *Plant Cell Rep.* 32, 1637–1646.
- Bailey, D.G., Dresser, G., and Arnold, J.M.O. (2013). Grapefruit–medication interactions: Forbidden fruit or avoidable consequences? *CMAJ Can. Med. Assoc. J.* 185, 309–316.
- Bajwa, S.J.S., and Haldar, R. (2015). Pain management following spinal surgeries: An appraisal of the available options. *J. Craniovertebral Junction Spine* 6, 105.
- Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., and Werck-Reichhart, D. (2011). Cytochromes P450. *Arab. Book Am. Soc. Plant Biol.* 9.
- Ballhorn, D.J., Kautz, S., Heil, M., and Hegeman, A.D. (2009). Analyzing plant defenses in nature. *Plant Signal. Behav.* 4, 743–745.
- Barto, E.K., and Cipollini, D. (2005). Testing the optimal defense theory and the growth-differentiation balance hypothesis in *Arabidopsis thaliana*. *Oecologia* 146, 169–178.
- Bartwal, A., Mall, R., Lohani, P., Guru, S.K., and Arora, S. (2013). Role of Secondary Metabolites and Brassinosteroids in Plant Defense Against Environmental Stresses. *J. Plant Growth Regul.* 32, 216–232.

- Batard, Y., Hehn, A., Nedelkina, S., Schalk, M., Pallett, K., Schaller, H., and Werck-Reichhart, D. (2000). Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems. *Arch. Biochem. Biophys.* 379, 161–169.
- Beauvoit, B., Belouah, I., Bertin, N., Cakpo, C.B., Colombié, S., Dai, Z., Gautier, H., Génard, M., Moing, A., Roch, L., *et al.* (2018). Putting primary metabolism into perspective to obtain better fruits. *Ann. Bot.* 122, 1–21.
- Becerra, J.X. (2007). The impact of herbivore–plant coevolution on plant community structure. *Proc. Natl. Acad. Sci.* 104, 7483–7488.
- Benovoy, D., and Drouin, G. (2006). Processed pseudogenes, processed genes, and spontaneous mutations in the *Arabidopsis* genome. *J. Mol. Evol.* 62, 511–522.
- Benvenuti, S., Bortolotti, E., and Maggini, R. (2016). Antioxidant power, anthocyanin content and organoleptic performance of edible flowers. *Sci. Hortic.* 199, 170–177.
- Berenbaum, M.R. (1990). Evolution of Specialization in Insect-Umbellifer Associations. *Annu. Rev. Entomol.* 35, 319–343.
- Berenbaum, M., and Feeny, P. (1981). Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? *Science* 212, 927–929.
- Berenbaum, M.R., Nitao, J.K., and Zangerl, A.R. (1991). Adaptive significance of furanocoumarin diversity in *Pastinaca sativa* (Apiaceae). *J. Chem. Ecol.* 17, 207–215.
- Bernays, E.A. (2017). *Insect-Plant Interactions* (1990) (CRC Press).
- Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *J. Biotechnol.* 124, 128–145.
- Bertea, C.M., Schalk, M., Karp, F., Maffei, M., and Croteau, R. (2001). Demonstration That Menthofuran Synthase of Mint (*Mentha*) Is a Cytochrome P450 Monooxygenase: Cloning, Functional Expression, and Characterization of the Responsible Gene. *Arch. Biochem. Biophys.* 390, 279–286.
- Bhattacharyya, M.K., and Ward, E.W.B. (1986). Expression of gene-specific and age-related resistance and the accumulation of glyceollin in soybean leaves infected with *Phytophthora megasperma* f. sp. *glycinea*. *Physiol. Mol. Plant Pathol.* 29, 105–113.
- Bonnelye, E., and Laudet, V. (1994). Les gènes chevauchants. *médecine/sciences* 10, 805.
- Bourgaud, F., Gravot, A., Milesi, S., and Gontier, E. (2001). Production of plant secondary metabolites: a historical perspective. *Plant Sci.* 161, 839–851.
- Bourgaud, F., Hehn, A., Larbat, R., Doerper, S., Gontier, E., Kellner, S., and Matern, U. (2006). Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem. Rev.* 5, 293–308.
- Bücherl, C.A., Bader, A., Westphal, A.H., Liptenok, S.P., and Borst, J.W. (2014). FRET-FLIM applications in plant systems. *Protoplasma* 251, 383–394.
- Cain, N., Darbyshire, S.J., Francis, A., Nurse, R.E., and Simard, M.-J. (2010). The Biology of Canadian weeds. 144. *Pastinaca sativa* L. *Can. J. Plant Sci.* 90, 217–240.

- Calla, B., Noble, K., Johnson, R.M., Walden, K.K.O., Schuler, M.A., Robertson, H.M., and Berenbaum, M.R. (2017). Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: Birth, death and adaptation. *Mol. Ecol.* *26*, 6021–6035.
- Caron, A.W., Archambault, J., and Massie, B. (1990). High-level recombinant protein production in bioreactors using the baculovirus-insect cell expression system. *Biotechnol. Bioeng.* *36*, 1133–1140.
- Cassinat, B., and Nicolas, M.H. (1994). Comparison of antibiotic combinations against penicillin-resistant pneumococci. *J. Antimicrob. Chemother.* *34*, 785–790.
- Cecchini, C., Coman, M.M., Cresci, A., Tirillini, B., Cristalli, G., Papa, F., Sagratini, G., Vittori, S., and Maggi, F. (2010). Essential oil from fruits and roots of *Ferulago campestris* (Besser) Grecescu (Apiaceae): composition and antioxidant and anti-Candida activity. *Flavour Fragr. J.* *25*, 493–502.
- Chan, K.X., Wirtz, M., Phua, S.Y., Estavillo, G.M., and Pogson, B.J. (2013). Balancing metabolites in drought: the sulfur assimilation conundrum. *Trends Plant Sci.* *18*, 18–29.
- Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* *4*, 258–268.
- Chhabra, D.N. (2013). Lac operon (Regulation of gene expression in Prokaryotes).
- Christensen, L.P., and Brandt, K. (2006). Bioactive polyacetylenes in food plants of the Apiaceae family: Occurrence, bioactivity and analysis. *J. Pharm. Biomed. Anal.* *41*, 683–693.
- Cistué, L., Ramos, A., Castillo, A.M., and Romagosa, I. (1994). Production of large number of doubled haploid plants from barley anthers pretreated with high concentrations of mannitol. *Plant Cell Rep.* *13*, 709–712.
- Clemens, S., and Weber, M. (2015). The essential role of coumarin secretion for Fe acquisition from alkaline soil. *Plant Signal Behav* *11*, e1114197.
- Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* *26*, 183–186.
- Conant, G.C., Birchler, J.A., and Pires, J.C. (2014). Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* *19*, 91–98.
- Cook, D.W., Burnham, M.L., Harmes, D.C., Stoll, D.R., and Rutan, S.C. (2017). Comparison of multivariate curve resolution strategies in quantitative LCxLC: Application to the quantification of furanocoumarins in apiaceous vegetables. *Anal. Chim. Acta* *961*, 49–58.
- Cooper, G.M. (2000). Chromosomes and Chromatin. *Cell Mol. Approach* 2nd Ed.
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* *16*, 10881–10890.
- Cremasco, M.A., and Mochi, V.T. (2013). Reaction of dissolved ozone in hydrogen peroxide produced during ozonization of an alkaline medium in a bubble column. *Acta Sci. Technol.* *36*, 81–85.

- Crosby, K.C., Pietraszewska-Bogiel, A., Gadella, T.W.J., and Winkel, B.S.J. (2011). Förster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes. *FEBS Lett.* *585*, 2193–2198.
- D’Amelia, V., Aversano, R., Ruggiero, A., Batelli, G., Appelhagen, I., Dinacci, C., Hill, L., Martin, C., and Carputo, D. (2018). Subfunctionalization of duplicate MYB genes in *Solanum commersonii* generated the cold-induced ScAN2 and the anthocyanin regulator ScAN1. *Plant Cell Environ.* *41*, 1038–1051.
- Dardalhon, M., de Massy, B., Nicolas, A., and Averbek, D. (1998). Mitotic recombination and localized DNA double-strand breaks are induced after 8-methoxypsoralen and UVA irradiation in *Saccharomyces cerevisiae*. *Curr. Genet.* *34*, 30–42.
- Didierjean, L., Gondet, L., Perkins, R., Lau, S.-M.C., Schaller, H., O’Keefe, D.P., and Werck-Reichhart, D. (2002). Engineering Herbicide Metabolism in Tobacco and Arabidopsis with CYP76B1, a Cytochrome P450 Enzyme from Jerusalem Artichoke. *Plant Physiol.* *130*, 179–189.
- Drouin, G., and Dover, G.A. (1990). Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. *J. Mol. Evol.* *31*, 132–150.
- Dueholm, B., Krieger, C., Drew, D., Olry, A., Kamo, T., Taboureau, O., Weitzel, C., Bourgaud, F., Hehn, A., and Simonsen, H.T. (2015). Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evol. Biol.* *15*, 122.
- Dugrand, A., Olry, A., Duval, T., Hehn, A., Froelicher, Y., and Bourgaud, F. (2013). Coumarin and Furanocoumarin Quantitation in Citrus Peel via Ultraperformance Liquid Chromatography Coupled with Mass Spectrometry (UPLC-MS). *J. Agric. Food Chem.* *61*, 10677–10684.
- Dutartre, L., Hilliou, F., and Feyereisen, R. (2012). Phylogenomics of the benzoxazinoid biosynthetic pathway of Poaceae: gene duplications and origin of the Bx cluster. *BMC Evol. Biol.* *12*, 64.
- Dytrtova, J.J., Straka, M., Belonoznikova, K., Jakl, M., and Ryslava, H. (2018). Does resveratrol retain its antioxidative properties in wine? Redox behaviour of resveratrol in the presence of Cu(II) and tebuconazole. *Food Chem.* *262*, 221–225.
- Ehrlich, P.R., and Raven, P.H. (1964). Butterflies and Plants: A Study in Coevolution. *Evolution* *18*, 586–608.
- Eisenbrand, G. (2007). Toxicological Assessment of Furocoumarins in Foodstuffs. *Mol. Nutr. Food Res.* *51*, 367–373.
- El-Sayed, M., and Verpoorte, R. (2007). Catharanthus terpenoid indole alkaloids: biosynthesis and regulation. *Phytochem. Rev.* *6*, 277–305.

- Farrow, S.C., Kamileen, M.O., Meades, J., Ameyaw, B., Xiao, Y., and O'Connor, S.E. (2018). Cytochrome P450 and O-methyltransferase catalyze the final steps in the biosynthesis of the anti-addictive alkaloid ibogaine from *Tabernanthe iboga*. *J. Biol. Chem.* *293*, 13821–13833.
- Ferrari, S. (2010). Biological elicitors of plant secondary metabolites: mode of action and use in the production of nutraceuticals. *Adv. Exp. Med. Biol.* *698*, 152–166.
- Field, B., and Osbourn, A.E. (2008). Metabolic Diversification--Independent Assembly of Operon-Like Gene Clusters in Different Plants. *Science* *320*, 543–547.
- Field, B., Fiston-Lavier, A.-S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A.E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 16116–16121.
- Fitzpatrick, M.C., Gove, A.D., Sanders, N.J., and Dunn, R.R. (2008). Climate change, plant migration, and range collapse in a global biodiversity hotspot: the *Banksia* (Proteaceae) of Western Australia. *Glob. Change Biol.* *14*, 1337–1352.
- Floss, H.-G., and Mothes, U. (1966). On the biosynthesis of furocoumarins in *Pimpinella magna*. *Phytochemistry* *5*, 161–169.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* *151*, 1531–1545.
- Frey, M. (1997). Analysis of a Chemical Plant Defense Mechanism in Grasses. *Science* *277*, 696–699.
- Gatehouse, J.A. (1997). Regulation of Gene Expression in Plants. In *Plant Biochemistry*, (Elsevier), 353–385.
- Geiduschek, E.P. (2010). An Introduction to Transcription and Gene Regulation. *J. Biol. Chem.* *285*, 25885–25892.
- Geriyol, P., Basavanneppa, H.B., and Dhananjaya, B.L. (2015). Protecting effect of caffeine against vinblastine (an anticancer drug) induced genotoxicity in mice. *Drug Chem. Toxicol.* *38*, 188–195.
- Glawischnig, E., Hansen, B.G., Olsen, C.E., and Halkier, B.A. (2004). Camalexin is synthesized from indole-3-acetaldoxime, a key branching point between primary and secondary metabolism in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 8245–8250.
- Gonzales-Vigil, E., Hufnagel, D.E., Kim, J., Last, R.L., and Barry, C.S. (2012). Evolution of TPS20-related terpene synthases influences chemical diversity in the glandular trichomes of the wild tomato relative *Solanum habrochaites*. *Plant J.* *71*, 921–935.
- Gorlov, I.P., Kimmel, M., and Amos, C.I. (2006). Strength of the purifying selection against different categories of the point mutations in the coding regions of the human genome. *Hum. Mol. Genet.* *15*, 1143–1150.



- Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* 267, 83–90.
- Gravot, A., Larbat, R., Hehn, A., Lièvre, K., Gontier, E., Goergen, J.-L., and Bourgaud, F. (2004). Cinnamic acid 4-hydroxylase mechanism-based inactivation by psoralen derivatives: cloning and characterization of a C4H from a psoralen producing plant—*Ruta graveolens*—exhibiting low sensitivity to psoralen inactivation. *Arch. Biochem. Biophys.* 422, 71–80.
- Green, D.M. (2017). Amphibian breeding phenology trends under climate change: predicting the past to forecast the future. *Glob. Change Biol.* 23, 646–656.
- Guo, F., Lei, J., Sun, Y., Chi, Y.H., Ge, F., Patil, B.S., Koiwa, H., Zeng, R., and Zhu-Salzman, K. (2012). Antagonistic Regulation, Yet Synergistic Defense: Effect of Bergapten and Protease Inhibitor on Development of Cowpea Bruchid *Callosobruchus maculatus*. *PLOS ONE* 7, e41877.
- Hajlaoui, H., Mighri, H., Noumi, E., Snoussi, M., Trabelsi, N., Ksouri, R., and Bakhrouf, A. (2010). Chemical composition and biological activities of Tunisian *Cuminum cyminum* L. essential oil: A high effectiveness against *Vibrio* spp. strains. *Food Chem. Toxicol.* 48, 2186–2192.
- Hamerski, D., and Matern, U. (1988a). Elicitor-Induced Biosynthesis of Psoralens in Ammi-Majus L Suspension-Cultures - Microsomal Conversion of Demethylsuberosin into (+)marmesin and Psoralen. *Eur. J. Biochem.* 171, 369–375.
- Hamerski, D., and Matern, U. (1988b). Biosynthesis of psoralens Psoralen Smonooxygenase activity from elicitor-treated Ammi majus cells. *FEBS Lett.* 239, 3.
- Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A., and Deisenhofer, J. (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Struct. Lond. Engl.* 1993 3, 41–62.
- Heinke, R., Franke, K., Michels, K., Wessjohann, L., Ali, N.A.A., and Schmidt, J. (2012). Analysis of furanocoumarins from Yemenite *Dorstenia* species by liquid chromatography/electrospray tandem mass spectrometry: Analysis of furanocoumarins. *J. Mass Spectrom.* 47, 7–22.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.* 42, 819–832.
- Hermes, D.A., and Mattson, W.J. (1992). The Dilemma of Plants: To Grow or Defend. *Q. Rev. Biol.* 67, 283–335.
- Höfer, R., Boachon, B., Renault, H., Gavira, C., Miesch, L., Iglesias, J., Ginglinger, J.-F., Allouche, L., Miesch, M., Grec, S., *et al.* (2014). Dual Function of the Cytochrome P450 CYP76 Family from *Arabidopsis thaliana* in the Metabolism of Monoterpenols and Phenylurea Herbicides1[W][OPEN]. *Plant Physiol.* 166, 1149–1161.

- Holopainen, J.K., Virjamo, V., Ghimire, R.P., Blande, J.D., Julkunen-Tiitto, R., and Kivimäenpää, M. (2018). Climate Change Effects on Secondary Compounds of Forest Trees in the Northern Hemisphere. *Front. Plant Sci.* 9, 1445.
- Hood, E.E., Gelvin, S.B., Melchers, L.S., and Hoekema, A. (1993). New *Agrobacterium* helper plasmids for gene transfer to plants. *Transgenic Res.* 2, 208–218.
- Hou, Q., Ufer, G., and Bartels, D. (2016). Lipid signalling in plant responses to abiotic stress. *Plant Cell Environ.* 39, 1029–1048.
- Hu, D., Liu, X.B., She, H.Z., Gao, Z., Ruan, R.W., Wu, D.Q., and Yi, Z.L. (2017). The lignin synthesis related genes and lodging resistance of *Fagopyrum esculentum*. *Biol. Plant.* 61, 138–146.
- Huang, C.-Y., Ju, D.-T., Chang, C.-F., Muralidhar Reddy, P., and Velmurugan, B.K. (2017). A review on the effects of current chemotherapy drugs and natural agents in treating non-small cell lung cancer. *BioMedicine* 7, 12–23.
- Hubert, J., Nuzillard, J.-M., and Renault, J.-H. (2017). Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem. Rev.* 16, 55–95.
- Hughes, A.L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B* 256, 119–124.
- Hughes, A.L. (2002). Adaptive evolution after gene duplication. *Trends Genet.* 18, 433–434.
- Hunault, G., Desmarest, P., and Manoir, J.D. (1989). *Foeniculum vulgare* : Cell Culture, Regeneration, and the Production of Anethole. In *Medicinal and Aromatic Plants II*, (Springer, Berlin, Heidelberg), pp. 185–212.
- Huot, B., Yao, J., Montgomery, B.L., and He, S.Y. (2014). Growth–Defense Tradeoffs in Plants: A Balancing Act to Optimize Fitness. *Mol. Plant* 7, 1267–1287.
- Hurst, L.D., Pál, C., and Lercher, M.J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* 5, 299–310.
- Hussain, T., Tan, B., Yin, Y., Blachier, F., Tossou, M.C.B., and Rahu, N. (2016). Oxidative Stress and Inflammation: What Polyphenols Can Do for Us? *Oxid. Med. Cell. Longev.* 2016.
- Ibrahim, T.A., El-Hela, A.A., El-Hefnawy, H.M., Al-Taweel, A.M., and Perveen, S. (2017). Chemical Composition and Antimicrobial Activities of Essential Oils of Some Coniferous Plants Cultivated in Egypt. *Iran. J. Pharm. Res. IJPR* 16, 328–337.
- Ikezawa, N., Iwasa, K., and Sato, F. (2009). CYP719A subfamily of cytochrome P450 oxygenases and isoquinoline alkaloid biosynthesis in *Eschscholzia californica*. *Plant Cell Rep.* 28, 123–133.
- Ioannou, I., Hafsa, I., Hamdi, S., Charbonnel, C., and Ghoul, M. (2012). Review of the effects of food processing and formulation on flavonol and anthocyanin behaviour. *J. Food Eng.* 111, 208–217.

- Itkin, M., Rogachev, I., Alkan, N., Rosenberg, T., Malitsky, S., Masini, L., Meir, S., Iijima, Y., Aoki, K., de Vos, R., *et al.* (2011). Glycoalkaloid metabolism is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell* 23, 4507–4525.
- Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R., *et al.* (2013). Biosynthesis of Antinutritional Alkaloids in Solanaceous Crops Is Mediated by Clustered Genes. *Science* 341, 175–179.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356.
- Jacob, F., Perrin, D., Sánchez, C., and Monod, J. (1960). L'opéron : groupe de gènes à expression coordonnée par un opérateur [C. R. Acad. Sci. Paris 250 (1960) 1727–1729]. *C. R. Biol.* 328, 514–520.
- Jana, S., and Shekhawat, G.S. (2010). *Anethum graveolens*: An Indian traditional medicinal herb and spice. *Pharmacogn. Rev.* 4, 179–184.
- Jian, W., He, D., and Song, S. (2016). Synthesis, Biological Evaluation, and Molecular Modeling Studies of New Oxadiazole-Stilbene Hybrids against Phytopathogenic Fungi. *Sci. Rep.* 6, 31045.
- Jin, J., Zhang, H., Zhang, J., Liu, P., Chen, X., Li, Z., Xu, Y., Lu, P., and Cao, P. (2017). Integrated transcriptomics and metabolomics analysis to characterize cold stress responses in *Nicotiana tabacum*. *BMC Genomics* 18.
- Jørgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjarnholt, N., Zagrobelny, M., Bak, S., and Møller, B.L. (2005). Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8, 280–291.
- Kai, K., Mizutani, M., Kawamura, N., Yamamoto, R., Tamai, M., Yamaguchi, H., Sakata, K., and Shimizu, B. (2008). Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. *Plant J.* 55, 989–999.
- Kaplan, I., Halitschke, R., Kessler, A., Sardanelli, S., and Denno, R.F. (2008). Constitutive and induced defenses to herbivory in above- and belowground plant tissues. *Ecology* 89, 392–406.
- Karamat, F., Olry, A., Doerper, S., Vialart, G., Ullmann, P., Werck-Reichhart, D., Bourgaud, F., and Hehn, A. (2012). CYP98A22, a phenolic ester 3'-hydroxylase specialized in the synthesis of chlorogenic acid, as a new tool for enhancing the furanocoumarin concentration in *Ruta graveolens*. *BMC Plant Biol.* 12, 152.
- Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., Hehn, A., Bourgaud, F., and Yazaki, K. (2014). A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *Plant J. Cell Mol. Biol.* 77, 627–638.
- Karasov, T.L., Chae, E., Herman, J.J., and Bergelson, J. (2017). Mechanisms to Mitigate the Trade-Off between Growth and Defense. *Plant Cell* 29, 666–680.

- Karp, F., Mihaliak, C.A., Harris, J.L., and Croteau, R. (1990). Monoterpene biosynthesis: specificity of the hydroxylations of (-)-limonene by enzyme preparations from peppermint (*Mentha piperita*), spearmint (*Mentha spicata*), and perilla (*Perilla frutescens*) leaves. *Arch. Biochem. Biophys.* 276, 219–226.
- Kawahata, M., Amari, S., Nishizawa, Y., and Akada, R. (1999). A positive selection for plasmid loss in *Saccharomyces cerevisiae* using galactose-inducible growth inhibitory sequences. *Yeast* 15, 1–10.
- Kim, S., Schlicke, H., Van Ree, K., Karvonen, K., Subramaniam, A., Richter, A., Grimm, B., and Braam, J. (2013). Arabidopsis Chlorophyll Biosynthesis: An Essential Balance between the Methylerythritol Phosphate and Tetrapyrrole Pathways. *Plant Cell* 25, 4984–4993.
- Klingenberg, M. (1958). Pigments of rat liver microsomes. *Arch. Biochem. Biophys.* 75, 376–386.
- Koppula, S., and Choi, D.K. (2011). *Cuminum cyminum* extract attenuates scopolamine-induced memory loss and stress-induced urinary biochemical changes in rats: A noninvasive biochemical approach. *Pharm. Biol.* 49, 702–708.
- Kosak, S.T., and Groudine, M. (2004). Form follows function: the genomic organization of cellular differentiation. *Genes Dev.* 18, 1371–1384.
- Kratz, J.M., Andrighetti-Fröhner, C.R., Leal, P.C., Nunes, R.J., Yunes, R.A., Trybala, E., Bergström, T., Barardi, C.R.M., and Simões, C.M.O. (2008). Evaluation of anti-HSV-2 activity of gallic acid and pentyl gallate. *Biol. Pharm. Bull.* 31, 903–907.
- Kräutler, B. (2008). Chlorophyll breakdown and chlorophyll catabolites in leaves and fruit. *Photochem. Photobiol. Sci. Off. J. Eur. Photochem. Assoc. Eur. Soc. Photobiol.* 7, 1114–1120.
- Kräutler, B., and Hörtensteiner, S. (2006). Chlorophyll Catabolites and the Biochemistry of Chlorophyll Breakdown. In *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*, B. Grimm, R.J. Porra, W. Rüdiger, and H. Scheer, eds. (Dordrecht: Springer Netherlands), pp. 237–260.
- Krieger, C., Roselli, S., Kellner-Thielmann, S., Galati, G., Schneider, B., Grosjean, J., Olry, A., Ritchie, D., Matern, U., Bourgaud, F., *et al.* (2018). The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants. *Front. Plant Sci.* 9.
- Kristensen, C., Morant, M., Olsen, C.E., Ekstrøm, C.T., Galbraith, D.W., Lindberg Møller, B., and Bak, S. (2005). Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1779–1784.
- Kruse, T., Ho, K., Yoo, H.-D., Johnson, T., Hippely, M., Park, J.-H., Flavell, R., and Bobzin, S. (2008). In Planta Biocatalysis Screen of P450s Identifies 8-Methoxypsoralen as a Substrate for the CYP82C Subfamily, Yielding Original Chemical Structures. *Chem. Biol.* 15, 149–156.
- Kumar, S. (2010). Engineering Cytochrome P450 Biocatalysts for Biotechnology, Medicine, and Bioremediation. *Expert Opin. Drug Metab. Toxicol.* 6, 115–131.

- Kumar, P.A., and Sharma, R.P. (1995). Codon Usage in Brassica Genes. *J. Plant Biochem. Biotechnol.* *4*, 113–115.
- Kumar, S., Kumari, R., and Pandey, R. (2015). New insight-guided approaches to detect, cure, prevent and eliminate malaria. *Protoplasma* *252*, 717–753.
- Kumari, M., Clarke, H.J., Small, I., and Siddique, K.H.M. (2009). Albinism in Plants: A Major Bottleneck in Wide Hybridization, Androgenesis and Doubled Haploid Culture. *Crit. Rev. Plant Sci.* *28*, 393–409.
- Küpper, A., Peter, F., Zöllner, P., Lorentz, L., Tranel, P.J., Beffa, R., and Gaines, T.A. (2018). Tembotrione detoxification in 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibitor-resistant Palmer amaranth (*Amaranthus palmeri* S. Wats.). *Pest Manag. Sci.* *74*, 2325–2334.
- Lakatos, L., Szittyá, G., Silhavy, D., and Burgyán, J. (2004). Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *EMBO J.* *23*, 876–884.
- Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., Bourgaud, F., and Matern, U. (2007). Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *J. Biol. Chem.* *282*, 542–554.
- Larbat, R., Hehn, A., Hans, J., Schneider, S., Jugdé, H., Schneider, B., Matern, U., and Bourgaud, F. (2009). Isolation and functional characterization of CYP71AJ4 encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. *J. Biol. Chem.* *284*, 4776–4785.
- Lattanzio, V. (2013). Phenolic Compounds: Introduction. In *Natural Products*, K.G. Ramawat, and J.-M. Mérillon, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 1543–1580.
- Laursen, T., Borch, J., Knudsen, C., Bavishi, K., Torta, F., Martens, H.J., Silvestro, D., Hatzakis, N.S., Wenk, M.R., Dafforn, T.R., *et al.* (2016). Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science* *354*, 890–893.
- Lee, J.H., Lee, J., Kim, H., Chae, W.B., Kim, S.-J., Lim, Y.P., and Oh, M.-H. (2018). Brassinosteroids regulate glucosinolate biosynthesis in *Arabidopsis thaliana*. *Physiol. Plant.* *163*, 450–458.
- Lee, S., Suh, S., Kim, S., Crain, R.C., Kwak, J.M., Nam, H.-G., and Lee, Y. (1997). Systemic elevation of phosphatidic acid and lysophospholipid levels in wounded plants. *Plant J.* *12*, 547–556.
- Li, H.-R., Habasi, M., Xie, L.-Z., and Aisa, H.A. (2014). Effect of chlorogenic acid on melanogenesis of B16 melanoma cells. *Mol. Basel Switz.* *19*, 12940–12948.
- Ludwig, A.A., Romeis, T., and Jones, J.D.G. (2004). CDPK-mediated signalling pathways: specificity and cross-talk. *J. Exp. Bot.* *55*, 181–188.
- Lynch, M., and Conery, J.S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science* *290*, 1151–1155.
- Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* *154*, 459–473.

- Makowska, K., and Oleszczuk, S. (2014). Albinism in barley androgenesis. *Plant Cell Rep.* 33, 385–392.
- Mameda, R., Waki, T., Kawai, Y., Takahashi, S., and Nakayama, T. (2018). Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of GmCHR5, which interacts with 2-hydroxyisoflavanone synthase. *Plant J.* 96, 56–74.
- Marchica, A., Lorenzini, G., Papini, R., Bernardi, R., Nali, C., and Pellegrini, E. (2019). Signalling molecules responsive to ozone-induced oxidative stress in *Salvia officinalis*. *Sci. Total Environ.* 657, 568–576.
- Margna, U. (1977). Control at the level of substrate supply—an alternative in the regulation of phenylpropanoid accumulation in plant cells. *Phytochemistry* 16, 419–426.
- Maruta, T., Noshi, M., Nakamura, M., Matsuda, S., Tamoi, M., Ishikawa, T., and Shigeoka, S. (2014). Ferulic acid 5-hydroxylase 1 is essential for expression of anthocyanin biosynthesis-associated genes and anthocyanin accumulation under photooxidative stress in *Arabidopsis*. *Plant Sci. Int. J. Exp. Plant Biol.* 219–220, 61–68.
- Mastelic, J., Politeo, O., Jerkovic, I., and Radosevic, N. (2005). Composition and Antimicrobial Activity of *Helichrysum italicum* Essential Oil and Its Terpene and Terpenoid Fractions. *Chem. Nat. Compd.* 41, 35–40.
- Matsuno, M., Compagnon, V., Schoch, G.A., Schmitt, M., Debayle, D., Bassard, J.-E., Pollet, B., Hehn, A., Heintz, D., Ullmann, P., *et al.* (2009). Evolution of a Novel Phenolic Pathway for Pollen Development. *Science* 325, 1688–1692.
- Mayer, K., Schüller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Düsterhöft, A., Stiekema, W., Entian, K.-D., Terryn, N., *et al.* (1999). Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402, 769–777.
- Mazza, C.A., Boccalandro, H.E., Giordano, C.V., Battista, D., Scopel, A.L., and Ballaré, C.L. (2000). Functional Significance and Induction by Solar Radiation of Ultraviolet-Absorbing Sunscreens in Field-Grown Soybean Crops. *Plant Physiol.* 122, 117–126.
- McCloud, E.S., Berenbaum, M.R., and Tuveson, R.W. (1992). Furanocoumarin content and phototoxicity of rough lemon (*Citrus jambhiri*) foliage exposed to enhanced ultraviolet-B (UVB) irradiation. *J. Chem. Ecol.* 18, 1125–1137.
- Meijer, H.J.G., and Munnik, T. (2003). Phospholipid-Based Signaling in Plants. *Annu. Rev. Plant Biol.* 54, 265–306.
- Melough, M.M., Cho, E., and Chun, O.K. (2018). Furocoumarins: A review of biochemical activities, dietary sources and intake, and potential health risks. *Food Chem. Toxicol.* 113, 99–107.
- Mendoza, N., and Silva, E.M.E. (2018). Introduction to Phytochemicals: Secondary Metabolites from Plants with Active Principles for Pharmacological Importance. *Phytochem. - Source Antioxid. Role Dis. Prev.*
- Meragelman, T.L., Tucker, K.D., McCloud, T.G., Cardellina, J.H., and Shoemaker, R.H. (2005). Antifungal flavonoids from *Hildebrandia barteri*. *J. Nat. Prod.* 68, 1790–1792.

- Milesi, S., Massot, B., Gontier, E., Bourgaud, F., and Guckert, A. (2001). *Ruta graveolens* L.: a promising species for the production of furanocoumarins. *Plant Sci.* *161*, 189–199.
- Mizutani, M., and Ohta, D. (1998). Two Isoforms of NADPH:Cytochrome P450 Reductase in *Arabidopsis thaliana*: Gene Structure, Heterologous Expression in Insect Cells, and Differential Regulation. *Plant Physiol.* *116*, 357–367.
- Mizutani, M., and Ohta, D. (2010). Diversification of P450 Genes During Land Plant Evolution. *Annu. Rev. Plant Biol.* *61*, 291–315.
- Mlotshwa, S., Pruss, G.J., Gao, Z., Mnguthini, N.L., Li, J., Chen, X., Bowman, L.H., and Vance, V. (2010). Transcriptional Silencing Induced by *Arabidopsis* T-DNA Mutants is Associated with 35S Promoter siRNAs and Requires Genes Involved in siRNA-mediated Chromatin Silencing. *Plant J. Cell Mol. Biol.* *64*, 699–704.
- Mohan, V., Borovsky, Y., Kamara, I., Zemach, H., and Paran, I. (2018). CaVIL1, a plant homeodomain gene that promotes flowering in pepper. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* *131*, 2639–2649.
- Montavon, T., and Duboule, D. (2013). Chromatin organization and global regulation of Hox gene clusters. *Philos. Trans. R. Soc. B Biol. Sci.* *368*.
- Morteza-Semnani, K., and Ghanbarimasir, Z. (2019). A review on traditional uses, phytochemistry and pharmacological activities of the genus *Ballota*. *J. Ethnopharmacol.* *233*, 197–217.
- Mugford, S.T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R.K., Papadopoulou, K., Melton, R., Philo, M., Sainsbury, F., *et al.* (2009). A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell* *21*, 2473–2484.
- Mugford, S.T., Louveau, T., Melton, R., Qi, X., Bakht, S., Hill, L., Tsurushima, T., Honkanen, S., Rosser, S.J., Lomonosoff, G.P., *et al.* (2013). Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell* *25*, 1078–1092.
- Müller, T., Ulrich, M., Ongania, K.-H., and Kräutler, B. (2007). Colorless Tetrapyrrolic Chlorophyll Catabolites Found in Ripening Fruit Are Effective Antioxidants. *Angew. Chem. Int. Ed.* *46*, 8699–8702.
- Munakata, R., Inoue, T., Koeduka, T., Sasaki, K., Tsurumaru, Y., Sugiyama, A., Uto, Y., Hori, H., Azuma, J., and Yazaki, K. (2012). Characterization of Coumarin-Specific Prenyltransferase Activities in *Citrus limon* Peel. *Biosci. Biotechnol. Biochem.* *76*, 1389–1393.
- Munakata, R., Inoue, T., Koeduka, T., Karamat, F., Olry, A., Sugiyama, A., Takanashi, K., Dugrand, A., Froelicher, Y., Tanaka, R., *et al.* (2014). Molecular Cloning and Characterization of a Geranyl Diphosphate-Specific Aromatic Prenyltransferase from Lemon. *Plant Physiol.* *166*, 80–90.
- Munakata, R., Olry, A., Karamat, F., Courdavault, V., Sugiyama, A., Date, Y., Krieger, C., Silie, P., Foureau, E., Papon, N., *et al.* (2016). Molecular evolution of parsnip (*Pastinaca*

*sativa*) membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytol.* *211*, 332–344.

Murthy, H.N., Lee, E.-J., and Paek, K.-Y. (2014). Production of secondary metabolites from cell and organ cultures: strategies and approaches for biomass improvement and metabolite accumulation. *Plant Cell Tissue Organ Cult. PCTOC* *118*, 1–16.

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., *et al.* (2014). The genome of *Eucalyptus grandis*. *Nature* *510*, 356–362.

Neal, J.J., and Wu, D. (1994). Inhibition of Insect Cytochromes P450 by Furanocoumarins. *Pestic. Biochem. Physiol.* *50*, 43–50.

Nelson, D.R. (1999). Cytochrome P450 and the individuality of species. *Arch. Biochem. Biophys.* *369*, 1–10.

Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalus, I.C., and Gotoh, O. (1993). The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol.* *12*, 1–51.

Nelson, D.R., Schuler, M.A., Paquette, S.M., Werck-Reichhart, D., and Bak, S. (2004). Comparative Genomics of Rice and Arabidopsis. Analysis of 727 Cytochrome P450 Genes and Pseudogenes from a Monocot and a Dicot. *Plant Physiol.* *135*, 756–772.

Nguyen, Q.-H., Talou, T., Cerny, M., Evon, P., and Merah, O. (2015). Oil and fatty acid accumulation during coriander (*Coriandrum sativum* L.) fruit ripening under organic cultivation. *Crop J.* *3*, 366–369.

Nikolic, N.C., Lazic, M.M., Karabegovic, I.T., Stojanovic, G.S., and Todorovic, Z.B. (2014). A Characterization of Content, Composition and Scavenging Capacity of Phenolic Compounds in Parsnip Roots of Various Weight. *Nat. Prod. Commun.* *9*, 811–814.

Nishiyama, I., and Motoyoshi, F. (1962). Cytogenetic Studies in *Avena*, X The Artificial Culture of Albino Sand Oats. *Jpn. J. Genet.* *37*, 427–440.

Niu, G., Wen, Z., Rupasinghe, S.G., Zeng, R.S., Berenbaum, M.R., and Schuler, M.A. (2008). Aflatoxin B1 detoxification by CYP321A1 in *Helicoverpa zea*. *Arch. Insect Biochem. Physiol.* *69*, 32–45.

Niu, Y., Bai, J., and Zheng, S. (2018). The Regulation and Function of Histone Methylation. *J. Plant Biol.* *61*, 347–357.

Nützmann, H.-W., and Osbourn, A. (2014). Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.* *26*, 91–99.

Nützmann, H.-W., Huang, A., and Osbourn, A. (2016). Plant metabolic clusters - from genetics to genomics. *New Phytol.* *211*, 771–789.

Ober, D. (2005). Seeing double: gene duplication and diversification in plant secondary metabolism. *Trends Plant Sci.* *10*, 444–449.



- Ober, D. (2010). Gene duplications and the time thereafter – examples from plant secondary metabolism. *Plant Biol.* 12, 570–577.
- Ohno, S. (1973). Ancient Linkage Groups and Frozen Accidents. *Nature* 244, 259–262.
- Ojala, T., Remes, S., Haansuu, P., Vuorela, H., Hiltunen, R., Haahtela, K., and Vuorela, P. (2000). Antimicrobial activity of some coumarin containing herbal plants growing in Finland. *J. Ethnopharmacol.* 73, 299–305.
- Okamoto, S., Taguchi, T., Ochi, K., and Ichinose, K. (2009). Biosynthesis of Actinorhodin and Related Antibiotics: Discovery of Alternative Routes for Quinone Formation Encoded in the act Gene Cluster. *Chem. Biol.* 16, 226–236.
- Oliva, A., Meepagala, K.M., Wedge, D.E., Harries, D., Hale, A.L., Aliotta, G., and Duke, S.O. (2003). Natural Fungicides from *Ruta graveolens* L. Leaves, Including a New Quinolone Alkaloid. *J. Agric. Food Chem.* 51, 890–896.
- Oliveira, D., S, A., Boiteux, L.S., Kormelink, R., and Resende, R.O. (2018). The Sw-5 Gene Cluster: Tomato Breeding and Research Toward Orthotospovirus Disease Control. *Front. Plant Sci.* 9.
- Omura, T., and Sato, R. (1964). The Carbon Monoxide-binding Pigment of Liver Microsomes I. Evidence for its hemoprotein nature. *J. Biol. Chem.* 239, 2370–2378.
- O'Neill, T., Johnson, J.A., Webster, D., and Gray, C.A. (2013). The Canadian medicinal plant *Heracleum maximum* contains antimycobacterial diynes and furanocoumarins. *J. Ethnopharmacol.* 147, 232–237.
- Osbourn, A. (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* 26, 449–457.
- Panchy, N., Lehti-Shiu, M.D., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* pp.00523.2016.
- Pandey, H., Pandey, P., Singh, S., Gupta, R., and Banerjee, S. (2015). Production of anti-cancer triterpene (betulinic acid) from callus cultures of different *Ocimum* species and its elicitation. *Protoplasma* 252, 647–655.
- Patel, S. (2015). Plant essential oils and allied volatile fractions as multifunctional additives in meat and fish-based food products: a review. *Food Addit. Contam. Part A* 32, 1049–1064.
- Pellegrini, E., Carucci, M.G., Campanella, A., Lorenzini, G., and Nali, C. (2011). Ozone stress in *Melissa officinalis* plants assessed by photosynthetic function. *Environ. Exp. Bot.* 73, 94–101.
- Peyret, H., and Lomonossoff, G.P. (2013). The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* 83, 51–58.
- Pichersky, E., Raguso, R.A., Lewinsohn, E., and Croteau, R. (1994). Floral Scent Production in *Clarkia* (Onagraceae) (I. Localization and Developmental Modulation of Monoterpene Emission and Linalool Synthase Activity). *Plant Physiol.* 106, 1533–1540.
- Pimenta Lange, M.J., and Lange, T. (2006). Gibberellin Biosynthesis and the Regulation of Plant Development. *Plant Biol.* 8, 281–290.

- Pitelka, L.F., and Group, P.M.W. (1997). Plant Migration and Climate Change: A more realistic portrait of plant migration is essential to predicting biological responses to global warming in a world drastically altered by human activity. *Am. Sci.* *85*, 464–473.
- Pollak, Y., Zelinger, E., and Raskina, O. (2018). Repetitive DNA in the Architecture, Repatterning, and Diversification of the Genome of *Aegilops speltoides* Tausch (Poaceae, Triticeae). *Front. Plant Sci.* *9*, 1779.
- Pompon, D., Louerat, B., Bronine, A., and Urban, P. (1996). [6] Yeast expression of animal and plant P450s in optimized redox environments. In *Methods in Enzymology*, (Elsevier), pp. 51–64.
- Pumplin, N., and Voinnet, O. (2013). RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nat. Rev. Microbiol.* *11*, 745–760.
- Pureswaran, D.S., Roques, A., and Battisti, A. (2018). Forest Insects and Climate Change. *Curr. For. Rep.* *4*, 35–50.
- Pushnik, J.C., Miller, G.W., and Manwaring, J.H. (1984). The role of iron in higher plant chlorophyll biosynthesis, maintenance and chloroplast biogenesis. *J. Plant Nutr.* *7*, 733–758.
- Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A. (2004). A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci.* *101*, 8233–8238.
- Qin, G., Gu, H., Ma, L., Peng, Y., Deng, X.W., Chen, Z., and Qu, L.-J. (2007). Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in *Arabidopsis* by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Res.* *17*, 471–482.
- Ramirez-Estrada, K., Vidal-Limon, H., Hidalgo, D., Moyano, E., Golenioswki, M., Cusidó, R., and Palazon, J. (2016). Elicitation, an Effective Strategy for the Biotechnological Production of Bioactive High-Added Value Compounds in Plant Cell Factories. *Molecules* *21*, 182.
- Reinold, S., and Hahlbrock, K. (1997). In Situ Localization of Phenylpropanoid Biosynthetic mRNAs and Proteins in Parsley (*Petroselinum crispum*). *Bot. Acta* *110*, 431–443.
- Rekka, E., Evdokimova, E., Eeckhoudt, S., Labar, G., and Calderon, P.B. (2002). Role of temperature on protein and mRNA cytochrome P450 3A (CYP3A) isozymes expression and midazolam oxidation by cultured rat precision-cut liver slices. *Biochem. Pharmacol.* *64*, 633–643.
- Renton, M., Childs, S., Standish, R., and Shackelford, N. (2013). Plant migration and persistence under climate change in fragmented landscapes: Does it depend on the key point of vulnerability within the lifecycle? *Ecol. Model.* *249*, 50–58.
- Reyes, L.F., Villarreal, J.E., and Cisneros-Zevallos, L. (2007). The increase in antioxidant capacity after wounding depends on the type of fruit or vegetable tissue. *Food Chem.* *101*, 1254–1262.
- Roselli, S., Olry, A., Vautrin, S., Coriton, O., Ritchie, D., Galati, G., Navrot, N., Krieger, C., Vialart, G., Bergès, H., *et al.* (2017). A bacterial artificial chromosome (BAC) genomic

approach reveals partial clustering of the furanocoumarin pathway genes in parsnip. *Plant J.* 89, 1119–1132.

Ruan, J.-X., Li, J.-X., Fang, X., Wang, L.-J., Hu, W.-L., Chen, X.-Y., and Yang, C.-Q. (2016). Isolation and Characterization of Three New Monoterpene Synthases from *Artemisia annua*. *Front. Plant Sci.* 7.

Rupasinghe, S.G., Wen, Z., Chiu, T.-L., and Schuler, M.A. (2007). *Helicoverpa zea* CYP6B8 and CYP321A1: different molecular solutions to the problem of metabolizing plant toxins and insecticides. *Protein Eng. Des. Sel. PEDS* 20, 615–624.

Ryu, S.B., and Wang, X. (1996). Activation of phospholipase D and the possible mechanism of activation in wound-induced lipid hydrolysis in castor bean leaves. *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* 1303, 243–250.

Sadowska-Bartos, I., and Bartosz, G. (2014). Effect of antioxidants supplementation on aging and longevity. *BioMed Res. Int.* 2014, 404680.

Saidi, N., Cherkaoui, S., Chlyah, A., and Chlyah, H. (1997). Embryo formation and regeneration in *Triticum turgidum* ssp. *durum* anther culture. *Plant Cell Tissue Organ Cult.* 51, 27–33.

Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A., and Orzaez, D. (2011). GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules. *PLOS ONE* 6, e21622.

Sasabe, M., Wen, Z., Berenbaum, M.R., and Schuler, M.A. (2004). Molecular analysis of CYP321A1, a novel cytochrome P450 involved in metabolism of plant allelochemicals (furanocoumarins) and insecticides (cypermethrin) in *Helicoverpa zea*. *Gene* 338, 163–175.

Sasaki, M., Meguro, F., Kumazawa, E., Fujita, H., Kakishima, H., and Sakata, T. (1988). Evidence for uptake of 8-methoxypsoralen and 5-methoxypsoralen by cellular nuclei. *Mutat. Res. Mol. Mech. Mutagen.* 197, 51–58.

Savatin, D.V., Gramegna, G., Modesti, V., and Cervone, F. (2014). Wounding in the plant tissue: the defense of a dangerous passage. *Front. Plant Sci.* 5.

Sawai, S., Shindo, T., Sato, S., Kaneko, T., Tabata, S., Ayabe, S., and Aoki, T. (2006). Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*. *Plant Sci.* 170, 247–257.

