



UNIVERSITÉ  
DE LORRAINE

BIBLIOTHÈQUES  
UNIVERSITAIRES

## 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 bibliothèque : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)  
*(Cette adresse ne permet pas de contacter les auteurs)*

## 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

École Doctorale "Sciences et Ingénierie des Ressources Naturelles"

Unité Mixte de Recherche 1121 Université de Lorraine-INRAE

Laboratoire Agronomie et Environnement

## Thèse

en vue de l'obtention du titre de

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

présentée par

**Alexandre Bouillé**

## Les furocoumarines : Cas d'évolution convergente et d'adaptation des plantes à leurs environnements

Thèse soutenue publiquement le 28 novembre 2024 sous la présidence de  
Pr. **Nathalie Guivarc'h** et présentée devant un jury composé de :

PR, <b>Alain Goossens</b>	VIB - University of Gand	Rapporteur (BE)
CR, <b>Gea Guerriero</b>	LIST	Rapportrice (LU)
PR, <b>Nathalie Guivarc'h</b>	Université de Tours	Examinateuse (FR)
PR, <b>Cinzia Comino</b>	University of Torino	Examinateuse (IT)
PR, <b>Alain Hehn</b>	Université de Lorraine	Directeur (FR)
CR, <b>Romain Larbat</b>	INRAe - IRHS	Codirecteur (FR)





# Remerciements

Je crois que ça y est, après 4 années d'intenses moments et d'une énergie sans faille, c'est la sanction, docteur. Si la thèse n'est qu'un passage, elle m'aura permis tant de rencontres, de partages et de liens indéfectibles. Il paraît que c'est l'occasion d'en parler.

Une thèse, ça ne sort pas de nulle part. En ce sens, je remercie particulièrement le Ministère français de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) et la région Grand Est qui ont co-financés mes travaux de recherche. Je remercie également, l'Université de Lorraine, l'École Doctorale Sciences et Ingénierie des Ressources Naturelles (SiReNA) et le Laboratoire Agronomie et Environnement (LAE) pour m'avoir accompagné tout au long de ce long chemin administratif et académique qu'est la thèse.

Je souhaite une bonne lecture à toute personne qui lira ce manuscrit. Je remercie sincèrement **Alain Goossens** (Professeur, Vlaams Instituut voor Biotechnologie (VIB), Université de Gand, Belgique) et **Gea Guerriero** (Principale associée en recherche et technologie, Luxembourg Institute of Science and Technology (LIST), Luxembourg) qui ont accepté d'examiner mon travail de recherche en tant que rapporteur.rice. De la même façon, je remercie sincèrement **Nathalie Guivarc'h** (Professeure, Laboratoire Biomolécules et Biotechnologies Végétales (BBV), Université de Tours, France) et **Cinzia Comino** (Professeure associée, Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA), University of Torino, Italie) pour leur évaluation de mon travail en tant qu'examinatrices. Enfin, je tiens à remercier **Arnaud Hecker** (Professeur, laboratoire Interactions Arbres-Microorganismes (IAM), Université de Lorraine, Nancy) et **Sébastien Besseau** (Maître de conférences, BBV, Université de Tours) pour leurs évaluations annuelles et leurs conseils avisés durant 3 ans dans le cadre de mes Comités annuels de Suivis Individuels de Thèse (CSIT).

J'ai beaucoup entendu dire que faire une thèse c'est devenir des extensions de ces encadrants. Si c'est le cas, je suis fier d'avoir pris un peu de vous deux **Alain Hehn** et **Romain Larbat** ! Si la passion pour la Science ne me fera jamais défaut, je suis très reconnaissant de la patience, la bienveillance et surtout la liberté que vous m'avez permise pour faire de cette thèse une aventure qui me ressemble. Vous avez réussi à m'insuffler les codes et la rigueur qui me faisait défaut et ça, ce ne s'était pas une mince affaire. S'il est toujours difficile d'appréhender la confiance, celle que vous m'avez apportée était palpable. Je tiens à vous dire que ça a été l'un des moteurs de mon acharnement dans des manips qui n'ont pas été simples. Avant de la commencer, je voyais la thèse comme un champ d'exploration, dont j'avais de fortes attentes. Vous avez rapidement compris que c'était vital pour moi de ne pas restreindre la Science à la Recherche et m'avez laissé aller au-delà de toutes mes exigences. C'est un cadeau qui restera pour moi une fierté. Enfin, vous l'entendrez sûrement pour la première fois, mais pour moi faire une thèse c'est avant tout apprendre à faire le deuil du deuil. À vos côtés j'ai pu apprendre à en faire de nombreux (ce qui est un élément positif, il faut le préciser), et j'en ressors plus avisé. Je garderais de cette aventure difficile, votre soutien sans faille qui a rendu accessible, le passe-droit à la vie que je souhaite mener, merci !

Parmi les aventures de ma thèse, celle de pouvoir creuser l'évolution de P450s en 3D en Angleterre reste inoubliable. Je suis très reconnaissant à **Janet Thornton** de m'avoir accueilli dans son équipe dans le très prestigieux centre européen de biologie moléculaire et de bioinformatique à Cambridge (EMBL-EBI) qu'elle a dirigée pendant de longues années. Être le dernier étudiant de l'une pour ne pas dire La pionnière de la biologie structurale computationnelle c'est un grand cadeau. Merci Janet, d'avoir été présente et si humble dans ma découverte de ce monde inespéré pour moi. Je garderais de précieux souvenirs de nos échanges très humains et d'une Science à la hauteur de ta sympathie. Merci également à **Rashmi Kumari** (postdoc) qui m'a accompagné dans ce projet, ainsi qu'à toute l'équipe du « Thornton group » **Ioannis G Riziotis, Raymund Hackett, Neera Borkakoti, Roman A. Laskowski, António Ribeiro, Marcia Aanhi Hasenahuer** et **Sophia Tsoka**. Je commence maintenant grâce à vous à comprendre ce qu'est la bioinformatique. Je tiens à remercier l'Université de Lorraine et son dispositif

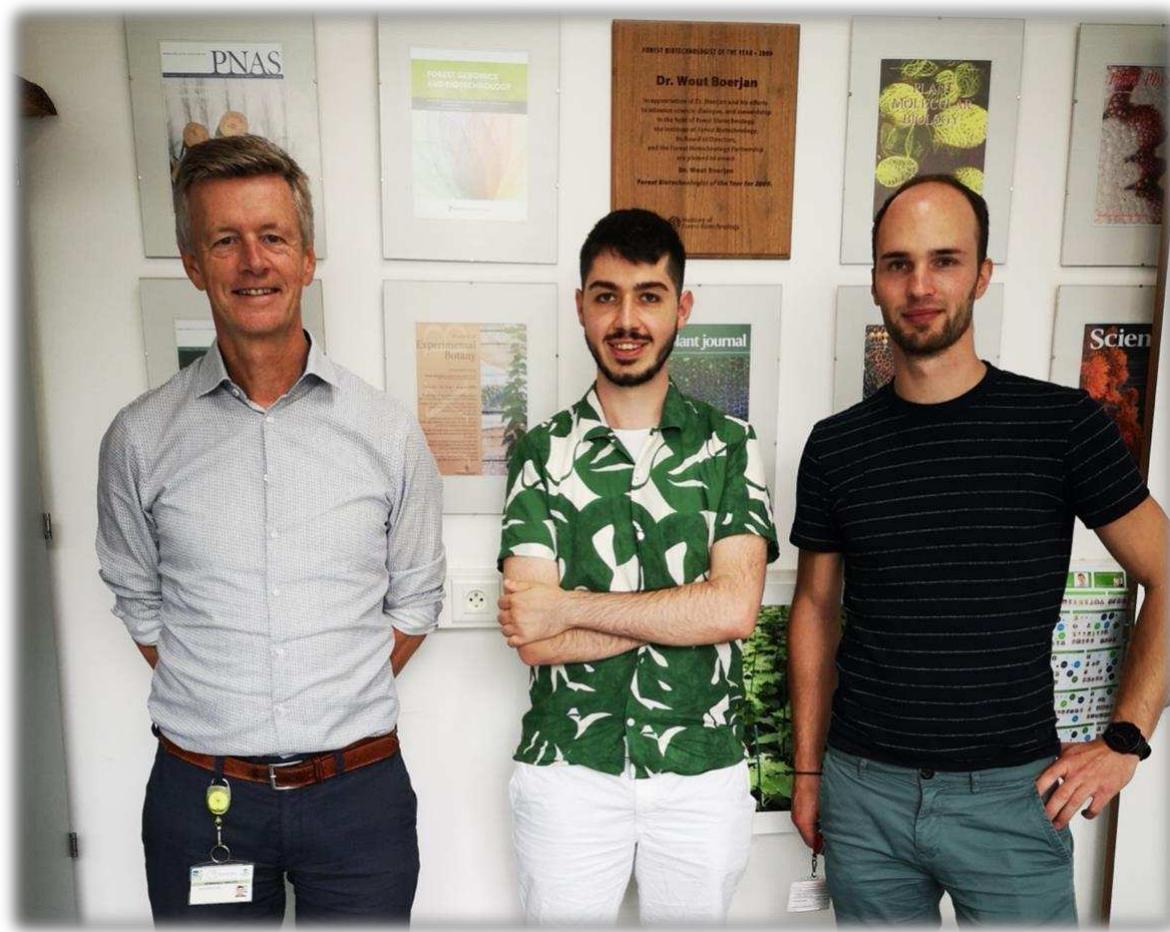
LUE DrEAM (Lorraine Université d'Excellence - Doctor, Explore and Achieve More) sans qui cette mobilité de 3 mois n'aurait pas pu exister.