Sayed-Ahmad, B., Talou, T., Saad, Z., Hijazi, A., and Merah, O. (2017). The Apiaceae: Ethnomedicinal family as source for industrial uses. *Ind. Crops Prod.* 109, 661–671.

Schiff, P.L. (2006). Ergot and Its Alkaloids. *Am. J. Pharm. Educ.* 70, 98.

Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., Dreher, K., Chavali, A.K., Nilo-Poyanco, R., Bernard, T., et al. (2017). Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiol.* 173, 2041–2059.

Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P., and Werck-Reichhart, D. (2001). CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J. Biol. Chem.* 276, 36566–36574.

- Schoch, G.A., Yano, J.K., Wester, M.R., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2004). Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J. Biol. Chem.* 279, 9497–9503.
- Schopfer, C.R., and Ebel, J. (1998). Identification of elicitor-induced cytochrome P450s of soybean (*Glycine max* L.) using differential display of mRNA. *Mol. Gen. Genet.* MGG 258, 315–322.
- Schuler, M.A. (1996). Plant Cytochrome P450 Monooxygenases. *Crit. Rev. Plant Sci.* 15, 235–284.
- Scott, J.G. (1996). Inhibitors of CYP6D1 in house fly microsomes. *Insect Biochem. Mol. Biol.* 26, 645–649.
- Seitz, C., Ameres, S., and Forkmann, G. (2007). Identification of the molecular basis for the functional difference between flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase. *FEBS Lett.* 581, 3429–3434.
- Serrano-Pérez, J.J., Merchán, M., and Serrano-Andrés, L. (2008). Photoreactivity of Furocoumarins and DNA in PUVA Therapy: Formation of Psoralen–Thymine Adducts. *J. Phys. Chem. B* 112, 14002–14010.
- Seyer, A., Einhorn, J., Brunelle, A., and Laprévote, O. (2010). Localization of Flavonoids in Seeds by Cluster Time-of-Flight Secondary Ion Mass Spectrometry Imaging. *Anal. Chem.* 82, 2326–2333.
- Sharma, S.K., and Brown, S.A. (1979). Affinity chromatography of *Ruta graveolens* L. *O*-methyltransferases. Studies demonstrating the potential of the technique in the mechanistic investigation of *O*-methyltransferases. *Can. J. Biochem.* 57, 986–994.
- Sircar, D., and Mitra, A. (2009). Accumulation of p-hydroxybenzoic acid in hairy roots of *Daucus carota* 2: Confirming biosynthetic steps through feeding of inhibitors and precursors. *J. Plant Physiol.* 166, 1370–1380.
- Siwinska, J., Siatkowska, K., Olry, A., Grosjean, J., Hehn, A., Bourgaud, F., Meharg, A.A., Carey, M., Lojkowska, E., and Ihnatowicz, A. (2018). Scopoletin 8-hydroxylase: a novel enzyme involved in coumarin biosynthesis and iron-deficiency responses in *Arabidopsis*. *J. Exp. Bot.* 69, 1735–1748.
- Song, H., Liu, J., Chen, T., and Nan, Z. (2018). Synonymous codon usage pattern in model legume *Medicago truncatula*. *J. Integr. Agric.* 17, 2074–2081.
- Sono, M., Roach, M.P., Coulter, E.D., and Dawson, J.H. (1996). Heme-Containing Oxygenases. *Chem. Rev.* 96, 2841–2888.
- Spicer, W.J., Franklin, J.C., and Spelman, D.W. (1984). Correlation of serum cefotaxime levels by HPLC with serum bacteriostatic and bactericidal levels. *J. Antimicrob. Chemother.* 14, 109–111.
- Stolt-Bergner, P., Benda, C., Bergbrede, T., Besir, H., Celie, P.H.N., Chang, C., Drechsel, D., Fischer, A., Geerlof, A., Giabbai, B., *et al.* (2018). Baculovirus-driven protein expression in insect cells: A benchmarking study. *J. Struct. Biol.* 203, 71–80.

- Sue, M., Nakamura, C., and Nomura, T. (2011). Dispersed Benzoxazinone Gene Cluster: Molecular Characterization and Chromosomal Localization of Glucosyltransferase and Glucosidase Genes in Wheat and Rye. *Plant Physiol.* *157*, 985–997.
- Sun, Z., Xu, C., Chen, S., Shi, Q., Wang, H., Wang, R., Song, Y., and Zeng, R. (2019). Exposure to Herbicides Prime P450-Mediated Detoxification of *Helicoverpa armigera* against Insecticide and Fungal Toxin. *Insects* *10*, 28.
- Taguchi, G., Fujikawa, S., Yazawa, T., Kodaira, R., Hayashida, N., Shimosaka, M., and Okazaki, M. (2000). Scopoletin uptake from culture medium and accumulation in the vacuoles after conversion to scopolin in 2,4-D-treated tobacco cells. *Plant Sci.* *151*, 153–161.
- Takaya, N., Suzuki, S., Kuwazaki, S., Shoun, H., Maruo, F., Yamaguchi, M., and Takeo, K. (1999). Cytochrome P450nor, a Novel Class of Mitochondrial Cytochrome P450 Involved in Nitrate Respiration in the Fungus *Fusarium oxysporum*. *Arch. Biochem. Biophys.* *372*, 340–346.
- Takos, A.M., and Rook, F. (2014). Plant-Specialized Metabolism and Its Genomic Organization in Biosynthetic Gene Clusters in *Lotus japonicus*. In *Lotus Japonicus Genome*, S. Tabata, and J. Stougaard, eds. (Berlin: Springer-Verlag Berlin), pp. 149–162.
- Tanaka, Y. (2006). Flower colour and cytochromes P450. *Phytochem. Rev.* *5*, 283–291.
- Tava, A., Pecetti, L., Ricci, M., Pagnotta, M.A., and Russi, L. (2007). Volatile compounds from leaves and flowers of *Bituminaria bituminosa* (L.) Stirt. (Fabaceae) from Italy. *Flavour Fragr. J.* *22*, 363–370.
- Thakur, M., and Sohal, B.S. (2013). Role of Elicitors in Inducing Resistance in Plants against Pathogen Infection: A Review.
- Thanaraj, T.A., and Argos, P. (1996). Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci. Publ. Protein Soc.* *5*, 1973–1983.
- Thorsoe, K.S., Bak, S., Olsen, C.E., Imberty, A., Breton, C., and Moller, B.L. (2005). Determination of catalytic key amino acids and UDP sugar donor specificity of the cyanohydrin glycosyltransferase UGT85B1 from *Sorghum bicolor*. Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. *Plant Physiol.* *139*, 664–673.
- Tintino, S.R., Neto, A.A. de C., Menezes, I.R.A., Oliveira, C.D. de M., and Coutinho, H.D.M. (2015). Antimicrobial Activity and Combined Effects on Antifungal and Antibacterial Drugs the Fruit of *Morinda citrifolia* L. *Acta Biol. Colomb.* *20*, 193–200.
- Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005). DNA Methylation Profiling Identifies CG Methylation Clusters in *Arabidopsis* Genes. *Curr. Biol.* *15*, 154–159.
- Tryjanowski, P., Rybacki, M., and Sparks, T. (2003). Changes in the first spawning dates of common frogs and common toads in western Poland in 1978–2002. *ANN ZOOL Fenn.* *40*, 6.
- Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., and Arita, M. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* *12*, 523–526.

- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihiro, D., Tanaka, W., Cajka, T., Saito, K., Fiehn, O., and Arita, M. (2016). Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* 88, 7946–7958.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* 141, 344–354.
- Tuomainen, J., Pellinen, R., Roy, S., Kiiskinen, M., Eloranta, T., Karjalainen, R., and Kangasjärvi, J. (1996). Ozone Affects Birch (*Betula pendula* Roth) Phenylpropanoid, Polyamine and Active Oxygen Detoxifying Pathways at Biochemical and Gene Expression Level. *J. Plant Physiol.* 148, 179–188.
- Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., *et al.* (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.
- Tuteja, N., and Mahajan, S. (2007). Calcium signaling network in plants: an overview. *Plant Signal. Behav.* 2, 79–85.
- Tzin, V., Malitsky, S., Aharoni, A., and Galili, G. (2009). Expression of a bacterial bi-functional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in *Arabidopsis*. *Plant J.* 60, 156–167.
- Umponstira, C., Warin, P., and Suckaluck, N. (2006). Physiological and biochemical responses of cowpea (*Vigna unguiculata* (L.) walp) to ozone. *Warasan Songkhla Nakharin Chabap Witthayasat Lae Technol.* 28, 861–869.
- Urban, P., Cullin, C., and Pompon, D. (1990). Maximizing the expression of mammalian cytochrome P-450 monooxygenase activities in yeast cells. *Biochimie* 72, 463–472.
- Vandelle, E., Vannozzi, A., Wong, D., Danzi, D., Digby, A.-M., Dal Santo, S., and Astegno, A. (2018). Identification, characterization, and expression analysis of calmodulin and calmodulin-like genes in grapevine (*Vitis vinifera*) reveal likely roles in stress responses. *Plant Physiol. Biochem.* 129, 221–237.
- Verpoorte, R., and Alfermann, A.W. (2000). *Metabolic engineering of plant secondary metabolism* (Dordrecht ; Boston, MA: Kluwer Academic Publishers).
- Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Lariat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., *et al.* (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants. *Plant J.* 70, 460–470.
- Vijayalekshmy, K.S., Menon, V.P., and Leelamma, S. (1992). Role of antibiotics in lipid peroxidation. *Indian J. Biochem. Biophys.* 29, 371–374.
- Vu, H.S., Shiva, S., Roth, M.R., Tamura, P., Zheng, L., Li, M., Sarowar, S., Honey, S., McElhiney, D., Hinkes, P., *et al.* (2014). Lipid changes after leaf wounding in *Arabidopsis thaliana* : expanded lipidomic data form the basis for lipid co-occurrence analysis. *Plant J.* 80, 728–743.

- Vu, H.S., Roston, R., Shiva, S., Hur, M., Wurtele, E.S., Wang, X., Shah, J., and Welti, R. (2015). Modifications of membrane lipids in response to wounding of *Arabidopsis thaliana* leaves. *Plant Signal. Behav.* *10*, e1056422.
- Walasek, M., Grzegorzczak, A., Malm, A., and Skalicka-Woźniak, K. (2015). Bioactivity-guided isolation of antimicrobial coumarins from *Heracleum mantegazzianum* Sommier & Levier (Apiaceae) fruits by high-performance counter-current chromatography. *Food Chem.* *186*, 133–138.
- Wang, M., and Lincoln, D.E. (2004). Effects of light intensity and artificial wounding on monoterpenoid production in *Myrica cerifera* from two different ecological habitats. *Can. J. Bot.* *82*, 1501–1508.
- Wang, C., Ding, D., Yan, R., Yu, X., Li, W., and Li, M. (2008). A novel bi-directional promoter cloned from melon and its activity in cucumber and tobacco. *J. Plant Biol.* *51*, 108–115.
- Wang, J., Liu, Y., Cai, Y., Zhang, F., Xia, G., and Xiang, F. (2010). Cloning and functional analysis of geraniol 10-hydroxylase, a cytochrome P450 from *Swertia mussotii* Franch. *Biosci. Biotechnol. Biochem.* *74*, 1583–1590.
- Wang, J.W., Zheng, L.P., Zhang, B., and Zou, T. (2009a). Stimulation of artemisinin synthesis by combined cerebroside and nitric oxide elicitation in *Artemisia annua* hairy roots. *Appl. Microbiol. Biotechnol.* *85*, 285–292.
- Wang, Q., Wan, L., Li, D., Zhu, L., Qian, M., and Deng, M. (2009b). Searching for bidirectional promoters in *Arabidopsis thaliana*. *BMC Bioinformatics* *10*, S29.
- Wang, R., Yan, Y., Zhu, M., Yang, M., Zhou, F., Chen, H., and Lin, Y. (2016). Isolation and Functional Characterization of Bidirectional Promoters in Rice. *Front. Plant Sci.* *7*.
- Webster, G.R., Teh, A.Y.-H., and Ma, J.K.-C. (2017). Synthetic gene design-The rationale for codon optimization and implications for molecular pharming in plants: Synthetic Gene Design. *Biotechnol. Bioeng.* *114*, 492–502.
- Wegel, E., Koumproglou, R., Shaw, P., and Osbourn, A. (2009). Cell Type-Specific Chromatin Decondensation of a Metabolic Gene Cluster in Oats. *Plant Cell* *21*, 3926–3936.
- Wendorff, H., and Matern, U. (1986). Differential response of cultured parsley cells to elicitors from two non-pathogenic strains of fungi. Microsomal conversion of (+)marmesin into psoralen. *Eur. J. Biochem.* *161*, 391–398.
- Wen-Sheng, X., Xiang-Jing, W., Tian-Rui, R., and Su-Qin, C. (2006). Purification of recombinant wheat cytochrome P450 monooxygenase expressed in yeast and its properties. *Protein Expr. Purif.* *45*, 54–59.
- Werck-Reichhart, D., and Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biol.* *1*, reviews3003.1-reviews3003.9.
- Weryszko-Chmielewska, E., and Chwil, M. (2017). Localisation of furanocoumarins in the tissues and on the surface of shoots of *Heracleum sosnowskyi*. *Botany* *95*, 1057–1070.

- Wester, M.R., Yano, J.K., Schoch, G.A., Yang, C., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2004). The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution. *J. Biol. Chem.* *279*, 35630–35637.
- Wickramasinghe, R.H., and Vilee, C.A. (1975). Early role during chemical evolution for cytochrome P450 in oxygen detoxification. *Nature* *256*, 509–511.
- Wiesner, M., Schreiner, M., and Zrenner, R. (2014). Functional identification of genes responsible for the biosynthesis of 1-methoxy-indol-3-ylmethyl-glucosinolate in *Brassica rapa* ssp. *chinensis*. *BMC Plant Biol.* *14*, 124.
- Wilderman, P.R., Xu, M., Jin, Y., Coates, R.M., and Peters, R.J. (2004). Identification of Syn-Pimara-7,15-Diene Synthase Reveals Functional Clustering of Terpene Synthases Involved in Rice Phytoalexin/Allelochemical Biosynthesis. *Plant Physiol.* *135*, 2098–2105.
- Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., and McRee, D.E. (2000a). Mammalian Microsomal Cytochrome P450 Monooxygenase: Structural Adaptations for Membrane Binding and Functional Diversity. *Mol. Cell* *5*, 121–131.
- Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., and McRee, D.E. (2000b). Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. *J. Inorg. Biochem.* *81*, 183–190.
- Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T.R., Li, Y., Meade, F., Teodor, R., Vaistij, F.E., *et al.* (2012). A *Papaver somniferum* 10-Gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine. *Science* *336*, 1704–1708.
- Wittstock, U., and Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. *Curr. Opin. Plant Biol.* *5*, 300–307.
- Wong, S., and Wolfe, K.H. (2005). Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat. Genet.* *37*, 777–782.
- Xia, J., Psychogios, N., Young, N., and Wishart, D.S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* *37*, W652.
- Xiao, K., Zhang, C., Harrison, M., and Wang, Z.-Y. (2005). Isolation and characterization of a novel plant promoter that directs strong constitutive expression of transgenes in plants. *Mol. Breed.* *15*, 221–231.
- Xu, M., Galhano, R., Wiemann, P., Bueno, E., Tiernan, M., Wu, W., Chung, I.-M., Gershenzon, J., Tudzynski, B., Sesma, A., *et al.* (2012). Genetic evidence for natural product mediated plant–plant allelopathy in rice. *New Phytol.* *193*, 570–575.
- Yang, C.-Q., Fang, X., Wu, X.-M., Mao, Y.-B., Wang, L.-J., and Chen, X.-Y. (2012). Transcriptional Regulation of Plant Secondary Metabolism. *J. Integr. Plant Biol.* *54*, 703–712.
- Yano, J.K., Hsu, M.-H., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2005). Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat. Struct. Mol. Biol.* *12*, 822–823.



- Yi, G., Sze, S.-H., and Thon, M.R. (2007). Identifying clusters of functionally related genes in genomes. *Bioinformatics* 23, 1053–1060.
- Zangerl, A.R. (1990). Furanocoumarin Induction in Wild Parsnip: Evidence for an Induced Defense against Herbivores. *Ecology* 71, 1926–1932.
- Zangerl, A.R., and Berenbaum, M.R. (1998). Damage-inducibility of primary and secondary metabolites in the wild parsnip (*Pastinaca sativa*). *Chemoecology* 8, 187–193.
- Zangerl, A.R., and Berenbaum, M.R. (2005). Increase in toxicity of an invasive weed after reassociation with its coevolved herbivore. *Proc. Natl. Acad. Sci.* 102, 15529–15532.
- Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298.
- Zhang, S., and Klessig, D.F. (2001). MAPK cascades in plant defense signaling. *Trends Plant Sci.* 6, 520–527.
- Zhang, M., Demeshko, Y., Dumbur, R., Iven, T., Feussner, I., Lebedov, G., Ganim, M., Barg, R., and Ben-Hayyim, G. (2018). Elevated  $\alpha$ -Linolenic Acid Content in Extra-plastidial Membranes of Tomato Accelerates Wound-Induced Jasmonate Generation and Improves Tolerance to the Herbivorous Insects *Heliothis peltigera* and *Spodoptera littoralis*. *J. Plant Growth Regul.* 38, 723–738.
- Zhang, Y., Wu, Y., Liu, Y., and Han, B. (2005). Computational identification of 69 retroposons in Arabidopsis. *Plant Physiol.* 138, 935–948.
- Zhang, Z., Moo-Young, M., and Chisti, Y. (1996). Plasmid stability in recombinant *Saccharomyces cerevisiae*. *Biotechnol. Adv.* 14, 401–435.
- Zhao, J., Davis, L.C., and Verpoorte, R. (2005). Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnol. Adv.* 23, 283–333.
- Zobel, A.M., and Brown, S.A. (1990). Dermatitis-inducing furanocoumarins on leaf surfaces of eight species of Rutaceous and Umbelliferous plants. *J. Chem. Ecol.* 16, 693–700.
- Zobel, A.M., and Brown, S.A. (1991a). Furanocoumarin concentrations in fruits and seeds of *Angelica archangelica*. *Environ. Exp. Bot.* 31, 447–452.
- Zobel, A.M., and Brown, S.A. (1991b). Psoralens on the surface of seeds of Rutaceae and fruits of Umbelliferae and Leguminosae. *Can. J. Bot.* 69, 485–488.



# **Annexes**





Article

# Untargeted metabolomics approach reveals diverse responses of *Pastinaca sativa* to ozone and wounding stresses

Gianni Galati<sup>1</sup>, Anthony Gandin<sup>2</sup>, Yves Jolivet<sup>2</sup>, Romain Lariat<sup>1,\*</sup> and Alain Hehn<sup>1</sup>

<sup>1</sup> INRA, I.AE, Université de Lorraine, F54000, Nancy, France; gianni.galati@univ-lorraine.fr (G.G.); alain.hehn@univ-lorraine.fr (A.H.)

<sup>2</sup> Université de Lorraine, AgroParisTech, INRA, UMR Silva, 54000 Nancy, France; anthony.gandin@univ-lorraine.fr (A.G.); yves.jolivet@univ-lorraine.fr (Y.J.)

\* Correspondence: romain.lariat@univ-lorraine.fr (R.L.)

Received: date; Accepted: date; Published: date

**Abstract:** Stresses such as wounding or atmospheric pollutant exposure have a significant impact on plant fitness. Since it has been widely described that the metabolome directly reflects plant physiological status, a way to assess this impact is to perform a global metabolomics analysis. In this study, we investigated the effect of two abiotic stresses (mechanical wounding and ozone exposure) on parsnip metabolic balance using a liquid chromatography-mass spectrometry-based untargeted metabolomics approach. For this purpose, parsnip leaves were submitted to an acute ozone exposure or were mechanically wounded and sampled 24, 48 and 72 h post-treatment. Multivariate and univariate statistical analyses highlighted numerous differentially accumulated metabolic features (DAMFs) as a function of time and treatment. Mechanical wounding enhanced a much higher response than ozone exposure. We found that the presence of coumarins and fatty acyls was upregulated in wounded leaves, while flavonoids are down-regulated in the same conditions. These results provide an overview of metabolic destabilization through differentially accumulated compounds and provide a better understanding of global plant metabolic changes in defence mechanisms.

**Keywords:** metabolomics; elicitation; parsnip; wounding; ozone

## 1. Introduction

During evolution, plants have adapted to stressful environmental conditions by developing multiple survival strategies. The most emblematic strategy might be the production of lignin, which allowed plants to grow erect, facilitated their photosynthesis and provided protection against fungi or insect attacks. However, less visible strategies occurring at different organization levels (molecular, cellular, and physiological) are also triggered by plants. These mechanisms have been extensively described in model plants and crops [1,2]. They are divided into several steps that are initiated by signal perception and the binding of an elicitor to a receptor [3]. In a second step, the signal is conveyed, especially through membrane depolarization and activation of kinase cascades [4]. The third step is characterized by the regulation of the expression of a large set of genes, resulting in the emergence of diverse physiological responses. Some of these steps are well documented, such as the modulation of photosynthetic activities driven by stomatal closure or metabolic modifications through the accumulation of enzymes and subsequent specialized metabolites [5,6]. These common general features have been reported in all plants, but additional specific responses were described for each plant species and, moreover, for each kind of stress. This is especially highlighted by the fine-tuned regulation of the stress gene network orchestrated by the interconnection between the main stress-responsive phytohormones (ethylene, abscisic acid, jasmonic acid and salicylic acid) [7].

It is evident that human activity increases the stressfulness of the environment in the biosphere. This stress can be directly associated with the accumulation of atmospheric pollutants or indirectly associated with global climate change. Among atmospheric pollutants, ozone ( $O_3$ ) is one of the most studied. It has been shown to negatively impact plant growth and development through the activation of oxidative stress responses and induction of programmed cell death [8]. To cope with such oxidative stress-related damage and excessive reactive oxygen species (ROS) production, plants have established a large panel of defence metabolites (*e.g.*, ascorbate-glutathione and polyphenols) or antioxidant enzymes (*e.g.*, catalases and peroxidases) [9]. However, because ROS also contribute to plant cell signalling, it is important for the plant to balance ROS concentration and avoid excessive ROS accumulation. The relative contribution of each component is still poorly understood. Other stresses related to global climate change were described and are increasingly expected in the near future, such as insect herbivory and modification of the geographic distribution of various pests [10]. Insect herbivory, notably through the wounding generated, impacts plant metabolism and stimulates defence reactions that could consist on modification of the concentration and distribution of primary and specialized metabolites throughout the plant. Metabolites are therefore products reflecting cellular functions, and their identification and relative concentration might be considered a good proxy for the physiological status of plants. The development of high-throughput metabolomics has allowed the identification and quantification of dozens of metabolites and might provide interesting insights into the perturbation generated by a stimulus on an organism [11,12].

The cultivated parsnip (*Pastinaca sativa*) is an Apiaceae closely related to carrot (*Daucus carota*), with a white pivotal root acting as a strong carbon sink. This plant is generally reported as a model plant for explaining coevolution between plant and insect pests. Studies conducted in the 1980s describe the emergence and evolution of a metabolic pathway leading to the production of linear and angular furanocoumarins [13]. These molecules are described as phytoalexins displaying phototoxic properties [14]. Upon absorbing UV light, furanocoumarins can establish covalent bonds with DNA or proteins, leading to physiological consequences such as blocking cell replication or inactivating enzymes [15]. In recent years, parsnip has become a model plant for the cloning and functional identification of genes involved in the furanocoumarin pathway [16–20]. Thus far, little is known regarding the phytochemistry of parsnip and its response to environmental stresses.

To determine the metabolic response of parsnip leaves to  $O_3$  and wounding stresses, we designed a liquid chromatography-mass spectrometry (U-HPLC-MS) untargeted metabolomics approach. Parsnip plants were subjected to acute  $O_3$  stress or mechanical wounding. The metabolite composition of the leaves was determined 24h, 48h, and 72 h after the initiation of the stress. The datasets generated were analysed by multivariate and univariate statistical approaches. Metabolite signals differentially accumulated between treated plants were identified and replaced in the corresponding metabolic pathways.

## 2. Material and methods

### 2.1. Stress application settings

Mechanical wounding (MW) was induced on all leaves of 8 uniformly sized 2-month-old parsnips cultured in a growth chamber at 21 °C/18 °C day/night and an irradiance of 380  $\mu\text{mol m}^{-2} \text{s}^{-1}$  for 16 h photoperiod as described by Roselli *et al* [18]. Concomitantly,  $O_3$  treatment was performed on 8 uniformly sized plants by exposure at 250 ppb for 6 h a day and for 3 days. After treatments, the leaves were collected, flash-frozen in liquid nitrogen and stored at -80 °C until use. The wounded and  $O_3$ -treated leaves were collected immediately before (*i.e.*, T0) and 24, 48, and 72 h (T1, T2 and T3) after wounding. Control parsnips were placed in the same conditions for MW and  $O_3$  exposure.

### 2.2. Metabolite extraction

The frozen parsnip leaves were crushed with a mortar and pestle. Eight hundred microliters of an 80 % methanol solution were mixed with 100 mg of powder independently collected for each

sample. After sonication (sweep mode, Elma s70 Elmasonic, Singen, Germany) for 10 minutes and a 20-minute-long centrifugation at 13 000 g, the supernatant was transferred into a new 2 mL microtube. A second extraction was performed on the pellet using the same conditions. The two supernatants were pooled and then dried at 40 °C using a speed vacuum (Concentrator plus, Eppendorf, Hambourg, Germany). The pellet was suspended in 100 µL of 80 % methanol containing 5 µM 4-methylumbelliferone as an internal standard.

### 2.3. U-HPLC-MS

The plant extracts were analysed by U-HPLC-MS (Prominence, Shimadzu, Kyoto, Japan). The device was equipped with a diode array detector SPD20A (Shimadzu, Kyoto, Japan) (PDA) and a single quadrupole mass spectrometer LCMS2020 (Shimadzu, Kyoto, Japan). The molecules were separated on a C18 column in inverse phase (ZORBAX Eclipse Plus 150 × 2.10 mm; particle size = 1.8 µm; Agilent Technologies, Santa Clara, CA, United States). The mobile phase consisted of 0.1 % formic acid in ultra-pure water (A) and 0.1 % formic acid in methanol (B). The molecules were eluted through a gradient: (A:B, v/v), (90:10) at 0 min, (80:20) at 1 min, (40:60) at 6 min, (10:90) at 10 min, (0:100) from 12 min to 16 min and (90:10) from 16 min to 20 min. The solvent flow was set at 0.2 mL/min and the injection volume was set at 5 µL. The compounds were detected using a UV detector (200 to 600 nm). They were further analysed with a mass spectrometer using a double ionization source (DUIS) set up in positive mode and by combination of electrospray ionization and chemical ionization with atmospheric pressure. The electrospray ionization received a tension of 4.5 kV, and the temperatures of the heating block, the entrance and the line of desolvation of the mass spectrometer were set at 400 °C, 350 °C and 250 °C, respectively.

### 2.4. High-resolution mass spectrometry

High-resolution mass spectrometry (HRMS) and tandem mass spectrometry (MS/MS) analyses were performed on selected samples. Forty microliters of the sample were analysed on an HPLC-MS LTQ-Orbitrap consisting of a liquid chromatography device (Dionex UltiMate® 3000, Thermo Scientific, Waltham, MA, United States) coupled with a quaternary solvent delivery pump and a Linear Ion Trap Mass Spectrometer (LTQ XL™, Thermo Scientific, Waltham, MA, United States) connected to an Orbitrap HRMS (Thermo Scientific, Waltham, MA, United States). The separation of the compounds was performed on a C18 column in inverse phase (Alltima™ 150 × 2.1 mm; size of particle = 5 µm, Alltech®). The mobile phase consisted of 0.1 % formic acid in ultra-pure water (A) and 0.1 % formic acid in methanol (B). The gradient was set up as follows: (A:B, v/v), (90:10) at 0 min, (43:57) at 20 min, (35:54) at 32 min, (10:90) at 38 min, and (90:10) at 40 min. The flow was set at 0.2 mL/min for a 40 min analysis. The tandem mass spectrometer was connected to a liquid chromatography system and a source of ionization ESI set in positive mode with a tension of 4.5 kV and a temperature of 300 °C. The flow of coaxial, auxiliary gases and barrier were adjusted in 40 and 10 arbitrary units (AU)/min, respectively. The voltage of the capillary of transfer, split lens and forehead (front) lens in the entrance of the source of ionization are set at 36 V, 44 V and 3.5 V. The analysis started from *m/z* ratios of 50 to 1000.

### 2.5. ROS analysis

Hydrogen peroxides were extracted from 100 mg FW of leaves ground in liquid nitrogen and 5 % PVPP and suspended in 0.1 % TCA. The sample was centrifuged at 18 000 g for 15 min at 4 °C. The supernatant was assayed in 5 mM potassium phosphate buffer (pH 7.0) and 0.75 M KI (according to [21]). The absorbance was immediately measured at 390 nm.

### 2.6. Metabolic data analysis

The U-HPLC-MS raw data were converted to .cdf files using the LabSolution software (Shimadzu, Japan) and were pre-processed using the XCMS Online platform (<https://xcmsonline.scripps.edu>). This pre-processing consisted of peak detection, retention time

correction and alignment of the molecules. Peak detection was performed with the CentWave method ( $\Delta m/z = 2.5$  ppm, peak width [5-50], S/N threshold = 20). Retention time correction was performed with the obiwarp method, and alignment was performed with the following parameters (minfrac=0.5, bw=5, mzwid=0.015). Analysis was performed by comparing the 8 replicates of each treatment at a given harvest time (T0, T1, T2, and T3). The pre-processed datasets, corresponding to 1592, 1077, 1091 and 1260 features in positive mode at T0, T1, T2 and T3, respectively, were imported into Microsoft Excel and converted to .csv files. Multivariate and univariate statistical analyses were performed with the MetaboAnalyst 4.0 online platform (<https://www.metaboanalyst.ca>). Prior to the statistical analyses, datasets were filtered based on the relative standard deviation (RSD/mean) to remove the features with low repeatability (40 % of the filtered features). Then, data were normalized with the Pareto method (each feature was mean centred and divided by the square of the standard deviation). Principal component analysis (PCA) was conducted on the filtered features, corresponding to 956, 641, 655 and 755 features for T0, T1, T2 and T3, respectively. A two-way analysis of variance (ANOVA,  $p < 0.001$ ) was performed to identify metabolic feature affected by treatment and sampling date. A one-way analysis of variance (ANOVA,  $p < 0.05$ , Fisher's least significant difference (LSD) post hoc test) was performed to identify the metabolic features significantly affected by  $O_3$  and MW stresses at specific sampling date.

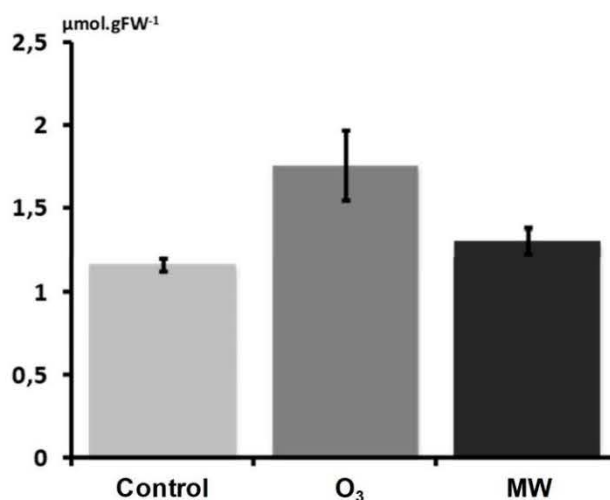
Metabolite identification was performed based on significantly affected metabolic features. Raw HPLC-MS LTQ-Orbitrap data files containing exact mass and mass fragmentation data were converted to .abf and analysed using MS-Dial (v.3.20) and MS-Finder (v.3.12) software. HRMS data allowed the determination of the molecular formula, while MS/MS data were used to propose molecular structures by comparison with 21 metabolomic databases, notably including UNPD, LipidMAPS, HMDB, MassBank, METLIN and ReSpect. When possible, the identification of metabolites was confirmed by comparison with authentic standards. This was the case for rutin, imperatorin, osthol, 8-geranyloxy-psoralen oleic acid, linolenic acid and LysoPC (16:0).

### 3. Results

#### 3.1. Impact of treatment on the oxidative status of parsnip leaves

To characterize the intensity of oxidative stresses, the hydrogen peroxide ( $H_2O_2$ ) concentration was assessed in parsnip leaves that were exposed to  $O_3$  or mechanically wounded (MW). To establish the sensitivity of parsnip to  $O_3$ , we exposed them to different concentrations of  $O_3$  (0, 80, 120, 150 and 250 ppb) for 6 h every day for 3 days. No visual symptoms were observed for any conditions. Therefore, the highest 250 ppb  $O_3$  concentration was used in subsequent experiments. After 3 days of  $O_3$  exposure, parsnip leaves did not show any chlorosis or early senescence symptoms.  $H_2O_2$  content was measured for the  $O_3$ , MW and control plants. No significant difference was identified, although a 1.5-fold increase could be observed between the  $O_3$ -treated and control plants (Figure 1).

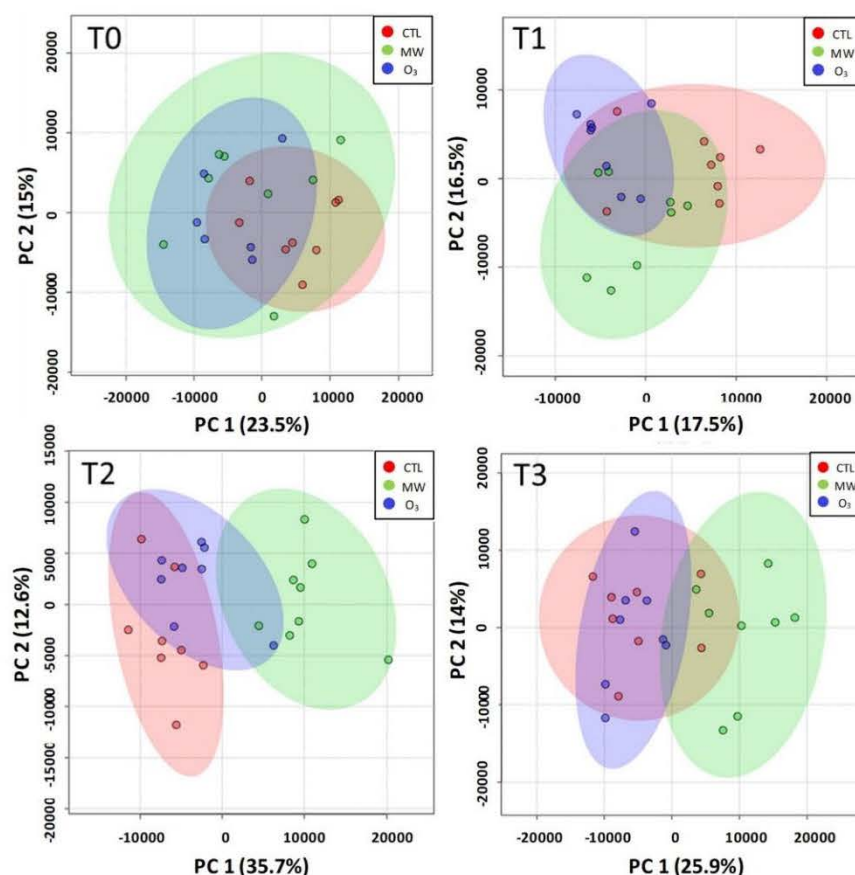




**Figure 1. H<sub>2</sub>O<sub>2</sub> assay in parsnip leaves.** The hydrogen peroxide measurement was performed 72 h after the first O<sub>3</sub> application (middle grey) and MW (dark grey) and at the same time as control plants (light grey). Vertical bars correspond to standard deviation. Kruskal-Wallis tests were performed with no significant results (p-value > 0.05 for each condition). Eight replicates per treatment were considered.

### 3.2. Global overview of the metabolomic profile evolution.

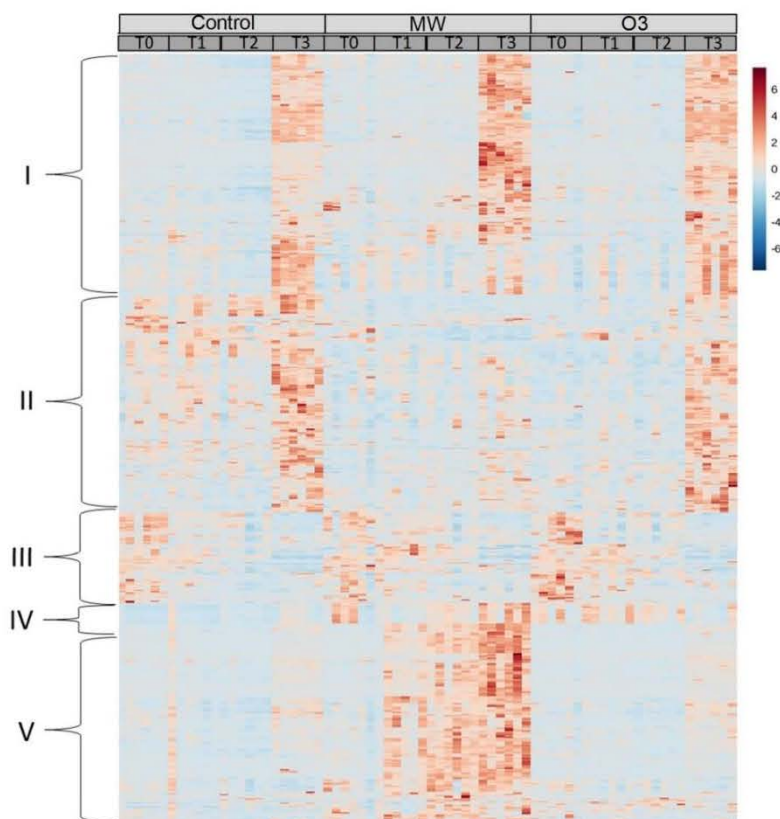
To determine the evolution of the global metabolic response in MW- and O<sub>3</sub>-treated parsnip leaves, we developed an untargeted metabolomics strategy. For this purpose, LC-MS analyses were performed on the hydro-methanolic extracts prepared from the leaves collected 24, 48 and 72 hours after the beginning of the stresses. Metabolic features characterized by a mass on charge ratio in positive mode ( $m/z$ ) and a retention time were detected and quantified in each sample (Supplemental data 1). In order to globally assess the impact of treatment on the parsnip leaf metabolome, we realized PCA on the metabolic features at each sampling date (Figure 2).



**Figure 2.** Principal component analysis (PCA) of the parsnip metabolomics datasets. PCA analyses at T0, T1 (24h), T2 (48h) and T3 (72h) after the application of the stresses. The red, green and blue areas and dots are related to control, mechanical wounding and ozone stress application, respectively.

From these PCA, it appears that the treatments, overlapping with the control at T0, tend to separate from it after T1. At T2, MW plants were clearly distinct from control plants whereas O<sub>3</sub> treatment was not completely separated from the control. At T3, O<sub>3</sub> treatment completely overlapped the control. Comparison of the PCA overtime suggested a transient metabolic response to both treatments with a maximum response at T2.

Metabolic features significantly affected by the treatment and the sampling time were identified by a two-way ANOVA ( $p < 0.001$ ). The heatmap constructed with the significant metabolic features (Figure 3) allows identifying clusters of metabolic features that are affected by the sampling time, either positively (cluster I, Figure 3) or negatively (cluster III, Figure 3). In addition, three main trends can be visualized regarding the impact of treatments: cluster II groups metabolic features that tend to decrease both under O<sub>3</sub> and MW, cluster IV groups few metabolic features accumulating under O<sub>3</sub> in comparison to control and cluster V groups metabolic features gradually accumulating under MW in comparison to both O<sub>3</sub> and control samples (Figure 3).



**Figure 3:** Heatmap plot of the 646 significant metabolic features determined by a two-way ANOVA on time and treatment ( $p < 0.001$ ) at T0, T1 (24h), T2 (48h) and T3 (72h) after the application of the stresses. The red colour indicates a high intensity level, whereas the blue colour indicates a lower intensity level for each metabolic feature. Each line refers to a single metabolic feature. Each row refers to a sample. Clusters of DAMFs are identified by the mark on the left. Each cluster is depicted in the result section.

The number of differentially accumulated metabolic features (DAMFs) between the two stressing treatment and the control were determined at each sampling date by a one-way ANOVA ( $p < 0.05$ ) and are summarized in the Table 1.

**Table 1.** Number of DAMFs for the O<sub>3</sub> and MW in comparison to the control. The differentially accumulated metabolic features are categorized according to the sampling time (T0 to T3), the treatment (O<sub>3</sub> and MW) and up- and downregulation. The DAMFs were identified from a one-way analysis of variance (ANOVA,  $p < 0.05$ , Fisher's Least significant difference (LSD) post hoc test). The detailed list of all metabolic features and DAMFs are described in Supplemental data 1.

	O <sub>3</sub> /Control		MW/Control	
	Down	Up	Down	Up
T0	1	1	1	0
T1	19	42	17	111
T2	19	6	41	304
T3	10	16	82	185

MW leads to a larger modification of the parsnip leaf metabolome than does the O<sub>3</sub> stress as regard by the number of DAMFs. The number of DAMFs increased as soon as T1. It was maximum at T2 for MW whereas it tended to decrease after T1 for the O<sub>3</sub> treatment. MW was characterized by a large majority of DAMFs over-accumulating whatever the sampling date.

### 3.3. Metabolite identification

Metabolite identification was performed on the DAMFs at T1, T2 and T3, which represented a dataset of 544 non-redundant entities. This dataset was filtered by subtracting fragment ions, and metabolic features with a small peak area or with no exploitable MS/MS data, leading to a final dataset of 245 DAMFs. High-resolution mass spectrometry and MS/MS fragmentation data led to the identification of 39 DAMFs (Table 2). The identified DAMFs were mostly related to MW rather than to O<sub>3</sub> exposure. According to the Classyfire metabolic annotation [22] the putatively identified metabolites were shared between 8 major classes: flavonoids, fatty acyls, coumarins and derivatives, indoles and derivatives, tetrapyrroles and derivatives, prenol lipids, glycerophospholipids and glycerolipids (Table 2). According to their response to the two stresses, the identified metabolites can be grouped in three categories. The first group (I) contains metabolites which accumulation is significantly reduced by O<sub>3</sub>, MW or both stresses. It comprises all the flavonoids identified, several fatty acyls and glycerolipids. The accumulation of these metabolites was significantly reduced under MW at T2 and T3. On the other hand, only a few of these metabolites were affected by O<sub>3</sub> at T1 and T2. The second group (II) includes metabolites which accumulation is massively increased under O<sub>3</sub> at T1 and T2 and under MW at T2 and T3. One of these metabolites (corresponding to the feature labelled M677T8, Table 2) was identified as a tetrapyrrole derivative based on its chemical formula and fragmentation pattern. Such metabolites, referred also as phyllobilins or nonfluorescent chlorophyll catabolites (NCC), were discovered three decades ago in senescent leaves or in maturing fruits [23]. NCC are characterized by a complex structure and a high diversity making them difficult to identify. However, recent mass spectrometry approaches contribute to their structural elucidation [24]. Based on these methods and by comparison with the mass fragmentation of M677T8, we assumed that 4 additional metabolites can be assigned as NCC (supplemental data 2).

The third group (III) contains metabolites which accumulation is exclusively induced under MW treatment. It includes compounds related to primary metabolism such as tryptophan, several fatty acids and glycerophospholipids. As expected, beside these molecules, secondary metabolites such as monoterpenoids, coumarins and furanocoumarins were also significantly induced. The accumulation of all these metabolites was increased up to T2 and slightly decreased at T3 in most cases.

**Table 2. Identification of the DAMFs.** The identified DAMI's are listed and categorized in three groups (I, II, III) as described in the results section. Group I contains DAMFs negatively affected by O<sub>3</sub>, MW or both; group II contains DAMFs over-accumulated under O<sub>3</sub> and MW treatment; group III contains DAMFs over-accumulating only under MW treatment. For each DAMI's, the fold change (FC) of treatment/control is given only when significant (p<0.001).