**Repas festif avec le Thornton group à l'EMBL-EBI à Hinxton, UK en 2023**

Une autre aventure incroyable aura été celle de pouvoir franchir les limites de la constitution de la lignine de mes tomates avec **Lennart Hoengenaert, Wout Boerjan** (VIB, Gand, Belgique) et **John Ralph** (Professeur, Department of Biochemistry, Wisconsin-Madison University, USA). Je remercie très sincèrement Lennart, avec qui la recherche devient tellement facile. Un grand merci à **Wout** pour son accueil et à toute l'équipe Bioenergy avec qui j'ai passé 2 semaines formidables dans la belle ville de Gand.

Un grand merci également à **John** pour nos échanges très enrichissants sur la RMN de la lignine. Merci au service de l’Institut National de la Recherche en Agronomie et Environnement (INRAe) sans qui cette mobilité n’aurait pas été possible, et surtout au soutien financier des services de la Direction Enseignement Supérieur, Sites et Europe (DESSE) et notamment à **Irina Vassileva**. Je remercie également les services INRAe pour l’expertise et l’accompagnement notamment dans les échanges de matériels biologiques dont j’ai pu bénéficier et notamment à **Jeremie Hagenmuller**.



**Wout et Lennart de l’équipe Bioenergy du VIB à Gand, Belgique en 2023**

Je me tourne maintenant vers les membres du LAE pour vous exprimer toute ma reconnaissance. Mine de rien, c'est 5% de ma vie que j'aurais passé avec vous, que ce soit pour des P450s, des tomates ou tout autre chose. Je tiens tout d'abord à remercier l'équipe des techs comme ils aiment se nommer. Merci à **Clément Charles** pour sa prose et son flegme de tout instant, **Aude Fauvet** pour sa résilience aux maux tomate,

**Jérémy Grosjean** pour ses analyses qui dépasserais presque le sport, **Julie Genestier** pour sa bienveillance décapante, à **Thamara Olivier** pour sa bienveillance et sa spontanéité qui marque plus que ce qu'elle ne peut l'imaginer, à **Jodie Thenard** pour sa vision réaliste de la recherche, à **David Marcolet** pour la joie que tu apportes aux moments des repas et à **Claude Gallois** alias Mr. bricolage. Je te remercie **Alexandre Olry** pour ta spontanéité, ta gentillesse, tes conseils et les myriades d'éléments auxquels tu as répondu présent. Je te remercie **Christophe Robin** pour nos échanges, le professionnalisme de tes conseils ainsi que pour ta présence sans faille. Tous ces éléments m'ont porté et apporté tant sur le plan scientifique, que pour la passion de la médiation scientifique qui nous anime tous les deux, mais également dans l'aventure politique et organisationnelle du Symposium National du Houblon qui n'a pas été de tout repos. Je te remercie **Sophie Slezack** pour ton aide dans l'approche de métabarcoding menée sur mes tomates. Je remercie également **Armand Guckert**, le fondateur du LAE pour nos échanges que j'ai particulièrement appréciés et qui dépassent le simple cadre de l'Histoire des Sciences. Merci à **Dominique Laurain Mattar**, **Alan Kerguteuil** et **Alice Michelot-Antalik** pour nos échanges aussi bien scientifiques qu'informels. Enfin merci à l'équipe de direction pour faire vivre le LAE dans toute son identité et sa richesse.

D'un point de vue scientifique, je remercie **Cloé Villard**, **Gianni Galati** et tous les autres chercheu.r.se.s de la lignée des furocoumarines au LAE, dont le premier **Frédéric Bourgaud**, pour avoir pavé de gènes et de savoir le chemin qui m'a conduit dans cette belle histoire. Je tiens également à te remercier **Ryosuke Munkata** pour ta bienveillance, tes conseils sur la transcriptomique du figuier et pour ta confiance. C'était un honneur de pouvoir former l'un de tes doctorants du Research Institute for Sustainable Humanosphere (RISH) de Kyoto sur la technologie GoldenBraid.