Class	Feature name	FC O <sub>3</sub> /control			FC MW/Control			Ion Adduct	RT (min)	[M-H] <sup>+</sup> m/z	Molecular Weight	Molecular formula	Assigned compound	Classify Class
I	M293T11	—	—	—	—	—	0.41	M+H <sub>2</sub> Na	10.56	293.0782	270.28	C16H14O4	Imperatorin	Coumarins and derivatives
I	M595T8	—	—	—	—	—	0.67	M+H	7.55	595.1646	594.52	C27H30O15	Kaempferol 3-rhamnoside-7-glucoside	Flavonoids
I	M611T6	—	—	—	—	—	0.48	M+H	6.43	611.1597	610.52	C27H30O16	Rutin	Flavonoids
I	M627T5	—	—	—	—	—	0.62	M+H	5.09	627.1542	626.52	C27H30O17	Quercetin 3,4-diglucoside	Flavonoids
I	M625T8	0.55	0.61	—	—	—	0.53	M+H	7.75	625.1777	624.54	C28H32O16	Myricetin 7-methyl ether 3,4'-di-O- $\alpha$ -L-rhamnopyranoside	Flavonoids
I	M757T6	—	0.48	—	—	—	0.74	M+H	6.46	757.2178	756.66	C33H40O20	Quercetin-3-O- $\alpha$ -L-rhamnopyranosyl(1-2)- $\beta$ -D-glucopyranoside-7-O- $\alpha$ -L-rhamnopyranoside	Flavonoids
I	M751T13	—	—	—	—	0.38	0.28	M+H <sub>2</sub> Na	13.37	751.4223	728.99	C43H68O9	CID 11104554	Fatty Acyls
I	M385T16	—	0.69	—	—	0.25	0.37	M+H	16.01	585.4473	584.87	C37H60O5	DG(20:5(5Z,8Z,11Z,14Z,17Z)/14:1(9Z)/0:0)	Glycerolipids
I	M383T15	—	—	—	—	0.55	—	M+H <sub>2</sub> Na	14.58	583.4348	560.84	C35H60O5	DG(14:1(9Z)/18:3(9Z,12Z,15Z)/0:0)	Glycerolipids
I	M769T16	—	—	—	—	0.28	0.44	M+H <sub>2</sub> Na	16.00	769.4824	747.01	C43H70O10	18:3/16:3-MGD	Glycerolipids
I	M783T15	—	—	—	—	0.48	0.37	M+H <sub>2</sub> Na	14.71	783.4601	760.99	C43H68O11	oxy phytodienoic acid/16:3-MGD	Glycerolipids
II	M677T8	3.04	2.20	0.44	—	2.10	1.63	M+H	8.43	677.2632	676.72	C38H36N4O8	(2S,2'(1R)-2'(1,2(2)-dicarboxy-8-ethenyl-2,7,12,18-tetramethyl-2,2(1)-dihydrobenzof[ <i>b</i> ]porphyrin-13,17-dipropionic acid	Tetrapyrroles and derivatives
II	M661T9	3.95	2.64	—	—	2.38	—	M+H	8.67	661.2835	660.76	C42H36N4O4	Putative chlorophyll catabolite	Tetrapyrroles and derivatives
II	M707T9	3.23	2.96	0.43	—	2.68	2.20	M+H	9.45	707.2927	706.74	C36H42N4O11	Putative chlorophyll catabolite	Tetrapyrroles and derivatives
II	M693T9	3.03	2.44	—	—	2.13	—	M+H	9.14	693.2769	692.71	C35H40N4O11	Putative chlorophyll catabolite	Tetrapyrroles and derivatives
II	M691T10	2.96	2.48	0.61	—	3.14	2.45	M+H	9.32	691.2606	690.70	C35H38N4O11	Putative chlorophyll catabolite	Tetrapyrroles and derivatives
III	M135T11	—	—	—	3.24	6.87	4.19	M+H	10.64	135.1161	134.22	C10H14	p-cymene	Prenol lipids
III	M151T11	—	—	—	2.57	3.94	2.01	M+H	10.74	151.1114	150.22	C10H14O	Thymol	Prenol lipids
III	M387T6	—	—	—	1.71	2.19	1.94	M+H <sub>2</sub> Na	5.60	387.0659	364.30	C17H16O9	Xanthoxol glucoside	Coumarins and derivatives
III	M205T4	—	—	—	—	1.83	—	M+H	4.20	205.0965	204.23	C11H12N2O2	Tryptophane	Indoles and derivatives
III	M197T7	—	—	—	—	1.81	—	M+H	6.86	197.1163	196.24	C11H16O3	4-(3-hydroxybutyl)-2-methoxyphenol	Phenols
III	M267T11	—	—	—	—	3.16	3.07	M+H <sub>2</sub> Na	10.73	267.1159	244.28	C15H16O3	Oethol	Coumarins and derivatives
III	M335T10	—	—	—	—	8.36	6.16	M+H	9.57	335.1266	334.36	C21H18O4	Anhydronotoptol derivative 1	Coumarins and derivatives
III	M335T12	—	—	—	2.25	10.19	2.49	M+H	12.01	335.1266	334.36	C21H18O4	Anhydronotoptol derivative 2	Coumarins and derivatives
III	M337T11	—	—	—	2.88	6.26	3.01	M+H	10.62	337.1427	336.38	C21H20O4	Anhydronotoptol	Coumarins and derivatives
III	M353T10	—	—	—	3.45	17.39	11.04	M+H	9.63	353.1359	352.38	C21H20O5	Lansumarin A isomer 1	Coumarins and derivatives
III	M353T11	—	—	—	—	7.87	2.30	M+H	10.72	353.1384	352.38	C21H20O5	Lansumarin A isomer 2	Coumarins and derivatives
III	M339T13	—	—	—	—	1.68	2.86	M+H	12.68	339.1586	338.39	C21H22O4	8-geranyloxy psoralen	Coumarins and derivatives
III	M613T15	—	—	—	2.74	6.22	—	M+H	14.66	613.4797	612.92	C39H64O5	DG(18:3(9Z,12Z,15Z)/18:3(9Z,12Z,15Z)/0:0)	Glycerolipids
III	M277T12	—	—	—	2.53	4.27	3.35	M+H	12.33	277.213	276.41	C18H28O2	Stearidonic acid	Fatty Acyls
III	M279T12	—	—	—	—	6.04	5.66	M+H	12.18	279.2312	278.43	C18H30O2	Linolenic acid	Fatty Acyls
III	M279T13	—	—	—	—	5.44	—	M+H	12.80	279.2312	278.43	C18H30O2	Linolenic acid	Fatty Acyls
III	M283T14	—	—	—	3.61	14.30	15.02	M+H	13.22	283.2624	282.46	C18H34O2	Oleic acid	Fatty Acyls
III	M295T12	—	—	—	3.01	17.50	21.44	M+H	11.81	295.2261	294.42	C18H30O3	13-HOTE	Fatty Acyls
III	M496T14	—	—	—	2.99	4.81	2.71	M+H	14.08	496.3385	495.63	C24H50NO7P	LysolPC(16:0)	Glycerophospholipids
III	M518T12	—	—	—	—	4.90	—	M+H	11.68	518.3201	517.63	C26H48NO7P	LysolC(18:3)	Glycerophospholipids
III	M534T12	—	—	—	3.33	7.21	5.25	M+H	11.66	534.3182	533.67	C27H52NO7P	LysolC(18:2/0:0)	Glycerophospholipids
III	M536T12	—	—	—	4.03	11.08	6.48	M+H	11.66	536.3339	535.65	C26H50NO8P	PC(16:1(9Z)/2:0)	Glycerophospholipids
III	M791T16	—	—	—	2.01	3.82	2.93	M+H	15.53	790.5586	790.06	C42H80NO10P	PS(18:0/18:1(9Z))	Glycerophospholipids

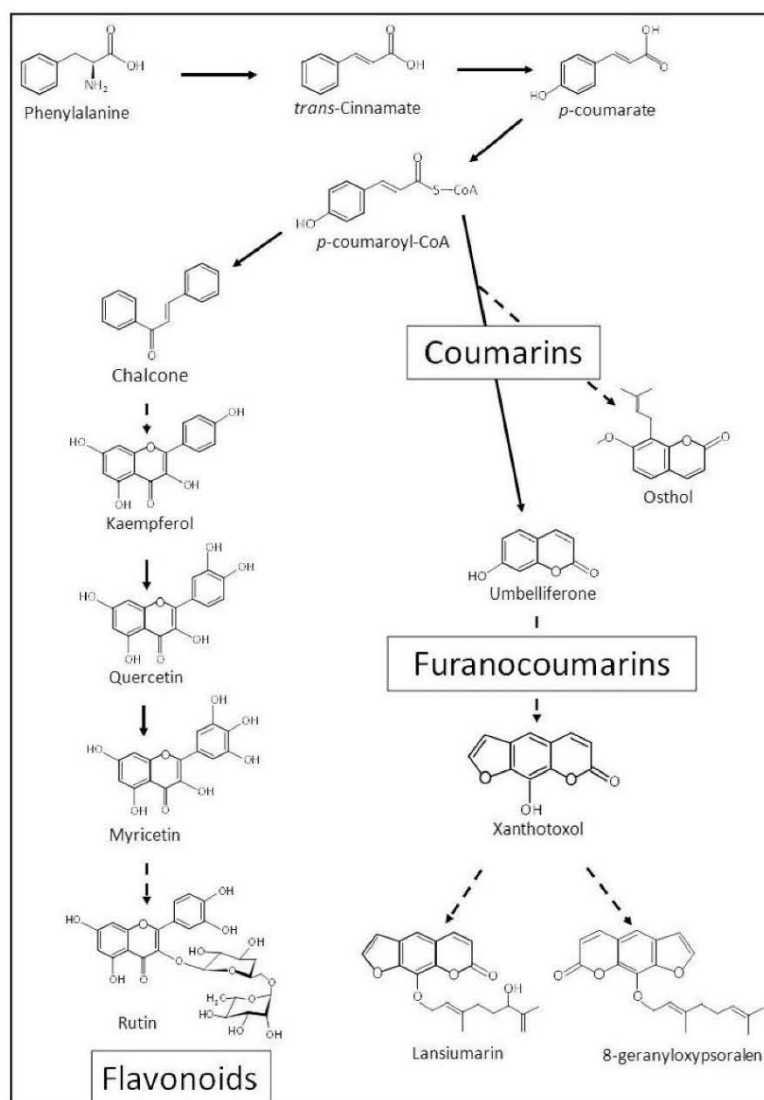


#### 4. Discussion

Plants have evolved several strategies to adapt to various environmental conditions. A way to investigate these strategies consists of applying different stresses to plants and assess the impact on their general aspect or on their metabolome. The O<sub>3</sub> stress that we used in our experimental conditions (250 ppb) can be considered as very severe compared to natural conditions (40 ppb on average on Earth) [25]. Such stress is generally reported to induce chlorotic or necrotic symptoms on leaves. In the case of parsnip, no visible symptom could be detected even after 72 h. Some species present an “avoiding” strategy, as reported in *Amaranthus palmeri* (Amaranthaceae), and limit O<sub>3</sub> influx to the leaf tissue in closing stomata [26]. Such a strategy usually contributes to minimizing oxidative stress and subsequent ROS accumulation. For parsnip, the mechanism might be different since we observed an increase in ROS concentration (Figure 1). The resistance of this plant to oxidative stress might rely on an alternative strategy and possibly on the large detoxifying arsenal present in plants. Part of this arsenal involves specialized metabolites, such as flavonoids, which display strong antioxidant properties. It has been reported that the synthesis of flavonoids is increased after O<sub>3</sub> exposure and closely related to ROS accumulation [27]. The concentration of flavonoids in parsnip leaves was globally not significantly different from that in the control with the exception of two flavonoids that were transiently affected. Specifically, these two flavonoids could contribute to manage O<sub>3</sub> response in parsnip through ROS scavenging. Alternatively, large pool of oxidized flavonoids issued from ROS scavenging might also be exported to the vacuole to maintain cell redox homeostasis or might be regenerated by peroxidases and/or reducing power [28]. In both cases, the flavonoid pool would be globally conserved and constant. O<sub>3</sub> exposure led also to an increase in NCC. This increase could be interpreted as a consequence of chlorophyll degradation in response to O<sub>3</sub> exposure, as such degradation has been largely shown in literature [29–31]. The role of NCCs in plant are still under investigation, however, some NCC were demonstrated to possess strong antioxidant properties [24]. The accumulation of NCC under O<sub>3</sub> exposure in parsnip may thus contribute to parsnip tolerance to O<sub>3</sub> through their antioxidant properties. In addition, parsnip tolerance to O<sub>3</sub> might also be supported by primary antioxidant molecules (e.g., ascorbate or glutathione) or detoxifying enzymes (e.g., catalase or peroxidase) and further investigations would be needed to evaluate their contributions.

Mechanical wounding significantly impacted the metabolic composition of parsnip leaves, as indicated by the accumulation of coumarins and furanocoumarins. This effect has previously been shown in roots of cultivated parsnips [18–20] and leaves of wild type parsnips [32,33], but interestingly, in these older studies, the panel of induced furanocoumarins was different. Indeed, we did not observe an induction of the synthesis of simple furanocoumarins, such as psoralen, xanthotoxin and isopimpinellin, but we noticed a significant increase in prenylated molecules such as 8-geranyloxy-psoralen. To our knowledge, this has not been reported in parsnip leaves so far.

Concomitantly with the furanocoumarin increase, we observed that the concentration of flavonoids decreased in the parsnip leaves submitted to mechanical wounding. This result is surprising since the flavonoid pathway is generally activated in response to biotic and abiotic stresses [34]. An explanation might be related to the biochemical relationship between the furanocoumarin and the flavonoid pathways. Both specialized metabolite pathways share a common precursor, *p*-coumaroyl-CoA (Figure 4). The induction of the synthesis of the prenylated furanocoumarin upon mechanical wounding may redirect the carbon flux to this pathway at the expense of the biosynthesis of flavonoids. Such metabolic competition for a common precursor has previously been demonstrated in genetically modified *Arabidopsis thaliana* expressing a bacterial bi-functional chorismate mutase/prephenate dehydratase [35].



**Figure 4.** Schematic diagram of putative *p*-coumaroyl-CoA competition between flavonoid and furanocoumarin pathways. Continuous arrows represent a one-step reaction, dotted arrows represent several enzymatic.

MW enhanced the accumulation of two monoterpenoids. This accumulation was significantly different from the control one day after the stress and reached an optimum after two days. Enhanced accumulation of such molecules in response to MW is well established since they were already reported to be involved in direct and indirect plant defense mechanisms in other plants such as *Myrica cerifera* and *Artemisia annua* [36,37].

In addition to specialized metabolites, MW also led to significant modifications in the lipid composition of the leaf extract with a global increase in fatty acyls and glycerophospholipids and a

decrease in glycerolipids. Such modifications were previously reported [38], and comprehensive mechanisms involving lipids in the signalling of plant response to abiotic stress have been reviewed. The modifications in the lipid profile occurring in parsnip leaves in response to MW are consistent with these previously described mechanisms. Hence, the accumulation of lysophospholipids (lysoPC (16:0), lysoPC (18:3), and lysoPC (18:2); Table 2) and fatty acids (stearidonic acid, linolenic acid, and oleic acid; Table 2) may be the consequence of the activation of phospholipases that cleave fatty acids from phospholipids [39]. The free fatty acids we detected also display a high degree of desaturation, which has previously been associated with plant responses to biotic and abiotic stresses and may be the consequence of the activation of fatty acid desaturases [40]. The accumulation of 13-HOTE, generated through the oxidation of linolenic acid by a lipoxygenase, and the corresponding accumulation of linolenic acid indicates that the oxylipin pathway was activated. This pathway notably leads to the synthesis of jasmonic acid, a well-known phytohormone involved in plant responses to biotic and abiotic stresses [41,42]. The decrease in glycerolipids, namely, diacylglycerols and monogalactosyldiacylglycerols, may be the consequence of the activation of phospholipases or may illustrate the degradation of plastids that are rich in glycerolipids. Such a decrease has also been observed in wounded *A. thaliana* [43,44].

In this study, we made evidence that plants do not have a single response to stress. The mechanisms involved are much more complex, but the approaches based on the analysis of the metabolome may help to understand them, especially by having an overview of the global network. This approach helped us to highlight an original response to stress in parsnip. Using metabolomics also provides solutions for going further and identifying the underlying molecular mechanisms. Based on different metabolic profiles of plant tissues, it will be possible to generate transcriptomic libraries that will help to identify genes related to the synthesis of target molecules.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Supplemental Table 1: Description of the datasets generated by the XCMS process and used for MetaboAnalyst statistical analyses. The list of DAMFs determined by ANOVA is provided. Supplemental Data 2: MS2 fragmentation of M677T8 and putative NCCs. A) The structure and fragmentation pattern of M677T8 identified as (2S,2(1)R)-2(1),2(2)-dicarboxy-8-ethenyl-2,7,12,18-tetramethyl-2,2(1)-dihydrobenzo[b]porphyrin-13,17-dipropanoic acid. B, C, D and E described the fragmentation pattern of M693T9, M691T10, M707T9 and M661T9. On each fragmentation pattern the mass-loss specific for NCCs is described (in blue) according to Roca et al., 2018.

**Acknowledgments:** The authors thank Clement Charles (LAE) for his technical support and Jérémie Goulmnik (LAE) for statistical analyses. This work was financially supported by the "Impact Biomolecules" project (ANR-15-IDEX-04-LUE Lorraine Université d'Excellence) (to A.H.).

**Author Contributions:** A.H, R.L. and G.G conceived the experiment and the study. G.G, R.L. and A.G performed the experiments and analysed the data. GG, A.G, Y.J, R.L. and A.H wrote the manuscript.

**Conflict of interest:** Authors declare that they have no conflict of interest.

## References

- Wahid, A.; Gelani, S.; Ashraf, M.; Foolad, M.R. Heat tolerance in plants: An overview. *Environ. Exp. Bot.* **2007**, *61*, 199–223.
- Walley, J.W.; Dehesh, K. Molecular Mechanisms Regulating Rapid Stress Signaling Networks in Arabidopsis. *J. Integr. Plant Biol.* **2010**, *52*, 354–359.
- Zhao, J.; Davis, L.C.; Verpoorte, R. Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnol. Adv.* **2005**, *23*, 283–333.
- Zhang, S.; Klessig, D.F. MAPK cascades in plant defense signaling. *Trends Plant Sci.* **2001**, *6*, 520–527.
- Ludwig, A.A.; Romeis, T.; Jones, J.D.G. CDPK-mediated signalling pathways: specificity and cross-talk. *J. Exp. Bot.* **2004**, *55*, 181–188.
- Vandelle, E.; Vannozzi, A.; Wong, D.; Danzi, D.; Digby, A.-M.; Dal Santo, S.; Astegno, A. Identification, characterization, and expression analysis of calmodulin and calmodulin-like genes in grapevine (*Vitis vinifera*) reveal likely roles in stress responses. *Plant Physiol. Biochem.* **2018**, *129*, 221–237.



- 364 7. Per, T.S.; Khan, M.I.R.; Anjum, N.A.; Masood, A.; Hussain, S.J.; Khan, N.A. Jasmonates in plants under  
365 abiotic stresses: Crosstalk with other phytohormones matters. *Environ. Exp. Bot.* **2018**, *145*, 104–120.
- 366 8. Baier, M.; Kandlbinder, A.; Goldack, D.; Dietz, K.J. Oxidative stress and ozone: perception, signalling and  
367 response. *Plant Cell Environ.* **2005**, *28*, 1012–1020.
- 368 9. Castagna, A.; Ranieri, A. Detoxification and repair process of ozone injury: From O<sub>3</sub> uptake to gene  
369 expression adjustment. *Environ. Pollut.* **2009**, *157*, 1461–1469.
- 370 10. Delcour, I.; Spanoghe, P.; Uyttendaele, M. Literature review: Impact of climate change on pesticide use.  
371 *Food Res. Int.* **2015**, *68*, 7–15.
- 372 11. Nakabayashi, R.; Saito, K. Integrated metabolomics for abiotic stress responses in plants. *Curr. Opin. Plant*  
373 *Biol.* **2015**, *24*, 10–16.
- 374 12. Shulaev, V.; Cortes, D.; Miller, G.; Mittler, R. Metabolomics for plant stress response. *Physiol. Plant.* **2008**,  
375 *132*, 199–208.
- 376 13. Berenbaum, M.; Feeny, P. Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a  
377 coevolutionary arms race? *Science* **1981**, *212*, 927–929.
- 378 14. Bourgaud, F.; Hehn, A.; Labat, R.; Doerper, S.; Gontier, E.; Kellner, S.; Matern, U. Biosynthesis of  
379 coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem. Rev.*  
380 **2006**, *5*, 293–308.
- 381 15. Smith, S.I.; Brodbelt, J.S. Rapid Characterization of Cross-links, Mono-adducts, and Non-covalent Binding  
382 of Psoralens to Deoxyoligonucleotides by LC-UV/ESI-MS and IRMPD Mass Spectrometry. *The Analyst*  
383 **2010**, *135*, 943–952.
- 384 16. Labat, R.; Kellner, S.; Specker, S.; Hehn, A.; Gontier, E.; Hans, J.; Bourgaud, F.; Matern, U. Molecular  
385 cloning and functional characterization of psoralen synthase, the first committed monooxygenase of  
386 furanocoumarin biosynthesis. *J. Biol. Chem.* **2007**, *282*, 542–554.
- 387 17. Labat, R.; Hehn, A.; Hans, J.; Schneider, S.; Jugdé, H.; Schneider, B.; Matern, U.; Bourgaud, F. Isolation and  
388 functional characterization of CYP71AJ4 encoding for the first P450 monooxygenase of angular  
389 furanocoumarin biosynthesis. *J. Biol. Chem.* **2009**, *284*, 4776–4785.
- 390 18. Roselli, S.; Olry, A.; Vautrin, S.; Coriton, O.; Ritchie, D.; Galati, G.; Navrot, N.; Krieger, C.; Vialart, G.;  
391 Bergès, H.; et al. A bacterial artificial chromosome (BAC) genomic approach reveals partial clustering of  
392 the furanocoumarin pathway genes in parsnip. *Plant J.* **2017**, *89*, 1119–1132.
- 393 19. Krieger, C.; Roselli, S.; Kellner-Thielmann, S.; Galati, G.; Schneider, B.; Grosjean, J.; Olry, A.; Ritchie, D.;  
394 Matern, U.; Bourgaud, F.; et al. The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of  
395 Coumarin Biosynthesis in Apiaceous Plants. *Front. Plant Sci.* **2018**, *9*.
- 396 20. Munakata, R.; Olry, A.; Karamat, F.; Courdavault, V.; Sugiyama, A.; Date, Y.; Krieger, C.; Silie, P.; Foureau,  
397 E.; Papon, N.; et al. Molecular evolution of parsnip (*Pastinaca sativa*) membrane-bound prenyltransferases  
398 for linear and/or angular furanocoumarin biosynthesis. *New Phytol.* **2016**, *211*, 332–344.
- 399 21. Junglee, S.; Urban, L.; Sallanon, H.; Lopez-Lauri, F. Optimized Assay for Hydrogen Peroxide  
400 Determination in Plant Tissue Using Potassium Iodide. *Am. J. Anal. Chem.* **2014**, *05*, 730.
- 401 22. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.;  
402 Subramanian, S.; Bolton, E.; et al. ClassyFire: automated chemical classification with a comprehensive,  
403 computable taxonomy. *J. Cheminformatics* **2016**, *8*, 61.
- 404 23. Kräutler, B. Chlorophyll breakdown and chlorophyll catabolites in leaves and fruit. *Photochem. Photobiol.*  
405 *Sci. Off. J. Eur. Photochem. Assoc. Eur. Soc. Photobiol.* **2008**, *7*, 1114–1120.
- 406 24. Müller, T.; Ulrich, M.; Ongania, K.-H.; Kräutler, B. Colorless Tetrapyrrolic Chlorophyll Catabolites Found  
407 in Ripening Fruit Are Effective Antioxidants. *Angew. Chem. Int. Ed.* **2007**, *46*, 8699–8702.
- 408 25. Oltmans, S.J.; Lefohn, A.S.; Shadwick, D.; Harris, J.M.; Scheel, H.E.; Galbally, I.; Tarasick, D.W.; Johnson,  
409 B.J.; Brunke, E.-G.; Claude, H.; et al. Recent tropospheric ozone changes – A pattern dominated by slow or  
410 no growth. *Atmos. Environ.* **2013**, *67*, 331–351.
- 411 26. Paudel, R.; Grantz, D.A.; Vu, H.-B.; Shrestha, A. Tolerance of Elevated Ozone and Water Stress in a  
412 California Population of Palmer Amaranth (*Amaranthus palmeri*). *Weed Sci.* **2016**, *64*, 276–284.
- 413 27. Gandin, A.; Davrinche, A.; Jolivet, Y. Deciphering the main determinants of O<sub>3</sub> tolerance in Euramerican  
414 poplar genotypes. *Sci. Total Environ.* **2019**, *656*, 681–690.
- 415 28. Takahama, U.; Oniki, T. A peroxidase/phenolics/ascorbate system can scavenge hydrogen peroxide in plant  
416 cells. *Physiol. Plant.* **1997**, *101*, 845–852.

29. Reich, P.B. Effects of Low Concentrations of O<sub>3</sub> on Net Photosynthesis, Dark Respiration, and Chlorophyll Contents in Aging Hybrid Poplar Leaves. **1983**, *73*, 6.
30. Pell, E.J.; Schlagnhauer, C.D.; Arteca, R.N. Ozone-induced oxidative stress: Mechanisms of action and reaction. *Physiol. Plant.* **1997**, *100*, 264–273.
31. Jolivet, Y.; Bagard, M.; Cabané, M.; Vaultier, M.-N.; Gandin, A.; Afif, D.; Dizengremel, P.; Le Thiec, D. Deciphering the ozone-induced changes in cellular processes: a prerequisite for ozone risk assessment at the tree and forest levels. *Ann. For. Sci.* **2016**, *73*, 923–943.
32. Zangerl, A.R. Furanocoumarin Induction in Wild Parsnip: Evidence for an Induced Defense against Herbivores. *Ecology* **1990**, *71*, 1926–1932.
33. Berenbaum, M.R.; Zangerl, A.R. Chemical phenotype matching between a plant and its insect herbivore. *Proc. Natl. Acad. Sci.* **1998**, *95*, 13743–13748.
34. Likic, S.; Rusak, G. Changes in Phenolic Compounds in Nicotiana Species as a Response to Wounding and Viral Infection. *J. Plant Pathol.* **2014**, *96*, 569–575.
35. Tzin, V.; Malitsky, S.; Aharoni, A.; Galili, G. Expression of a bacterial bi-functional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in Arabidopsis. *Plant J.* **2009**, *60*, 156–167.
36. Ruan, J.-X.; Li, J.-X.; Fang, X.; Wang, L.-J.; Hu, W.-L.; Chen, X.-Y.; Yang, C.-Q. Isolation and Characterization of Three New Monoterpene Synthases from *Artemisia annua*. *Front. Plant Sci.* **2016**, *7*.
37. Wang, M.; Lincoln, D.E. Effects of light intensity and artificial wounding on monoterpene production in *Myrica cerifera* from two different ecological habitats. *Can. J. Bot.* **2004**, *82*, 1501–1508.
38. Lee, S.; Suh, S.; Kim, S.; Crain, R.C.; Kwak, J.M.; Nam, H.-G.; Lee, Y. Systemic elevation of phosphatidic acid and lysophospholipid levels in wounded plants. *Plant J.* **1997**, *12*, 547–556.
39. Ryu, S.B.; Wang, X. Activation of phospholipase D and the possible mechanism of activation in wound-induced lipid hydrolysis in castor bean leaves. *Biochim. Biophys. Acta BBA - Lipids Lipid Metab.* **1996**, *1303*, 243–250.
40. Zhang, M.; Demeshko, Y.; Dumbur, R.; Iven, T.; Feussner, I.; Lebedev, G.; Ganim, M.; Barg, R.; Ben-Hayyim, G. Elevated  $\alpha$ -Linolenic Acid Content in Extra-plastidial Membranes of Tomato Accelerates Wound-Induced Jasmonate Generation and Improves Tolerance to the Herbivorous Insects *Heliothis peltigera* and *Spodoptera littoralis*. *J. Plant Growth Regul.* **2018**.
41. Marchica, A.; Lorenzini, G.; Papini, R.; Bernardi, R.; Nali, C.; Pellegrini, E. Signalling molecules responsive to ozone-induced oxidative stress in *Salvia officinalis*. *Sci. Total Environ.* **2019**, *657*, 568–576.
42. Savatin, D.V.; Gramegna, G.; Modesti, V.; Cervone, F. Wounding in the plant tissue: the defense of a dangerous passage. *Front. Plant Sci.* **2014**, *5*.
43. Vu, H.S.; Shiva, S.; Roth, M.R.; Tamura, P.; Zheng, L.; Li, M.; Sarowar, S.; Honey, S.; McElhiney, D.; Hinkes, P.; et al. Lipid changes after leaf wounding in *Arabidopsis thaliana*: expanded lipidomic data form the basis for lipid co-occurrence analysis. *Plant J.* **2014**, *80*, 728–743.
44. Vu, H.S.; Roston, R.; Shiva, S.; Hur, M.; Wurtele, E.S.; Wang, X.; Shah, J.; Welti, R. Modifications of membrane lipids in response to wounding of *Arabidopsis thaliana* leaves. *Plant Signal. Behav.* **2015**, *10*.

## A bacterial artificial chromosome (BAC) genomic approach reveals partial clustering of the furanocoumarin pathway genes in parsnip

Sandro Roselli<sup>1,2</sup>, Alexandre Olry<sup>1,2</sup>, Sonia Vautrin<sup>3</sup>, Olivier Coriton<sup>4</sup>, Dave Ritchie<sup>5</sup>, Gianni Galati<sup>1,2</sup>, Nicolas Navrot<sup>6</sup>, Célia Krieger<sup>1,2</sup>, Guilhem Vialart<sup>1,2</sup>, Hélène Bergès<sup>3</sup>, Frédéric Bourgaud<sup>1,2</sup> and Alain Hehn<sup>1,2,\*</sup> 

<sup>1</sup>Laboratoire Agronomie et Environnement, INRA UMR1121, 2 avenue de la forêt de Haye TSA 40602, 54518 Vandœuvre-lès-Nancy, France,

<sup>2</sup>Laboratoire Agronomie et Environnement, Université de Lorraine UMR1121, 2 avenue de la forêt de Haye – TSA 40602, 54518 Vandœuvre-lès-Nancy, France,

<sup>3</sup>Centre National de Ressources Génomiques Végétales – INRA – 24 Chemin de Borde Rouge – Auzeville CS 52627, 31326 Castanet Tolosan Cedex, France,

<sup>4</sup>Plate-Forme de Cytogénétique Moléculaire – UMR1349 IGEPP INRA – Agrocampus Ouest-Université de Rennes 1, 35653 Le Rheu, France,

<sup>5</sup>INRIA Nancy Grand Est, 615 rue du Jardin Botanique, 54600 Villers-lès-Nancy, France, and

<sup>6</sup>Institut de biologie moléculaire des plantes – UPR2357 CNRS, Université de Strasbourg, 12 rue du Général Zimmer – 67084 Strasbourg Cedex, France

Received 29 February 2016; revised 2 December 2016; accepted 6 December 2016; published online 10 December 2016.

\*For correspondence (e-mail alain.hehn@univ-lorraine.fr).

### SUMMARY

Furanocoumarins are specialized metabolites that are involved in the defense of plants against phytophagous insects. The molecular and functional characterization of the genes involved in their biosynthetic pathway is only partially complete. Many recent reports have described gene clusters responsible for the biosynthesis of specialized metabolites in plants. To investigate possible co-localization of the genes involved in the furanocoumarin pathway, we sequenced parsnip BAC clones spanning two different gene loci. We found that two genes previously identified in this pathway, *CYP71AJ3* and *CYP71AJ4*, were located on the same BAC, whereas a third gene, *PsPT1*, belonged to a different BAC clone. Chromosome mapping using fluorescence *in situ* hybridization (FISH) indicated that *PsPT1* and the *CYP71AJ3-CYP71AJ4* clusters are located on two different chromosomes. Sequencing the BAC clone harboring *PsPT1* led to the identification of a gene encoding an Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenase (*PsDIOX*) situated in the neighborhood of *PsPT1* and confirmed the occurrence of a second gene cluster involved in the furanocoumarin pathway. This enzyme metabolizes *p*-coumaroyl CoA, leading exclusively to the synthesis of umbelliferone, an important intermediate compound in furanocoumarin synthesis. This work provides an insight into the genomic organization of genes from the furanocoumarin biosynthesis pathway organized in more than one gene cluster. It also confirms that the screening of a genomic library and the sequencing of BAC clones represent a valuable tool to identify genes involved in biosynthetic pathways dedicated to specialized metabolite synthesis.

**Keywords:** furanocoumarin, P450, cluster, prenyltransferase,  $\alpha$ -ketoglutarate-dependent dioxygenase, *Pastinaca sativa*, bacterial artificial chromosome, bacterial artificial chromosome-fluorescence *in situ* hybridization.

### INTRODUCTION

Furanocoumarins are specialized metabolites involved in plant defense against phytophagous insects (Beier and Oertli, 1983; Beier, 1990). These molecules were mainly reported in Fabaceae (*Psoralea cinerea*), Moraceae (*Ficus*

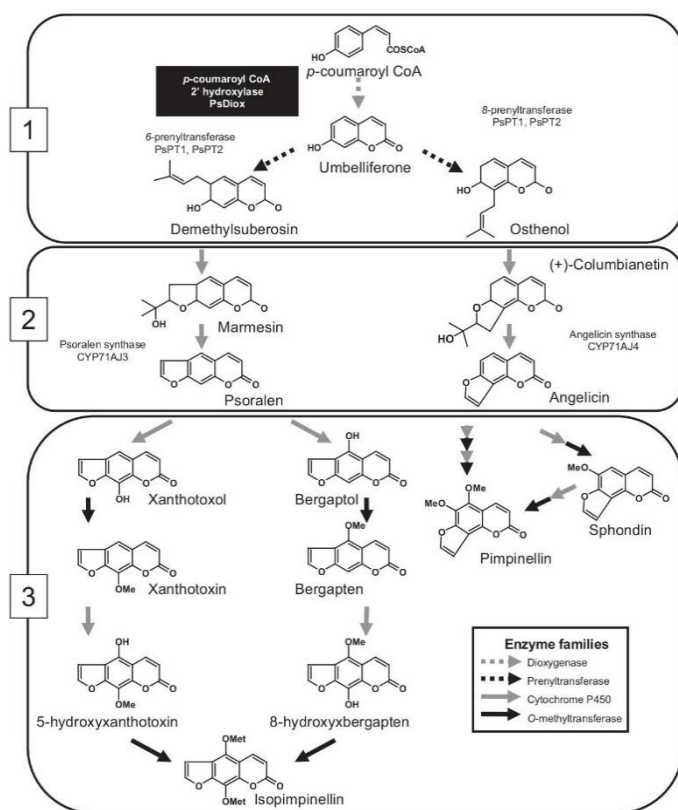
*carica*), Rutaceae (*Citrus bergamia*, *Ruta graveolens*) and Apiaceae (*Pastinaca sativa*, *Petroselinum crispum*) (Seigler, 1998; Bourgaud *et al.*, 2006). Two structural isomers of furanocoumarins were described based on the position of

the furan ring on the coumarin core molecule, leading to either linear or angular isomers (Ellis and Brown, 1974) (Figure 1). While linear forms have been reported in all plants that synthesize furanocoumarins, angular forms have only been detected in Fabaceae, Apiaceae and Moraceae (Bourgaud *et al.*, 2014). The appearance of angular furanocoumarins has been described as an arms race between plants and herbivores (Berenbaum and Feeny, 1981; Berenbaum, 1983). Although these compounds have been widely studied for decades, the furanocoumarin pathway has not been fully characterized at the molecular level and several genes remain unidentified. In this study we focused on *P. sativa* as a model plant for which four genes encoding enzymes related to furanocoumarin biosynthesis have already been reported (Figure 1). Two prenyltransferases (PsPT1 and PsPT2) were recently shown to catalyze the prenylation of umbelliferone on C6 or C8, yielding demethylsuberosin (DMS) and osthenol, respectively (Munakata *et al.*, 2016). These enzymes have the same activity as an orthologous enzyme previously identified in parsley (Karamat *et al.*, 2014) but are preferentially

involved in the synthesis of linear or angular isomers. The two other enzymes described thus far in *P. sativa* belong to the cytochrome P450 superfamily. CYP71AJ3 (psoralen synthase) is responsible for the conversion of marmesin to psoralen (Larbat *et al.*, 2007), while CYP71AJ4 (angelicin synthase) catalyzes the conversion of columbianetin to angelicin (Larbat *et al.*, 2009). These two enzymes have an overall protein sequence identity of 70% and are assumed to be paralogs derived from a common ancestor (Larbat *et al.*, 2009).

Recent studies have indicated that genes involved in the biosynthetic pathways of some defense compounds are co-localized as functional clusters in plant genomes (Kliebenstein and Osbourn, 2012; Nützmann and Osbourn, 2014; Nützmann *et al.*, 2016). This type of organization for functionally related genes would provide at least three major advantages for the plant:

- (i) A simplified transfer of a group of genes from one genome to another, hence guaranteeing the transfer of the whole biosynthesis pathway from one plant to another.



**Figure 1.** Schematic representation of a simplified linear and angular furanocoumarin biosynthesis pathway in parsnip.

Different enzyme types are involved in furanocoumarin biosynthesis pathways. The steps catalyzed by enzymes from the same family are represented by arrows as follows: dioxigenase is represented by a dashed grey arrow, prenyltransferases are represented by black dashed arrows, cytochrome P450 members are represented by full grey arrows and O-methyltransferases are represented by full black arrows.

Upstream, intermediate and downstream parts of the furanocoumarin pathway are surrounded by frames 1, 2 and 3, respectively. Functionally characterized enzymes are boxed in white, PsDIOX (*Pastinaca sativa*  $\alpha$ -ketoglutarate-dependent dioxigenase) is boxed in black.



- (ii) A lower risk of interrupting synthesis in the middle of the pathway, which could enhance the concentration of harmful intermediate compounds in the plant itself (Osborn, 2010; Boycheva *et al.*, 2014).
- (iii) An easier co-regulation of gene expression (Nützmann *et al.*, 2016; Yu *et al.*, 2016). Therefore, sequencing the genomic regions surrounding previously described genes could be a valuable tool to identify other genes involved in the same biosynthetic pathway (Osborn, 2010).

In the present study, we investigated the possibility that the furanocoumarin biosynthetic pathway might be organized as a cluster. Indeed, a striking feature of the plant gene clusters identified to date is the presence of P450-encoding genes, which are known to be involved in the adaptation of plants to their environment (Boycheva *et al.*, 2014; Kawai *et al.*, 2014). As previously stated, most steps of this pathway are catalyzed by cytochrome P450 enzymes (Hamerski and Matern, 1988). The role of this enzyme family in the adaptation of plants to their environment has been widely reported in the literature. Such evolution may be the result of the individual duplication, neo-functionalization and positive selection of genes (Edger *et al.*, 2015), and some studies have shown that these coding sequences are organized in gene clusters (Osborn, 2010). While studying furanocoumarin biosynthesis, Larbat *et al.* (2007, 2009) noted that the genes encoding the enzymes involved in the synthesis of psoralen, a linear furanocoumarin, and angelicin (the corresponding angular counterpart) are paralogous cytochrome P450s and that *CP71AJ4* likely originated through a gene duplication event followed by neo-functionalization. These structurally related coding sequences may be grouped in the parsnip genome.

To investigate a clustered organization, it is necessary to assess the physical genomic environment of the genes encoding enzymes involved in the biosynthesis pathway. Although plant genomic or transcriptomic databases are readily available due to next-generation sequencing technologies, no genomic information concerning parsnip is currently available. An alternative approach to whole genome data mining is the combinatorial use of BAC library screening together with fluorescence *in situ* hybridization (FISH). These complementary approaches are suited to investigate the physical localization of genes throughout the genome in only three consecutive steps:

- (i) Identification of genome fragments harboring already functionally characterized genes is performed through the direct screening of a BAC library.
- (ii) Sequencing of 'positive' BAC clones provides high-quality sequences of the genomic regions of interest, and the analysis of these sequences provides the genomic organization of the surrounding genomic region of a restricted area of 100 kb up to more than 1 Mb.

- (iii) FISH methods using BAC clones as probes pinpoint genomic regions directly onto a species' genome organization as suggested by Iovene and collaborators (Iovene *et al.*, 2008).

In this study, we investigated the genomic organization of the furanocoumarin pathway in parsnip. We first describe the construction of BAC library. Using this library together with FISH methods revealed that three genes already described (*CYP71AJ3*, *CYP71AJ4* and *PsPT1*) and involved in the furanocoumarin biosynthetic pathway are located on two different chromosomes. The sequencing of the corresponding BACs led us to identify a gene coding for an Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenase that specifically metabolizes *p*-coumaroyl CoA to produce umbelliferone, making this enzyme an ideal candidate to provide umbelliferone to the furanocoumarin pathway situated downstream.

## RESULTS

### BAC library construction and characterization

The *HindIII* restriction enzyme was used because all of the genes already characterized (*CYP71AJ3*, *CYP71AJ4* and *PsPT1*) lack the corresponding restriction site. The BAC library constructed from parsnip genomic DNA consisted of 59 904 clones arrayed into 156 384-well microtiter plates. These BAC clones were produced in two steps from the same ligation product; an initial pool of 41 472 clones was completed by another pool of 18 432 clones. To evaluate the average insert size of the library, DNA was isolated from 425 randomly selected BAC clones, digested with a rare cutter *NotI*, and analyzed using pulsed-field gel electrophoresis (PFGE). All fragments generated by *NotI* digestion contained the 7.5-kb plndigoBAC-5 vector band and various insert fragments. Analysis of the insert sizes from these BAC clones indicated a mean size of 106 kb (see Figure S1 for a subset of 43 BAC clones). Because the haploid DNA content of *Pastinaca sativa* is 1.73 pg, its genome size was estimated to be 1691 Mb (Bennett and Leitch, 2012). Considering these data, the coverage of our parsnip BAC library named Pas-B-Dlg is approximately 3.5× genome equivalents.

### Screening with probes targeting genes involved in the furanocoumarin pathway and identification of BAC clones of interest

To identify BAC clones harboring previously characterized coding sequences involved in furanocoumarin synthesis, specific nucleotide probes targeting *CYP71AJ3*, *CYP71AJ4* and *PsPT1* were designed. Screening using *CYP71AJ3* and *CYP71AJ4* labelled probes on the first part of the library (~2.6× equivalents of the estimated genome size) led to the identification of 15 and 22 BAC clones, respectively, of which 13 were common to both probes. *CYP71AJ3* and

*CYP71AJ4* share 70% sequence homologies (Larbat *et al.*, 2009). Therefore, to assess the hybridization specificity of the probes, we performed a qPCR amplification on each of the 13 selected BAC clones, and we compared the resulting melting curves with those obtained using parsnip genomic DNA and plasmid DNA harboring genes of interest as controls. This enabled us to identify only 27B24 and 102B12 as real positive candidates. Sequencing of the PCR products obtained on 27B24 and 102B12 confirmed that both amplicons were 100% identical to the targeted sequences described in the literature (Larbat *et al.*, 2007, 2009). The presence of 11 false-positive BAC clones (among the 13 selected) suggested the existence of other *CYP71AJ* genes (or even *CYP71* genes) sharing high homologies with *CYP71AJ3* and *CYP71AJ4*. Following digestion with *NotI*, the sizes of 27B24 and 102B12 were estimated to be 105 and 110 kb, respectively which is consistent with the library average insert size estimation of 106 kb.

A screening using *PsPT1* as probe was performed on the 59 904 BAC clones constituting the whole BAC library. Only four positive BAC clones were detected, and all of them were individually validated by qPCR, comforting a high specificity of hybridization. PCR products corresponding to part of *PsPT1* gene were sequenced to validate that the clones carried the targeted sequence. Sequencing of the corresponding PCR products revealed 100% identity with the *PsPT1* sequence for three BAC clones (67N6, 92M14 and 112I2) (Munakata *et al.*, 2016). In the fourth clone (57P4), 2 SNPs (single-nucleotide polymorphisms) were detected. Clones 67N6, 92M14 and 112I2 were digested with *NotI* to estimate the insert size of the corresponding BAC clones. The largest insert containing the *PsPT1* gene was found in BAC clone 67N6 and had an estimated size of 129 kb. Inserts from clones 92M14 and 112I2 were estimated to be 115 kb and 97 kb, respectively. Two BAC clone 67N6 and 112I2 were selected for full-insert sequencing.

#### BAC clone sequencing and physical mapping

Bacterial artificial chromosome clones containing the genes of interest were sequenced at a minimum of 31× coverage, and the raw sequences obtained for each BAC clone were assembled individually. The assembly lengths for BAC clones 27B24 (12 contigs) and 102B12 (6 contigs) that harbor the *CYP71AJ3* and *CYP71AJ4* genes, were 98.6 and 104.8 kb, respectively (Figure S2a). BAC clone 67N6, which carries the *PsPT1* locus, was assembled into 13 large contigs, and the total assembly length was 110.6 kb. BAC clone 112I2 has been sequenced using the PacBio method and was assembled into one unique contig of 97.6 kb (Figure S2b). All assembly sizes were consistent with the previously estimated average insert sizes.

BAC-by-BAC sequence alignment allowed partial orientation and organization of the overlapping BAC sequences

(Figure S2). The total overlapping length between BAC clones 27B24 and 102B12 was 50.8 kb, with an average identity of 98.3%. These results led us to assume that the two BAC clones carried sequences that belong to two homologous chromosomes. Indeed, *P. sativa* L. subsp. *sativa* 'Demi Long de Guernesey' is a diploid heterozygous parsnip species ( $2n = 22$  chromosomes), and the use of gene-specific probes could not distinguish between these chromosomes. However, based on the BAC clone sequences and the alignment start and end positions, we were able to estimate the total size of the *CYP71AJ3/CYP71AJ4* genomic region sequenced to be 117 kb (104.8 kb corresponding to BAC 102B12 plus 12.1 kb corresponding to the M13 region of 27B24 (Figure S2a).

The sequences of BAC clones 67N6 and 112I2 had a total of 50.1 kb sequence overlap, with an average identity of 99.9% (Figure S2b), demonstrating that they physically overlapped and originated from the same chromosome. The total size of this physical map of the *PsPT1* clone is estimated to be 158 kb.

#### Fluorescence *in situ* hybridization

Previously, Wegel *et al.* (2009) have used DNA-FISH to visualize adjacent genes within a biosynthetic gene cluster in oat. To determine whether the different BAC clones carrying genes involved in furanocoumarin biosynthesis are localized in the same genomic area, the clones were used as molecular probes in BAC-FISH to identify their respective chromosomal localizations in *P. sativa*.

A preliminary analysis of the sequence performed with the RepeatMasker tool (<http://www.repeatmasker.org>) using the *Arabidopsis thaliana* genome revealed that the BAC clones carried repetitive elements (11% in 27B24 and 67N6 and 8% in 102B12) that might be responsible for the dispersion of hybridization signals. Therefore, to reduce the background, we used a competitive *C<sub>0</sub>t* repetitive-DNA fraction that acts as a 'blocker' to reduce the non-specific binding.

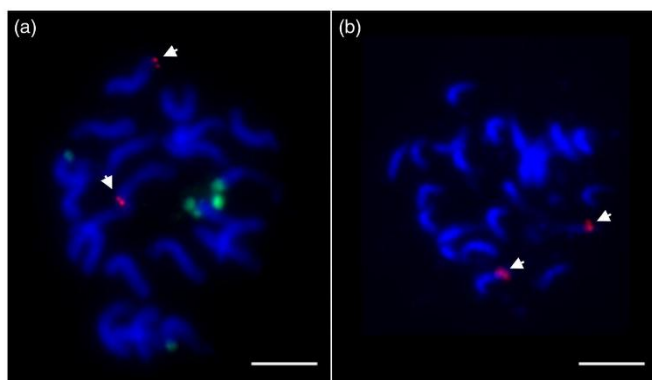
The identified BAC clones harboring *PsPT1* (i.e. 67N6) or *CYP71AJ3/CYP71AJ4* (i.e. 27B24 and 102B12) were mapped separately on chromosomes to ensure that the hybridization produced a signal. Hybridization of a probe specific to the 67N6 clone harboring the *PsPT1* gene gave single-locus signals that were observed in the telomeric regions of one pair of chromosomes (Figure 2a, red dots). The efficiency of the hybridization was assessed using a 45S rDNA probe as a control for which mapping to two pairs of chromosomes was expected (Figure 2a, green dots) (Iovene *et al.*, 2008). As for 67N6, hybridization of specific probes targeting the BAC clones carrying *CYP71AJ* genes, 27B24 and 102B12 yielded single-locus signals that were detected at the ends of a chromosome pair (Figure 2b). Finally, probes specific to the BAC clones carrying the *CYP71AJ3-CYP71AJ4* cluster (102B12) and the *PsPT1* gene (67N6)



**Figure 2.** Double fluorescence *in situ* hybridization (FISH).

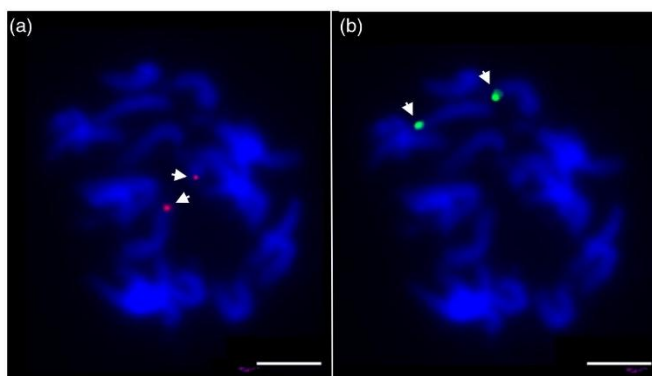
(a) Double fluorescence *in situ* hybridization (FISH) combining both BAC 67N6 (containing *PsPT1*) probe (red) and 45S rDNA probe (green) was carried out on mitotic metaphase chromosomes of *P. sativa*.

(b) FISH using BAC 27B24 (containing *CYP71AJ3* and *CYP71AJ4*) probe (red) was performed on mitotic metaphase chromosomes of *P. sativa*. The BAC signals are indicated by arrows. Chromosomes were counterstained with DAPI (blue). The scale bar represents 5  $\mu$ m. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)].



**Figure 3.** Double FISH.

(a, b) Double FISH combining both BAC 102B12 (containing *CYP71AJ3* and *CYP71AJ4*) probe (red) (a) and BAC 67N6 (containing *PsPT1*) probe (green) (b) was carried out on the same mitotic metaphase chromosomes of *P. sativa*. Chromosomes were counterstained with DAPI (blue). The BAC signals are indicated by arrows. The scale bar represents 5  $\mu$ m. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)].



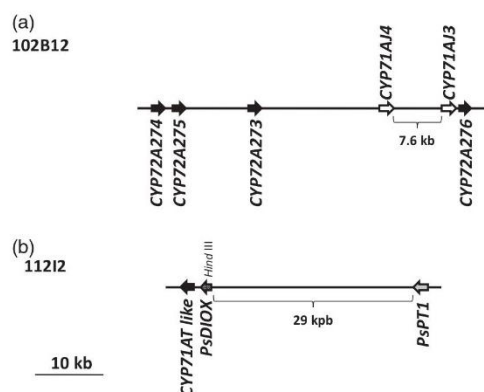
were simultaneously used in BAC-FISH to assess whether the corresponding genes co-localized on the same chromosome. This clearly demonstrated that both hybridizations were separately detected on a different pair of chromosomes (Figure 3a, b), indicating that these three genes do not belong to a single cluster physically related on the same chromosome.

#### Identification of genes potentially related to furanocoumarin biosynthesis

The assembled sequences of the different BAC clones (27B24 and 102B12 for *CYP71AJ3*-*CYP71AJ4*; 67N6 and 112I2 for *PsPT1*) were annotated to explore their coding sequences content. We searched for gene families belonging to cytochrome P450s, Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenases, prenyltransferases and methyltransferases, which have already been reported to play roles in furanocoumarin synthesis in higher plants (Hehmann *et al.*, 2004; Larbat *et al.*, 2007, 2009; Vialart *et al.*, 2012; Munakata *et al.*, 2016). Additionally, we extended our investigation to

other proteins such as transporters and transcription factors (Table S2). The analysis of BAC 27B24 and 102B12 confirmed the presence of both *CYP71AJ3* and *CYP71AJ4*. These two genes are separated by 6548 and 7594 bp on BAC clones 27B24 and 102B12, respectively and are in the same orientation (Figure 4a). In addition to these two *CYP71AJ* genes, four additional genes encoding putative enzymes from the CYP72A subfamily (namely *CYP72A273*, *CYP72A274*, *CYP72A275* and *CYP72A276*, Accession N°KY081680, KY081681, KY081682 and KY081683 respectively) were identified. One of these genes (*CYP72A275*) might be a pseudogene because it includes a frameshift mutation caused by an insertion of five nucleotides (CTAAT).

The *PsPT1* gene sequence was identified on clone 67N6 and 112I2. An extensive analysis of all the contigs revealed the presence of a putative gene encoding an Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenase (*PsDIOX*, Accession N°KY081678) sharing 56% identity with the *p*-coumaroyl CoA 2'-hydroxylase (*C2'H*) isolated from *Ruta graveolens* (Vialart *et al.*, 2012). Although only the first 382 nucleotides



**Figure 4.** Schematic representation of the genomic regions with furanocoumarin biosynthetic gene clusters in parsnip. (a) Genomic fragment of interest isolated from BAC clone 102B12 (78 kb). *CYP71AJ* genes are represented by white arrows. The genes encoding P450 enzymes from the *CYP72A* subfamily are represented by black arrows. (b) Genomic fragment of interest isolated from BAC clone 112I2 (97 kb). The *PsPT1* gene is represented by a grey arrow. *PsDIOX* gene encoding an  $\alpha$ -ketoglutarate-dependent dioxygenase is indicated by a black arrow with white dots. The *CYP71AT*-like gene is represented by a black arrow. The scale bar is indicated. A *HindIII* restriction site is localized within the *PsDIOX* coding sequence.

of the gene were identified on BAC clone 67N6, a full-length *PsDIOX* was found on BAC clone 112I2 (Figure 4b). Sequence analysis of clone 112I2 confirmed that *PsPT1* and *PsDIOX* are in the same orientation and that the genes are separated by 28986 bp (Figure 4b). The fact that only a partial *PsDIOX* fragment was found on BAC clone 67N6 is reasonable because a *HindIII* cleavage site was identified in the *PsDIOX* gene sequence at nucleotide position 384 of the coding sequence. An additional cytochrome P450 (Accession N° KY081679) coding sequence has also been identified on BAC112I2 and is located 850 nt upstream of *PsDIOX*. Its genomic sequence is 1692 nt long and includes one predicted intron. The translated protein is 72% homologous to *CYP71AT* subfamily of cytochrome P450s (Table S2 and Figure S4).

#### Functional characterization of the *CYP71AT*, *CYP72A* and *PsDIOX* genes

To have a more precise idea about the involvement of these genes in the synthesis of furanocoumarins *in planta* we assessed a possible correlation between the gene expression level and the concentration of these molecules in different organs of young parsnips. These analyzes highlighted that the concentration of most furanocoumarins (with the exception of pimpinellin) is higher in the leaves than in roots and significantly higher than the concentrations observed in the stems (Figure 5a). These results are

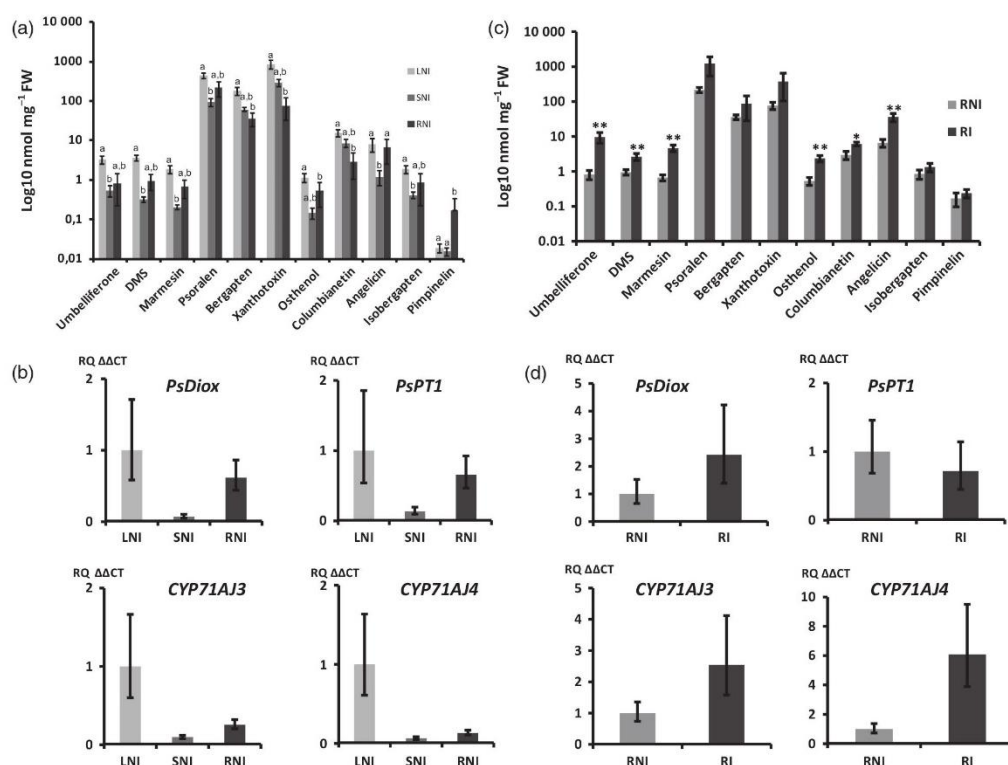
consistent with the expression levels of the different gene we tested: high expression in leaves, lower expression in roots and much lower expression in stems (Figure 5b). In addition, the synthesis of furanocoumarins has been described to be inducible (Larbat *et al.*, 2009; Munakata *et al.*, 2016). When roots are mechanically wounded we found a significant increase of the concentration of all the furanocoumarins (Figure 5c) consistent with a noticeable enhancement of the expression of genes encoding *PsDIOX*, *CYP71AJ3* and *CYP71AJ4* (Figure 5d).

To investigate their biochemical function we started to clone the different candidate genes and express the corresponding proteins in heterologous expression systems. Although *CYP71AT* specific primers led us to amplify the genomic sequence, we never could amplify the coding sequence from cDNA prepared from roots and leaves of *in vitro* plants or soil cultured plants. This gene might not be expressed in our experimental conditions.

All *CYP72As* could be amplified successfully using cDNA prepared from mRNA as template. However, for *CYP72A275*, as expected from the BAC sequence analysis, we always highlighted a 5-nt insertion that introduced a frameshift in the predicted coding sequence, probably leading to the synthesis of a non-functional expression product. Similarly, *CYP72A276* always included an intron that was never spliced (negative and positive controls were performed to assess the absence of the remaining genomic DNA). Therefore, we assumed that *CYP72A275* and *CYP72A276* are likely to be pseudogenes. To perform functional characterization, *CYP72A273* and *CYP72A274* were expressed in yeast. Resulting microsomes were incubated with 11 different coumarins and furanocoumarins (Figure S3a) and four iridoids (loganic acid, loganin, 7-deoxyloganin and secologanin) (Figure S3b) (Miettinen *et al.*, 2014). These substrates were not metabolized.

To identify the function of the putative Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenase, we first compared the protein sequence to four other enzymes whose activities were already described as related to the coumarin metabolism. These analyses highlighted that the parsnip enzyme shared 1) 61% amino acids (AA) identity with F6'H1 from *Arabidopsis thaliana* and 62.2% with Ib1-2 from *Ipomoea batatas* which specifically transform feruloyl CoA into scopoletin (Kai *et al.*, 2008; Matsumoto *et al.*, 2012) and 2) 58% AA identity with C2'H from *R. graveolens* or 59% with Ib2-1 from *Ipomoea batatas*. These two latter enzymes show a substrate promiscuity for feruloyl CoA and *p*-coumaroyl CoA responsible for the formation of both scopoletin and umbelliferone (Matsumoto *et al.*, 2012; Vialart *et al.*, 2012). A more detailed sequence comparison showed that 161 amino acids (Figure 6, yellow boxes) were strictly conserved in the four Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenases. This analysis, however, does not indicate a clear affiliation of *PsDIOX* with one activity or





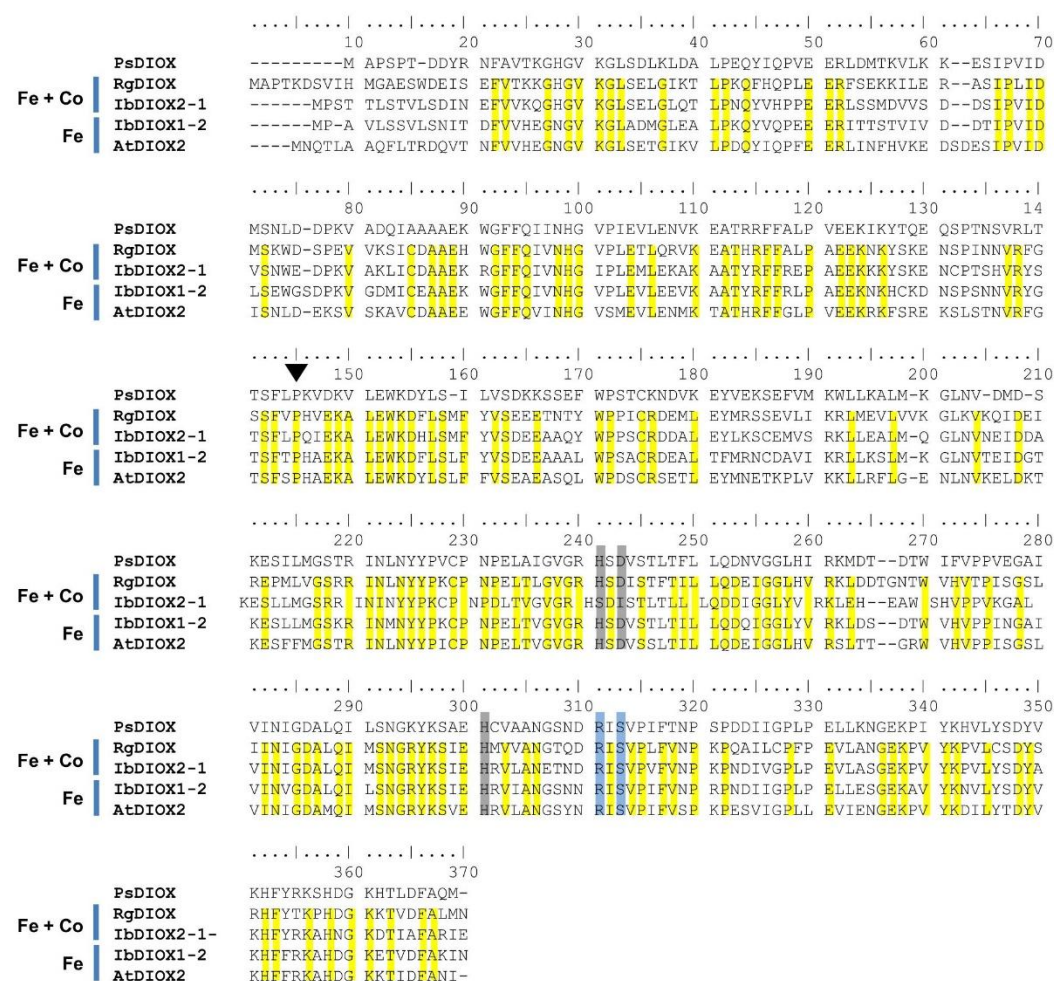
**Figure 5.** Spatio-temporal accumulation of FCs in parsnip and corresponding expression pattern of *PsDIOX*, *PsPT1*, *CYP71AJ3* and *CYP71AJ4*. (a, b) (a) Assessment of the FCs content and (b) mRNA expression pattern of *PsDIOX*, *PsPT1*, *CYP71AJ3* and *CYP71AJ4* in parsnip leaves non-induced (LNI), stems non-induced (SNI) and Roots non-induced (RNI). Error bars represent the standard error for six replicates. (c, d) (c) Assessment of the FCs content and (d) mRNA expression pattern of *PsDIOX*, *PsPT1*, *CYP71AJ3* and *CYP71AJ4* in parsnip non-induced roots (light grey) and induced roots (dark grey). The stars represent the significant differences of the results between the two conditions according to the Wilcoxon test (\* means a *P*-value < 0.05, \*\* means a *P*-value < 0.01). Values represented by the same letter did not differ significantly (Kruskal–Wallis test, *P*-value < 0.05).

another. Therefore, we performed a second analysis based on 3D model structure alignments using the Kpax tool (Ritchie *et al.*, 2012). The models mainly differ in the conformations of their C-terminal loops, which enter the active site (Figure S4). In these models, the most invasive loop is observed in *PsDIOX*, which likely endows this enzyme with distinct substrate specificity compared to the previously described enzymes. However, as the feruloyl CoA 6-hydroxylase structure (DOI: 10.2210/pdb4xae/pdb) that was used as the modeling template lacks the C-terminal residues, these predictions must be further confirmed by experimental approaches. We finally chose to express the protein in a bacterial system as described for the 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* (Vialart *et al.*, 2012) in order to realize an experimental functional characterization. After purification, the His-tagged protein was incubated with four different hydroxycinnamic acids and

their corresponding coenzyme A esters (Figure S3c). Only *p*-coumaroyl CoA was transformed in product which has been identified as umbelliferone, the precursor of linear and angular furanocoumarins (Figure S5), highlighting the exclusive function of this enzyme for the synthesis of umbelliferone. The optimal experimental conditions were assessed (pH = 8.14 and  $T_{opt}$  = 13.4°C; Figure S5) and the calculated enzymatic constants obtained in these conditions ( $K_m$  = 201 ± 59 μM,  $V_{max}$  = 44 ± 7 pmol min<sup>-1</sup> per μg; Figure S6).

## DISCUSSION

The operon-like gene clusters encoding genes involved in biosynthetic pathways of specialized metabolites in plants, which were unknown a few years ago, are now widely investigated (Osborn *et al.*, 2012; Boycheva *et al.*, 2014; Nützmann *et al.*, 2016). Based on such a model, we



**Figure 6.** ClustalW peptide alignment of PsDIOX with previously characterized dioxygenases (RgDIOX GenBank ID: ADV77970.1, Ib1-1DIOX GenBank ID: AB636149, Ib2-1DIOX GenBank ID: AB636153, AtDIOX2 GenBank ID: AAS49108.1). The conserved regions binding to Fe(II) ions are highlighted in grey, and the conserved regions of dioxygenases are highlighted in blue. A black triangle indicates the position of the conserved proline residue (position 145). Amino acids conserved in enzymes catalyzing hydroxylation of feruloyl CoA and p-coumaroyl CoA are highlighted in yellow. Fe indicates feruloyl CoA, and Co indicates p-coumaroyl CoA. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)].

hypothesized that the genes involved in the furanocoumarin biosynthetic pathway may be co-localized in the parsnip genome. If this assumption could be confirmed, then knowledge of the sequences surrounding already known genes would help to identify all genes involved in the synthesis of these molecules. To assess the genomic organization of the furanocoumarin biosynthetic pathway, we chose to construct a BAC library, a strategy which has been described as a valuable resource to explore genomic regions carrying genes of interest (Hamberger *et al.*, 2009; Anistoroaei *et al.*, 2011). We

constructed a 3.5× genome equivalent BAC library of *P. sativa*. This library was further screened with functionally described genes (*CYP71AJ3*, *CYP71AJ4* and *PsPT1*) (Larbat *et al.*, 2007, 2009; Munakata *et al.*, 2016) and used to identify four positive BAC clones. These BAC clones were sequenced with two different technologies: 454-Roche (27B24, 102B12, 67N6) and PacBio (112I2). The resulting contigs were assembled and showed that *CYP71AJ3* and *CYP71AJ4* were separated by approximately 7 kb, whereas *PsPT1* and *PsDIOX* were separated by 29 kb. The use of sequencing technologies providing short output sequences



(reads) led to high-quality assemblies covering 117 kb with only four scaffolds for the *CYP71AJ3/CYP71AJ4* locus and 158 kb with only 12 contigs for the *PsDIOX/PsPT1* locus. This assembling is greatly facilitated and improved using the PacBio sequencing approach by providing a unique assembly sequence of BAC clone 112I2. If the clusters are enlarged by additional screenings, whole-locus PacBio resequencing could be an alternative way to obtain the complete sequence data and to ensure that no genes would be lost in the annotation steps.

A detailed analysis of the BAC clones harboring the *CYP71AJ3* and *CYP71AJ4* genes led us to identify four additional cytochrome P450 genes. These genes all belong to the same CYP72A P450 subfamily and are likely the result of gene duplication events because two of them are not functional in plants. Gene duplication followed by neofunctionalization and positive selection has been observed and demonstrated in the glucosinolate biosynthesis pathway (Edger *et al.*, 2015). It is possible that an ancestral CYP72A gene was duplicated (likely *CYP72A273* or *CYP72A274*) and that the new copies of the genes are not functional. CYP72A enzymes are involved in the synthesis of monoterpene indole alkaloids (Irmeler *et al.*, 2000; Miettinen *et al.*, 2014) and triterpenoid saponins (Seki *et al.*, 2011). Irmeler *et al.* (2000) demonstrated that CYP72A1 was able to catalyze an oxidative ring cleavage, converting loganin into secologanin in iridoid metabolism (Figure 3b). This reaction shares similarities with the synthesis reactions of marmesin or columbianetin in the furanocoumarin biosynthesis pathway (Figure 1). Unfortunately, the enzymatic reaction with marmesin did not lead to the synthesis of any new product. It is however possible that *in planta* this reaction occurs with a transient intermediate molecule as substrate, such as a glycosylated DMS or osthenol, although these molecules have not been observed in parsnip. The function of these CYP72A remains unknown so far and will need a broader screening with additional candidate substrates. In addition, it is likely that this lack of metabolism is related to non-recognition of the substrates by the enzyme, we cannot exclude that it is related to the lack of production of the recombinant protein in yeast.

The sequence analysis of the genomic fragment bearing the *PsPT1* gene enabled the discovery of *PsDIOX* that was unknown prior to this study. An *in silico* comparison highlights that this enzyme shares over 160 AA (43% of the sequence) with four previously functionally characterized Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenases. The remaining 57% of the sequence are more or less different from other Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenases involved in the coumarin metabolism making it a putative specific enzyme dedicated to furanocoumarin synthesis in parsnip.

The *in silico* analysis revealed no particular subcellular addressing signal which is consistent with other Fe(II)

$\alpha$ -ketoglutarate-dependent dioxygenases (Kawai *et al.*, 2014). This raises more generally the question of the intracellular synthesis pathway of these molecules. Umbelliferone, the product formed by *PsDIOX*, is produced in the cytoplasm and is subsequently prenylated by a prenyltransferase which is located in the plastidic membranes (Munakata *et al.*, 2016). The prenylated molecule is then redirected to cytochrome P450 enzymes which are located in the ER (Larbat *et al.*, 2007, 2009). Although these enzymes are located in different subcellular compartments they might be spatially co-located in the cell. Such direct contact has already been described between the ER and plastidic membranes by Andersson *et al.* (2007) and Ginglinger *et al.* (2013).

The differences with other Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenases were also highlighted by a structure comparison based on protein models that show that *PsDIOX* has a C-terminal loop that might be involved in distinct substrate specificity compared to the other enzymes. These predicted differences are confirmed by *in vitro* characterization experiments which make evidence that it specifically catalyzes the transformation of *p*-coumaroyl CoA, with umbelliferone as a unique product. The identification of this enzyme contributes to the understanding of the molecular basis of furanocoumarin synthesis in parsnip and is clearly different from the pathway described in *R. graveolens* (Vialart *et al.*, 2012), where a bifunctional dioxygenase able to transform *p*-coumaroyl CoA into umbelliferone and also feruloyl CoA into scopoletin has been described. Here in *P. sativa*, *PsDIOX* is exclusively committed to the conversion of *p*-coumaroyl CoA into umbelliferone, the immediate precursor of all furanocoumarins, which might accommodate better with the furanocoumarin pathway than the previously described dioxygenase from *R. graveolens*. This difference between *P. sativa* and Rutaceae is reinforced by an analysis of the genomic organization of orthologous genes in *Citrus clementina*, another plant producing furanocoumarins. A local sequence alignment (BLAST) analysis using *PsDIOX* and *PsPT1* as probe was done on the citrus genome (<https://www.citrusgenomedb.org>). This data mining showed that several putative Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenases with more than 56% identity with *PsDIOX* are located on scaffold 31 and 44 of the genome. A similar search done with *PsPT1* highlighted a single scaffold (scaffold 39) which is harboring an orthologous gene sharing only less than 40% identity with *PsPT1*. Even if these genes display any activity in the synthesis of furanocoumarins in citrus (which has to be demonstrated) they are not physically close to each other. It would be interesting now to investigate the substrate specificity of other Apiaceae and Rutaceae orthologous dioxygenases to check if this reflects a general rule or not. This study might also be extended to other furanocoumarin producer plants such as Moraceae and Fabaceae (Bourgaud *et al.*, 2014).

We recently also identified a second prenyltransferase (PsPT2) involved in the synthesis of DMS and osthenol in parsnip. It is quite probable that localizing the *PsPT2* gene in our BAC library with a specific PsPT2 probe would lead to the discovery of other BACs involved in the synthesis of furanocoumarins. These BACs might be close to the BACs identified here, and contribute to extending one of the two clusters presented in Results. Alternatively, *PsPT2* might be localized elsewhere, giving a new cluster. Since the number of enzymatic steps involved in the synthesis of xanthotoxin, a major end-product from the furanocoumarin pathway starting from *p*-coumaroyl-CoA, is presumably seven, we can anticipate a cascade of new genes discovered from our BAC-oriented strategy.

The identification of two genomic regions carrying furanocoumarin biosynthetic genes in parsnip led us to investigate their respective localizations on the chromosomes. Indeed, gene clusters described thus far in specialized plant metabolism can span over 100 000 kb (Osborn, 2010). To overcome this possible issue, we used a BAC-FISH approach to map the respective chromosomal localizations of corresponding gene clusters in parsnip. We found evidence that the genes involved in furanocoumarin biosynthesis are located at least on two separate chromosomes in *P. sativa*. This element rules out the hypothesis of a single cluster containing all of the genes involved in furanocoumarin synthesis. It is however interesting to note that the *PsPT1*-*PsDIOX* cluster contains enzymes involved in the upstream biosynthetic pathway leading to coumarin synthesis (umbelliferone, DMS and osthenol; Figure 1), whereas the second cluster contains P450 genes dedicated to the intermediate pathway (Figure 1) specifically involved in the furanocoumarin synthesis. So far, we cannot exclude the possibility that the genes involved in the synthesis of the downstream furanocoumarin derivatives might be physically located at a different position from the first two loci identified in this work (Figure 1) or even not grouped but spread all over the genome. We intend to investigate this matter further by: (i) screening other BAC clones using new probes, such as *PsPT2*, another gene involved in the synthesis of osthenol (Munakata *et al.*, 2016) and also with other identified P450 genes (Dueholm *et al.*, 2015); and (ii) identifying and sequencing the clones surrounding the already identified BAC sequences.

Parallel analyzes of furanocoumarin concentration and gene expression level *in planta* is consistent with the involvement of the enzymes in the synthesis of these molecules. Leaves are the organs containing the highest concentrations of furanocoumarins in comparison with the stems and roots. This observation is in line with the physiological role of these molecules that are produced by plants in response to an attack by herbivores (McCloud *et al.*, 1992). Our experimental data provides evidence that the different genes investigated are more expressed in these organs than

in the other parts of the plant. The correlation between these genes and the synthesis of furanocoumarins is reinforced in wounded roots where a significant increase of both furanocoumarins and gene expression could be observed. The only gene which is not subjected to any modification at the expression level is the prenyltransferase PsPT1. This result differs from the one described by Munakata *et al.* (2016) who stimulated *in vitro* plants with methyljasmonate. Since the experimental conditions are quite different with ours, we can assume that the regulation signals might be different. However this needs deeper investigation.

The results described in this study confirmed the potential value of a BAC-driven approach for the discovery of genes involved in a given specialized pathway (Nützmann *et al.*, 2016). We validate the BAC strategy to decipher the furanocoumarin pathway, which may then be applied to any secondary pathway. Such a BAC strategy is greatly underused as compared to transcriptomic approaches for finding candidate genes. Until now there is no information available in the literature on the chromosomal organization of the furanocoumarin pathway and we demonstrate that this pathway is, at least, organized in two gene clusters in parsnip. In addition we have also functionally characterized *PsDIOX* as a new gene discovered by the application of our BAC approach. In the future, the BAC tool should be used to further explore the regulatory elements that control the furanocoumarin pathway in higher plants. For this purpose, promoters of the different coding sequences will be identified and characterized.

## EXPERIMENTAL PROCEDURE

### Plant material

Parsnips (*Pastinaca sativa* L. subsp. *sativa* 'Demi Long de Guernesey') were cultivated as described by Munakata *et al.* (2016). Leaves were collected from plantlets after 3 to 5 days in darkness, flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until use.

### Construction of BAC libraries

High-molecular-weight (HMW) DNA was prepared from young parsnip leaves as described by Peterson *et al.* (2000) and modified as described by Gonthier *et al.* (2010). Agarose embedded HMW DNA was partially digested with *HindIII* (New England Biolabs, Ipswich, MA, USA), subjected to two size selection steps by PFGE using a Bio-Rad CHEF Mapper system (Bio-Rad Laboratories, Hercules, CA, USA), and ligated into the pIndigoBAC-5 *HindIII*-Cloning Ready vector (Epicenter Biotechnologies, Madison, WI, USA). Pulsed-field migration programs, electrophoresis buffer, and ligation desalting conditions were based on Chalhoub *et al.* (2010). The insert size of the BAC clones was assessed using the *Not I* restriction enzyme and analyzed by PFGE. DNA fingerprinting was performed after an *EcoRI* digestion.

### Construction and radiolabeling of probes

Radiolabeled probes ( $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ ) were designed to screen the genomic BAC library of parsnip for the presence of genes involved in furanocoumarin synthesis. DNA fragments were PCR-amplified



from parsnip genomic DNA using specific primers (Table S1) targeting *CYP71AJ3*, *CYP71AJ4* and *PsPT1* providing 525, 552 and 420 bp long probes, respectively. For PCR amplification the reaction was set at 95°C for 5 min prior to 30 cycles of 10 sec denaturation at 95°C followed by 15 sec of annealing at 51°C or 53°C (*CYP71AJ3*, *CYP71AJ4* or *PsPT1*, respectively) and a 30 sec extension at 72°C. The reaction was finished with an additional 10 min extension at 72°C. The probes were sequenced prior to the screening of the BAC library.

### Screening of the BAC library

High-density colony filters were prepared with the genomic BAC library of *P. sativa* using a robotic workstation QPix2 XT (Molecular Devices, Sunnyvale, CA, USA). BAC clones were spotted in duplicate using a 6 × 6 pattern onto 22 × 22 cm Immobilon-Ny+ filters (Molecular Devices). The whole BAC library comprising 59 904 BAC clones was represented by a set of two filters, containing 41 472 and 18 432 clones, spotted in duplicate. After incubation at 37°C for 17 h, filters were processed as follows:

- (i) Denaturation on Whatman paper soaked with a solution of 0.5 M NaOH and 1.5 M NaCl for 4 min at room temperature and for 10 min at 100°C.
- (ii) Neutralization on a Whatman paper soaked with 1 M Tris-HCl pH 7.4 and 1.5 M NaCl for 10 min, incubation in a solution of 0.25 mg ml<sup>-1</sup> proteinase K (Sigma-Aldrich, St. Louis, MO, USA) for 45 min at 37°C and baking for 45 min at 80°C.
- (iii) UV fixation on a Biolink 254 nm crosslinker (Thermo Fisher Scientific, Waltham, MA, USA) with an energy of 120 000 µJ.

Probe radiolabeling and filter hybridization were performed as described in Gonthier *et al.* (2010). Hybridized filters were imaged with a Storm 860 PhosphorImager (GE Healthcare, Little Chalfont, UK), and analyses were performed using HDFR software (Incogen, Williamsburg, VA, USA). Positive BAC clones detected by hybridization were validated individually by quantitative PCR (qPCR) amplification using the primer pairs used for probe synthesis and genomic and plasmid DNA harboring genes of interest as controls. qPCR amplified products were visualized using agarose gel electrophoresis and sequenced to confirm the specific amplification of targeted genes. BAC-end sequencing was used to compare and discriminate identified BAC clones.

### Sequencing and assembling

**Roche-454.** Bacterial artificial chromosome clones 27B24, 102B12 and 67N6 were sequenced in two distinct pools of 10 BAC clones. One microgram of each individual BAC DNA was used to prepare individual tagged libraries using the GS FLX Titanium Rapid Library Preparation Kit (Roche, Brandford, CT, USA). All of the libraries were pooled to proceed with the emulsion PCR. After amplification, the beads binding the DNA were loaded on a GS FLX+ sequencer flow cell according to the manufacturer's instructions.

To eliminate low-quality, highly repeated and contaminated reads, the raw reads were cleaned with PyroCleaner (Jerôme *et al.*, 2011) and filtered against the *E. coli* complete genome sequence. Then, a standard assembly was performed on the remaining reads (Newbler 2.9) with the vs option for screening the vector.

**PacBio RS II.** Two micrograms of BAC clone 112I02 were pooled with 11 other BAC clones to obtain a total amount of 24 µg. One library was generated using the standard Pacific Biosciences library preparation protocol for 10-kb libraries. This library was sequenced in one PacBio RS II SMRT Cell using the P4

polymerase in combination with C2 chemistry (GATC Biotech, Mulhouse, France).

Assembly of the PacBio RS II reads was performed following the HGAP workflow (<https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP>). The SMRT<sup>®</sup> Analysis (v2.2.0) software suite was used for HGAP implementation. Reads were first aligned using BLASR against '*E. coli* str. K12 substr. DH10B, complete genome'. Identified *E. coli* reads and low-quality reads (read quality < 0.80 and read length < 500 bp) were removed from the data used for the BAC clone sequence assembly. Vector sequences were trimmed during the assembly process. Each BAC assembly was individualized by matching their back-end sequences (BES) on the ends of the assembled sequences with BLAST.

### Sequence analysis and gene annotation

A threshold of 95% identity was used to determine relevant contig alignments for BAC heterozygous clones 27B24/102B12, whereas a threshold of 99% identity was used for BAC clones 67N6 and 112I02, which originated from the same chromosome. Only the best hits were conserved for each contig (i.e. highest identity % and longest alignment length). The alignment identity was calculated by dividing the total number of base matches by the total length of the alignment (including substitutions, insertions, and deletions). Genes were annotated on the final assemblies using the gene prediction software EUGENE (Foissac *et al.*, 2008).

Repetitive content in the BAC clone sequences was identified using the web program RepeatMasker (<http://www.repeatmasker.org/>) with *Arabidopsis thaliana* selected as the DNA source.

### Chromosome preparation

Root tips 0.5–1.5 cm length were treated in the dark with 0.04% 8-hydroxyquinoline for 2 h at 4°C followed by 2 h at room temperature to allow them to progress to metaphase and were then fixed in 3:1 ethanol-glacial acetic acid for 12 h at 4°C and stored in 70% ethanol at –20°C until use. After washing in 0.01 M enzyme buffer (citric acid-sodium citrate pH 4.5) for 15 min, the roots were digested in a solution of 5% Onozuka R-10 cellulase (Sigma) and 1% Y23 pectolyase (Sigma) at 37°C for 15 min. The root tips were then washed carefully with distilled water for 30 min. A root tip was transferred to a slide and macerated with a drop of 3:1 fixation solution using fine-pointed forceps. After air-drying, slides with good metaphase chromosome spreads were stored at –20°C.

### BAC-FISH

Preparation of slides and hybridization using bacterial artificial chromosome-fluorescent *in situ* hybridization (BAC-FISH) was carried out according to procedures detailed in Suay *et al.* (2014). BAC clones 27B24, 67N6 and 102B12 were labelled by random priming with biotin-14-dUTP (ThermoFisher Scientific, Waltham, MA, USA). The ribosomal probes used in this study were 45S rDNA (pTa 71) (Gerlach and Bedbrook, 1979) which contained a 9-kb *EcoRI* fragment of rDNA repeat unit (18S-5.8S-26S genes and spacers) isolated from *Triticum aestivum*. The pTa 71 and 67N6 probes were labelled with Alexa-488 dUTP by random priming. C<sub>0</sub>t-100 fraction of *Pastinaca sativa* genomic DNA to block labelled repetitive sequences in BAC clones from hybridization to the chromosomes was prepared according to Zwick *et al.* (1997) and Szinay *et al.* (2008). Biotinylated probe was immunodetected by Texas Red avidin DCS (Vector Laboratories, Burlingame, CA, USA) and the signal was amplified with biotinylated anti-avidin D (Vector Laboratories). The chromosomes were mounted and counterstained in Vectashield (Vector Laboratories) containing

2.5 µg ml<sup>-1</sup> 4',6-diamidino-2-phenylindole (DAPI). Fluorescence images were captured using a CoolSnap HQ camera (Photometrics, Tucson, AZ, USA) on an Axioplan 2 microscope (Zeiss, Oberkochen, Germany) and analysed using MetaVue™ (Universal Imaging Corporation, Downingtown, PA, USA).

#### Real-time quantitative PCR of *CYP71AJ3*, *CYP71AJ4*, *PsPT1* and *PsDIOX* mRNA transcripts

Total mRNA extraction from plants cultivated in soil was performed according to Munakata *et al.* (2016). Remaining DNA was digested using Amplification Grade DNase kit (Sigma-Aldrich, Saint-Louis, MO, USA) with twice more mRNA and three times more different solutions. The cDNA was synthesized with 100 ng of extracted mRNA using High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, USA). The cDNA was diluted five-fold before real-time quantification (StepOnePlus, Applied Biosystems, Foster City, CA, USA). The real-time quantification was performed using SYBR Green Master Mix: SYBR *Pre-mix Ex Taq II* (Tli RNase H Plus) kit (Clontech, Mountain View, CA, USA).

The primers used for real-time quantification of *PsPT1*, *PsDIOX*, *CYP71AJ3*, *CYP71AJ4* and *ribulose-1,5-biphosphate carboxylase/oxygenase* mRNA are listed in supplemental Table S1. *Ribulose-1,5-biphosphate carboxylase/oxygenase* gene was used as a reference. Efficiency of the primers was checked. The 2<sup>-ΔΔCt</sup> method was used for the quantification of the transcripts expressed in different tissues and compared with the expression levels in induced and non-induced *in vitro* plants. The results are the means of six independent replications. Error bars represent standard deviations.

#### Coumarin and furanocoumarin extraction and LC-MS analysis

Leaves, stems and roots of *P. sativa* plantlets were grinded in a mortar with a pestle. Protocol to prepare the samples and analyze the furanocoumarin content by UPLC/MS was derived from Munakata *et al.* (2016). Briefly, frozen powder (100 mg) was collected and weighted for each sample. Eight hundred microlitre of 80% methanol was added to the powder and the mixture was homogenized by vortexing. After 10 min sonication (Elma s70 Elmasonic, sweep mode) and a 20 min long centrifugation at 15 000 g, the supernatant was collected and transferred into new microtubes. The pellet was submitted to a second extraction. The methanolic extract (1.6 ml) was dried (Eppendorf Concentrator plus) at 40°C for 15 h. The resulting pellet was finally resuspended with 100 µl of methanol 80%, vortexed and sonicated 10 min. For UPLC quantification analyses, a standard curve with all the furanocoumarins (1, 5, 20, 30 µM) was prepared. 4-Methylumbelliferone at a concentration of 5 µM was used as an internal standard for mass quantification.

#### Cloning and heterologous expression of *PsDIOX*, *CYP72A23* and *CYP72A24*

Total RNAs were extracted from mechanically wounded parsnip leaves (6 h post wounding) with the RNeasy plant extraction kit (Qiagen, Hilden, Germany) and treated with DNase I to avoid any genomic DNA contamination. cDNA was generated using High Capacity cDNA Reverse Transcriptase (Applied Biosystems) and random primers. The full-length cDNA sequences were PCR-amplified using a high-fidelity DNA polymerase (PrimeSTAR®Max; Clontech), using the primers described in supplemental Table S1. The PCR conditions were as follows: 5 min at 98°C, 35 cycles (10 sec at 98°C, 10 sec at 52°C or 55°C, 2 min at 72°C), and

a final 5 min extension step at 72°C. PCR products were cloned using a pCR®8/GW/TOPO® TA Cloning® Kit (ThermoFisher Scientific), and sequenced. Specific primers were used to amplify the coding sequences corresponding to five genes and each PCR product was sequenced (Table S1).

The amplified *PsDIOX* coding sequence was subcloned into the pET28b expression vector. The recombinant plasmid was introduced into *E. coli* Rosetta 2 bacteria. Bacterial culture, His-tagged protein purification and functional characterization were performed as described by Vialart *et al.* (2012) excepting that protein induction by 1 mM IPTG was performed for 4 h at 37°C. Substrate synthesis and enzymatic characterization were performed according to Vialart *et al.* (2012). The *CYP72A* coding sequences were subcloned in pYEDP60 and the corresponding proteins were produced in the *S. cerevisiae* strain WAT21 (Pompon *et al.*, 1996). Microsome preparation and enzyme assays were conducted as described previously (Larbat *et al.*, 2007, 2009). The catalytic parameters are the means of three independent replicates, and errors represent standard deviations. The kinetic parameters were calculated using the SigmaPlot software program (Systat Software Inc., London, UK).

#### Bioinformatics

The protein sequence comparison was performed with ClustalW software. Protein structure comparison was performed using the Kpax protein structure alignment tool (Ritchie *et al.* (2012), <http://kpax.loria.fr/>).

#### ACKNOWLEDGEMENTS

SR was funded by the European funding program INTERREG IV (SaarLorBioTech). The author would like to acknowledge Sébastien Carrere from the Laboratory of Plant-Microbe Interactions- INRA for the bioinformatics annotation performed on the parsnip BACs, the CATI BBIC (<https://bbic.toulouse.inra.fr>) for according access to the Genome Browser and the Gentoul Bioinformatic Platform Genotoul (<http://bioinfo.genotoul.fr/>) for Inter-Pro annotations. The authors would also like to acknowledge Jérémy Grosjean and Clément Charles for their technical help, and Dr David Nelson for annotation of cytochrome P450s. There are no conflicts of interest.

#### AUTHOR CONTRIBUTIONS

SR, AO, NN and GV cloned the different genes and performed the biochemical characterizations of the corresponding enzymes. SR, SV and HB constructed the BAC library and performed the screening. OC performed the FISH experiments. DR completed the *in silico* modeling. SR and CK performed the *in silico* data mining. SR, GG and AO performed the real-time PCR experiments. SR, FB and AH planned and designed the experiments and wrote the manuscript.

#### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Subset of 43 random BAC clones DNA from parsnip BAC library Pas-B-Dlg digested with *NotI* and analyzed with pulsed-field gel electrophoresis (PFGE) for insert size estimation.

**Figure S2.** Physical maps of (a) BAC 27B24 and 102B12 harboring *CYP71AJ3* and *CYP71AJ4* and (b) BAC 11212 and 67N6 harboring *PsPT1*.



**Figure S3.** Compounds used as potential substrates in the enzymatic screening experiments.

**Figure S4.** 3D model structure alignments of dioxygenases using the Kpax tool.

**Figure S5.** *In vitro* metabolism of *p*-coumaroyl-CoA by PsDIOX.

**Figure S6.** Enzymatic characterization of PsDIOX.

**Table S1.** Primer sequences used in this study.

**Table S2.** List of genes identified on sequenced BAC clones.

## REFERENCES

- Andersson, M.X., Goksor, M. and Sandelius, A.S. (2007) Optical manipulation reveals strong attracting forces at membrane contact sites between endoplasmic reticulum and chloroplasts. *J. Biol. Chem.* **282**, 1170–1174.
- Anistoroaei, R., Ten Hallers, B., Nefedov, M., Christensen, K. and de Jong, P. (2011) Construction of an American mink Bacterial Artificial Chromosome (BAC) library and sequencing candidate genes important for the fur industry. *BMC Genom.* **12**, 354.
- Beier, R.C. (1990) Natural pesticides and bioactive components in foods. *Rev. Environ. Contam. Toxicol.* **113**, 47–137.
- Beier, R.C. and Oertli, E.H. (1983) Psoralen and other linear furanocoumarins as phytoalexins in celery. *Phytochemistry*, **22**, 2595–2597.
- Bennett, M. and Leitch, I. (2012) Plant DNA C-values database. (<http://www.kew.org/cvalues/>)
- Berenbaum, M. (1983) Coumarins and Caterpillars: a case for coevolution. *Evolution*, **37**, 163–179.
- Berenbaum, M. and Feeny, P. (1981) Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? *Science*, **212**, 927–929.
- Bourgaud, F., Hehn, A., Larbat, R., Doerper, S., Gontier, E., Kellner, S. and Matern, U. (2006) Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem. Rev.* **5**, 293–308.
- Bourgaud, F., Olry, A. and Hehn, A. (2014) Recent advances in molecular genetics of furanocoumarin synthesis in higher plants. In *Recent Advances in Redox Active Plant and Microbial Products* (Jacob, C., Kirsch, G., Slusarenko, A., Winyard, P.G. and Burkholz, T., eds). Dordrecht: Springer Science+Business Media, pp. 363–375.
- Boycheva, S., Daviet, L., Wolfender, J.L. and Fitzpatrick, T.B. (2014) The rise of operon-like gene clusters in plants. *Trends Plant Sci.* **19**, 447–459.
- Chalhoub, B., Belcram, H. and Caboche, M. (2010) Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. *Trends Plant Sci.* **3**, 225.
- Dueholm, B., Krieger, C., Drew, D., Olry, A., Kamo, T., Taboureaux, O., Weitzel, C., Bourgaud, F., Hehn, A. and Simonsen, H. (2015) Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evol. Biol.* **15**, 122.
- Edger, P.P., Heide-Fischer, H.M., Bekaert, M.L. et al. (2015) The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl Acad. Sci. USA*, **112**, 8362–8366.
- Ellis, B.E. and Brown, S.A. (1974) Isolation of dimethylallylpyrophosphate: umbelliferone dimethylallyltransferase from *Ruta graveolens*. *Can. J. Biochem.* **52**, 734–738.
- Foissac, S., Gouzy, J., Rombauts, S., Mathé, C., Amselem, J., Sterck, L., Veau d'Peer, Y., Rouzé, P. and Schiex, T. (2008) Genome annotation in plants and fungi: Eugène as a model platform. *Curr. Bioinform.* **3**, 87–97.
- Gerlach, W. and Bedbrook, J. (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Res.* **7**, 1869–1885.
- Ginglinger, J.F., Boachon, B., Hofer, R. et al. (2013) Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in *Arabidopsis* flowers. *Plant Cell*, **25**, 4640–4657.
- Gonthier, L., Bellec, A., Blassiau, C., Prat, E., Helmstetter, N., Rambaud, C., Huss, B., Hendriks, T., Bergès, H. and Quillet, M.C. (2010) Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). *BMC Res. Notes*, **3**, 225.
- Hamberger, B., Hall, D., Yuen, M., Oddy, C., Hamberger, B., Keeling, C.I., Ritland, C., Ritland, K. and Bohlmann, J. (2009) Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defence reveal insights into a conifer genome. *BMC Plant Biol.* **9**, 106.
- Hamerski, D. and Matern, U. (1988) Biosynthesis of psoralens. Psoralen 5-monooxygenase activity from elicitor-treated *Ammi majus* cells. *FEBS Lett.* **239**, 263–265.
- Hermann, M., Lukacin, R., Ekiert, H. and Matern, U. (2004) Furanocoumarin biosynthesis in *Ammi majus* L.: cloning of bergapton O-methyltransferase. *Eur. J. Biochem.* **271**, 932–940.
- Iovene, M., Grzebelus, E., Carpato, D., Jiang, J. and Simon, P.W. (2008) Major cytogenetic landmarks and karyotype analysis in *Daucus carota* and other Apiaceae. *Am. J. Bot.* **95**, 783–804.
- Irmeler, S., Schroder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U. and Schroder, J. (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797–804.
- Jérôme, M., Noirot, C. and Klopp, C. (2011) Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool. *BMC Res. Notes*, **4**, 149–149.
- Kai, K., Mizutani, M., Kawamura, N., Yamamoto, R., Tamai, M., Yamaguchi, H., Sakata, K. and Shimizu, B.I. (2008) Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. *Plant J.* **55**, 989–999.
- Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., Hehn, A., Bourgaud, F. and Yazaki, K. (2014) A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *Plant J.* **77**, 627–638.
- Kawai, Y., Ono, E. and Mizutani, M. (2014) Expansion of specialized metabolism-related superfamily genes via whole genome duplications during angiosperm evolution. *Plant Biotechnol.* **31**, 579–584.
- Kliebenstein, D.J. and Osbourn, A. (2012) Making new molecules - evolution of pathways for novel metabolites in plants. *Curr. Opin. Plant Biol.* **15**, 415–423.
- Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., Bourgaud, F. and Matern, U. (2007) Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *J. Biol. Chem.* **282**, 542–554.
- Larbat, R., Hehn, A., Hans, J., Schneider, S., Jugdé, H., Schneider, B., Matern, U. and Bourgaud, F. (2009) Isolation and functional characterization of CYP71AJ4 encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. *J. Biol. Chem.* **284**, 4776–4785.
- Matsumoto, S., Mizutani, M., Sakata, K. and Shimizu, B. (2012) Molecular cloning and functional analysis of the ortho-hydroxylases of *p*-coumaroyl coenzyme A/feruloyl coenzyme A involved in formation of umbelliferone and scopoletin in sweet potato, *Ipomoea batatas* (L.) Lam. *Phytochemistry*, **74**, 49–57.
- McCloud, E.S., Berenbaum, M.R. and Tuveson, R.W. (1992) Furanocoumarin Content and Phototoxicity of Rough Lemon (*Citrus jambhiri*) Foliage Exposed to Enhanced Ultraviolet-B (Uvb) Irradiation. *J. Chem. Ecol.* **18**, 1125–1137.
- Miettinen, K., Dong, L.M., Navrot, N. et al. (2014) The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 11.
- Munakata, R., Olry, A., Karamat, F. et al. (2016) Molecular evolution of parsnip membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytol.* **211**, 332–344.
- Nützmann, H.W. and Osbourn, A. (2014) Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.* **26**, 91–99.
- Nützmann, H.W., Huang, A.C. and Osbourn, A. (2016) Plant metabolic clusters - from genetics to genomics. *New Phytol.* **211**, 771–789.
- Osbourn, A. (2010) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* **26**, 449–457.
- Osbourn, A., Papadopoulos, K.K., Qi, X.Q., Field, B. and Wegel, E. (2012) Finding and analyzing plant metabolic gene clusters. In *Natural Product Biosynthesis by Microorganisms and Plants*. Pt C. San Diego: Elsevier Academic Press Inc, pp. 113–138.
- Peterson, D.G., Tomkins, J.P., Frisch, D.A., Wing, R.A. and Paterson, A.H. (2000) Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. *J. Agric. Genomics*, **5**, 1–100.
- Pompon, D., Louerat, B., Bronnie, A. and Urban, P. (1996) Yeast expression of animal and plant P450s in optimized redox environments. *Methods Enzymol.* **272**, 51–64.

© 2016 The Authors

*The Plant Journal* © 2016 John Wiley & Sons Ltd, *The Plant Journal*, (2017), **89**, 1119–1132

- Ritchie, D., Ghoorah, A., Mavridis, L. and Venkatraman, V. (2012) Fast protein structure alignment using Gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics*, **28**, 3278–3281.
- Seigler, D. (1998) *Plant Secondary Metabolism*. New York: Springer Science+Business Media.
- Seki, H., Sawai, S., Ohyama, K. *et al.* (2011) Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell*, **23**, 4112–4123.
- Suay, L., Zhang, D., Eber, F. *et al.* (2014) Crossover rate between homologous chromosomes and interference are regulated by the addition of specific unpaired chromosomes in Brassica. *New Phytol.* **201**(2), 645–656.
- Szinay, D., Chang, S., Khrustaleva, L., Peters, S., Schijlen, E., Bai, Y., Stiekema, W., Van Ham, R., de Jong, H. and Klein Lankhorst, R. (2008) High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. *Plant J.* **56**, 627–637.
- Vialart, G., Hehn, A., Olry, A. *et al.* (2012) A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl coA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants. *Plant J.* **70**, 460–470.
- Wegel, E., Koumproglou, R., Shaw, P. and Osbourn, A. (2009) Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. *Plant Cell*, **21**, 3926–3936.
- Yu, N., Nuttmann, H.W., MacDonald, J.T. *et al.* (2016) Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res.* **44**, 2255–2265.
- Zwick, M.S., Hanson, R.E., Islam-Faridi, N., Stelly, D.M., Wing, R.A., Price, H.J. and McKnight, T.D. (1997) A rapid procedure for the isolation of C<sub>0</sub>t-1 DNA from plants. *Genome*, **40**, 138–142.





# The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants

Célia Krieger<sup>1†</sup>, Sandro Roselli<sup>1†</sup>, Sandra Kellner-Thielmann<sup>2†</sup>, Gianni Galati<sup>1</sup>,  
Bernd Schneider<sup>3</sup>, Jérémy Grosjean<sup>1</sup>, Alexandre Olry<sup>1</sup>, David Ritchie<sup>4</sup>, Ulrich Matern<sup>2</sup>,  
Frédéric Bourgaud<sup>5</sup> and Alain Hehn<sup>1\*</sup>

<sup>1</sup> Laboratoire Agronomie et Environnement, Institut National de la Recherche Agronomique, Université de Lorraine, Nancy, France, <sup>2</sup> Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Marburg, Germany, <sup>3</sup> Max Planck Institute for Chemical Ecology, Jena, Germany, <sup>4</sup> INRIA Nancy, Grand-Est Research Centre, Laboratoire Lorrain De Recherche En Informatique Et Ses Applications, Nancy, France, <sup>5</sup> Plant Advanced Technologies, Nancy, France

## OPEN ACCESS

### Edited by:

Danièle Werck,  
Centre National de la Recherche  
Scientifique (CNRS), France

### Reviewed by:

Alain Tissier,  
Leibniz-Institut für Pflanzenbiochemie  
(IPB), Germany  
May Berenbaum,  
Illinois Rocstar, University of Illinois at  
Urbana-Champaign, United States

### \*Correspondence:

Alain Hehn  
alain.hehn@univ-lorraine.fr

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Plant Metabolism  
and Chemodiversity,  
a section of the journal  
Frontiers in Plant Science

**Received:** 19 March 2018

**Accepted:** 28 May 2018

**Published:** 19 June 2018

### Citation:

Krieger C, Roselli S,  
Kellner-Thielmann S, Galati G,  
Schneider B, Grosjean J, Olry A,  
Ritchie D, Matern U, Bourgaud F and  
Hehn A (2018) The CYP71AZ P450  
Subfamily: A Driving Factor  
for the Diversification of Coumarin  
Biosynthesis in Apiaceous Plants.  
Front. Plant Sci. 9:820.  
doi: 10.3389/fpls.2018.00820

The production of coumarins and furanocoumarins (FCs) in higher plants is widely considered a model illustration of the adaptation of plants to their environment. In this report, we show that the multiplication of cytochrome P450 variants within the CYP71AZ subfamily has contributed to the diversification of these molecules. Multiple copies of genes encoding this enzyme family are found in Apiaceae, and their phylogenetic analysis suggests that they have different functions within these plants. CYP71AZ1 from *Ammi majus* and CYP71AZ3, 4, and 6 from *Pastinaca sativa* were functionally characterized. While CYP71AZ3 merely hydroxylated esculetin, the other enzymes accepted both simple coumarins and FCs. Superimposing *in silico* models of these enzymes led to the identification of different conformations of three regions in the enzyme active site. These sequences were subsequently utilized to mutate CYP71AZ4 to resemble CYP71AZ3. The swapping of these regions lead to significantly modified substrate specificity. Simultaneous mutations of all three regions shifted the specificity of CYP71AZ4 to that of CYP71AZ3, exclusively accepting esculetin. This approach may explain the evolution of this cytochrome P450 family regarding the appearance of FCs in parsnip and possibly in the Apiaceae.

**Keywords:** coumarin, cytochrome P450, furanocoumarin, *Pastinaca sativa*, psoralen, specialized metabolism

## INTRODUCTION

Plants have developed a wide range of strategies to take advantage of their environment and to adapt to many conditions by methods such as the production of defense molecules. This process has been studied for many decades (Fraenkel, 1959), but the molecular characterization of the enzymes responsible for the synthesis of the defense molecule arsenal primarily started in the 1990s (Funk and Croteau, 1993; Werck-Reichhart et al., 1997; Glawischnig et al., 1999). Beside dioxygenases (Kawai et al., 2014), the cytochrome P450s (P450s) are the most diversified enzyme family in plants (Nelson et al., 2004; Nelson and Werck-Reichhart, 2011) and are estimated to represent up to 1% of the annotated genes in plant genomes (Mizutani and Ohta, 2010). Because of

their diversity, these enzymes are considered major players in boosting metabolic pathway activity and therefore defense molecules (Glawischneg et al., 1999; Irmisch et al., 2014). Thanks to the development of next-generation sequencing methods, the number of transcriptomic databases has exploded. However, although the number of annotated P450 sequences is increasing day by day, the precise P450 functions often remain unknown. For example, more than 70% of the 245 P450s in the *Arabidopsis* genome, which is smaller than many other plant genomes, are orphans at the biochemical level (Mizutani and Sato, 2012).

Furanocoumarins (FCs) are molecules involved in defense reactions against attacks by herbivores (Berenbaum et al., 1991; Li et al., 2004). They have been detected in only a few plant families, and these families are not phylogenetically related. Linear (psoralens) and angular FCs may have arisen by convergent evolution (Bourgaud et al., 2014) and are distinguished by the position of the furan-ring on the coumarin core molecule (Figure 1). Plants first acquired the ability to produce linear isomers, and the formation of the angular analogs evolved later (Berenbaum, 2002). No plant that can synthesize only angular molecules has been described. Elicitor treatment of *A. majus* cell cultures increased the synthesis of FCs and the conversion of psoralen to 5-hydroxypsoralen (bergaptol) (Hamerski and Matern, 1988a), and the use of radiolabeled precursors assigned the enzymatic conversion of demethyl [3-<sup>14</sup>C]suberosin into labeled (+) marmesin (Hamerski and Matern, 1988b) to P450s. However, only a few of these enzymes have been isolated and identified. The first P450 entity that was characterized at the molecular level catalyzes the synthesis of psoralen from (+)-marmesin and was assigned to the CYP71AJ P450 subfamily (Larbat et al., 2007). A variant in parsnip named CYP71AJ4 and sharing 70% amino acid identity with psoralen synthases did not accept marmesin as a substrate and was the first P450 shown to be involved in angular FC biosynthesis, converting (+)-columbianetin to angelicin (Larbat et al., 2009) (Figure 1). The assignment of two different P450s to the same subfamily but restricted to only one biosynthetic pathway each suggested that these enzymes evolved from a common ancestor (Larbat et al., 2009). A CYP71AJ-targeted search in various Apiaceae transcriptomic databases identified 36 partial or full-length paralogous genes distributed in 11 different apiaceous plants that either produce or lack FCs (Dueholm et al., 2015). A phylogenetic analysis of 19 of these CYP71AJ sequences revealed a clustering of 5 enzymes that have been described to be involved in the synthesis of FCs (Larbat et al., 2007, 2009; Dueholm et al., 2015). The functions of the 14 remaining enzymes have not been elucidated. This analysis suggests that the CYP71AJ P450 family might have evolved to synthesize these toxic defense molecules. An ancestral version of these enzymes might have been identified in plants that do not produce FCs (Dueholm et al., 2015). The spread of this gene family probably occurred through gene duplication associated with neofunctionalization. Such a mechanism has been described, for example, in *Arabidopsis* and for the glucosinolate biosynthesis pathway (Kliebenstein et al., 2001; Benderoth et al., 2006; Kliebenstein and Osbourn, 2012; Edger et al., 2015).

Here, we describe the identification and the functional characterization of several members of the CYP71AZ P450

subfamily. CYP71AZ1 was isolated by using a differential display RT-PCR approach applied to *A. majus* *in vitro* cultured cells that were treated with *Phytophthora sojae* extracts. CYP71AZ1 shares 33.8% identity with the psoralen synthase CYP71AJ3. BLAST searches performed on transcriptomic and genomic databases led to the identification of numerous paralogous genes that interestingly (and except one) were restricted to Apiaceae. In this report, we focused on four genes that were identified in parsnip and share 63 to 85% identity with each other. An *in vivo* investigation revealed that these four genes have different expression patterns in wounded parsnip roots. An *in vitro* functional characterization using proteins expressed in yeast showed that these enzymes are involved in the synthesis of both FCs and coumarins. A molecular shuffling of regions that change the catalytic sites of CYP71AZ4 allowed us to transform a FC/coumarin non-specific enzyme to a coumarin-specific enzyme. This study provides evidence that this multigenic P450 subfamily contributed to the diversification of coumarins and provides new insight into the evolution of higher plants.

## MATERIALS AND METHODS

### *Ammi majus* Culture and Isolation of CYP71AZ1

Total RNA and differential amplification of 3'-fragments of P450 cDNAs from elicited or water-treated (control) *A. majus* cell cultures was done as already described previously (Larbat et al., 2007).

### Parsnip Culture

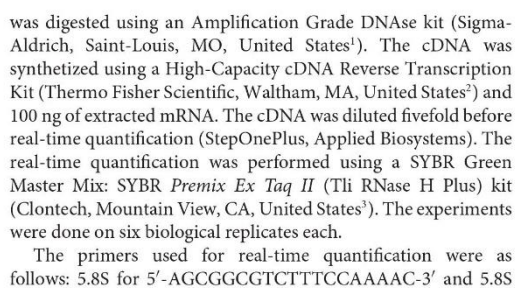
Parsnips (*Pastinaca sativa* L. subsp. *sativa* "Demi Long de Guernesey") were cultivated for 2 months as described by Munakata et al. (2016). Leaves and roots were separated from six plants used as controls and another six that endured the mechanical elicitation treatment as described by Roselli et al. (2016). Seeds were germinated in a growth chamber in soil under regulated climatic conditions, namely, temperature: 20/24°C; hygrometry: 40/70%; and light: 16 h/8 h day/night. Then, germinated seeds were transplanted into soil pots at the two leaf stage for an additional 2 months growth with the same climatic conditions.

### Wounding Induction

Plants were removed from the soil, and the roots were carefully washed with distilled water and dried. Wounding was done uniformly with a metal tip on the surfaces of the leaves, roots, and stems. Wounded tissues were frozen in liquid nitrogen and crushed with a mortar and a pestle prior to polyphenol extraction.

### Real-Time Quantitative PCR of the Target Gene Transcripts

Total RNA extraction from plants cultivated in soil was performed according the E.Z.N.A. Plant RNA kit protocol (Omega Bio-Tek, Norcross, GA, United States). Remaining DNA



rev 5'-ATATCTCGGCTCTCGCATCG-3' for the house-keeping gene 5.8S; CYP71A3rev 5'-TGACGACCGTAGAGCTTGCA-3' and CYP71A23rev 5'-ATTGGGAACCACTGCTGG-3' for CYP71A23; CYP71A4rev 5'-GGAAAAAGAAATCCAA CGAATG-3' and CYP71A4rev 5'-TGTTGCTGCCTTTT TTGTTTC-3' for CYP71A24; CYP71A26rev 5'-GTTTGGG CGATGACAGCA-3' and CYP71A26rev 5'-TGTCTTCCTG GGCTTTCTTCA-3' for CYP71A26; and CYP71A13rev 5'-AA CAATAGCGACGGCTTTGG-3' and CYP71A13rev 5'-TTAAAA TCCGACGCGCATG-3' for CYP71A13.

The CYP71AZ1 open reading frame was PCR-amplified using a high-fidelity DNA polymerase (PrimeSTAR Max;



Clontech) and primers encoding an additional 6xHis tag at the 3' end (CYP71AZ1 for 5'-ATGCAGATGGATGCAGTAGTATCCTTCTTATTC-3' and CYP71AZ1rev 5'-TTAGTGGTGA TGGTGATGATGTGAAACATATATG-3'). PCR conditions were as follows: 5 min at 98°C, 30 cycles (10 s at 98°C, 10 s at 55°C, and 120 s at 72°C), and a final 5 min extension step at 72°C. PCR products were purified, cloned using a pCR®8/GW/TOPO® TA Cloning® Kit (Invitrogen®), and sequenced. The resulting PCR product cloned in pCR®8/GW/TOPO® was sequenced and then subcloned in the yeast expression vector pYedP60-GW (Dueholm et al., 2015) using LR clonase II (Invitrogen).

### Cloning of CYP71AZ3, CYP71AZ4, and CYP71AZ6 Open Reading Frames

Total RNA was extracted from the leaves of parsnip 6 h after mechanical wounding (Larbat et al., 2009; Roselli et al., 2016) with the RNeasy plant extraction kit (Qiagen). The cDNA was generated using a High-Capacity RNA-to-cDNA system (Applied Biosystems) and random primers. The full-length cDNA sequence was then PCR-amplified using Platinum® Taq DNA Polymerase High Fidelity (Invitrogen) and primers with additional restriction sites and a 6xHis tag sequence at the 3' end: CYP71AZ3 for 5'-GGATCCATGGAGCCAGTATTTCTCTTTC-3', CYP71AZ3rev 5'-GAATTCTTTAGTGGTGATGGTGATGTGGAACATATATTTTATTAGG-3', CYP71AZ4 for 5'-GGTACCATGGATCCAGCAGCTATC-3', CYP71AZ4rev 5'-GAATTCTTTAGTGGTGATGGTGATGTGGATGTACATATATTTTATTAGG-3', CYP71AZ6 for 5'-ATGGATCCAGTAGTTATCTTTCTTGCTTGCCTT-3', and CYP71AZ6rev 5'-TTATGGACATATATATTTTATTAGGTCGAATGTAG-3'. PCR conditions were as follows: 2 min at 94°C, 35 cycles (30 s at 94°C, 30 s at 55°C, and 2 min at 68°C), and a final 10 min extension step at 68°C. PCR products were purified, cloned using a pCR®8/GW/TOPO® TA Cloning® Kit (Invitrogen), and sequenced. The resulting PCR products cloned in pCR®8/GW/TOPO® were sequenced and finally subcloned in the yeast expression vector pYedP60-GW (Dueholm et al., 2015) using LR clonase II (Invitrogen).

### Expression in a Yeast System

Recombinant plasmids and an empty plasmid (negative control) were used to transform the *S. cerevisiae* strain WAT11, which expresses the *Arabidopsis thaliana* NADPH P450 reductase 1 (Pompon et al., 1996), except for pYedP60-GW::CYP71AZ1, which was used to transform the *S. cerevisiae* strain WAT21, which expresses the *A. thaliana* NADPH P450 reductase 2. Microsome preparation and enzyme assays were conducted as described previously (Larbat et al., 2007). The molecules tested are listed in **Supplementary Figure S1**. The catalytic parameters are the means of three independent replicates, and errors represent standard deviations. The kinetic parameters were calculated using the SigmaPlot 12 program (Systat Software Inc.).

### Metabolite Analysis

The roots of 2-month-old *P. sativa* var. Demi Long de Guernesey were harvested, frozen in liquid nitrogen and ground in a mortar

with a pestle. The protocol to prepare the samples and analyze the FC content by ultrahigh performance liquid chromatography (UHPLC)/MS was derived from Munakata et al. (2016).

### LTQ Orbitrap

The analytical column was a Grace-Discovery Alltima C18 (150 mm × 4.6 mm, 5 μm) reversed-phase column. The mobile phases used for the separation consisted of distilled water with 0.1% formic acid (A) and methanol with 0.1% formic acid (B). The samples were analyzed in the following gradient mode: from 0 to 1 min, 10%B, increasing B concentration to 70% at 20 min and 100% at 22 min. The mobile phase was held constant at 100%B between 22 and 25 min and then back at beginning conditions during 45 min. All samples were injected with a volume of 25 μL, and the mobile phase flow rate was kept at 0.2 mL/min. The HPLC/MS system was composed of a quaternary solvent delivery pump and linear ion trap mass spectrometer (LTQ-MS, Thermo Electron Corporation, Waltham, MA, United States) coupled with an Orbitrap HRMS (Thermo Electron Corporation, Waltham, MA, United States). The LTQ-Orbitrap was equipped with an electrospray ionization (ESI) interface operating in positive ion mode. The orbitrap was used in scan mode between 80 and 400 m/z, and the 10 most intense selected masses were analyzed in MS/MS using LTQ fragmentation.

### Ultrahigh Performance Liquid Chromatography-MS

The analytical column was a Phenomenex Kinetex XB-C18 (150 mm × 2.1 mm, 2.6 μm) reversed-phase column. The mobile phases used for the separation consisted of distilled water with 0.1% formic acid (A) and methanol with 0.1% formic acid (B). The samples were analyzed in gradient mode as follows (A/B; v/v): 90:10 at 0 min, 80:20 at 0.74 min, 40:60 at 5.88 min, 10:90 at 10 min, 0:100 between 12 and 16 min, and 90:10 from 16.01 to 20 min. All samples were injected with a volume of 3 μL, and the mobile phase flow rate was kept at 0.2 mL/min. The column was kept at 40°C during the runtime. The UHPLC/MS system was composed of a Shimadzu Nexera UHPLC (pump, automatic sampler, and PDA analyzer, Shimadzu Corp, Kyoto, Japan) coupled with a Shimadzu LCMS2020 single quadrupole mass detector (Shimadzu Corp., Kyoto, Japan). MS was used with an ESI interface operating in positive ion mode.

### Nuclear Magnetic Resonance Spectroscopy

<sup>1</sup>H NMR, <sup>1</sup>H-<sup>1</sup>H COSY, and HSQC spectra were recorded on an Avance III HD 700 NMR spectrometer (Bruker-Biospin, Karlsruhe, Germany) equipped with a 1.7 mm TCI microcryoprobe. Spectra were measured at 300 K and referenced to the residual solvent signals of MeOH- $\delta_4$  ( $\delta_H$  3.31 and  $\delta_C$  49.15) and acetone- $\delta_6$  ( $\delta_H$  2.05 and  $\delta_C$  29.92). Coupling constants are in Hertz (Hz). Data acquisition and processing were accomplished using TopSpin 3.2. Standard pulse programs as implemented in TopSpin were used for data acquisition.

## Modeling

A library of 1,023 three-dimensional (3D) structures of CYP450 domains (Pfam code PF00067) was extracted from the KBDock database (Ghoorah et al., 2014), which had been built from a June 2013 snapshot of the Protein Databank (PDB). A sequence-based search of these structures using version 3.0 of the Kpax program (Ritchie et al., 2012) yielded the PDB structure 4EJI (P450 2A13) as the closest sequence homolog to CYP71AZ (39% sequence identity). Four similar structures (PDB codes 3QU8, 3T3Z, 3TK3, and 2NNJ) having no more than 90% sequence similarity to 4EJI were then extracted from the library using a structure-based search in Kpax. A multiple structure alignment of the first selected structures was built with Kpax, and the resulting multiple sequence alignment profile was used as input to the MODELLER (v9.13) homology modeling program (Sali and Blundell, 1993). Ten structural models were built using default MODELLER parameters, and the structure with the best DOPE score was retained as the final model of the CYP71AZ structure.

## In Silico Data Mining

The presence of various CYP71AZs was investigated, and their sequences were collected in transcriptomic and genomic databases [*Angelica archangelica* (onekp), *Apium graveolens* (SRR1023730 and DRR003696), *Bupleurum chinense* (SRR1002958), *Bupleurum scorzonrifolium* (SRR1002959), *Centella asiatica* (Phytometasyn), *Cimicifuga racemosa* (onekp), *Coriandrum sativum* (SRR1700630, SRR1700819, and SRR1700873), *Daucus carota* (SRR187755-1<sup>4</sup>), *Oenanthe javanica* (SRR1119291), *Petroselinum crispum* (SRR1030614), *P. sativa* (Dueholm et al., 2015; Roselli et al., 2016), *Thapsia garganica* (Henrik Toft Simonsen), and *Conium maculatum* (Heiko Rischer and (Hotti et al., 2015))]. Sequences are available in **Supplementary Figure S10**. The single reads of Illumina and 454 databases were assembled using the CLC genomic workbench (Qiagen).

The protein sequences were compared using Clustal X, and phylogenetic trees were constructed using the maximum likelihood (ML) methods included in MEGA7.

## RESULTS

### Identification of a Novel P450 Subfamily: CYP71AZ

Various studies have reported the pivotal role of P450 enzymes in the synthesis of FCs (Hamerski and Matern, 1988a,b; Larbat et al., 2007, 2009). To identify new genes involved in the synthesis of these molecules, we used a differential expression strategy that has been described previously (Larbat et al., 2007). A reverse transcription differential display approach using primers targeting a conserved P450 sequence identified a partial gene that was expressed more in *A. majus* cell

cultures elicited with crude *Phytophthora* cell wall elicitors than non-elicited cells. The full-length sequence was cloned using rapid amplification of cDNA ends (RACE) experiments, and the deduced protein sequence was submitted to the international P450 nomenclature committee and assigned as CYP71AZ1, the first member of a novel P450 subfamily (GenBank ABO32529.1).

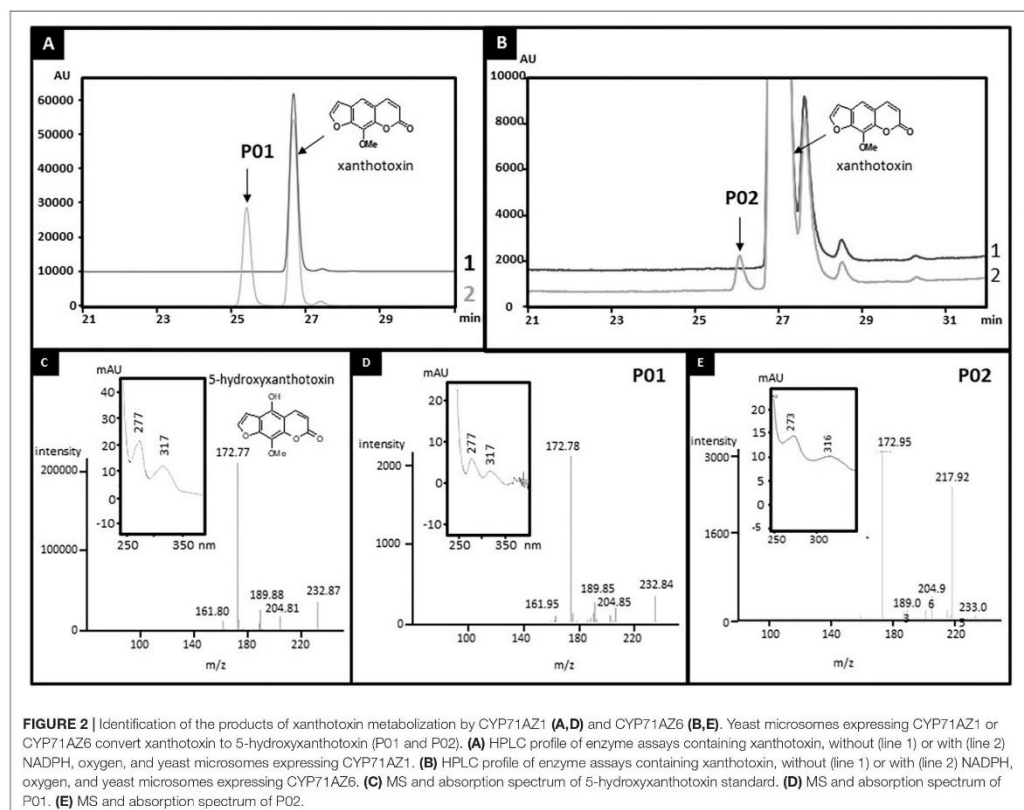
CYP71AZ1 was expressed in a heterologous yeast expression system (Pompon et al., 1996) for functional characterization. A metabolic screening done with microsomes prepared from the yeast cells overexpressing this enzyme in the presence of FC pathway intermediates (**Figure 1**) showed the transformation of xanthotoxin into a single product, P01 (**Figure 2A**). Comparison with a set of standard compounds revealed that the elution time in HPLC and the UV spectrum of P01 matched those of 5-hydroxyxanthotoxin (**Figure 2C**), which is a precursor of isopimpinellin. The identity of the molecule was further confirmed by mass spectrometric analysis (**Figure 2D**). The apparent affinity for xanthotoxin ( $K_m = 13.1 \pm 3.2 \mu\text{M}$ ) is similar to the substrate affinities of most P450s involved in the synthesis of secondary metabolites. The isolation of CYP71AZ1 paved the way to a new class of hydroxylases relevant to the metabolism of FCs.

### CYP71AZ: A Multigenic P450 Subfamily

Databases were searched for the presence of CYP71AZ paralogs to assess the occurrence and diversity of this P450 subfamily. Coding sequences sharing high homologies with CYP71AZ1 are fairly restricted to a few spermatophytes and, except for *C. racemosa* (Ranunculaceae), were specifically reported in apiaceous plants such as *A. archangelica*, *P. sativa*, *D. carota*, *Heracleum lanatum*, *A. graveolens*, *B. chinense*, *B. scorzonrifolium*, *C. asiatica*, *C. sativum*, *O. javanica*, *P. crispum*, *C. maculatum*, and *T. garganica*. This search revealed 17 partial and full-length sequences sharing more than 55% identity and therefore belonging to the CYP71AZ subfamily (Werck-Reichhart et al., 2000). Among these 17 genes, 14 were identified in Apiaceae members known to produce FCs. However, some of them were also identified in transcriptomes of plant species that do not produce FCs, i.e., *Thapsia* spp. (Apiaceae). This result suggested that the substrate specificities of CYP71AZ-related enzymes might not be restricted to xanthotoxin (as for CYP71AZ1) and could instead provide a wider range of functions. This assumption is reinforced through a phylogenetic analysis of P450s belonging to Clan71, which shows that 10 CYP71AZ genes, including CYP71AZ1, are grouped in a cluster that were only identified in FC-producing plants (**Figure 3**). The genes identified in plants that do not produce FCs are localized out of this group. To investigate the blooming of this cluster, we further focused on *P. sativa*, which harbors four different genes. Three of them are included in the cluster and one is outside. CYP71AZ3, 4, 5, and 6 share 81.1, 71.1, 76.6, and 85.4% protein sequence identity with *A. majus* CYP71AZ1, respectively (**Supplementary Table S1**). GenBank entries for CYP71AZ3, 4, 5, and 6 are MH000218, MH000219, MH000220, and MH000221, respectively.

<sup>4</sup><http://apiaceae.njau.edu.cn>





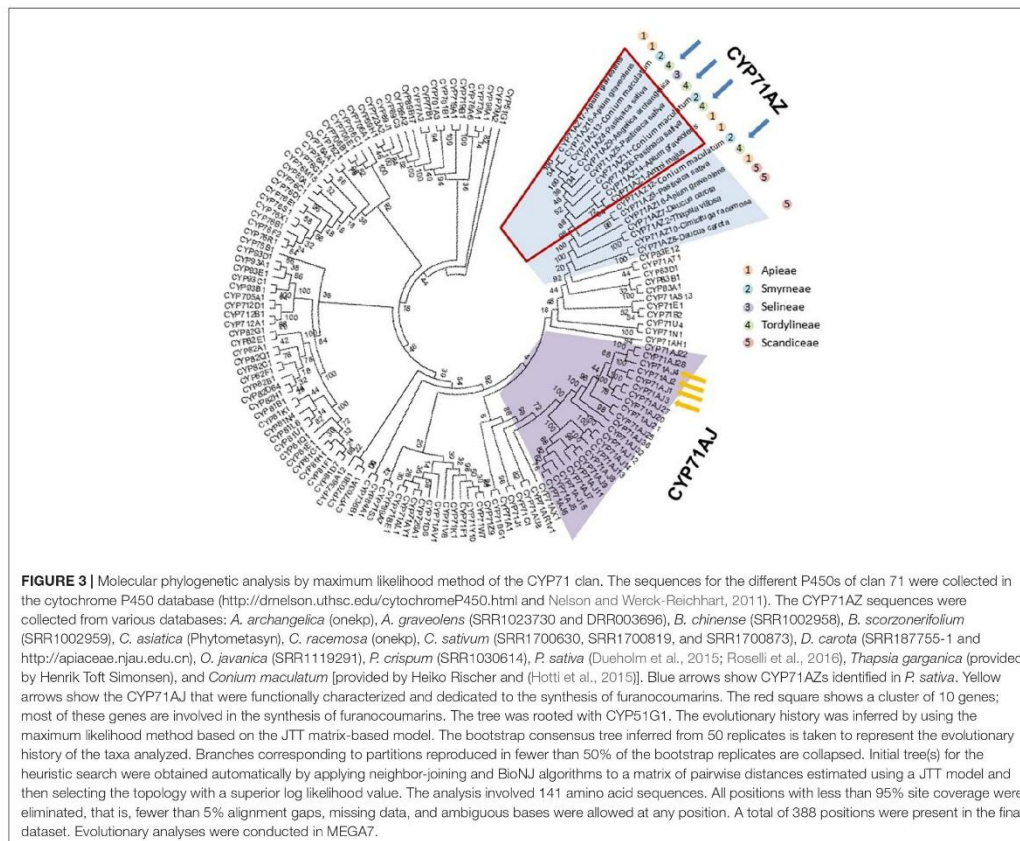
### In Situ Expression of CYP71AZ in Parsnip Plants Suggests Their Role in the Synthesis of FCs

The metabolism of FCs in plants is influenced by environmental factors. Assessing the relationship between the concentration of FCs and the expression level of these different genes in *P. sativa* plants upon mechanical wounding was therefore relevant. In agreement with previously reported data (Roselli et al., 2016), the global FC concentration was higher in wounded roots than in unwounded roots (Figure 4A). The expression levels of the 4 CYP71AZ mRNAs were assessed and compared to the expression level of CYP71AJ3, which is involved in the synthesis of FCs in *P. sativa* (Larbat et al., 2007). RT-qPCR experiments (Figure 4B) highlighted different patterns of transcription for the four genes but these differences were not statistically relevant. The differential expression level of CYP71AJ3 was already demonstrated by Roselli et al. (2016). The tendency of the expression profiles of CYP71AZ4 and CYP71AZ6 matched that of CYP71J3. In contrast, the expression pattern of CYP71AZ3 was not modified in wounded plants. Although the coding

sequence of CYP71AZ5 is characterized in a genomic library (Roselli et al., 2016), we failed to amplify a full-length coding sequence corresponding to this gene. We concluded that this gene is not expressed in *P. sativa* tissues under our experimental conditions. Although not absolute proof, the diversity of the expression pattern of CYP71AZ genes suggests different roles in plant metabolism for the corresponding enzymes.

### Functional Characterization of CYP71AZs

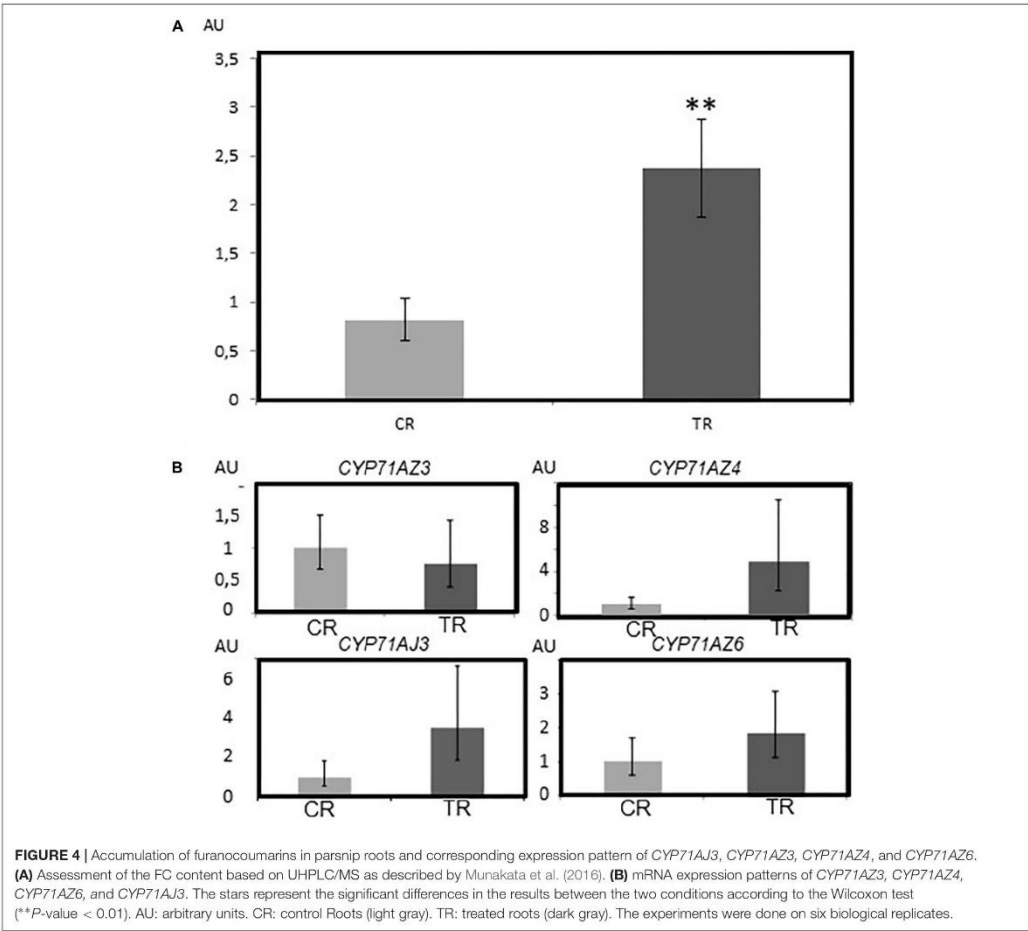
CYP71AZ3, CYP71AZ4, and CYP71AZ6 were expressed in yeast cells, and their enzymatic activities were assayed using different intermediates of the FC pathway as substrates (Figure 1). CYP71AZ6 displayed the same enzymatic activity as CYP71AZ1, transforming xanthotoxin to 5-hydroxyxanthotoxin (product P02) with an apparent  $K_m$  of  $6.3 \pm 1.2 \mu M$  (Figures 2B,E and Table 1). This result was expected since CYP71AZ1 and CYP71AZ6 are very close from a phylogenetic point of view (Figure 3). As for CYP71AZ4, the enzymatic data provided evidence that this prefers psoralen as a substrate (Figure 5A),



with a substrate affinity ( $K_m$ ) of  $11.3 \pm 2.2 \mu\text{M}$  (Table 1), and mass spectrometric evidence confirmed hydroxylation at the C8 position of the coumarin core structure occurred from xanthotoxol (P03; Figures 5B,C). Taken together, these first elements amend our knowledge of the psoralen metabolism: CYP71AZ4 catalyzes the step preceding *O*-methylation, and CYP71AZ1 and 6 perform the subsequent hydroxylation en route to isopimpinellin (Figure 1).

In this first set of assays with FC substrates, we could not ascribe any enzyme activity to CYP71AZ3. The metabolic screening for substrates was therefore extended to simple coumarins (Supplementary Figure S1). Some coumarins are precursors and represent the core of FCs. This additional screening was done on all the CYP71AZs. The extended screening revealed that at least five coumarins beside psoralen served as substrates for CYP71AZ4 (Supplementary Figures S2–S4), although the kinetic constants determined *in vitro* for some of these compounds were higher than commonly expected for P450s and appear to preclude an *in situ*

role as substrates. 6-Methoxycoumarin was converted to two products (Supplementary Figure S2A), one of which (P04) was identified as scopoletin (Supplementary Figures S2B,C,  $K_m = 9.5 \pm 1.5 \mu\text{M}$ , Table 1), while the other (P05) remained unassigned but showed a maximal UV absorption shift at shorter wave length than its precursor (Supplementary Figure S2D,  $K_m = 44.8 \pm 16.3 \mu\text{M}$ , Table 1). The CYP71AZ4-encoded hydroxylase slowly converted scopoletin to a product (P06) identified as fraxetin (Supplementary Figures S3A–C,  $K_m = 762.2 \pm 586.4 \mu\text{M}$ , Table 1). Moreover, 3-methyl-7-methoxycoumarin and 7-methoxycoumarin also served as substrates for CYP71AZ4, yielding product P07 (Supplementary Figure S4A) and P08 (Supplementary Figure S4B), respectively, which were recognized as 8-hydroxy-3-methyl-7-methoxycoumarin (Supplementary Figures S4C, S5) and daphnetin-7-methyl ether (Supplementary Figures S4D,E), respectively. However, the substrate affinities for both compounds are fairly high ( $K_m = 159.9 \pm 41.1$  and  $72.5 \pm 15.2 \mu\text{M}$ , Table 1) and probably

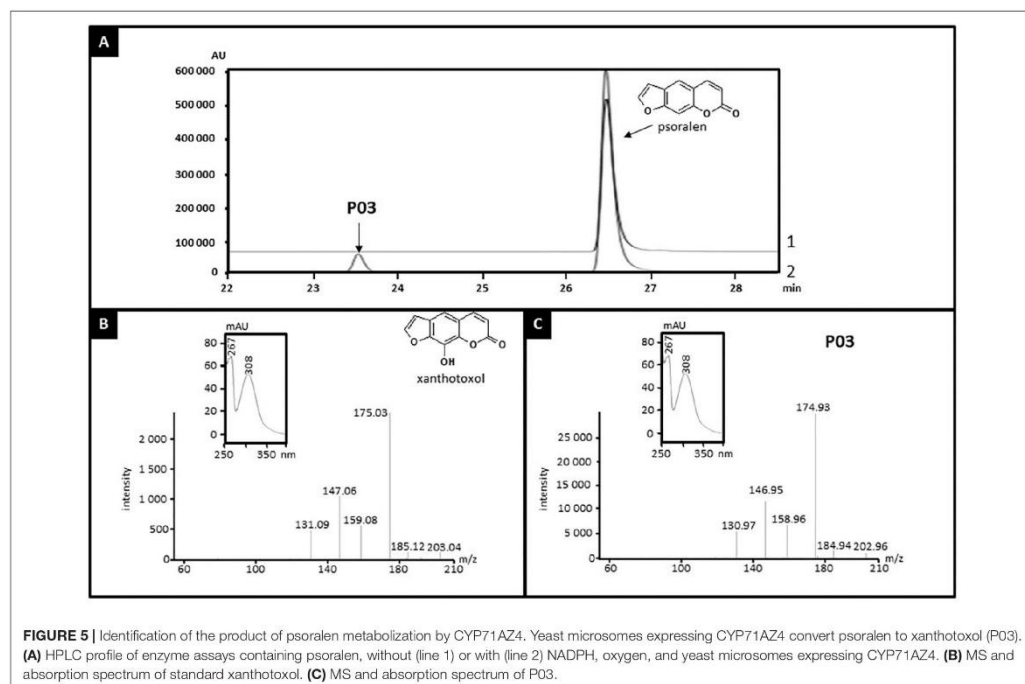


**TABLE 1 |** Metabolization pattern and kinetic parameters for CYP71AZs.

Substrate	Product	Apparent affinity (μM)			
		CYP71AZ1	CYP71AZ3	CYP71AZ4	CYP71AZ6
Xanthotoxin	5-hydroxyxanthotoxin	P01: 13.1 ± 3.2	N.D.	N.D.	P02: 6.3 ± 1.2
Psoralen	Xanthotoxol	N.D.	N.D.	P03: 11.3 ± 2.2	N.D.
6-Methoxycoumarin	Scopoletin	P11: N.Q.	N.D.	P04: 9.5 ± 1.5	P12: N.Q.
	<i>m/z</i> = +16	N.D.	N.D.	P05: 44.8 ± 16.3	N.D.
Scopoletin	Fraxetin	N.D.	N.D.	P06: 762.2 ± 586.4	N.D.
7-Methoxy-3-methylcoumarin	<i>m/z</i> = +16	N.D.	N.D.	P07: 159.9 ± 41.1	N.D.
7-Methoxycoumarin	Daphnetin-7-methylether	P09: N.Q.	N.D.	P08: 72.5 ± 15.2	P10: N.Q.
Esculetin	<i>m/z</i> = +16	N.D.	P13: 248.6 ± 51.9	N.D.	N.D.

Reactions were set up in KPi buffer (pH 7.4) at 28°C for 10 min and stopped by the addition of 10 μL of 0.1 M HCl. Products were quantified after HPLC separation at 330 nm. N.D., not detected; N.Q., not quantifiable.





**FIGURE 5 |** Identification of the product of psoralen metabolism by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert psoralen to xanthotoxol (P03). **(A)** HPLC profile of enzyme assays containing psoralen, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. **(B)** MS and absorption spectrum of standard xanthotoxol. **(C)** MS and absorption spectrum of P03.

also preclude any *in situ* relevance of these *in vitro* data. The mass spectrum of P07 showed a molecular ion peak at  $m/z$  207  $[M-H]^+$  (Supplementary Figure S4C), which agrees with a molecular formula of  $C_{11}H_{10}O_4$ , indicating 16 mass units more than the mass of the substrate, 3-methyl-7-methoxycoumarin. Thus, in accordance with incubation with CYP71AZ4, the product P07 was hypothesized to possess an additional oxygen atom. Subsequent NMR analysis showed that P07 was 8-hydroxy-3-methyl-7-methoxycoumarin (Supplementary Figure S5).

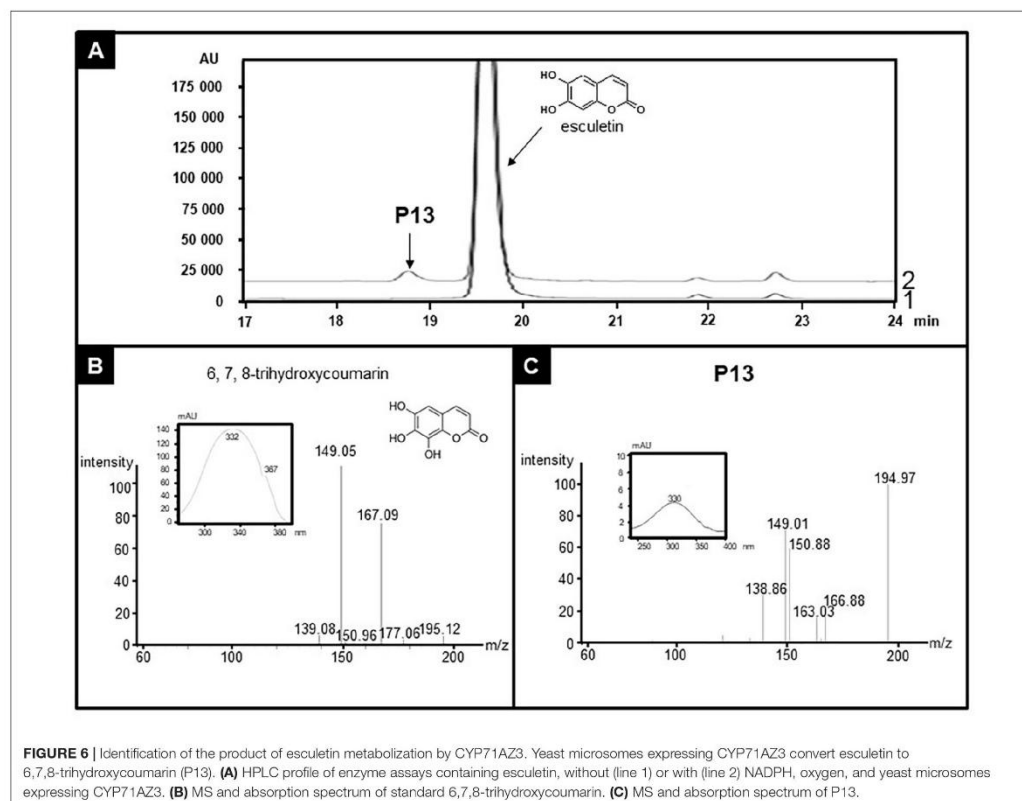
Extended enzyme assays also disclosed some side activities for CYP71AZ1 and CYP71AZ6. In addition to xanthotoxin, both enzymes hydroxylated 7-methoxycoumarin with very poor efficiencies to 5-hydroxy-7-methoxycoumarin (P09 or P10, Supplementary Figures S6A–E). Furthermore, both enzymes also slowly metabolized 6-methoxycoumarin to products P11 and P12 (Supplementary Figures S7A,C). We were not able to measure the catalytic parameters for this set of experiments with 7-methoxycoumarin and 6-methoxycoumarin as substrates. The products share their MS profiles (Supplementary Figures S7B,D) but could not be identified by NMR because of their limited availability. Finally, the screening also revealed a substrate for CYP71AZ3, the gene that did not respond to the wounding of *P. sativa* plants. This enzyme exclusively accepted esculetin among all the compounds included in the assays (Figure 6A) and formed (Table 1) 6,7,8-trihydroxycoumarin with a poor turnover

rate ( $K_m = 248.6 \pm 51.9 \mu M$ ). This product was identified by absorption spectroscopy as well as MS analyses (Figures 6B,C and Supplementary Figure S8). The mass spectrum of P13 showed a molecular ion peak at  $m/z$  195  $[M-H]^+$  (Figure 6C), which is in agreement with a molecular formula of  $C_9H_6O_5$ , indicating 16 mass units more than the mass of the substrate, esculetin. Thus, in accordance with incubation with CYP71AZ3, the product P13 was hypothesized to possess an additional oxygen atom. The structure of 6,7,8-trihydroxycoumarin was further confirmed by NMR (Supplementary Figures S8, S9).

The enzymatic and analytical data underline the significance of the CYP71AZ subfamily of enzymes for the overall synthesis of coumarins and FCs (Table 1) and adds several pieces to the FC pathway puzzle. Surprisingly, CYP71AZ4 displayed a broad range of substrate specificities *in vitro* and may be involved in multiple branches of the coumarin biosynthesis *in situ*. This result is noteworthy because the other CYP71AZs catalyze more well-defined steps in these pathways.

### Identification of Prominent Regions in the Catalytic Sites of CYP71AZ4 and CYP71AZ3

Alignments of the four complete CYP71AZ peptide sequences identified from parsnip display a moderate degree of conservation (Figure 7). To better understand the

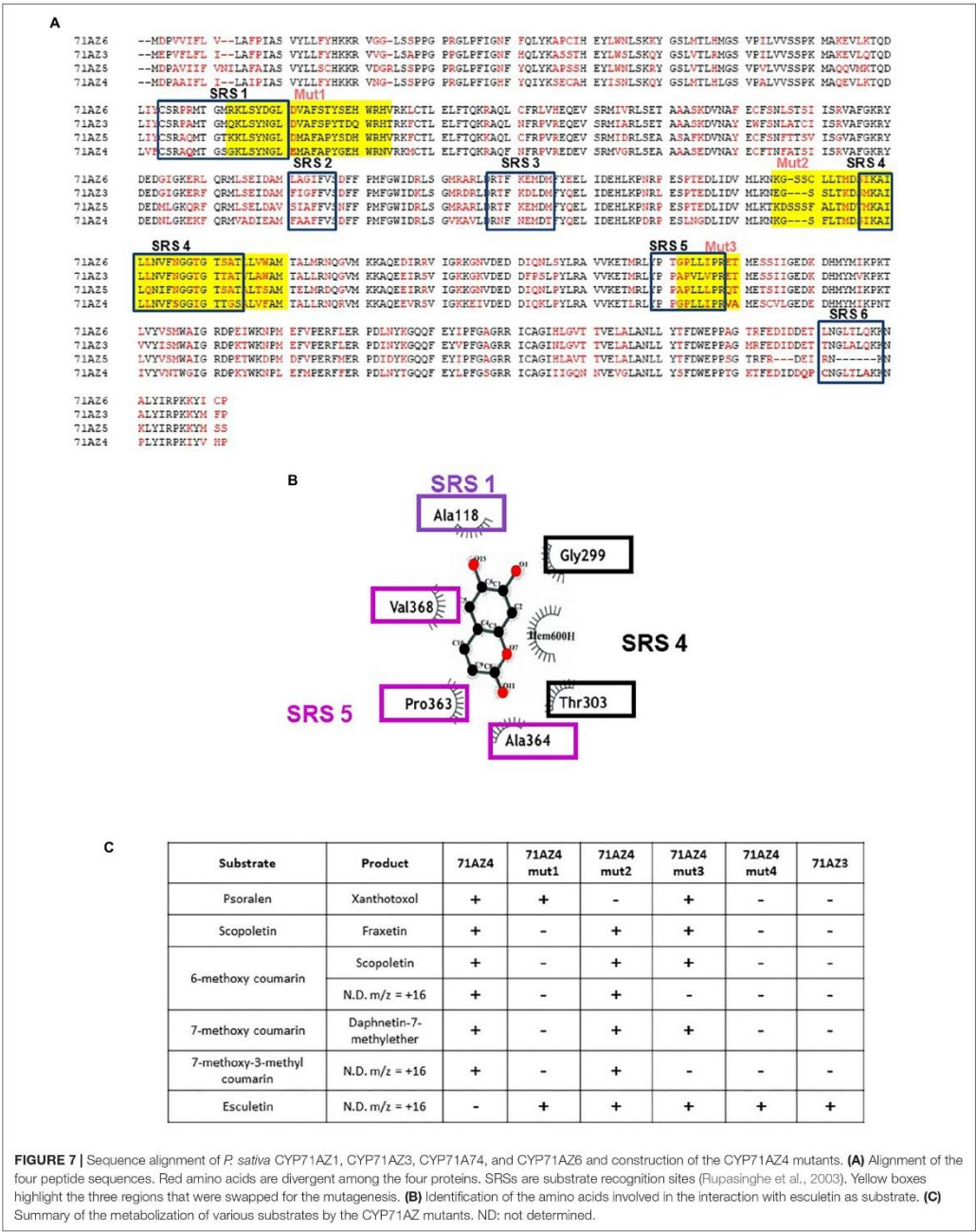


structure/activity relationship of these enzymes, we compared the highly specific for coumarin enzyme CYP71AZ3 and the broadly specific for coumarin/FC enzyme CYP71AZ4. An *in silico* model of CYP71AZ3 was constructed with esculetin fitted in the active site. This model was superimposed with sequences of the other CYP71AZs using the “Kpax” software (Ritchie et al., 2012) and led us to identify several regions that clearly differ in their active site conformations. These regions include substrate recognition sites (SRSs) (Rupasinghe et al., 2003) SRS1, SRS4 (alpha helix), and SRS5 (beta turn) (Figure 7B). Consequently, three CYP71AZ4 mutants were constructed (mut1-3, Figure 7C) to corroborate the involvement of these regions in the differential substrate specificities of CYP71AZ4 and CYP71AZ3. In these mutants, either one of these regions of wild-type CYP71AZ4 was replaced by the corresponding CYP71AZ3 peptide sequences (Nucleotide sequence 318-386 for mutant 1; 837-936 for mutant 2; and 1089-1115 for mutant 3). A fourth mutant was also constructed, hosting the three modifications in a single protein. The enzymatic characterization of the mutants expressed in yeast revealed four different substrate specificities (Figure 7C). The swapping of the mut2 sequence most likely affected the

volume of the active site pocket since psoralen was no longer metabolized. The replacement of SRS1 (mut1) narrowed the substrate specificity to solely target esculetin (in addition to psoralen), which is not even a substrate for wild-type CYP71AZ4. The exchange of SRS5 (mut3) caused an intermediate type of substrate specificity. Finally, the simultaneous swapping of the three regions (mut4) resulted in a mutant CYP71AZ4 whose specificity was identical to that of CYP71AZ3 (Figure 7C). In summary, the data provided unequivocal evidence that the SRS regions of CYP71AZs are intimately involved in the specificities of these enzymes and confirm the relevance commonly assigned to SRSs in cytochrome P450s.

## DISCUSSION

The adaptation of plants to changed environmental conditions often requires the creation of modified or novel enzymatic capabilities, and the P450 superfamily is particularly susceptible to such evolutionary processes as confirmed by the vast array of reactions catalyzed by this multigene family (Mizutani and Ohta,





2010; Mizutani and Sato, 2012). This adaptation is illustrated, for example, by the accumulation of FCs in umbelliferous plants infested with herbivorous insects (Berenbaum, 2001, 2002; Jogesh et al., 2015). The production and variation of FCs under these conditions were interpreted as the outcome of an “arms race” between plants and insects. The biosynthetic pathway leading to FCs has been studied for many years in Apiaceae and was shown in the 1980s to involve the sequential action of several P450s (Hamerski and Matern, 1988a,b). The first gene of this enzyme category, CYP71AJ1, was identified in 2007 (Larbat et al., 2007).

An intimate cooperation of P450s from wheat or maize has been demonstrated (Jonczyk et al., 2008), converting indole-3-glycerol phosphate (from the L-tryptophan pathway) to benzoxazinones by several variants of a single P450 subfamily (CYP71D). Such variants have likely evolved from one ancestor. The same mechanism was suspected to lead to the synthesis of FCs. This hypothesis initially gained support by the functional characterization of CYP71AJ4 (Larbat et al., 2009) catalyzing the synthesis of angelicin which is the analogous reaction in the pathway branching to angular FCs (Figure 2). Nevertheless, further studies of diversification revealed the presence of CYP71AJ paralogs in some Apiaceae species that do not produce FCs (Dueholm et al., 2015). In addition, no other CYP71AJs were related to the synthesis of FCs. With the discovery of the blooming CYP71AZ P450 subfamily, the hypothesis that only a few number of P450 families are specialized in the synthesis of FCs and coumarins had been revived. Since simple coumarins are abundantly present in spermatophytes, and FCs are present in only a few of them, the plants may have gained new CYP71AZ functionalities during evolution concomitant with or after those of the CYP71AJ family.

The study and functional characterization of five different CYP71AZs reported here provides evidence that each of them has its own evolutionary history. CYP71AZ3 is not overexpressed upon wounding of *P. sativa* plants. Our experiments show that this enzyme does not metabolize FCs but rather accepts a coumarin as substrate, although the catalytic properties determined *in vitro* suggest that esculetin is not the physiological substrate. Therefore, to identify the enzyme real substrate, the screening must be extended to other coumarins. CYP71AZ1 and 6 are very close to each other and have a restricted substrate specificity dedicated to the synthesis of an FC, which might reflect that they were selected with the appearance of the pathway and are now almost specialized for the synthesis of xanthotoxol. These reactions are catalyzed with reasonable  $K_m$  values. The evolution of this cytochrome P450 subfamily might therefore be close to that of the CYP71AJ subfamily. We can assume that the CYP71AZ and CYP71AJ P450 subfamily were originally involved in the synthesis of some common metabolites and evolved to be able to participate to the synthesis of FCs.

The emergence of the branching pathway to linear and angular FCs with CYP71AJs as the pivotal enzymes (Larbat et al., 2007, 2009; Dueholm et al., 2015) was a major breakthrough for the defense of these plants. CYP71AZ4 might thus be allocated to an intermediate stage to convert not only simple coumarins but also FCs. Regarding coumarins, CYP71AZ4 was

shown to convert 6-methoxycoumarin to scopoletin *in vitro* at a physiological acceptable  $K_m$  (9.5  $\mu$ M), and scopoletin can be further converted to fraxetin. However, this second reaction required an extraordinary high scopoletin concentration ( $K_m = 762.2 \mu$ M). CYP71AZ4 also accepts psoralen as a substrate ( $K_m = 11.3 \mu$ M, Table 1). Since the  $K_m$  values are close to each other and in an acceptable physiological range, CYP71AZ4 may be an enzyme on its way to becoming specialized in the evolutionary scheme.

Finally, CYP71AZ5 has no assigned function. The fact that this coding sequence is not transcribed and could not be cloned under our experimental conditions could reflect a pseudogenization or a gene duplication where the neofunctionalization has not yet occurred. From the evolutionary point of view, this gene might be close to CYP71AZ4, which is consistent with the phylogenetic tree. Thus, gene duplication and neofunctionalization appear to have occurred at several stages of the evolution of the CYP71AZ subfamily of genes and in close association with the evolution of CYP71AJ sequences. This evolution can be mimicked with targeted mutagenesis. Our sequence-swapping experiments showed that modifications required for modifying one enzyme into another are not limited to simple amino acid changes (mut1-3) but also involve modification of the tertiary structure of the protein (mut 4). Such a directed evolutionary strategy could be used to develop enzymes with new functionalities, as it has already been described elsewhere (Park et al., 2009).

A screening of various databases such as OneKP<sup>5</sup>, The European Nucleotide Archive<sup>6</sup>, phytometasyn<sup>7</sup> using CYP71AZ1 as target makes evidence that only two plant families harbored paralogous genes with significantly high enough homologies to belong to the same P450 subfamily. Intriguingly, no CYP71AZ could be identified in other plants producing FCs such as Rutaceae (citrus for example) or Moraceae (Ficus for example). Although we could not test all the plants belonging to these plant species because of the absence of available databases, it might be possible that the synthesis of these molecules in other plants is realized by other genes and is in favor of a convergent evolution as suggested previously (Berenbaum, 1983; Bourgaud et al., 2014; Munakata et al., 2016).

All the CYP71AZ paralogous genes, highlighted in our transcriptomic screening, were identified in Apiaceae except CYP71AZ10, which was identified in *C. racemosa*. Interestingly, Moon et al. (2011) shows that this plant is able to produce isoimperatorin. An additional search done through the Dictionary of Natural Products<sup>8</sup> also showed the presence some molecules which are structurally related to FC such as (norvisnagin: 4-hydroxy-7-methylfuro[3,2-g]chromen-5-one (chromene) and angelicin: 4-hydroxy-7-(hydroxymethyl)-2-(2-hydroxypropan-2-yl)-2,3-dihydrofuro[3,2-g]chromen-5-one). It might be possible that CYP71AZ10 could be involved

<sup>5</sup><http://onekp.com>

<sup>6</sup><https://www.ebi.ac.uk/ena>

<sup>7</sup><https://bioinformatics.tugraz.at/phytometasyn>

<sup>8</sup><http://dnp.chemnetbase.com>

in the synthesis of these molecules. A deeper analysis of new transcriptomic data will probably help, in the near future to get more tracks for identifying the function of this enzyme or other new paralogues of CYP71AZ.

To identify the function of other CYP71AZ, the emergence and publication of many metabolomics studies will help to identify potential substrates. We show that CYP71AZ4 was able to metabolize a few coumarins. Among the molecule tested, only two of them could be identified in parsnip extracts (fraxetin and osthol, **Supplementary Figure S1**). These result makes evidence that this enzyme is obviously not highly specific. An extensive screening should be realized based on phytochemicals studies. For example, very few reports showed the presence of several coumarins in the roots of *T. garganica* and are present in very low amounts. Among these molecules, various authors described scopoletin and 6-methoxy-7-geranyloxycoumarin (Larsen and Sandberg, 1970; Rubal et al., 2007). CYP71AZ2 may be involved in the synthesis of this coumarin.

Plants produce a plethora of metabolites that are required for their growth and propagation as well as for regulatory needs and adaptation to environmental conditions. Among the 370,000 predicted vascular plants, only a small part has been studied at the chemical level (Lughadha et al., 2016). According to the Royal Botanical Kew Garden, 28,000 plants are useful for medicines and so far 300,000 specialized metabolites were identified throughout the plant diversity (Hubert et al., 2017). Here, we provide a new example of the importance of cytochrome P450s, which have already been identified to be deeply involved in this diversification (Nelson and Werck-Reichhart, 2011), and suggest how they can be involved in the appearance of new molecules to meet further challenges for an ecological function (Pichersky and Gang, 2000; Lewinsohn and Gijzen, 2009).

## AUTHOR CONTRIBUTIONS

CK and SR cloned the genes (CYP71AZ3, 4, 6, and AZ4mut1-4), expressed all the proteins in yeast and realized the kinetic characterization. SK-T did the *A. majus* cell cultures and identified and isolated CYP71AZ1. GG did the induction of parsnip and the RT-qPCR experiments for CYP71AZs and CYP71AZ3. BS did the NMR analyses. JG did the LC/manuscript analyses. AO did the biochemical analyses and wrote the manuscript. DR performed the modeling and Kpax experiments and edited the manuscript. UM led the work done on the *A. majus* cell cultures and edited the manuscript. FB and AH led the work done on the CYP71AZ characterization. AH led the whole project and wrote the manuscript.

## FUNDING

SR was funded by the SAARLORBIOTEC research project (INTERREG).

## ACKNOWLEDGMENTS

The authors thank Heiko Rischert and Henrik Toft Simonsen for providing the sequences from *T. garganica* and *C. maculatum*. They acknowledge Clément Charles for his technical support. The authors Dr. D. Nelson for the assignment of CYP71AZ subfamily.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00820/full#supplementary-material>

**FIGURE S1** | Name, molecular weight, chemical structure, and code access number of the molecules used for the *in vitro* screening of CYP71AZs.

**FIGURE S2** | Identification of the products of 6-methoxycoumarin metabolism (A) by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert 6-methoxycoumarin to P04 and P05. (A) HPLC profile of enzyme assays containing 6-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. (B) MS and absorption spectrum of standard scopoletin. (C) MS and absorption spectrum of P04. (D) MS and absorption spectrum of P05.

**FIGURE S3** | Identification of the product of scopoletin metabolism by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert scopoletin to P06. (A) HPLC profile of enzyme assays containing scopoletin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. (B) MS and absorption spectrum of standard Fraxetin. (C) MS and absorption spectrum of P06.

**FIGURE S4** | Identification of the products of 3-methyl-7-methoxycoumarin (A) and 7-methoxycoumarin (B) metabolism by CYP71AZ4. Yeast microsomes expressing CYP71AZ4 convert 3-methyl-7-methoxycoumarin to P07 and 7-methoxycoumarin to P08. (A) HPLC profile of enzyme assays containing 3-methyl-7-methoxycoumarin, NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. (B) HPLC profile of enzyme assays containing 7-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ4. (C) MS and absorption spectrum of P07. (D) MS and absorption spectrum of standard daphnetin 7-methyl ether. (E) MS and absorption spectrum of P08.

**FIGURE S5** |  $^1\text{H}$  NMR spectra (700 MHz, acetone- $d_6$ ) of a sample containing 8-hydroxy-3-methyl-7-methoxy coumarin (P07). (A) Full spectrum. (B) Extended region showing aromatic and double bond signals of P07 (marked with green arrows). (C) Extended region showing the methoxy signal of P07 (marked with green arrow).  $^1\text{H}$  NMR (700 MHz, acetone- $d_6$ ):  $\delta$  7.03 (1H, J = 8.2 Hz, H-5), 6.98 (1H, d, J = 8.2 Hz, H-6), 7.65 (1H, s, H-4), 3.92 (1H, s, 7-OCH $_3$ ), signal of 3-CH $_3$  not detected due to overlap with matrix signals.  $^{13}\text{C}$  NMR (175 MHz, acetone- $d_6$ , signal extracted from HSQC spectrum):  $\delta$  56.6 (7-OCH $_3$ ) HSQC cross signals of other carbon atoms were not detected due to limited amount of material. The  $^1\text{H}$  NMR spectrum of P07 (700 MHz, acetone- $d_6$ ) displays doublets (J = 8.2 Hz) of an AB spin system comprising H-5 ( $\delta$  7.03) and H-6 ( $\delta$  6.98) of the aromatic ring in this figure. The proximity of H-5 and H-6 was confirmed by a strong roof effect of the two  $^1\text{H}$  NMR signals and a correlation signal in the  $^1\text{H}$ ,  $^1\text{H}$  COSY spectrum. The singlet at  $\delta$  3.92 (3H) and an HSQC correlation signal  $\delta_{\text{H}}$  3.92/ $\delta_{\text{C}}$  56.6 are assignable to the methoxy group (7-OCH $_3$ ). The signal of the methyl group protons at C-3 remained obscured due to overlap with matrix signals. Since the presence of the H-5/H-6 AB spin system ruled out hydroxylation at C-5 and C-6, positions 4 and 8 remained possibilities for the new hydroxyl group. Thus, the singlet of the isolated aromatic proton resonating at  $\delta$  7.65 (1H) had to be assigned to either to H-4 or H-8. A proton at C-8, corresponding to hydroxylation at C-4, could be disregarded because an *m*-coupling with H-6 was missing and the chemical shift of H-8 would be expected near  $\delta$  6.5. The latter value is largely different from the observed chemical shift of  $\delta$  7.65, which is consistent with the



value expected for H-4. Finally, hydroxylation at the 2-methyl group could also be excluded because the  $^1\text{H}$  NMR signal of a prochiral hydroxymethylene group with an expected chemical shift of approximately  $\delta$  4.5 was absent from the spectrum. From these data, P07 was identified as 8-hydroxy-3-methyl-7-methoxycoumarin.

**FIGURE S6** | Identification of the products of 7-methoxycoumarin metabolism by CYP71AZ1 (**A and D**) and CYP71AZ6 (**B and E**). Yeast microsomes expressing CYP71AZ1 and CYP71AZ6 convert 7-methoxycoumarin to P09 and P10, respectively. **(A)** HPLC profile of enzyme assays containing 7-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ1. **(B)** HPLC profile of enzyme assays containing 7-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ1. **(C)** MS and absorption spectrum of 5-hydroxy-7-methoxycoumarin. **(D)** MS and absorption spectrum of P09. **(E)** MS and absorption spectrum of P10.

**FIGURE S7** | Identification of the products of 6-methoxycoumarin metabolism by CYP71AZ1 (**A and B**) and CYP71AZ6 (**C and D**). Yeast microsomes expressing CYP71AZ1 and CYP71AZ6 convert 6-methoxycoumarin to P11 and P12, respectively. **(A)** HPLC profile of enzyme assays containing 6-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ1. **(B)** MS and absorption spectrum of P11. **(C)** HPLC profile of enzyme assays containing 6-methoxycoumarin, without (line 1) or with (line 2) NADPH, oxygen, and yeast microsomes expressing CYP71AZ6. **(D)** MS and absorption spectrum of P12.

**FIGURE S8** |  $^1\text{H}$  NMR spectra (700 MHz,  $\text{MeOH-}d_4$ ) of a sample containing 6,7,8-trihydroxycoumarin (P13). **(A)** Full spectrum. **(B)** Extended region showing

signals of P13 (marked with green arrows).  $^1\text{H}$  NMR (700 MHz,  $\text{MeOH-}d_4$ ):  $\delta$  7.86 (1H,  $J = 8.2$  Hz, H-4), 6.23 (1H, d,  $J = 8.2$  Hz, H-3), 6.82 (1H, s, H-5).  $^{13}\text{C}$  NMR (175 MHz,  $\text{MeOH-}d_4$ , signals extracted from HSQC spectrum):  $\delta$  112.6 (C-4), 146.0 (C-5), 103.6 (C-6). The  $^1\text{H}$  NMR spectrum of P13 (700 MHz,  $\text{MeOH-}d_4$ ) showed two doublets (at  $\delta$  7.86 and 6.23), which according to their chemical shifts and coupling constant of  $J = 8.2$  Hz were assigned to the Z-configured double bond protons H-4 and H-5. A  $^1\text{H}$ ,  $^1\text{H}$ -COSY spectrum confirmed the AX spin system of these two protons. Hence, the new hydroxyl group cannot be located in the pyrone ring but must be attached to either position 5 or 8 of the aromatic ring. The signal of the remaining proton was detected at  $\delta$  7.68, and assignment to H-6 was achieved by comparing the chemical shift with reported data of 6,7,8-trihydroxycoumarin (Kayser and Kolodziej, 1995; Yasuda et al., 2006). Furthermore, the  $^{13}\text{C}$  chemical shifts of the protonated carbon atoms C-3, C-4, and C-5 were extracted from the HSQC spectrum of P13 (Supplementary Figure S9) and agreed with reported data (Kayser and Kolodziej, 1995; Yasuda et al., 2006) of 6,7,8-trihydroxycoumarin.

**FIGURE S9** | Partial HSQC NMR spectrum (700 MHz,  $\text{MeOH-}d_4$ ) of a sample containing 6,7,8-trihydroxycoumarin (P13). Top:  $^1\text{H}$  NMR spectrum.  $^1\text{H}$  NMR signals and  $^1\text{H}/^{13}\text{C}$  correlation signals of P13 are marked with green arrows.

**FIGURE S10** | Sequences of the different CYP71AZs collected in databases.

**TABLE S1** | Sequence homologies between different CYP71AZs from *A. majus* (CYP71AZ1) and *P. sativa* (AZ3, AZ4, AZ5, and AZ6). Comparison of the full length sequences are in white squares. Comparison restricted to the six substrate recognition sites (SRS) are in gray squares.

## REFERENCES

- Benderoth, M., Textor, S., Windsor, A. J., Mitchell-Olds, T., Gershenzon, J., and Kroymann, J. (2006). Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9118–9123. doi: 10.1073/pnas.0601738103
- Berenbaum, M. (1983). Coumarins and caterpillars: a case for evolution. *Evol. Int. J. Org. Evol.* 37, 163–179. doi: 10.1111/j.1558-5646.1983.tb05524.x
- Berenbaum, M. (2001). Chemical mediation of coevolution: phylogenetic evidence for Apiaceae and associates. *Ann. Mo. Bot. Gard.* 88, 45–59. doi: 10.2307/2666131
- Berenbaum, M. (2002). Postgenomic chemical ecology: from genetic code to ecological interactions. *J. Chem. Ecol.* 28, 873–896. doi: 10.1023/A:1015260931034
- Berenbaum, M., Nitao, J., and Zangerl, A. (1991). Adaptive significance of furanocoumarin diversity in *Pastinaca sativa* (Apiaceae). *J. Chem. Ecol.* 17, 207–215. doi: 10.1007/BF00994434
- Bourgau, F., Olry, A., and Hehn, A. (2014). "Recent advances in molecular genetics of furanocoumarin synthesis in higher plants," in *Recent Advances in Redox Active Plant and Microbial Products*, eds C. Jacob, G. Kirsch, A. Slusarenko, P. Winyard, and T. Burkholz (Dordrecht: Springer), 363–375. doi: 10.1007/978-94-017-8953-0\_14
- Dueholm, B., Krieger, C., Drew, D., Olry, A., Kamo, T., Taboureau, O., et al. (2015). Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71A1 subfamily. *BMC Evol. Biol.* 15:122. doi: 10.1186/s12862-015-0396-z
- Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112
- Fraenkel, G. S. (1959). Reason D'être of secondary plant substances. *Science* 129, 1466–1470. doi: 10.1126/science.129.3361.1466
- Funk, C., and Croteau, R. (1993). Induction and characterization of a Cytochrome-P-450-dependent camphor hydroxylase in tissue-cultures of common sage (*Salvia-officinalis*). *Plant Physiol.* 101, 1231–1237. doi: 10.1104/pp.101.4.1231
- Ghoorah, A. W., Smail-Tabbone, M., Devignes, M.-D., and Ritchie, D. W. (2014). KBDock 2013: a spatial classification of 3D protein domain family interactions. *Nucleic Acids Res.* 42, 389–395. doi: 10.1093/nar/gkt1199
- Glawischnig, E., Grun, S., Frey, M., and Gierl, A. (1999). Cytochrome P450 monooxygenases of DIBOA biosynthesis: specificity and conservation among grasses. *Phytochemistry* 50, 925–930. doi: 10.1016/S0031-9422(98)00318-5
- Hamerski, D., and Matern, U. (1988a). Biosynthesis of psoralens. Psoralen 5-monooxygenase activity from elicitor-treated *Ammi majus* cells. *FEBS Lett.* 239, 263–265.
- Hamerski, D., and Matern, U. (1988b). Elicitor-induced biosynthesis of psoralens in *Ammi majus* L. suspension cultures. Microsomal conversion of demethylsuberosin into (+) marmesin and psoralen. *Eur. J. Biochem.* 171, 369–375.
- Hotti, H., Seppanen-Laakso, T., Arvas, M., Teeri, T., and Rischer, H. (2015). Polyketide synthases from poison hemlock (*Conium maculatum* L.). *FEBS J.* 282, 4141–4156. doi: 10.1111/febs.13410
- Hubert, J., Nuzillard, J.-M., and Renault, J.-H. (2017). Dereplication strategies in natural product research: how many tools and methodologies behind the same concept? *Phytochem. Rev.* 16, 55–95. doi: 10.1007/s11101-015-9448-7
- Irmisch, S., McCormick, A. C., Guenther, J., Schmidt, A., Boeckler, G. A., Gershenzon, J., et al. (2014). Herbivore-induced poplar cytochrome P450 enzymes of the CYP71 family convert aldoximes to nitriles which repel a generalist caterpillar. *Plant J.* 80, 1095–1107. doi: 10.1111/tpj.12711
- Jogesh, T., Stanley, M. C., and Berenbaum, M. R. (2015). Evolution of tolerance in an invasive weed after reassociation with its specialist herbivore. *J. Evol. Biol.* 27, 2334–2346. doi: 10.1111/jeb.12469
- Jonczyk, R., Schmidt, H., Osterrieder, A., Fiessmann, A., Schullehner, K., Haslbeck, M., et al. (2008). Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of Bx6 and Bx7. *Plant Physiol.* 146, 1053–1063. doi: 10.1104/pp.107.111237
- Kawai, Y., Ono, E., and Mizutani, M. (2014). Expansion of specialized metabolism-related superfamily genes via whole genome duplications during angiosperm evolution. *Plant Biotechnol.* 31, 579–584. doi: 10.5511/plantbiotechnology.14.0901a
- Kayser, O., and Kolodziej, H. (1995). Highly oxygenated coumarins from *Pelargonium sidoides*. *Phytochemistry* 39, 1181–1185. doi: 10.1016/0031-9422(95)00166-5
- Kliebenstein, D. J., Lambrix, V. M., Reichelt, M., Gershenzon, J., and Mitchell-Olds, T. (2001). Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in arabidopsis. *Plant Cell* 13, 681–693. doi: 10.1105/tpc.13.3.681

- Kliebenstein, D. J., and Osbourn, A. (2012). Making new molecules - evolution of pathways for novel metabolites in plants. *Curr. Opin. Plant Biol.* 15, 415–423. doi: 10.1016/j.pbi.2012.05.005
- Larbat, R., Hehn, A., Hans, J., Schneider, S., Jugdé, H., Schneider, B., et al. (2009). Isolation and functional characterization of CYP71A4 encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. *J. Biol. Chem.* 284, 4776–4785. doi: 10.1074/jbc.M807351200
- Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., et al. (2007). Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *J. Biol. Chem.* 282, 542–554. doi: 10.1074/jbc.M604762200
- Larsen, P. K., and Sandberg, F. (1970). Constituents of umbelliferous plants. XV. Coumarins from *Thapsia garganica* L. The structure of new coumarin. *Acta Chem. Scand.* 24:1113. doi: 10.3891/acta.chem.scand.24-1113
- Lewinsohn, E., and Gijzen, M. (2009). Phytochemical diversity: the sounds of silent metabolism. *Plant Sci.* 176, 161–169. doi: 10.1016/j.plantsci.2008.09.018
- Li, W., Zangerl, A. R., Schuler, M. A., and Berenbaum, M. R. (2004). Characterization and evolution of furanocoumarin-inducible cytochrome P450s in the parsnip webworm, *Depressaria pastinacella*. *Insect Mol. Biol.* 13, 603–613. doi: 10.1111/j.0962-1075.2004.00518.x
- Lughadha, E. N., Govaerts, R., Belyaeva, I., Black, N., Lindon, H., Allkin, R., et al. (2016). Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* 272, 82–88. doi: 10.11646/phytotaxa.272.1.5
- Mizutani, M., and Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61, 291–315. doi: 10.1146/annurev-arplant-042809-112305
- Mizutani, M., and Sato, F. (2012). Unusual P450 reactions in plant secondary metabolism. *Arch. Biochem. Biophys.* 507, 194–203. doi: 10.1016/j.abb.2010.09.026
- Moon, L., Ha, Y. M., Jang, H. J., Kim, H. S., Jun, M. S., Kim, Y. M., et al. (2011). Isoimperatorin, cimisiside E and 23-O-acetylshengmanol-3-xyloside from *Cimicifuga rhizome* inhibit TNF- $\alpha$ -induced VCAM-1 expression in human endothelial cells: involvement of PPAR- $\gamma$  upregulation and PI3K, ERK1/2, and PKC signal pathways. *J. Ethnopharmacol.* 133, 336–344. doi: 10.1016/j.jep.2010.10.004
- Munakata, R., Olry, A., Karamat, F., Courdavault, V., Sugiyama, A., Date, Y., et al. (2016). Molecular evolution of parsnip membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytol.* 211, 332–344. doi: 10.1111/nph.13899
- Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-3113.2011.04529.x
- Nelson, D. R., Schuler, M. A., Paquette, S. M., Werck-Reichhart, D., and Bak, S. (2004). Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiol.* 135, 756–772. doi: 10.1104/pp.104.039826
- Park, S.-H., Park, H.-Y., Sohng, J. K., Lee, H. C., Liou, K., Yoon, Y. J., et al. (2009). Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.* 102, 988–994. doi: 10.1002/bit.22150
- Pichersky, E., and Gang, D. R. (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* 5, 439–445. doi: 10.1016/S1360-1385(00)01741-6
- Pompon, D., Louerat, B., Bronnie, A., and Urban, P. (1996). Yeast expression of Animal and Plant P450s in optimized redox environments. *Methods Enzymol.* 272, 51–64. doi: 10.1016/S0076-6879(96)72008-6
- Ritchie, D., Ghoorah, A., Mavridis, L., and Venkatraman, V. (2012). Fast protein structure alignment using gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics* 28, 3278–3281. doi: 10.1093/bioinformatics/bts618
- Roselli, S., Olry, A., Vautrin, S., Coriton, O., Ritchie, D., Navrot, N., et al. (2016). Genomic approach to decipher specialized plant pathways: a case study with furanocoumarins in parsnip. *Plant J.* 89, 1119–1132. doi: 10.1111/tpj.13450
- Rubal, J. J., Moreno-Dorado, F. J., Guerra, F. M., Jorge, Z. D., Saouf, A., Aksira, M., et al. (2007). A pyran-2-one and four meroterpenoids from *Thapsia transtagana* and their implication in the biosynthesis of transtaganolides. *Phytochemistry* 68, 2480–2486. doi: 10.1016/j.phytochem.2007.06.023
- Rupasinghe, S., Baudry, J., and Schuler, M. A. (2003). Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. *Protein Eng.* 16, 721–731. doi: 10.1093/protein/gzg094
- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/jmbi.1993.1626
- Werck-Reichhart, D., Hehn, A., and Didierjean, L. (2000). Cytochromes P450 for engineering herbicide tolerance. *Trends Plant Sci.* 5, 116–123. doi: 10.1016/S1360-1385(00)01567-3
- Werck-Reichhart, D., Schalk, M., Nedelkina, S., Batard, Y., and Durst, F. (1997). Modification of critical residues in the active site of a plant P450: CYP73A1. *FASEB J.* 11, A793–A793.
- Yasuda, T., Fukui, M., Nakazawa, T., Hoshikawa, A., and Ohsawa, K. (2006). Metabolic fate of fraxin administered orally to rats. *J. Nat. Prod.* 69, 755–757. doi: 10.1021/np0580412

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Krieger, Roselli, Kellner-Thielmann, Galati, Schneider, Grosjean, Olry, Ritchie, Matern, Bourgaud and Hehn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

#### Annexe 4 : Séquences codantes des différents gènes connus de la voie de biosynthèse des furocoumarines chez le panais

##### ➤ *PsDiox*

ATGGCTCCATCTCCCACTGATGATTACAGAAATTTTGCAGTAACAAAAGGGCACGGAGTAAAGGGTCTCT  
CCGATCTGAAGCTCGACGCTTTACCAGAGCAATATATTCAACCTGTGGAGGAGCGATTAGACATGACAAA  
AGTCCTGAAGAAAGAGTCCATACCTGTGATTGACATGTCCAATTTGGATGATCCGAAGGTGGCGGATCAG  
ATTGCGGCTGCTGCTGAAAAGTGGGGATTCTTTCAGATTATTAATCATGGAGTCCCCATTGAGGTTCTTG  
AGAATGTCAAGGAAGCTACACGTAGATTCTTTCGCTTGCCTGTGGAGGAGAAGATCAAGTACACCCAAGA  
ACAGTCTCCTACAAATTCTGTTAGGCTTACAACAAGCTTTCTTCCAAAAGTAGACAAAAGTTCTTGAGTGG  
AAAGATTACCTTAGTATCCTTGTCTGATAAGAAGTCTTCGGAATTTTGGCCATCTACGTGCAAGAATG  
ATGTAAAGGAATATGTTGAGAAAGTCTGAGTTTGTCAAGTGGCTGCTAAAAGCCCTGATGAAGGGGT  
AAATGTGGACATGGACTCAAAGAATCAATTCTAATGGGCTCAACCAGAATCAACCTCAACTATTATCCT  
GTTTGTCCAAACCCTGAGCTTGCAATTGGAGTGGGACGCCATTCTGACGTCTCCACACTCACATTTCTGC  
TTCAAGACAACGTCGGAGGATTACACATCAGAAAGATGGACACAGATACTTGGAATTTTCGTGCCTCCTGT  
AGAGGGAGCAATAGTAATTAACATTGGTGATGCTCTTCAAATCTTGAGTAATGGTAAATACAAAAGTGCT  
GAGCACTGTGTTGCAGCCAATGGAAGCAACGACAGAATTTCCGGTGCCAATCTTTACTAACCCTAGTCCGG  
ATGATATTATTGGCCCGCTACCTGAGCTGCTCAAGAATGGAGAAAAACCAATATACAAACACGTTCTGTA  
TTCGATTATGTGAAACATTTCTACAGGAAATCTCACGATGGGAAGCACACCCTTGATTTGCGACAGATG  
TGA

##### ➤ *PsPT1*

ATGGCTCAAACAATTATGCATTCACGATTATCGTCCGGTTTCTTACATCTTCAACGAGACAAAAGGCTTTC  
GTACACTTCCAACGCAAAGAAGACATGCTAAAGTAGTGAACGGAGACCAAGAATTTGCTTTCAGGGTGGT  
TTCATGCGATAAAAATTTAGATTCAACAAAGAATTTTAGCGGCAGTTGTGAAAAGCCAATTAGAACACAC  
ACAAACAAACTTTTACAAACAATTAGTGCCACATCAGACAGAGAGGCTATTATACAGCCCAAAGATGATT  
ACGAAGCACCATGGCAGAATACTTTACGGAGAAAATGGGATGCATTTTGCACGTTTGGTCGTCCATACTC  
TGCCATCTGCACAATTATTGGAATAAGTTCGGTTTCTCTGCTGCCCTTAACTTCCGTTAAAGATTTCTCG  
GCTCCATATTTTGTGGGATTACTACAGGCATTGATCCCATTTCTTGTGCCAACATCTATACCTCGGGGA  
TAAATCAACTTGTGATGTGGACATAGACAAAATAACAAGCCTTACCTGCCACTGGTTTCGGGAGAATT  
TTCCTTGGGCGAAGGGGAGAGCAATTGTCTCAGCACTGGCTTTTATGTGCCTGGCCGTGGGAATCCTATCT  
CATTCTACACCATTGTTTGTGGGGTCTTGTCTATTTTTTAATTGGAACCGCATATTCTGTTGAGCTAC  
CTCTACTTAGATGGAACAAACACCGGCCATGGCTGCATTTTCTATGGCCGGCCTAATGGGACTCACGAT  
CCAGCCTGCTGTCTTCTATCACATCCAGAATGCACTCGGTAAACCAATGGTATTTTCAAAGACAGTGGCC  
TTTGCTACCATTTTCTCAGTGTCTTTCAGCTGTTCTTGAGCGATCAAGGACGTACCTGATGTTGAAG  
GAGACACAGCGTTCGGCAACCGAACATTCAAGTGTGAGATATGGCCAAGAGAAAAGTCTTTTCTGTCTGTCT  
AAATATCCTGCTGCTTGCCTACGGGTTTGTGTGGTGGTAGGAGCTTCATCCTCGTTCCTATTATGCAAG  
ATTGTATCTGTGATGGGTCACACTACACTTGCCTCCCTACTTTTGCTTCGAGCAAAATCTACTAATCCGA  
AAGATCCTGAATCAACGCAGTCCTTTTACATGTTTTTATTAAAGCTATTGTACGCGGAGTATGTGCTAAT  
TCATTTTCATGCGCTGA



➤ *PsPT2*

ATGACTCAGACACTTATGCATTACGCTTCTCCTCCGGTTTCTTACATCTTCAACGACGAGAAAGATCAG  
GCTTTCTTACATCATTTCCAACCCGGAGAAGACATGCTACAATATTGAATAAAGACAAAGAATTGACTGT  
CAGGGTGGTTTCTTGCGATAAAATTTTGGATTCAACAAACAATTCTAGCAGAAAGTTGTGAAAAGCCAATT  
AATAGAACAAACACAAGCACGCTTTTACAAACACTTGGTGCCTCATCAGAGGGTGAGGTTATTATACAAC  
CCAAAGCTGAATACGAAGAAACATGGGATAGTATTTTCTGGAGAAAATGGGATGCCTTTGTGACGTTTGG  
TCGACCATATTCTGTCCTTGGCTCAATTATTGGAATAAGTTCAGTTTCTCTACTGCCCTTAACCTCAGTC  
GGAGATTTCTCACTTGCAGTTTTTGTGGGATTTCGTACAGGCATTGATCCCGTTTGTGTGCGAACATTT  
ATGCCTCAGGAATAAATCAAGTGGTTGATGTGGAAATAGACAAAATTAACAAGCCCTATCTTCCACTGGT  
TTCTGGAGATTTTTCTATGGGGGAAGGGAAAGCAGTTGTCTCAGCTACGGGCTTTTTGTGCTTGGCCATG  
ACAATCATGTTTGGGTCTCTACCATTTGTTTCTGGGAGTTCTTGGATATTTTCTTTATGCAACTGCATATT  
CTGTGGAGCTACCTTTCCCTTAGATGGAAAACAAAACCGTTTATGGCTGCATTAGTATGGCTGGTCTAAT  
GGGACTCACCATTACGCCTTCTGTCTACTATCACATCCAAAACGTGCTTGGTAGACCAATGGTTTTGACA  
AAGCCAGTGGTCTTCGCTACCAGTTTCATCAGTGTCTTTTTCAGCTGTTCTTGCAATGATCAAGGACTTAC  
CTGATGTTGAAGGAGACAGAGCGTTGGGCAACCTAACATTAGTGTAGATATGGCCAAGAGAAAAGTCTT  
TAATATCTGTGTAGGTATCATGCTGGCTGCCTATGCATCTGCTGTGGTGACAGGATCTTTCTCCTCGCTT  
CTAATATGCAAGCTTGTCTGTGATTGGTCACACTGCACCTTGCCCTTCTCCTTATGCTTCGAGCAAAAT  
CTATTGATGTAAATGATCCTGAATCAACGCAGTCCTTTTACATGTTTGCTTTTCAGCTATTATATGCAGA  
GTATGTGCTTATTCATTTCATACGCTGA

➤ *CYP71A3*

TATTTCTTTCCATTGTTTCTTGTGACAAATATTTCTGTACAAGTGGTTAGTTAAGAAAACACCCTCAAAAA  
ATCTGCCACCTTCTCCACCAAGATTACCCATTATTGGGAACCTTCATCAAATTGGTCCAGACCCGCAAAT  
TTCACTCAGAGACTTGGCTCGGGAGTATGGTCCAGTAATGCACCTTAAGTTTGAAGTGTACCTGTTCTT  
GTTGTTTCTTCAGCTGATGGGGCTCGGGAGATTTTCAAAACCCATGATCTGGTTTTTCGCAGACAGGCCTT  
ATTGAGTGTAGCCAATAGAATCTTCTATAACGGAAGGGACATGGTGTGCTCGTTACACTGAATATTG  
GAGACAAGTAAAGAGTACATGTGTTACTCAACTCCTAAGTGTCAAAAGGGTTCAATCTTTTCACAATGTG  
AGAGAAGAAGAAGTTGCTCTGTTACTCGATAATATTGAAAATTCTAAATCCAAAGTAATAAATTTAAGCG  
AAATGTTAATCGAATAACTGGAACGTAGTCTGCAGGGCTGCACTAGGCAGTGGTTATAATGTGGATTG  
ATACAAATCACTACTTCTGCAGATAATGGATATGTTAGGCTATTCCCGCAGCATAGAGGATTTTTTTCCG  
TCGCTCGGTTGGGTTGATTGGATCACTGGATTGAAGGGAAAGGTGGAGAAAGCAGCCAATGGGGTTGATG  
CTTTTCTGGAAGGTGTTCTTAAAAATCATACAAATCCAAGTACTAGTAGTGAAACAAAGACTTTGTATC  
CATTTTGCTTGAGATTCAAGAGGCAGATGCTGGCTCTTCTATGGATAAAGAATGTATCAAATCTCTTATC  
TGGGATATGTTGGGTGCTGGAACGAAACAATAGCGACGGCTTTGGAGTGGACAATAGGAGCGCTGATTA  
AAAGTCCGGACGCCATGTCCAAATTGCAGAAAGAGGTGAGGGAAATAGGCAAAGGCAATCAAGGATAGA  
AGAGGGTGATCTAGTCAAAATGGACTACCTGAAAGCAGTAATGAAAGAGAGTATGCGACTGTACTTTACA  
GCCCCACTTCTAGTTCCCAGAGAAGCGAGGCAGGACGTGAAATTTATGGGGTATGACATAAAATCAGGAA  
CACAAGTATTGATAAATGCATGGGCAATTGCAAGAGACCCTTCATCATGGGACAATCCAGAGGAGTTCCG  
GCCTGAGAGGTTCTTGAATAGTCCTATTGATTACAAAGGCTTTAACTATGAGTATATTCCCTTTGGAGCT  
GGTCGGAGGGGTTGTCTGGGATCCAGTTTGCAATCTCTGTTAATGAGCTTGTAGTAGCAAATGTTGTCA  
ACAAGTTTAATTTGAATTGCCTGATGGGAAAAGATTGGAGGAAATGGACATGACTGCTTCCACTGGCAT  
TACTTTTCACAAAAATC

➤ CYP71AJ4

TATTTCTTTTCATTGTTTCTTGTCACAATATTTCTGTACAAGTGGTTAGCTAAGAAAAACACCCTCAAAAA  
ATCTGCCACCTTCTCCACCAAGATTACCCATTATTGGAAACCTTCATCAAATTGGTCCAGACCTGCACAT  
TTCACCTAGAGACCTGGCTCGAAAGTATGGTCCATTAATGCAACTTCAGCTTGGACGTATACCTGTTCTT  
GTTGTTTCTTCAGCTGAGGCAACTCGGGAGGTTTTGAAAACCTCACGATGTGGTTTTCTCCCAACGACCTA  
TTACAAGTGCCATCGATAAACTCTGCTATAAGGGAAGGGACGTGGCGTTTTCTCGTTATAGTGAATATTG  
GAGACAAGTAAGGAGTACATGTGTTACTCAGCTCCTAAGTAACAGTAGGGTTCATTTCGTTTCACAATATC  
AGAGAAGAAGAAGTTGCTCTTTTAATCCAAAATATTGAAAATTCTGCTTCGGAAGTTATAAATTTAGGCG  
AGCAATTAATCCAATAACCAGAAATGTAGTCTGCAGGGTTTCAGTAGGAAGCGAATATCTAAGTGGTCA  
TAAAGGGAAATTATACCAGAAATTACTTGCCGAAGTAACTGAGATGTTAGCTTATACCTACAGCATAGGG  
GATTTTATTCCATTGCTCGGTTGGGTTGATTGGCTCAGTGGATCCAAGGCAAAGGTGGAGAAAACGGCCA  
AAGAAGTTGATGCTTTTCTGGAAGGTGCTCTTAGAGATCATATTAAGTATGGCATCGAATAAGGGTTC  
TGCAAACGACGACTTCTTATCCATTTTCTTGAGATTCGAGAGGCAGATGCTGGCTCTACTCTTGATGAA  
GAATGCATCAAAGCTATTGTCTGGGATATGATCCTTGGTGGAACCTGAAACAACATCTACAACCTTTGGAGT  
GGATAGTAGCAGCGATCATAAAAAATCCGGACGTCATGTTCAAATTGCAGAAAGAGGTGAGAGAGATAGG  
AAAAGGCAAATCAAAGATAGAAGAGGTTGATCTAGTCAAAATGAACTACCTGAAAGCAGTAATGAAAGAG  
AGTATGCGATTGTACATTACAGCATTCTACTTCCCAGAGAAGCAAAGCAGGATGTGAAATTAATGGGGT  
ATGACATAAGCTCAGGGACACAAGTACTGATAAATACATGGGAACTGCAAGAGACCCTTCATTGTGGGA  
CAACCCTGAGGAGTTCCGGCCTGAGAGGTTCTTGAATAGTCCTATTGATTACAAAGGCCTTCACTATGAG  
TATCTTCCGTTTGGAGGTGGTCGGAGGGGTTGCCCTGGTATCCAGTTTGCTATGGCTGTTAATGAGCTTG  
CAGTAGCAAATGTTGTATACAAGTTTGATTTCAAAATGCCTGATGGAGAAAGATTTCAGGATTTGGACAT  
GTCTGGAGTCCCTGGCATTAGTCTTTATCGAAAATA

## Annexe 5: Liste des amorces utilisés en RTqPCR

Nom	Sens	Séquence	Nom	Sens	Séquence
5,8s	+	AGCGGCGTCTTTCCAAAAC	GGFd14	+	GCA GGC CGC AGA ATG TG
5,8s	-	CGATGCGAGAGCCGAGATAT	GGRv14	-	GCA CTT GCT CTG CCG TAG TTG
CYP71AJ3	+	AACAATAGCGACGGCTTTGG	GGFd15	+	AGT GAA GGA GGC AAT GAG ATT GA
CYP71AJ3	-	CATGGCGTCCGACTTTTAA	GGRv15	-	TCA ATG GTT TCA TGA GGC AGA A
CYP71AJ4	+	CATTGCTCGGTTGGGTGA	GGFd26	+	GGA CAT GGA TGA GAC AAA TGC A
CYP71AJ4	-	TGGCCGTTTCTCCACCTT	GGRv26	-	CCG TGG CAA TTA GGA AAA GG
PsPT2	+	CTGCACTTGCCTTCCTCCTT	GGFd29	+	TGA CTG AAG GGC CTG GAC TT
PsPT2	-	CGTTGATTCAGGATCATTTACATCA	GGRv29	-	CGG CTG GCA TCT TAC TAG CAA
PsPT1	+	CGA AGC ACC ATG GCA GAATA	GGFd33	+	GGA TGA TGA CGA GGC AAA GC
PsPT1	-	GAC CAA ACG TGC AAA ATGCA	GGRv33	-	TCC GGC CAT AGA CAA AAC CT
PsDiox	+	GTGCCCAAACCCTGAGCTT	GGFd34	+	CAC CAC ATG AAG CAA TGG AAG A
PsDiox	-	TGTGAGTGTGGAGACGTCAGAAT	GGRv34	-	TGC CGG CTG AAA TGT GAA
GGFd1	+	TCC ACC CTC CAG TCC CAT TAC	GGFd35	+	GCG ACT CAG GTC CTG CAC TT
GGRv1	-	CCT CCA ATA ATG CGT GTT TCT G	GGRv35	-	TTC CCA GTT CAA ACC CAT GAA
GGFd8	+	AGA GCA TGC GAT TGT ACT TGA CA	GGFd41	+	GGT GCT ACA TCC CAA GAT TTC G
GGRv8	-	TTT ACG TCC TGC CTT GCT TCT C	GGRv41	-	ACCAAA CAA GTT GCC TTA ACA ACA
GGFd10	+	AGG ATG TTA AAG TTA TGG GCT ACG A	GGFd44	+	AACCAT GTT GGG AGG AAT AGA CA
GGRv10	-	CCC CAA GCG TTG ATG ATG AC	GGRv44	-	CGC CTG CAA GTA GAC CAA GTT

## Annexe 6: Alignements des séquences du gène *PsDiox* (A) et *PsPT1* (B) natives du panais avec les séquences recodées pour le clonage par GoldenBraid

**A**

1	ATGGCTCCAT	CTCCCAC	TGATTACAG	AATTTTGAG	TAACAAAGG	GCACGAGTA	AAGGGTCTCT	CGATCTGAA	GCTCGACGCT	TTACCAGAGC	AATATATTCA	ACCTGTGGAG	GAGCGATTAG
2	ATGGCTCCAT	CTCCTACTGA	TGATTACAG	AACCTTCGCTG	TTACAAAGGG	ACATGGTGT	AAAGGACTTT	CAGATCTTAA	ATTGSAATGCT	TTGCCAGAGC	AATACATTCA	ACCTGTGTAA	GAGAGGCTTG
Consensus	ATGGCTCCAT	CTCCCAC	TGATTACAG	AACCTTCGAG	TAACAAAGG	GCACGAGTA	AAAGGACTCT	CAGATCTGAA	acTcGACGCT	TtACCAGAGC	AATACATTCA	ACCTGTGTAA	GAGAGGCTTG
131													260
1	ACATGACAAA	AGTCTGAAG	AAAGAGTCCA	TACCTGTGAT	TGACATGTCC	AATTTGAGTG	ATCCGAAGGT	GCGGATCAG	ATTGCGGCTG	CTGCTGAAAA	GTGGGGATTG	TTTCAGATTA	TTAATCATGG
2	ATATGACTAA	GGTTTGAAG	AAAGAGTCCA	TCCAGTTAT	TGATATGTCA	AATCTGTGAT	ATCTTAAAGT	TGCTGATCAA	ATTGCTGCTG	CTGCTGAAAA	GTGGGGATTG	TTCCAAATTA	TTAATCATGG
Consensus	ACATGACAAA	agTcTGAAG	AAAGAGTCCA	TaCCAAGTgAT	TGAcATGTcCA	AATcTgGATG	ATCCgAAAGT	gCGgATCAcA	ATTGcGcTG	CTGCTGAAAA	GTGGGGATTG	TtCcAAATTA	TTAATCATGG
261													390
1	AGTCCCCATT	GAGGTTCTTG	AGAATGTCAA	GSAAGCTACA	CGTAGATTCT	TTGCTTTGCC	TGTGGAGGAG	AAGATCAAGT	ACACCCAAGA	ACAGTCTCCT	ACAAATCTCG	TTAGGCTTAC	AACAAGCTTT
2	TGTTCCAATC	GAGGTTCTTG	AGAATGTCAA	AGAGGCTACT	AGAAGATTCT	TCGCTTTGCC	TGTGGAGAG	AAGATTAAAT	ACACTCAAGA	ACAATCTCCA	ACAAACTCAG	TTAGACTTAC	TACATCTTTT
Consensus	agTCCaATcA	GAGGTTCTTG	AGAATGTCAA	agAGGCTACA	agAGATTCTc	TcGcTtTGCC	TGTGGAGAG	AAGATcAAAT	ACACcCAAGA	ACAATCTCCA	ACAAACTCAG	TTAGACTTAC	aCAAGcTTT
391													520
1	CTTCCAAAA	TAGACAAAGT	TCTTGAGTGG	AAAGATTACC	TTAGTATCCT	TGTTTCTGAT	AAGAAATCTT	CGGAATTTTG	GCCATCTACG	TGCAAGAATG	ATGTAAAGGA	ATATGTGAG	AAGTCTGAGT
2	CTTCCTAAGG	TTGATAAGT	TTTGAGTGG	AAAGATTACC	TTTCTATCTT	GTTTTCAAGT	AAGAAATCTT	CGAATTTTG	GCCATCTACT	TGTAAAGATG	ATGTAAAGA	ATACGTGAG	AAGTCAAGT
Consensus	CTTCCAAAA	TaGAcAAAGT	TcTgGAGTGG	AAAGATTACC	TtAGTATCCT	gTTTTCAGT	AAGAAATCTT	cGGAATTTTG	GCCATCTACG	TGcAAGAATG	ATGTAAAGGA	ATACGTGAG	AAGTCAAGT
521													650
1	TGTGATGAA	GTGGCTGCTA	AAAGCCCTGA	TGAAGGGGTT	AAATGTGGAC	ATGGAATCAA	AAGAATCAAT	TCTAATGGG	TCAACAGAA	TCAACCTCAA	CTATTATCCT	GTGTTGCCAA	ACCTGAGCT
2	TGCTATTGAA	GTGGCTTTTG	AAAGCTCTTA	TGAAGGGATT	GAATGTTGAT	ATGGAATCTA	AGGAGTCAAT	CCTTATGGG	TCAACAGAA	TTAATTGGA	TTATTACCTC	GTGTTGCCAA	ATCTGAGCT
Consensus	TGTGATGAA	GTGGCTGcTA	AAAGcCTcTA	TGAAGGGATt	aAATGTgGAc	ATGGAATCtAA	aAGaATCAAT	cCTAATGGc	TCAACAGAA	TcAAcCTcAA	cTATTAcCTC	GTGTTGCCAA	ACCTGAGCT
651													780
1	TGCAATTGGA	GTGGAGCGCC	ATTCTGACGT	CTCCACACTC	ACATTCTGCG	TTCAAGACAA	CGTCCAGGGA	TTACACATCA	GAAAGATGGA	CACAGATACG	TGGATTTTGC	TGCTCTCTGT	AGAGGGAGCA
2	TGCTATTGGA	GTGGTAGGCG	ATTCTGATGT	TTCAACTCTT	ACATTCTCTT	TGCAAGATAA	CGTGGAGGTT	TTGCATATTA	GAAAGATGGA	TACTGATACA	TGGATTTTGG	TTCCACCTGT	TGAGGGAGCT
Consensus	TGCAATTGGA	GTGGAGcCGC	ATTCTGAcGT	cTCAcAcTc	ACATTCTGcG	TgCAAGAcAA	CGTcCAGGGA	TtAcAcATcA	GAAAGATGGA	cAcAGATAcA	TGGATTTTcG	TgCcAcCTGT	agAGGGAGCA
781													910
1	ATAGTAATTA	ACATTGGTGA	TGCTCTTCAA	ATCTTGAGTA	ATGGTAAATA	CAAAAGTGCT	GAGCATCTGT	TTGACGCCAA	TGGAGCAAC	GACAGAATTT	CGGTGCCAAT	CTTTACTAAC	CCTAGTCCGG
2	ATGTTAATTA	ATATCGGAGA	TGCTCTTCAA	ATCTTGCTTA	ATGGTAAATA	TAAGTCAAGCT	GAACATTGTG	TTGCTGCTAA	TGGTTCTAAC	GATAGAATCT	CAGTTTCCAT	TTTTACTAAC	CCATCTCTGT
Consensus	ATAGTAAATTA	AcATcGGAGa	TGCTCTTCAA	ATCTTGAcTA	ATGGTAAATA	cAAAGcAGCT	GAACATtGTG	TTGcAGCcAA	TGGaaccCAAC	GAcAGAATcT	cAGTtGCCAAT	cTTTAcTAAC	cCAATcCTGT
911													1040
1	ATGATATTAT	TGGCCCGCTA	CCTGAGCTGC	TCAAGAAATGG	AGAAAAACCA	ATATACAAC	ACGTTCTGTA	TTGCAATTAT	GTGAACATT	TCTACAGGAA	ATCTCACGAT	GGAAGACCA	CCCTGTGATT
2	ATGATATTAT	TGGACCACTT	CCTGAGCTTT	TGAAGAAATGG	TGAAAAAGCT	ATATACAAGC	ATGTTTGTGA	CTCTGATTAC	GTTAAGCATT	TCTACAGGAA	ATCAACATGAT	GGAAGACCA	CATTGTGATT
Consensus	ATGATATTAT	TGGAcCAcTA	cCTGAGcTGC	TcAAGAAATGG	agAAAAcCCA	ATATAcAAc	AcGTTcTGTGA	cTcGcATTAc	GtGAaCATT	TcTAcAGGAA	ATCAcAGAT	GGAAGAcCA	CacTgGATT
1041													1053
1	CGCACAGATG	TGA											
2	TGCTCAATG	TGA											
Consensus	cGCACaATG	TGA											

**B**

1	ATGGCTCAAA	CAATTATGCA	TTACAGATTA	TCGTCGGTT	TCTTACATCT	TCAACGAGAC	AAAGGCTTTC	GTACACTTTC	AACGCAAGA	AGCATGCTA	AAGTAGTAA	CGGAGACCAA	GAATTTGCTT
Consensus	ATGGCTCAAA	CAATTATGCA	TTAcAGATTA	TcGTCAGGAT	TcTtAcATCT	TCAAGAGAC	AAAGGcTTTA	GgACTtTTC	AACAcAAGA	AGcATGCTA	AAGTAgTAA	CGGAGAcCAA	GAATTTGCTT
1	TCAGGGTGGT	TTATGCGAT	AAAAATTAG	ATTCAACAA	GAATTTTAC	GGCAGTTGTG	AAAAAGCAAT	TAGAACAC	ACAAACAAAC	TTTTCAAA	AATTAGTGC	ACATCAGACA	GAGAGGCTAT
2	TTAGGGTGGT	TTATGCGAT	AAAAATTAG	ATTCAACAA	GAATTTTAC	GGCAGTTGTG	AAAAAGCAAT	TAGAACAC	ACAAACAAAC	TTTTCAAA	AATTAGTGC	ACATCAGACA	GAGAGGCTAT
Consensus	TcAGGGTGGT	TTATGCGAT	AAaATcTAg	ATTCAAcAA	GAATTTcAcA	GgCAGTtGcG	AAaAGCcAAT	cAGAACAcAc	ACAAAcAAc	TTTTCAAA	aATTAcGcC	AcATcAGAcA	GaGAGcCTAT
1	TATACAGCCC	AAAGATGATT	ACGAAGCACC	ATGGCAGAAT	ACTTTACGGA	GAAAAAGGGA	TGCAATTTTG	ACGTTTGGTC	GTCACACTC	TGCACTCTG	ACAATTATTG	GAATAAGTTC	GGTTTCTCTG
2	TATTCACCAA	AAAGATGATT	ATGAAGCTCC	TTGGCAAAAT	ACTTTAGAAA	GAAAAAGGGA	TGCTTTTGTG	ACATTTGGAA	GACCACTACT	AGCTATCTG	ACATTATTG	GTATCTCTCT	AGTTTCTCTT
Consensus	TATAcCAcAA	AAaGATGATT	AcGAAGcAcC	aTGGCAaAAT	AcTcTaaGAA	GaaaaATGGGA	TGCAATTTTGc	AcATTTGGaa	GAcCACTACTC	agCcATCTGc	AcAATTATTG	GaATaAcTTC	aGTTTCTCTG
1	CTGCGCTTAA	CTTCGGTAA	AGATTCTCTG	GCTCCATATT	TTGTGGGATT	ACTACAGGCA	TTGATCCCAT	TTCTTTGTGC	CAACATCTAT	ACCTCGGGGA	TAAATCAACT	TGTTGATGTG	GACATAGACA
2	TTGCACTTGA	CATCAGTAA	GAATTTTCT	GCTCCTTATT	TTGTGGGACT	TTTGCAAGCT	TTGATCCCTT	TTCTTTGTGC	TAACATCTAT	ACTTCAGGTA	TAAATCAACT	TGTTGATGTT	GATATCGATA
Consensus	cTGCATTAa	CaTcGTTAA	aGATTtCTG	GCTCCATATT	TTGTgGGAcT	acTAcAGcCA	TtGATCCcAT	TtCTTTGTGc	cAACATCTAT	AcCTAcGGGA	TaAATCAACT	TGTTGATGTG	GAcATgAGAc
1	AAATAAACAA	GCTTACCTG	CCACTGGTTT	CGGAGAAAT	TTCTTGGGC	GAGGGAGAG	CAATTGTCTC	AGCACTGGCT	TTTATGTGCC	TGGCGGTGG	AATCCTATCT	CATTCTACAC	CATTGTTTGT
2	AGATTAATAA	GCCATATTG	CCCTTGTGTT	CAGGAGAAAT	TTCTTGGGA	GAGGGAGAG	CAATTGTCTC	TGCTTTGGCT	TTTATGTGCC	TTGCTGTTGG	TATTTGTACA	CATTCTACAC	CATTGTTTGT
Consensus	AaATaAACAA	GcCATAcCTG	CcAcTgGTTT	CAGGAGAAAT	TtCTTGGGA	GAGGGAGAG	CAATTGTcTC	agCACTGGCT	TTTATGTGCC	TgGcGTgGG	AATcCTATcA	CATTCTAcAC	CacTgTtCGT
1	CGGGGTTCTT	GTCTATTTT	TAAATGGAAC	CGCATATTCT	GTGAGCTAC	CTCTACTTAC	ATGGAACAA	AAACCGGCA	TGGCTGCATT	TTCTATGGC	GGCCTAATGG	GACTCAGAT	CCAGCCTGCT
2	TGGAGTTTGG	GTCTATTTT	TTATCGGTAC	AGCTTACTCA	GTGGAATTGC	CACCTTTTGA	GTGGAGACT	AAACCTGCTA	TGGCTGCTT	TTCTATGGC	GGATTGAGG	GTCTTACTAT	TCAACAGCT
Consensus	cGGGTTcTg	GtCTATTTTc	TaATcGGAAC	agCATAcTCA	GtTGAAcTAc	CaCTAcTgAG	aTGGAAcAA	AAACCGcCA	TGGCTGcATT	TtCTATGGCc	GGAcTAATGG	GAcTcAGAT	cCAcCAcCT
1	GTCTTCTATC	ACATCCAGAA	TGCATCGGT	AAACCAATGG	TATTTTCAAA	GACAGTGGCC	TTTGTACCA	TTTTCTCAG	TGCTTTTGA	GCTGTTCTTG	GAGCATCAA	GGACGTACCT	GATGTTGAAG
2	GTCTTCTATC	ATATCCAAAA	CGCTCTTGA	AAGCCTATGG	TTTTCTCTAA	GACTGTTGCT	TTGCTACAA	TTTTCTTTTC	TGTTTTCGCT	GCTGTTTGG	GTGCTATTAA	GGATGTTCTC	GATGTTGAAG
Consensus	GtCTTCTATc	AcATCCAAaA	cGcATcGGa	AAAcCAATGG	TaTTTcTCAa	GAcAGTgGCc	TtTGCTAcAa	TTTTCTTcAc	TGtTtTCGcA	GcTGtTtGcA	GAGcATcAA	GGAcGTAcCT	GATGTTGAAG
1	GAGACACAGC	GTTCGGCAAC	GGAACATTCT	GTGTGAGATA	TGGCCAAGAG	AAAGTCTTTT	CTGCTGTCT	AAATATCCTG	CTGCTTGCTC	ACGGGTTTGC	TGTGGTGGTA	GGAAGCTTCA	CCTCTGCTCT
2	GAGATACCTG	TTTCGGTAA	AGAACATTTT	CAGTTAGGTA	TGGTCAAGAG	AAGGTTTTCT	CTGTTGTGCT	TAAATATCCT	TTGCTTGCTC	ACGGATTTCG	TGTTGTGTT	GGGCTCTCTC	CATCTTTTCT
Consensus	GAGAcAcAGC	gTTCGGcAAC	aGAACATTcA	caGTgAGATa	TGGcCAAGAG	AAAGTcTtCT	cTGcTGTCTc	aAaATcCTG	cTGCTTGcCT	AcGGaTTTGC	TGTgTgTGA	GGAcGTAcCT	CaTcGTtCTCT
1	ATTATGCAAG	ATTGATCTG	TGATGGGTCA	CACACACTT	GCCTCCCTAC	TTTTGCTTGC	AGCAAAATCT	ACTAATCCGA	AAGATCCTGA	ATCAACGAG	TCCTTTTACA	TGTTTTTAT	TAAAGTATTG
2	CTTTTGAAG	ATTGATCTG	TTATGGGTCA	TACTACATT	GCTCTATTGC	TTTTGCTTAG	GGCTAGCTC	ACTAATCCAA	AGGATCCTGA	ATCAACACAA	TCCTTTTACA	TGTTTTTAT	TAAAGTATTG
Consensus	acTATGCAAG	ATTGATcCTG	TgATGGGTCA	cACTAcACTg	GcCTcAcTAc	TTTTGCTTAg	agCAAAATcT	ACTAATCCaA	AAGATCCTGA	ATCAAcAcAA	TcCTtTAcA	TGTTTcTATt	TAAAcTATg
1	TACGCGAGT	ATGCTGTAAT	TCATTTTATG	CGCTGA									
2	TATGCTGAAT	ATGTTTGTAT	TCATTTTATG	AGGTGA									
Consensus	TAcGCGAGAT	ATGcTgTAAT	TcATTTtATG	agCtGA									

**Annexe 7 : Séquences des 22 contigs issus du tri de la banque RNAseq (454) amplifiés par PCR à partir d'ADN génomique de panais**

➤ Contig C1

AGTATGGATAATGTCAGAACTCATAAAAAATCCATCTGTAATGAAGAAAGCACGAGAAGA  
AGTGAGAGGCGTTGTCAAGGGAAAAGGAGTATTCGTACATGAAAGCGATCTTCCAAAAC  
TGAGTATCTAAAGATGATTATAAAAGAGGCGCTAAGACTCCACCCTCCAGTCCCATTACT  
ACTCCCTCGTGAACTACAGAAACACGCATTATTGGAGGTTATGATATTCCAGCCAAAAC  
AAGGGTGTTTCATCAATGCCACTGCAATATCTATGGATCCAGAACTGTGGAAAGATCCAGA  
GGAGTTCAAACCAGAAAGATTCTTGAATAGTTCCATTGATTTCAGAGGCCAGCACTTTGA  
GTTATTACCGTTTGGGGCTGGTCGGAGGGGTTGTCCAGGAATCAATTTTGCAGGGCTCAT  
AGTAGAGCTTGCACTTGCTAATCTGTTGCTCAGCTTTGATTGGACA

➤ Contig C8

AAGAATGAGGTCAGAGAAGTAGGCATAGGCAAATCCAAAATAGAGGACGGTGATTTAGTC  
AAGATGAACTATCTGAAAGCAGTATTGAAAGAGAGCATGCGATTGTACTTGACAGCTCCA  
CTGCTACTTCCTAGAGAAGCAAGGCAGGACGTAAACGTAATGGGGTATGACATTAAAGCT  
AGGACACAAGTACTTATAAATGCCTGGGCAATTGACAAGAGATCCTTCTCTGTGGGAAAA  
CCCA

➤ Contig C10

ACGTTACATCCACCACTTCCGTTATTAGTCCCACGAACTGCAAGACAGGATGTTAAAGT  
TATGGGCTACGATATTGCAGCTGGAACATGGTCATCATCAACGCTTGGGGAATCGGAAG  
GGACCCTAGTTCATGGGATGAACCGGATGAGTTTAGGCCAGAAAGTTCTTGAATTCGAC  
AATAGATTACAGAGGTCATGACTTCAGTTGATCCCATTGGTGCTGGGAGGAGGGGTTG  
TCCCGGAATAAGTTTTGCCATGTCAAATTACGAGTATGTGTTGGCAAATCTTCTGCATAA  
GTTTGATTGGAACTTCCTAATGGAATGAAGAGTGATGAGCTTAATATGGATGAGGGCAC  
TGGTCTTACTTGTCTAGAAAAGTCCCTCTACTTGTAGTTGCAACTCCAATTTCTTGTTG  
AGCAGACTGTCTTTTATTCTTCTATCATGTGATGCAGTTCGGGTCCTCTCTGTTTCAGGT  
CCATTTAATGATTTTACAAAATAAAGGTCAAGAATACAAAGATAATTAAATAGWAMWWAA  
AAAAAAAAACAAAAAAAAAAAAA

➤ Contig C12

ACAGCTACAGTCCTAGAATGGGCAATGTCGGAACCTTGAGACATCCATCTGCTATGAAT  
ACTGTGCAAACGGAGATCAGAGAAGTTCTTGCTAGCAGACCTGATATAACAGACGATGAT  
TTGGAGAAAATGCAGTATCTGAGAGCAGTGATCAAAGAACTCTCCGTCTGCATCCACCA  
ATTCCGTTGTTGGTTCCAAGATCAGCACGCGAAGATGTAGAAAGTAAATGGCTACGACAT  
TGCTGCTGGAACAATGGTTATAATTAATGCCTGGGCAATAGGAAGAGACCCTGCAACATG  
GGATGAACCAGAGGAGTTTCAGCCAGAGAGGTTCTTGAATTCCTCTATTGATTTCAAAGG  
ACAAGATTTTCAACTGATTCCCTTTGGAGCTGGCAGAAGAGGCTGCCCTGGAATCGCTTT  
TTCCATGGCCACCAATGAATTTGTGTTGGCAAATCTTCTGCATAAATTTGACTGGAAGTT  
GCCTGACGGTAGAAACGGGGAGGATTTAGACATGAGTGAAAGGCCTAGTTTAACAATCCA  
AAGAAAAGTTCCTCTCCTTGCTGTGCGCAACTTCATGTTGCTGCTAGTGTAACGTAAAAA  
ATACCAAGATTGTCTTCAGCTCCATCACCTTAAGGTAGCTTTTGGAGTTTGGATGTTCTG

ATGAAAGCTTGATGTTGTGGAAGTTGTCTTTGTTGGTTGAATAAAATATGATTCCATCGA  
AAAAACAGACCGCATATCTGTTTAATGGAACGATCAATTGTAGTACGAATAATATATGATT  
CATTTAGGGTCATCCGCTATTTCGCAGTGAGTGACAGGCC

➤ Contig C13

AAAAARARAAAATTTGAAAACCTTGCCTCCAGGTCCTCCTCAGCTTCCCATCATAGGCAAC  
CTCCATCAACTTGGAACCTGCCTCATGTTTCTCTGTACAAGCTGTCCCAGGAATATGGC  
CCTGTCATGCACGTAAACCTAGGCCAAGTCCCTGCTCTGGTCGTACGTCCCCAGAGATG  
GCAAAAGAGGTGCTCAAAGTTCAAGACACAAAGTGCTGCAATAGGCCTGATTCTTATGGT  
ATGAGAAAATTGTCGTACAACCGAAAGGACATATCTTTTAGTCCTTATGGTGATTACTGG  
CGAGAGATGCGTAAGCTTTGTGTTATAGAACTTTTACTGTCAAGAGAGTTTCGTTCAATT  
CAACATGTTAGAGATCGAGAGATTGCTAAGTTTGTAAACAGTCTCTCAAAGAAGCTTTA  
GACCCTAAGAGTAAATTCATYCCCCCTTGATACAAAAATATACTCACTTGCAAAAGACATT  
ACTTGTGAGGTTGCTTTTGGCACTAATTTGAACGGGAGAACTTTAGGGAAACAGAAATA  
CGAAAGACTTTACTGGATGCCCTAACAGTGATTAGTGGATTTTGTGCAGCAGATTTCTTC  
CCATATTATGGATGGATAATTGATATATTAACAGGGTTTCGCCAGAAGCTGAGAAAGTGC  
TATCTTGTGTTGATAAAATTTATGAGACTGTTATACAAGAACATCTTAGAACCTTGTAG  
GCCGGGTTTCGCAGTGAGTGACAGGCC

➤ Contig C14

ATCGAGAAACCACTGAAAAATGTGTGGTAAGTGGCTATGATATCGAAGCTAAAACATGTG  
TATACGTAAACACATATGCTATAGGAAGGGATCCTGAATGCTGGGAAAAATCCAGACGAGT  
TTATACCCGAAAGATTCATTAATAGTAGTATTGATTTTAGAGGACACGATTTTGAGCTGA  
TACCATTTCGGAGCAGGCCGCAGAATGTGTCCAGGGATATCAATGGGAGCAACTACGGCAG  
AGCAAGTGCTTGCTAATCTTCTTTACTCTTTAACTGGGAAATGCCTCCCGGTAAATATA  
GGGAAGATATGGACATGGCCACGCTTCCTGGTTTAAACGCGCACAGAAGAATCATCTTT  
GTCTTGTCCTCCAAAATAATTCATAATACATGTGCTAGTTTGGATTATTATATGATCAGT  
ATGTTTACTGAGAGCATCTCCGATGATTGTCACCTCTGAAGGGATGCCAACTTTGCCACA  
TWMKYARWRATRTKAYARRCYATAAGTCTCTCTCTTTTCGTAATTAACAAAAAAAAAAAA  
AAAAAAA

➤ Contig C15

GGACCATGACTTTATTAATTAAGAACCCACAAAGATGAAGATAGTACGACAAGAAGTCA  
GAGAGTTAATTGSAGWWGAAAGGATTTGTAGACGAAAACGATGTTCAAAGCTTGATTAT  
CTTAAAGCCGTAGTGAAGGAGGCAATGAGATTGAACCCCCCTGTTCCACTTCTTCTGCCT  
CATGAAACCATTGATAAATGTGTTATAAGTGGCTACCGAATTGAAGCTAAACCTCGCGTA  
TACATAAATGTTTATGCTATCGGAAGGGATCCCGAATTTTATGAAAATCCTGAAGAGTTT  
ATCCCCGAAAGATTCATGAATAGTAGTATTGATTTTAAAGGCCAGCATTTTCRAGCTGATG  
CCATTTGGAACAGGACGGAGAATTTGTCCAGGGATGTCAATGGGAGTGGCTACGACAGAG  
CTCGCGCTAGCTAATCTTCTCTACTCTTTAACTGGGAATTGCCTCCCGGYAAGAAGAAG  
GAAGATATTGACATGGCTGGTCTGCCTGGYGTGTAGTGCAGAAGAAGAATMATCTCYGT  
CTTGTMCCAGAARTTATAAACYAAGTATCACAGTCAWCACWTGCACATATWWATKTATKT  
STTATKAACAHATTWTGTGTTTTGTTAACAGACTATTCTGGCTATGTTTGGATCATTGTC  
ACAGTATGTTTTGTGTAATGTATTTGTAG

➤ Contig C16

ATTGAGATTGAACCCCCCTGCTCCACTTCTTGTTCCTCGAGAAACAATTGAAAAATGTGA  
GATTAATGGGTACGAAATCGAAGCTAAACACTTGTATACGTCAACGCATATGCAATAGG  
AAAGGATCCCGAATCTTGGGAAAAATCCCAACGAGTTTTTACCCGAAAGATTCATGAATAG  
TAGCATCGACTTTAAAGGGCAGGATTTTGAGCTATTACCATTTGGGGTAGGCCGGAGAAT  
GTGTCCAGGACTAGTAATGGGAGTGGCTACCACGGAGCTAGCACTTGCTAATCTTCTCTA  
CTCTTTTAACTGGGAATTGCCTCCCGGTAAGAAGGTGGAAGATATTGACATGGCTCCCCT  
ACCTGGTCTAACAATGCACAAGAAAAATCATCTTTGTCTTGTCCCAGAAATCATTAATTA  
AGTATCACTAGATATTTAAATGTGTTATGTGTCATGCATCAGTTGGATTTACACAAGAAA  
CACCAAGTGTAATGTAGTGCGTGTGGTTTTTANATTCAGAGTATGTTATGTGTAATAATT  
TTGTTATAAATCAATGTAGCAGAATGTTCTAGTAAGTAAAAAAAAAAAAAAAAAAAAATAAA

➤ Contig C18

GTGTAAAAATTTGCAAGATATCATGGCAGTTAAGTTAAACTAAGATTTTTACGACATTCT  
AACTTGTAGCCTTCAACTAGTAAGTCCGTTTCAGATCCTAGTTATTGCCAAATATTGTTT  
GATTAAAGCTGTCCTTGTCTGCTAGATAGGATCAAAAAATATCTGAAACCATTTTCCCAG  
TGCAATGATTTTCAACCACCAAAGAGAAATGTAAGTCACTAGTTTTTTGTTGACGAACTAAACC  
AAGTCATCTTTAAGAAGTTCTATGGACCTGCATCAGTAAAGAAGATAGTCACACAACCAT  
CGAAAATGCCTGCTTTTTTCTCAGATAATAGAGCAACTAGGAAATTCACCGCTATTAAAAAC  
CCTTCGTTCTCTCASTTGAATTTTATCTTCAACGATCTTCCTCATCAAACGACTCTTTG  
CAACTTCAGTTAGCCAATAAAAAACTACCACCTTCTCCGTCGAGAGTCCCAATCATCGGA  
CACTTGCATAAATTAGGCAGATATCCGCATCTCTCACTACCAAAGTTGTCACAGAAATA  
TGGACCGATCATGCTGCTACGCCTTGGTAGCGTACCGACACTCGTTAGTCTACTTCTGCT  
GAGGCGGCTCAAGAGATCATGAAAACATGATTTAGCATTTGCAGATAGGCCTAGAT

➤ Contig C20

ACAGCAGCTTTTGGGAAAAAAGTTGATGAATGGGGGAAAGGCAGAAAGTTATGGATGCTA  
ATAAAGGAGACAGAGGAATTACTTGGCCGATTTGATGTTGGAACCTTTACTCCATGGCTT  
GCATGGATCAATACTGTTTCAGGTTTCAATGCCAAAGTAAAAACATTGCAAAAGAGAAC  
GATGAATTTTTCGAACGGATTGTGGAAGAGCACATAAATGCTTCAACAAGAAAGAGAAA  
GGGGGTGAAGAAAATTAAGAAACAGAGGTTTTTCTCGATATCTTGCTTGGAGCACAAGA  
GGATGCCATCGGTGGCTTTCATATTGATAGAGATGGCATCAAAGCTCTGATCCTGGATAT  
GAT

➤ Contig C25

AAATAAGCATAAAAAATCTTTCCTTTGCTGTGCTTATATACATTTCCCTTCCTAATTAAT  
CCTTCTTAGTTAAATCTATACTCTTCTTCACATCACCTTGCATGCGCATTACATATCAAT  
GGAATTCCCATCCCCCTTCTCTATTATATCTTCGCTCATCTTTATTCTTGTCATATTCCA  
ATTGGTTACGAAATCGAAACGAAAAGCCAAATCTAATCTTCCTCCAGGGCCATGGAAACT  
GCCAATCTTGGGGAACCTTACTCCAAGTAGCTGGCAAAATTCCGCACCTGGGTCTCAGAGA  
ATTAGCAAAAAATATATGGACCTCTGATGCACCTACAACCTAGGTGAAATTTCCGCAATTGT  
TATTTCCAATCCAAGAGTTGCCAAAGAAGTACTAAGAACGAATGATCTTTGCTTCGCTGA  
CAGGCCAGAATTGATACTAGGCACAATCATTTTAGCCAATTGTGCGTGACATAGTCTTGGC  
TATCTATGGCGATTATTGGAGACAAATGCGAAAAATTTGTACATTGGAGCTTTTAAGTTC  
GCATAAAGTAAGATCGTTTTCGTTTCATTCCGGGAGGATGAGGCGTGGCAACTAATTCAGTC  
CATCAAATCATCCTCAGGATTACCAGTTAATATTAGTGACAAGGTTTCCAAAGTGTCGAA

TGCCATAACTTGTAGATCCACAATTGGAGAGAAATGCAAGTACCAAGATGAACTTATAGA  
GCTAGTTGAAGAAATAGCTCACTGGGGTTTCGGGCTTTTTTCATGGCGGATTTGTTTCGTTT  
CGTAAAATTTCTTCCCCAGCTTACGGGAATGAAGCCTGCATTAAAGAAGATAAGGAGCAA  
GCTTGATGTTATCTTTGATAGTATTATCAAGGAGCACAAAGGAAAGGTTGAATAAGAGGAA  
AGAAGGAACTGAATATGCCAATGAAGAGGAAGATTTGGTTGATGTTCTTTTGAGAATTAA  
CGAGAGCCAGCGCCTAGAATTTCCAATATCATCCGGTGATATCCAAGGCCTCGTTTTGGA  
CATGTTTACCGCCGGAAGTATACGCTCTTCGGCGGTACTAGAATGGGCAATGTCGGAATT  
AATGAAACATCCGAGAGTAATGAAAAAGGTGCAATCTGAAGTCCGAGAGGTAGTGAAAGG  
GAAAGAAAGAATAGAAGAAACAGATATACAAAAATGAGTTATCTAAAACCTTGATGTCAA  
AGAAACCCTACGCTTACATGCTCCTGTTCCATTACTACTACCAAGAGAATGTAGAAAAAGA  
ATGCGAGGTAGATGGATACACTATTCTGTTGGAACCAAAGTCATCGTAAATGCGTGGGC  
GATCGGAAGAGATCCTGAATAGTTGGGCTGATTCTGAGAGTTTTATACCAGAGAGATTTCG  
AGAACAGTTCAATTGATTATATCGGAGCTAATTTGAGTTTATTCCGTTTGGCGCGGGTC  
GGAGAATGTGTGCTGGAATCTCATTGGTGTAGCTACTGTTGAGCTTCCTCTAGCTCAAA  
TGCTTTACAGTTTTGATTGGAAGCTTCCTAATGAAATGAAGCCAGAAGATTGGACATGG  
ATGAGACAAATGCAGCAACATCCAAGAGGAAAAATAATCTTTATCTATTAGCCACCGAAT  
ATCCAATTGAAGTTGTGGAATAAGATCAAGAAAAATAATRACAATCATGGMCTSGVAGTA  
AATCAAGTACTAGYCGTTTTMGCAAGTGAAGTACAKGYCGATGTTGTTTTAAATAATTTGT  
ACAACCTTATGATGTTGCTTTTATGGTCTTTACGACTAGGTCTGTTGCAAATTCATCA  
GATTCAGACTTGTCTGTTGTAAGAGGATGGCGTAACATCTCGGTCTCGTAATGAAGTAAG  
ATTTAAATTGCGGAGATAAATACGSAWAATAGGTAATTATAATTTAAAAANAAAAAA  
AATAAA

➤ Contig C26

TGATTGGAAGCTGCCTAATGATCTGAAGCCGGAAGATTTGGACATGGATGAGACAAATGC  
AGCTACTAGCAAGAGAAAGAACAACCTTTTCTAATTGCCACGGATTATCCGACTGAATC  
TGTCATTAATTTTTTTTTTATTTTTAAAGTCATACAGTCACGTTAGGAAATAAAGTATT  
GTAGCAAGGATACGTAAGTGTGAGATTTGTACGTAAGTTGAAATTTATAAGTTTGGTTGCAG  
TTTGTGAGCAAACTGTTGATCAAGTTGAAAAATCTCAACTGATTTGCTCTTCGACCTCT  
TCGTCATATTGTGAGTTCAAAAAGGCGCATTTTGGTTGATGTGTTTTAAGAAATACCCAG  
GGACCATGAATGTGAGAAACAGATACTCGGATCTCTGTTTCCAAATGAATACTTGTAAG  
TAACCGGTGAAGAATCATATTTATAATTTAAATAAAAAAAAAAAAAA

➤ Contig C29

GCGGACTGCAAAGTGGAAGGTTATTCCATCCCAGAAGGCACTATGCTCATTGTGAATCTA  
TGGGCAATTCAGAATGACCCMAAAATATGGGCAGAACCAAGCAGGTTTATCCCAGAAAGG  
TTTGAGGGCTTTGAAGGAACCAAGAGACGGGTACAAGTTTATGCCATTTGGGGCGGGAAGG  
AGAGCCTGTCCAGGTGAAGGTCTAGCAATGCGAATGATAGGGTTAGTTTTGGGGTCACTT  
CTTCAATGCTTTAACTGGGAAAGGGTTGAAAAGAATTAGTTGATATGACTGAAGGGCCT  
GGACTTACCATGCCAAAGCTCCTACCTTTGCTAGTAAGATGCCAGCCGCGTGTGACAATG  
GCAAGCCTACTAGATCAAATTTAAACAAATTGTGCTATTCAAGTATTGCCATTGCGATAT  
TGGAAGAAAGTTGCACTAAATTTGATCGTCCAAGTTGAAGCTTGAGTGTAACATTAAGG  
TCAGCTACTAAGGCGAAGACCAATATATGAACACTGCATTTGATTGTAAATGTTTTAGTA  
GTATAATTTGCATGAAAAATCGCAAGTTGCTTTAGGTTTGTTAATGGATTTGTATACTGT  
TTCTTATTGTGTTGTTTGGAAACAACTGTGGAACAATCTGGGAATATACTCCTGTGACC  
AAAAGTGAAGATAATTCAGAAGTTCAGAAGTTGGCAAGTATTGCTTTGTTCCATGATTGTTG



ATAAACTTAAATGGTTTTTGAATATGGATTGACATGCTTTAAAGAACTAAAAAAAAAAAA  
AAAAATAAA

➤ Contig C32

GCAATAGGTTTTGATATTTGCTGGATCAGATACATCAGGTGTCACAATTGAATGGGCAAT  
GTCTCTTCTACTGAATCATCCACACGTGTTGAAGAAAAGCTAGAGACGAGATAGATACTGT  
AATTGGTCAAGATCGCCTTGTTGAAGAGTCAGATCTATCTAAATTACATTATCTCCATAA  
TATTATTTTGGAGACGTTCCGGTTATTTCCACCTGCACCTTTATTATTGCCGCACCAGGC  
ATCAGCTGACTGCAAGGTTGGTGGCTATGATATTCCAGCTGGCACCTTGTATTGGTTAA  
TGCCTGGTGCATTACAGGAATCCCACGGTGTGGGATGATCCGACAAGCTTTAAACCAGA  
GAGATTTGAAGGCATAGAAGTTGAAACACACAAGTTGATGCCGTTTGAATGGGCAGGAG  
GTCTTGTCGGGGGAATGGACTAGCTCATCGTGTGGTGGGCTTGGTTTTAGCGTCCTTGAT  
ACAATGCTTTGAATGGGGAGCGAATTAGTGATGAGAAAATTGATTTAGCTAGAAGGAAAG  
GACTTACGATGACCTAAAGCTGAGCCACTAGAAGCTATGTAGTAAGACTCGCAGTGTCAT  
AGAACAAAGTTAACACCTAAGCTACCCAGAGTGTTAGAACAATACTTATTAGCTTTGACA  
TTCAATTTCTCTGTTTTTAATAAACTGATCAACTCACATACTACCAATATATGGTTTN  
CTCAATTTTTATATGTAGCTTAGTGATACTAAACAAAAAAAAAAAAAAAAAAAA

➤ Contig C33

GAGGTGAATCAGTTGCTGCACAGATTATCTACAAAACACGCGAAATGAGTTTGAAAAAGT  
AGTGTGAAACCAAACCTTGATTGATCTGTCTTCAACATCATAATGAGGATGATTGCAGG  
AAGAAGATACTACGGGGAAGTTGTGGATGATGACGAGGCAAAGCACGTGCGGGATATAAT  
TGAGGAGGTTTTGTCTATGGCCGGAGCTTCATATCCTGGCGATTTCTACCATTTTTAAG  
TTGGATTGATTACAAGAATTTAAGAAGCGTGCAATTTGTATTGTTAAGAAAATGGATAG  
CCTTTTGCAAGGGATAATTGAAGAGCATAAAGATGGTGAGGGTAGAACTCCAATGGTTG  
GTCATTTGCTTTCTCTTCAGAAATCGGAGCCTGATTCTTACAGTGATGTCATTATTAAG  
GTCTCATGGTGGTACGTACACTTTATTAGCCAATTAATAAGAGTATTAATGTTGTAAAT  
TTAAACGCAAGAAAGAAGAATAAGACCGCACATAATCGAAATGATTTAGTTGAGATGCT  
ATCCAAGTCCAATAGAGCTAAATTAGTTTCAGTAGGCTAGAGCGATCAGGATCAATGGTG  
AGAAGGGAAGGTTAATGAACATAAGACTAAAAATAGGGCTAGACTAAGATTGATGTACGA  
GGGAACATAGAAATCACAYCAACCTTTAACAGAACTGTTTCAGATTATGGAGTTACATT  
AACTATGTAAGTATCACTTGTATCATATCCNGGTATTGTATAACGGATGCCCCCTTCT  
CTGTAGAAATCAGAAAGCTAATAGTTGTACACAGACACAAT

➤ Contig C34

GCCACATGCTAAAAAAGTTCAAGATGAATTGAAAAGCACGTGCGAAGAGATCGCCAAG  
TAAATGAATCAGATATGAAGAATCTACCTTACTTACGAGCAATTGTCAAGGAAACCCTGC  
GATTATACCCAGCTGGACCACTTGCTGCACCACATGAAGCAATGGAAGATTGCACTCTGG  
CAGGCTTTCACATTTAGCCGGCACACACTTATTTGTAAACCTCTCGAAGTTGCAACGTG  
ACCCTAGCATTTGGTCCAACCCTCTAGAGTTCCAACCTGAGAGGTTTCTCGAAAAACATG  
TGGATATTGACATGTTTGGGAATAATTTTGAACATAACCATTTGGTTTCGGAAGGAGGG  
CATGCCCTGGAATCCCACTTGCACTCCAGGTGCTGCACTTGACACTGGCTCGGTTCCCTC  
ATGGATTTGAGCTTGGAAGTGTTCGGACTTGCCATTGATATGACCGAAAGTGCAGGCC  
TTAGTAACCCAAAAGCAACCACTGGAAGTTACCTTCAGGCCTCGACTGGTTGCTTCTC  
TCTATGTATGAAACAGAGCTTTCATCTACAAGTTCTGCATGTTTAGCTAAGCTACAAAGG

CAGAGTACTTCACGTCCTTTATACCAGATGTAACAAGTAAATTAGTTTGACAATAATACC  
ACTATATTACTGCTAGAACAAAAAAAAAAAAAAAA

➤ Contig C35

AAATGGACAAGTATGTAGGAAGAGATCGCCAAGTAAAGGAATCTGATGTGAAGAATCTAA  
CCTACTTGCAAGCCATTGTCAAGGAGACACTGCGATTGTATCCATCAGCACCCTATCAC  
TCCAACACGAAGCAATGGAGGATTGCACAGTGGCAGGATTCAATGTTCTGAAGGAACAC  
GGCTACTTGTGAACCTCTGGAAGATGCATCGTGACCCGAAAGTTTGGTCTGATCCTCTAG  
AATTTCAACCAGACAGGTTCTGACAAAAACATGTTAATGTGGACATTTGGGGGCAGAAT  
TTTGAGCTTTTGCCATTCCGGATCAGGACGAAGGTCATGCCCTGGGATCACATTCGCGACT  
CAGGTCCTGCACTTGACGCTTGCTCAATTGATTCATGGGTTTGAAGTGGGAACCGTCTTA  
AACTCACCCATTGATATGACCGAAAGTGCAGGAATTACTAACCCAAAAGCAACTCCACTG  
GAGGTTAACTCAGACCACGACTGCCTCCTGCAGTCTACCTATAAACCGGCATCCCCTTC  
TAGTTGTAACCTGTAATTATGTTATGTGTAATGTAAGCAACACAAGTGGTTTCTGTATG  
AGGCTAGTTCTATGTGATGGAAGTATAGGTTAAATAAAAAAAAAAAAAAAAA

➤ Contig C38

GAAACGAAATGGCTACCAGCATTCAACTGCAAGATATCATCCTGTCTTGCCCTGGATTTT  
TTTCTACAATTTTTTTATGGAGAATTTTCAGCACCTATGTCACAAGGAAGAACAACCTGTT  
GCAGACCACCAGAACCAGCCGAAAATGGCCGTTTATCGGTACCTTCATCTTCTTGAG  
CAAGCAAGATATTACATCATACTGGGAGACATGGCGGATAAATACGGGCCAATCTTCT  
CTTTAACTTGGAATTAACAGAACTTTTGTGATCA

➤ Contig C39

ATAACTTATAGAAATGGAATTTAGCAGTCAGCCTCTAGTGATTGTACTCTGTGGCCTTTT  
TTGTTTTGTTACATTTGTATGGAGGATTTTCAGCTTGCAAGCAACAAAGA  
ACAACTCTCCCCCTCAACCTGCTGGTGCATGGCCAATTATCGGTACCTTCATCTTTTAG  
GAGGCAACAAGCTGTTACATCATGTTTTCGGAGAAATGACAGATAAATATGGACCTGTCT  
TCTTTTTGCGACGAGGAGTTCACAAAACCTCTTGTTGGTAAATAGCTGGGAAGTTGCAAAAG  
AATGTTTTACAGTTCAGGACAAAGTCTTGCCTCCCATCCAAAATCTTTGCCCTGGAGT  
TCATGGGATATGACCAAGGTATGATTGGATTCTTCTTATGGAGAAAAGTGGCGTGATC  
TTCGTAAGCTAGTAATGGTTGAAATGCTATGCAACAGCAGACTTGAAAAGCTCAAGCATA  
TTCCGGAATCAGAAGTCAATTTCTTTATCAAGGGGCTTTATGACCTATGGATAAGCAAAG  
GAGCAGGAAATTCCATGCCTGTGGTAGAACTGAAAGAGAGATTTGGAGACCTAACACGA  
ATATAGTTGTGAGAATGGTTAGTAGGAAAGCAATTATTAGCTACTTCTGAGGAGTCTT  
CGCAGTGAGTGACAGGCC

➤ Contig C41

AGAAAGGCATAACCAATCTCTATCAATATGTAAGTCTCTCTTTGAAACAGCTGTA  
ATATTGTTTTCTTTTTGCTTCTTCTGCTACTATCTGCTATATCTTGCGGTCATGGACA  
CCATATGACTAAATAAACTTACAAAAGCCTCGACAAGCCTACCCGGAAGCTGGAGGGT  
GCATGGCCTATAATTGGCCACTTCCATCTGTTATGTGGCTCAAAGTTGCCTCATGTTGTG  
TTGGGGTCTATGGCAGAAAAATGTGGACCGATTTTCGCAATTAACTCGGTTGCTAGAA  
AGCGATGGTAGTAAGTGAAGGTAGCTGAAGAGTGTTATTCATCCAACGATAAGGT  
CTTCTTGAACCGCCCTTCAAAGCAKTGGAGTTGACAGGCCTAATGGCGTATAACTATGC

CATGTCTGGTCTTGGTCCATATGGAYCTTATTGGCAAAAAATGCGCAAGATTACACGAGA  
GCTTCTAAGTGTAACATATTGAGATGTTGGAACATGTTGGAATATCTGAAATAAGGA  
ATCCACGGAAGATATATATGAGTTTTGGTTAAAAAATGAAAGTTGTACGGAACCAGATGG  
TGACTATTGATATGAATCAATGGTTTGGATCCTTGACAATGAACATAGTTCTGAGAATGA  
TTGTGGGGCAACGATATAACTGGAATGAACGAAAAGGGTATTACATTTTTACAAAGTCTA  
CATAAGTATTTAATTTTCTAGCTACAGGTTGGTAGCTTTGTCTTGAGAGATTATGTTTCT  
CTTTAAGATGGTTGGACTGGGGAGACTATGAGAAGACCATGCGGAAGGCCGCAAAGGAA  
ACTGATTGTATACTTGGTGGATGGTTACGAGACCACAAACAGAGAAGAAAACTGAAAGAT  
GAACAACACGATATCATGGACATGATGCTCTCCAAAATTGATGGTGCTACATCCCAAGAT  
TTCGAAGGCTTTGACCCCGACGCTGTTGTTAAGGCAACTTGTTTGGTATGCTCTGCCTTT  
GTTTCGTCAATTGTTAATGATCCTGAATCTTTATATGTTTAAATACCTAAATATGATTGTAC  
ATATAGTAACGTCTTAACCTTATTCCGTTCTTTTCTTACTA

➤ Contig C42

GATTGGTCCCTACCACTTTACTCATTTTTCTACTTTTTTTTATTAATTTACATTTTTCTT  
AAGTCCGTGTGTCCCCGAAATAAGGTAAACAACGTGGGGGACAGAGGGAGTATTAAAG  
AAAGTTGTTGAATGGATGTTAGSSATATATATGTAAACACAATCCTCTAGGGACTTAT  
CTACTATCCACTCTSCACTCTCTTACAKCGATWSTTTGTTTAACTAACATGGATCTGTCT  
CCCAGTWTCTCCTTGCATTTTCTTTGTTTTTCCCTTTTTCTYATTGCTTACTWTCAATTC  
AAAAGATATGGTGAAACACTGTCCACYGCRAGCTCCTCAACTYCCTGGAGCATTGCCC  
TTCRTWGGCCACYTACATCTCTTAAGGGGTCAATACCCCTTGCCCCGAATCCTCGGAAAC  
AAGGCTGATAAATTTGGCCCCRTWTTCTCTMTCMGACTCGGACACCAYTATGGAGTGGTG  
RTTAGCAGWTGGGAGATGGTGAAAGAGTGTTTCACYACCAATGACAGTAATTTTTGCRAC  
TAGGCCGACTATGTTCTKTGTCCAAATACTTGGGATATGACTATGCAGCCTTTGCGTATT  
TCCCCTTATKSGCMRTGAKTGRCRSGATGTTTCGTAAGCTTGTCATCTTGAGCTCTTAGC  
CAGCCACCAGWTCGCARTGAGTCACAGGCC

➤ Contig C44

CTGATTAGTAACCCTCATGTGCTTGAGAAGGCTAGAGATGAACTGAACAACCATGTTGGG  
AGGAATAGACAAGTCGAAGAGTCCGACATTAAGAACTTGGTCTACTTGCAGGCGATTTTT  
AAAGAATCTATGCGTTTATATCCTGTTTCATTACTACTACCACCTCGTGAATCCATAGAA  
GACTGTGTTATCAGCGGTTACAATGTCCCTAAGGATACACGATTAATAGTTAATCTGTCA  
AAGATCCACCGTGATCCTCAAGTATGGCCTAATCCTGATGAGTTTCAACCGGAGAGGTTT  
TTGACTAGCCATCAAAATGTTGATGTTAAAGGCAATCACTATGAGTTGTTGCCATTTGGA  
AGTGGGAGAAGAAGGTGTCCTGGAATGTTTCTTGGGCTAAGACTTATGCAGCTGGCTTTA  
GCAAATCTGATCCACCGATTGAGTTTGAGAAAAGCAGATGATGAACCAGATGATACTAGT  
GCAACAGTTGGACTAACTAATTCTATAGCCCAATGTAAAATTCTCATACAAACCAAGCAT  
ATCTGTAGAGAAATAAGTTTTAGACAACTCTGGAAGTTAAGATTAAGTGCACACATTA  
TGTATCAATGCTACTAGATGTAATGTTAATAATTAGACG

## Annexe 8: Liste des amorces utilisés en PCR

Nom	Sens	Utilisation	Séquence
CYP71A3	+	Amplification PCR de <i>CYP71A3</i>	ATG AAG ATG CTG GAA CAG AAT CCC CTG TAC CTG T
CYP71A3	-		TCA AAC ATG TGG TGT GGC AAC CAC AAA GAT AGG A
CYP71AT A	+	Amplification PCR de la séquence codante de <i>CYP71AT(A)</i>	ATGATTATTACTTA TGCTAAA TCTGAAGAAAACAAAA TGATGGCT
CYP71AT A	-		CTGTTCCGGCAAAAAATAGGTTCTGCATCAACATTATA ACTGCTTTAATGTTATTTAAGGTGAAATCAATTGAAG CTTCAATTGATTTACCTTAAATAACA TTAAGCAGT TATAATGTTGATGCAGAACCTATTTTGGCCGGAACAG CTAACAACAATTTCTGGGTTGGAGGCAAA G
CYP71AT B	+	Amplification PCR de la séquence codante de <i>CYP71AT(B)</i>	ATG ATT ATT ACT TAT GCT AAA TCT GAA GAA AAC
CYP71AT B	-		CTA ATG GTG ATG GTG ATG GTG ACA ACA ATT TCT GGG TTG GAG G
CYP71AT	+	Reconstruction de la séquence codante de <i>CYP71AT</i>	ATGAAGTGGCTTTTGAAAGCTCTTATGAAGGGAT
CYP71AT	-		CATCAATCCACTTGCTTTGAAGACGTGGT
GGDIOX+PT1F	+	Amplification PCR de la construction [35s-PsDiox-tNos:35s-PsPT1-tNos]	ATGAGTGTATGATTTTTTGACAACTCTTTATCATTGTCAT
GGDIOX+PT1R	-		ATGATGATGATGATGATGGTATATGTTTATTTGCTTTGAATGGCA TTGCTTGCTA
Contig C1	+	Amplification PCR du contig 1 ( <i>CYP71A9</i> )	ATGCAGTTATCAGCATGGCTGCAACCT
Contig C1	-		ATGATGATGATGATGATGTTAATGAGCCTGAAACAGAGAGACC
Contig C10	+	Amplification PCR du contig 10 (MF5)	ATGCACAACTCAGACTCAAATCATCAATAAAAA TG
Contig C10	-		ATGATGATGATGATGATGTTATCGCAAA TTTATTTCTGGATACAA TCAAGTCCTGTAT
Contig C14	+	Amplification PCR du contig 12 ( <i>CYP71A8</i> )	ATGGAACCTCCGTCCTCTCCCA
Contig C14	-		ATGATGATGATGATGATGTCAGTAACCAAGATTGATAAAATTAATAACGTACGTTC
Contig C26	+	Amplification PCR du contig 26 ( <i>CYP71D98</i> )	T
Contig C26	-		

## Annexe 9: Séquences codantes des 5 candidats issues de l'analyse de la banque RNAseq Illumina

### ➤ Contig C1

ATGAGTGTATGATTTTTTTGACAATCTTCTTTATCATTGTCATCCCTATATTCCTGACAATTATTAACATGAA  
CAAAGCAGACTTCCTCCTGGGCCTAAGAACTTCCATTGGAATCTTCATCAGCTTGGTCAATTTCTC  
ATCGAACACTTCAAGGTCCTTCGAGAAAACATGGACCCCTCATGTTTCTTCAGCTCGGTTCTGTTCCAACCTCT  
AGTAGTCTCATCAGCAGCAATGGCCAGAGAAATTTGCAAGACACATGACTTAGTTTTCTCTAGCCGTCCTAGT  
TTGTGTGCCGTAAATAAGTTGGGGTACAAGAGCTCAACTATTACTTTTGCTCCCTATGGAGAGTACTGGAGGC  
AAATACGAAAGATTGCACTTGTAGAGCTACTAAGTGTAAGCAGAGTTCAATCTTTTGAGGTTATTCGGGAACA  
AGAAGTTGCATTTATGATAGAAAATGTGGATATTTCTTCATCCGTTTGATCAATTTAAGGGAATTGATGCTCT  
TGCTAGCTAATAATATTGTGCTTAGAGTTGTATTTAGTATGAAAGGGAATAATTACGGTGATAAAAATGTAAAG  
AGTGAGTTTGATGAAATCTCCGCGAAACACAAGATTTATTAGGCATGGTCAATATCGCGGACTATCTCCAT  
GGATGGGGTGGTTTAACAAGTTGAATGGAGTTGAAGCGAGATTAGATAAGAATTTCCGAGAAGTGGATAGCT  
TTTATGACAGAGCGATCCAAGAGCATCGGGAGTGCCCTACGAGGACCATAACAGATGAGCATGAAGACTCTTG  
TTGATGTATTGCTTCGAGTTGAGCTGATCCAAATCAAGATATTGCACTCACTGATGATGATGAAGGCTGT  
TTTAACGACATGTTTATCGCTGGAACAGATACTTCTGCAGCAACAATAGTATGGATAATGTCAGAAGTCAATA  
AAAAATCCATCTGTAATGAAGAAAGCACGAGAAGAAGTGAGAGGCGTTGTCAAGGGAAAAGGAGTATTCGTA  
CATGAAAGCGATCTTCCAAAACCTTGAGTATCTAAAGATGATTATAAAAGAGGCGCTAAGACTCCACCCTCCAG  
TCCCATTACTACTCCCTCGTGAAACTACAGAAACACGCATTATTGGAGGTTATGATATTCAGCCAAAACAAG  
GGTGTTCATCAATGCCACTGCAATATCTATGGATCCAGAAGTGTGGGAAGATCCAGAGGAGTTCATACCAGA  
GAGATTCTTGAATAGTTCCATTGATTTGAGAGGCCAGCACTTTGAGTTATTACCGTTTGGGGCTGGTCGGAG  
GGGTTGTCCAGGAATCAATTTTGAGGGCTCATAGTAGAGCTTGCACTTGCTAATCTGTTGCTCAGCTTTGAT  
TGGACACTGCCTGATGGAATAAATGCCGAAGACATTGATATGGAAGAGGCTGTTGGACTTACTGTCCACAAA  
AAACTCCTCTCCGTCTTGTTGCATCTAAGCCATATCTGTAGTCATGCATGTGCATCAGAAGAAGCAACTAGC  
AAGCAATGCCATTCAAAGCAAATAAAACATATAC

➤ Contig C10

ATGCAGTTATCAGCATGGCTGCAACCTCTCATTATCTTTCTCATACCCTTTTTCTTAACTATCTTCGTATTTAAA  
CCATATTTTGCTACTAAACCTGGTAAAAACCTCCCACCCTCTCCGCCAAAGCTTCCTATCATTGGCCATTTGC  
ACAAACTGGGAAGCTTACCACACCGTCCGCTCCTCAGCTTGTCTAGAAAAATATGGACCACTCATGCTACTTCA  
CCTTGGAAGCATTCCGACTCTAGTTGTTTCATCTGCTGAGGCAGCTCAAGAGATCATGAAAACTCACGATGTC  
AAGTTTGCGGACAGGGTTTGAACATATGCTAATGAGAAGCTTCTTTATTGTTATAAGGATGTGTCTATGGCAC  
CTTATGGCGACTACTGGAGAAAAATGAAAAGCTTTTGTGTTGTCCAGCTTTTGAGTAACAAAAGGGTTGAAGA  
TATGGGAACGATAAGAGAGGAAGAACTGCTTTTCTACTGAAAAATATTCTAGACTCTAATTCTTCGGTTGTGA  
ACTTGAGTGAAGTGAATCACACTCACATTCAATCTTGTTCCTACAGCAGCTTTTGGGAAAAAAGTTGATGAA  
TGGGGGAAAGGCAGAAAGTTATGGATGCTAATAAAGGAGACGGAGGAATTACTTGGCCGATTTGATGTTGGA  
ACCTTTACTCCATGGCTTGCATGGATCAATACTGTTTCAGGTTTCAATGCCAAAGTAAAAAACATTGCAAAAGA  
GAACGATGAATTTTTCGAACGGATTGTGGAAGAGCACATAAATGCTTCAACAAGAAAGAAAGGCGGTGA  
GAAAATTAAGAAACAGAGGATTTTCTCGATATCTTGCTTGGAGCACAAAGAGGATGCCATCGGTGGCTTTTCAT  
ATTGATAGAGATGGCATCAAAGCTCTGATCCTGGATATGATTGCTGGTGAAGTGCACACTACCTACACAGCTC  
TAGAATGGACAATGACGGAACTTTTAAGGCACCCGAAAGCCATGAAGGAAGTGCAAAATGAGATAAGGAATA  
TTTGCAAGGAAAAATCAACATATCTTATCCAGATGACTTGGAAAACATGCACACTACTTGAAAGCAGTGATTGAA  
GAGGCATACGGTTACATCCACCACCTCCGTTATTAGTCCCACGAACTGCAAGACAGGATGTTAAAGTTATG  
GGCTACGATATTGCAGCTGGAAGTATGGTCATCATCAACGCTTGGGGAATCGGAAGGGACCCTACTTCATGG  
GATGAACCGGATGAGTTTAGGCCAGAAAGGTTCTTGAATTCGACAATAGATTACAGAGGTCATGACTTCCAGT  
TGATCCCATTTGGTGCTGGGAGGAGGGGTTGTCCCGGAATAAGTTTGGCATGTCAAATTACGAGTATGTGT  
TGGCAAATCTTCTGCACAAGTTTGATTGGAACTTCTAATGGAATGAAGAGTGATGAGCTTAATATGGATGA  
GGGCACTGGTCTTACTTGTATAGAAAAGTCCCTCTACTTGTAGTTGCAACTCCAATTTCTTGTAAAGCAGAC  
TGTCTTTTATTCTTCTATCATGTGATGCAGTTCGGGTCTCTCTGTTTCAGGTCCATTTAATGA

➤ Contig 14

ATGCACAAACTCAGACTCAAACCTCATCAATAAAAAATGGAAGTACTACTTGTCTTGTGCAATCACTCTTCCTC  
TGTTTTTGCTCCTACTTAAAAATTACAATGCCAGCAAGTCGAATCGGCACGCACCAGGTCCCCATGGCCTCC  
CGTTGATCGGAAACATGCACCAGTTCGACAGCTTAAACACTCACCTCTACTTGCACCAGCTCTCCAAAGAGTA  
TGGTCTCTCATTTCTTTGAAGTTAGGCTCAGTTAAATCCTAGTGGTTGCCTCTGTGAGCGCCGCGAAAGAG  
GTCTTCAAATTCATGATCTTTGTTTTTGAATAGGCCCTCCTCAGTTGCCATCCAGAAGCTGTCTTACAACAA  
GCTAGGCTTTAGTTTCGCGCCATATGATGAGTACTGGAGGAGCATGAGGAAATCTCTGCTCTGCATCTCTTT  
AGTGCTAAGAGGTTGAGGTCAATTCACCTATTCGCGATGAAGAAATTGCACGAATGGTCAGAACGATAGGT  
GATGCAGCTGCAGCTGATTCCGGGGTAGTAGATTTGAGCAAACTTTTATGACTTTGACAAGCTCGATTATAT  
TTAGAATCACTTTTGGGAAAAGATATGATGATCAAGAAGATTGTCCAGAGGATGGGAATAATATGAGGAGTCA  
ATTTCAATTGGCTACTTGTGAAACTCAGGCCAACGTTGTAAGTTTCTTTCTTACAGATTATTTTCTCTAATGG  
GTCGATTTCTTGATAGAGTAAGCGGTGCATGGGCTCGGTTGGAAGAGAGTTTTAGCCAAACAGATGCTTTTTA  
TCAACAACCTCATCGATGAACATCTTCATGCATCCACTGTATCTACTACTGGAGACTGCAGCATCCTTGACATC  
TTGCTTGAGATGAAAAAGGATTCCCCTGTTTACACGTTTGATCACGTCAAAGCAATATTAATGAATATTATAGT  
TGCCGCAACAGATACAACCTGCAGCAGCTATAACTTGGGCCATGACTTTGTTAATCAAGAATCCCGCATCCATG  
AAAAAAGTGCAACAAGAGGTTAGAGATCTAGCCCAGACAGAAAGGATTTGTGATGAAGACGACATCCAAAAA  
CTTGTATCTCAAAGCAGTCGTGAAGGAGGCGATGCGATTGCACCCCCCTGCTCCAATTCTGCATCGAGAA  
ACCACTGAAAAATGTGTGGTAAGTGGCTATGACATCGAAGCTAAAACATGTGTTTACGTGAACACGTATGCTA  
TCGGAAGGGATCCTGAATGTTGGGACAATCCAGACGAGTCTTGCCCGAAAGATTTATTGAAAGCAGTATTG  
ATTTTAGAGGGCATGATTTTGAAGTGTATACCATTCGGAGCAGGGCGGAGAATATGTCCAGGAATATCAATGG  
GAGCTGCTACCACAGAGCTGGTTCTTGCTAATCTTCTCTACTCTTTTGAAGTGGGAACTGCCTGCGGGTAAAAA  
TAGGGAAGATATTGACATGGCCACGAAGCCTGGCTTAACCGTGCACAAGAAGAACCATCTCTGTCTTGTACC  
TAAAATTAGTGATCAAGTATCACGCATGCTTATCTAAAGCTGTGATGCATTGTGATGGGTTGGGCCATGCAAT  
GATGCTAGTTCACCTCATGTAATATATGGCATACAGGACTTGATTGTATCCAGAAATAATTTGCGATAA

➤ Contig 26

ATGGAAC TTCCGTCCCCCTTCTCCATTTTATCTTCTCTCATATTCGTCATATTTTTCGTCTCTTTACTCAAAAAA  
TCGAGACGAAAAATCGAAGTCGAATCTCCCTCCCGGGCCACGGAAGCTACCATTGATAGGAAACATGCACCAA  
GTCGCTGGAAAAATTCGCGACCTTGTTCTCAGAGAATTAGCACAGAATGCGGCCCTCTAATGCACTTGCGAG  
CTCGGCGAAATATCGGCGATTGTTATATCGAATCCACGGGTGGCTAAAGAAGTGCTCAAGACGAATGATCTT  
GCATGCGCCGATAGGCCCGAATTGATTTTAACTATTATTTGAAAACTGCCGTGACATAGTCTTGGCTC  
TGTACGGTGATTACTGGAGACAGATGCGGAAGATTGTACACTGGAGCTTCTGAGTGCGAACAAAGTAAGGT  
CTTTTCGTGCCATTGCGCAGGATGAGACATGGCAGCTACATCAATCGATCAGATCATCGATAACAATCAGGAT  
CACTAGTTAATATTAGTGACTTGGTTGGTAAATTGGCTAATGCCATAACGTGTAGGTCTACAATTGGAGAGAA  
ATGCAAGTACCAAGATGAGCTTGTGGAGTGTGTTGAGGAGATTGCTCATCTTGGTTCCGGGTTTTTATGGC  
GGATATGTTTCGGTCCATAAAATTCCTTCCTCTGATTACCGGGATGAAGCCCGCGTTGGAGAGAATTAGGCG  
GAAGCTTGATGTTATCTTTGATTATATTATCAAGGAGCATGAGCAAAAATTGAGGAACCGGAAAGAAGGATCT  
GAAGTTGTGGCTGAAGAGGAGGACTTGGTCGATGTTCTTCTGAGAATTAACCAGACTCAGCGCCTTGAGTTT  
CCCATTTCTTCGAATGACATCCAAGGCCTTGTTCCTGATATGTTTACGGCTGGAACCGATACCTCTTCAGCAG  
TGCTCGAATGGACAATGTCCGAATAATGAAAAACCCGAAAGTAATGAAGAAGCTGCAAGCCGAAGTGAGGG  
AAGTAGGGAACGGAAAAAGAACGAATAGACGAAACAGATATACCAAGAATGAGCTACTTAAAGCTGGTGGTCA  
AAGAAGCCTTACGTTTACATGCACCAGTTCCATTGTTGCTACCAAGAGAATGTGCAAAAGAATGTGAAATCGA  
TGTTTACACAATCCCGTAGGAACCAAGGTCATCGTAAACGCGTGGGCAATCGGAAGAGACCCCGAGTACT  
GGCCTAATGCCGACAGTTTCATACCCGAGAGATTTCGAGAACAGTTCGGTGGATTATACAGGGGCTAATTACG  
AGTTTCATTCCGTTTGGCGCTGGCCGAAGAATGTGTGCTGGAATCTCGTTCCGGGGTAGCTACCGTTGAGCTTC  
CGCTAGCTCAATTAGTCTATTCTTCGATTGGAAGCTCCCTAATGAAATGAAGCCGGAGGATTTGGACATGGA  
TGAGACTAATGCAGCTACGAGCAAGAGGAAGAATAATCTATTGCTAATTGCTACGGATTATCCGATTTCGATCC  
GTGAATTAATTTTAAATTGAAGTCTTGTCTACTGTCTACGTTAGGAAATAAATTATAGCAGAGAGTTGCCGA  
GAGTTTATTGTTGTTCAATTTACATGTGTAATATCTGTCTACTTAATTATTATTCACCAGTAACGAGTAAGGTG  
CAGAGCTGAGGAACGTACGTATTAATAATTTATCAATCTGGTTACTGA

➤ Contig 35

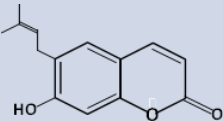


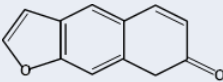



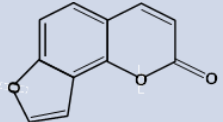


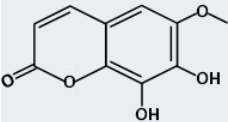


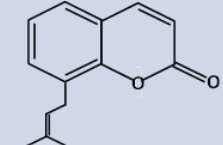


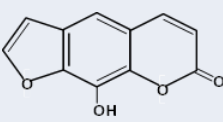


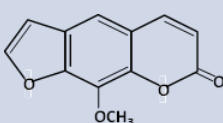


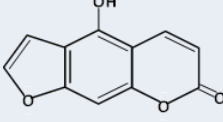


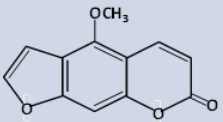


ATGGGTATTTACCTTCAAATGCAAGATATCATACTGTTTTCCCTTGATTTTTTTCTACTCTTATTTTATGGAGA  
ATTTTATAGTACCTATGTCATAAGAAAGAAGACTTGCTCCGGACCACCAGAACCAGCTGGAAGATGGCCATTAA  
TCGGTCACCTCCATCTTCTTGAGGAAGCAAGATACTGCATCATATACTTGGAGACATGGCTGACGAGTACG  
GGCCAATCTTCTCTCTGAACCTGGGAATCAACAAAACCTGTTGTGATCACTAGTTGGGAAGTTGCGAAAGAATG  
TTTTACAACACAAGACAGGGTTTTTGCCACTCGTCCGAAATCTGTTGTTGGACAAGTAGTAGGATACAACAGT  
AGAGTCATGATATTTTCAACAATATGGAGCATACTGGCGTGAGATGCGTAAGCTTGCAATAATTGAGCTTCTAT  
CCAATCGCAGACTGGATATGCTCAAACATGTCCGGGAATCAGAGGTGAATCTTTTCATCAAGGAGCTGTATG  
AGCAATGGTCAGCCAATGGGAATGGAAGCAAAGTTGTGGTGGAAATGATGAAAAGGTTTGGAGATCTGACAA  
CAAATATCGTGGTCAGGACAGTGGCAGGTAAAAAGTATTCTGGTACTGGTGTACATGGAAATGAAGAGTCGA  
GGCAATTTTCAAGGCCATGGCAGAATTCATGCATTTAGGGGGGTTGTTAATGGTTTCTGATGCACTCCCT  
ACTTGGTTGGATAGATACAGTTAAAGGATGCAAAGGCCAAAATGAAGAAGACTGCAGAAGAGATTGATCATATA  
CTTGGAAGCTGTTTAAAGGAGCATCAGCAGAAAAAGAACAAATATCAGCAATAATCATTCCGAGGACGACTTCA  
TTTATGTTATGCTGTCTGCCATGGACGGGAACCAAGTTTCTGGAATTGACACCGATACTGCCATCAAGGGCA  
CTTGTTTAAAGTCTTATCTTAGGTGGTTATGATACGACATCAGCCACACTTATGTGGGCGCTCTCACTATTACTA  
AACAACCGTCATGTGTTGAAGAAAGCACAAGATGAAATGGACCAGTACGTGGGAAGGGATCGCCAGGTAAA  
GGAATCAGATGTGAAGAATTTAACTTATTTGCAAGCCATTGTCAAGGAAACACTGCGATTATATCCAGCAGCA  
CCACTATCAGTCCAGCACAAAGCAATGGCAGATTGCACGGTGGCTGGATTCAACATTCCAGCAGGCACACG  
GTTAGTTGTGAACCTCTGGAAGATGCATCGTGACCCAAAAGTTTGGTCTGATCCTCTAGAGTTTCAACCAGAA  
AGGTTCTGCAAAAAACATATTAATGTGGACATATGGGGTCAGAATTTGAGCTCTTACCATTGGATCAGGAC  
GGAGGTCATGCCCTGGTATCACATTCGCGATGCAGGTCTCCACTTGACACTTGCGCAATTGCTTCATGGGT  
TTGAAGTGGGAACTGTCTTAGACTCGTCCATTGATATGACTGAAAGTTTCAAGGATTACCGACCCAAGAGCAAC  
GCCACTGGAGGTTACTCTGACCCCGGACTGCCTCCTGCAGTCTACCAATAAAACATCACCTTTAGTAGCTT  
AATCTTCCCTCGGAGATCCCGAGTTGTACGACTTTTATCTATCAAATAAGGTGCAAATATGATCCTTTGTATCA  
CTTGAAGCAATAATATTTAAATACTAAATGAAGCTCTATAATTGAACACTGTTTTAAAAAAAAAAAAA

## Annexe 10: Séquence codante du gène *CYP71AT*

### ➤ *CYP71AT*

ATGATTATTACTTATGCTAAATCTGAAGAAAACAAAATGATGGCTTTCTTAACTTATCTCTATCTTGT  
TGTTGGTGTTCCTGTACTACTCATACTTGTTTTTCCCAAACCTTCTCGGTAACATAACATCGAAACTG  
CCACATCCACCAGGTCCGCGGGGACTTCCGTTTATCGGAAACTTGCATCAGCTTGATATCTCCAT  
CCTTCCACACCAACTATGGAGCCTTTCAAAACAATATGGTCCAATGTTGTATCTCAGGCTTGGTTA  
TTTACCAGCTCTTATAATCTCTTCAACGGAAATGGCTAAACAATTGTTGAAATCGCATGACTTGATT  
TTCTGTAGCAGGCCTAACTTCTTTGGGTTAAGAAAAGTTTCGTACGATGGCACTGATATTGCTTTT  
GCTCCGTATAATGATTATTGGAAGGAAATGAGAAAGATTCTCACCGTTCATGTCTTTAGCTACAAA  
AGAGTACAGTCCTTCCGTTCGATTTCGCGAAGAAGAAGTCTTTCGTATGATTAAACGAATTTCTGAA  
AAGGCTTCAGAAAATGAATTGTTAAATTTGAGTGACATCATGATGCCACTTACAAACACTATAACTT  
GTAGACTTGCTTTTGGAAAGAAGTTTGAAGAAGTTGAAATGAAAAGATTTGAAGGGTACTAAAAA  
GATTAGGGCTTGTAATGGCAACTTTCTACTTTTCTGACCATTTTCCTTTGTTTGGATGGCTTGATAG  
GTTTACTGGATCTTCTGCTCGGCTGGAAAGAAGTTTCAATGACATGGATCTGTTCTATCAAGAACT  
AATTGATGAACATCTCAGCCCGTGTAGGCCATGTTTCGATGGACGGGGATGTCATTGACATCTTAC  
TTCAGTTAAAGAAGGATGGCCAGATATCTTCAATTGATTTTCACCTTAAATAACATTAAAGCAGTTAT  
AATGTTGATGCAGAACCTATTTTTTGC CGGAACAGAGTCAACAGCAAACACAATAATATGGATCAT  
GACAAGCCTGATAAAGAACCCGGGAGTGATGCGCAAAGCACAGAAGAAGTGAGGAATATGATG  
CATGGAAAAGAGAGAATATATGAAGAAGATCTTCAAAATTTAAAGCTTCCTTATCTAGAAGCAGTA  
ATCAAGGAAACCATGAGAATGTACCCCATTTCTCCCTTATTCCGCGAGAATCGTTGGATAGGTT  
GGTTGTAAATGAATATTTAATCGAATCTAAAACCTATAGTCTATGTGAATGTTTTGGCTATTCATAGG  
GATCCTAAGGTTTGGGAAAATCCAGATGAATTTTTGCCTGAGAGATTCCTCGACAATGACATTCAA  
TATAAAGGACACGATTTTGAATATATTCCATTTGGGACTGGTCGCAGGATTTGCCAGGAATGAAT  
ATGGCTGTTACGACCCTGGAGCTAGTTATTTCAAACCTTCTATACTCTTTCGATTGGGAGCTGCCT  
CCTGGTATAAAGAAGGAAGACATAGACACTGAAGATGGCCATGTGGATGGTCATGGTGTAACATT  
GCATAAGAAATACCCGCTTTGCCTCCAACCCAGAAATTGTTGTTAG

**Annexe 11 : Liste des substrats utilisés pour le criblage métabolique de CYP71AT, de C10 et de CYP71AZ4.**

Substrat	Structure	Poids moléculaire (g/mol)	CYP71AZ4	CYP71AT	C10 (MFS)
DMS		230,263			
Psoralène		186,16			
Angélicine		186,16			
Fraxétine		208,169			
Osthénol		230,263			
Xanthotoxol		202,16			
Xanthotoxine		216,189			
Bergaptol		202,165			
Bergaptène		216,19			





## Etude de la synthèse des furocoumarines chez le panais par des approches d'ingénierie métabolique et de multi-omique

Les plantes sont soumises durant leur vie à de nombreux stress environnementaux. Face à ces contraintes, les végétaux ont développé au cours de l'évolution différentes stratégies. La plus emblématique est la mise en place du métabolisme spécialisé, représenté par une grande diversité chimique et fonctionnelle. Bien que ce métabolisme soit de plus en plus étudié ces dernières années, de nombreuses lacunes persistent à son propos, liées notamment (i) à la complexité des modifications métaboliques engendrées par la perception de stress, (ii) aux coûts et avantages que ces métabolites imputent à la plante les accumulant, et (iii) aux voies métaboliques menant à cette diversité de composés. Pour appréhender ces différentes problématiques, nous avons adopté une stratégie combinant des approches de phytochimie, de biologie moléculaire et de génétique. Dans un premier temps, nous avons étudié les changements métaboliques globaux engendrés par l'application de deux stress environnementaux, l'ozone et la blessure mécanique, sur une plante modèle au laboratoire, le panais, en fonction du temps. Les résultats de ces travaux nous ont permis d'identifier 40 métabolites différenciellement accumulés dans ces conditions, dont certaines furocoumarines. Par la suite, nous avons focalisé notre étude sur ces molécules en évaluant leurs profils d'accumulation, en condition de stress par blessures mécaniques, par la biais d'analyses différentielles. A partir de ces données, nous avons initié la recherche et l'identification de gènes candidats potentiellement impliqués dans cette voie à partir de plusieurs banques transcriptomiques et génomiques de panais. La fonction des gènes sélectionnés a été évalué par des approches d'expression hétérologue dans la levure. En parallèle de ces travaux, nous avons développé une stratégie destinée à mieux comprendre le coût métabolique de la synthèse de métabolites spécialisés. Pour ce faire, nous avons adapté aux furocoumarines une technique de clonage multigénique permettant de transférer dans une plante, et en une seule opération, plusieurs gènes impliqués dans la même voie de biosynthèse. Cette méthode nous a permis d'initier la génération de lignées stables ayant intégré les deux premiers gènes de la voie. Ces plantes seront comparées à des plantes sauvages et permettront ainsi d'étudier les coûts métaboliques et physiologiques de l'introduction de cette nouvelle voie de biosynthèse ainsi que ses bénéfices en termes de défense de la plante.

**Mots clés:** Furocoumarines, *Pastinaca sativa*, métabolomique, blessures mécaniques, ozone, voie de biosynthèse, clonage multigénique, coût métabolique, gènes candidats.

## Study of furocoumarin synthesis in parsnip using metabolic engineering and multi-omic approaches

Plants are subjected to many environmental stresses during their life. Faced with these constraints, plants have developed different strategies during their evolution. The most emblematic is the establishment of a specialized metabolism, represented by a great chemical and functional diversity. Although this metabolism has been studied more and more in recent years, many gaps remain, related in particular (i) to the complexity of the metabolomic changes generated by the perception of stress, (ii) to the costs and benefits that these metabolites impute to the producing plant, and (iii) to the metabolic pathways leading to the diversity of compounds. To cope with these different issues, we adopted a strategy combining approaches of phytochemistry, molecular biology and genetics. First, we studied global metabolic changes caused by the application of two environmental stresses, ozone and mechanical wounding, on parsnip. The obtained results allowed us to identify 40 metabolites differentially accumulated under these conditions, including some furocoumarins. Subsequently, we focused our study on these molecules by evaluating their accumulation profiles under mechanical wounding stress condition, using differential analyzes. From this data, we initiated the search and identification of candidate genes potentially involved in this pathway based on transcriptomic and genomic parsnip libraries analyses. The function of the selected genes was evaluated by heterologous expression approach in yeast. In parallel to this work, we have developed a strategy to better understand the metabolic cost of specialized metabolites synthesis. To do this, we have adapted a multigene cloning method to furocoumarines, allowing to transfer several genes involved in the same pathway in a plant, in a single operation. This method allowed us to initiate the generation of stable lines having integrated the first two genes of the pathway. These plants will be compared to wild plants and will thus allow to study the metabolic and physiological costs of the introduction of this new biosynthetic pathway and its benefits in terms of plant defense.

**Keywords:** Furanocoumarins, *Pastinaca sativa*, metabolomic, mechanical wounding, ozone, biosynthesis pathway, multigenic cloning, metabolic costs, candidates genes.