Collègues du RISH de Kyoto au colloque Phytofactories au Luxembourg en 2023

Le LAE c'est aussi et surtout des amis ! Un immense merci à toutes les personnes qui sont passées dans ces murs lorsque j'y étais. Tout particulièrement, je tiens à remercier **Omar Hafidi** pour tous les moments que l'on a passé ces 4 années ensemble, ce n'est pas rien ! Merci à toi Florent Ducrocq membre honoraire du bureau où tout a commencé. Une pensée particulière pour **Andréïna Nuttens** ancienne postdoc du LAE qui a profondément marqué mon passage dans ce bureau. Une pensée également pour **Jessica Amaral** avec qui on a passé de sacrés moments pour l'expression de P450s en levure, mais pas que. Un grand merci à **Camille Gay, Stéphane Pinck, Eugène Maurey, Élisabeth Villard, Nicolas Bras, Nicolas Kahn, Florent Magot, Corentin Bossé, Andy Laprie, Loïc Girardeau, Lucile Auger, Sausthène Evrad, Emma Soulé, Éric Stell, Komla Amegan, Guillaume Bodjrenou, Tania Akponikpe, Abdelhak Rouabah, Floriane Colas** et tous les autres. Il y a des amitiés qui perdurent longtemps et des expériences que l'on n'est pas prêt d'oublier. Un grand merci aux collègues de la filiale Cellengo et Plant Advanced Technologies (PAT) pour leur soutien, leurs conseils et leur

expertise qui m'a été d'un grand secours tout au long de ma thèse. Je pense notamment à **Cindy Signe, Sissi Miguel, Ludivine Hocq** et **Frédéric Bourgaud**. Enfin et surtout, je tiens de tout cœur à remercier **Marwa Roumani**, sans qui cette thèse n'aurait clairement pas été la même. Tu m'as guidé et m'as aidé à prendre la confiance qui devait m'être légitime, je suis très fier d'avoir pu me construire à tes côtés !



**Une partie de l'équipe Métabolisme Spécialisé du LAE en 2022**

Si la thèse est une expérience scientifique, quoi de plus essentiel que de partager de façon accessible les savoirs et les vécus. En ce sens, je suis très reconnaissant d'avoir pu faire mes débuts en enseignement durant une année dans la spécialisation biotechnologie de l'École Nationale Supérieure en Agronomie et Industries Alimentaires (ENSAIA) en tant qu'Attaché Temporaire d'Enseignement et de Recherche (ATER). Je remercie notamment **Alain Hehn** de m'avoir permis d'accéder à ces fonctions comme je le souhaitais. Je remercie chaleureusement **Stéphane Delaunay**, pour son accompagnement durant cette année qui m'a très largement permis de me sentir à ma place dans cette mission nouvelle. Enfin, je remercie tous les étudiants de la Spé Biotech et notamment les étudiant.e.s qui ont participé à mes projets tutorés, je pense notamment à **Juliette Braganti-Coral, Camille Chassin, Jeanne Couderc, Bleuenn Suard, Valentin Mrugalla, Ségolène Bouche** et **Damien Mergoux**. Je suis très heureux

d'avoir pu contribuer très modestement à votre devenir. Il est évident que je le suivrais avec attention.



**Promotion 2022-2023 de la spé Biotech à l'ENSAIA, Vandoeuvre-lès-Nancy**

Au-delà des cours, je suis très reconnaissant des dispositifs de grandes ampleurs et de la confiance que nos institutions portent sur nous les jeunes chercheu.r.se.s. En ce sens je tenais à remercier le comité d'organisation de l'édition 2022 du concours international Ma Thèse en 180s (MT180s) et notamment à **Catherine Flauder**. Lors de cette édition, je me suis qualifié jusqu'en finale régionale. Un moment inoubliable ! La vidéo : (<https://ultv.univ-lorraine.fr/video/11687-ma THESE-en-180-secondes-2022-alexandre-bouille-les-furocoumarines-cas-devolution-convergente-et-dadaptation-des-plantes-a-leur-environnement/>).

Je tiens également à remercier **Meriem Fournier** qui, dans le cadre de ces fonctions de présidente de centre INRAe Grand-Est Nancy, m'a invité pour communiquer sur mes projets de recherches lors du 50<sup>e</sup> anniversaire du Centre. Un moment inoubliable, où j'ai

pu échanger avec les pionniers de l'agroforesterie et des symbioses mycorhiziennes comme **François le Tacon** et où j'ai même eu le droit à une parodie improvisée de mes recherches par une compagnie de théâtre. Je remercie également **Laetitia Nori** de la Vigie de L'eau qui nous a embarqués avec plusieurs collègues doctorants dans une parenthèse de médiation scientifique à Vittel, que ce soit auprès de collégiens et du grand public. Une aventure de médiation portée par le projet international de médiation scientifique Experimentarium.



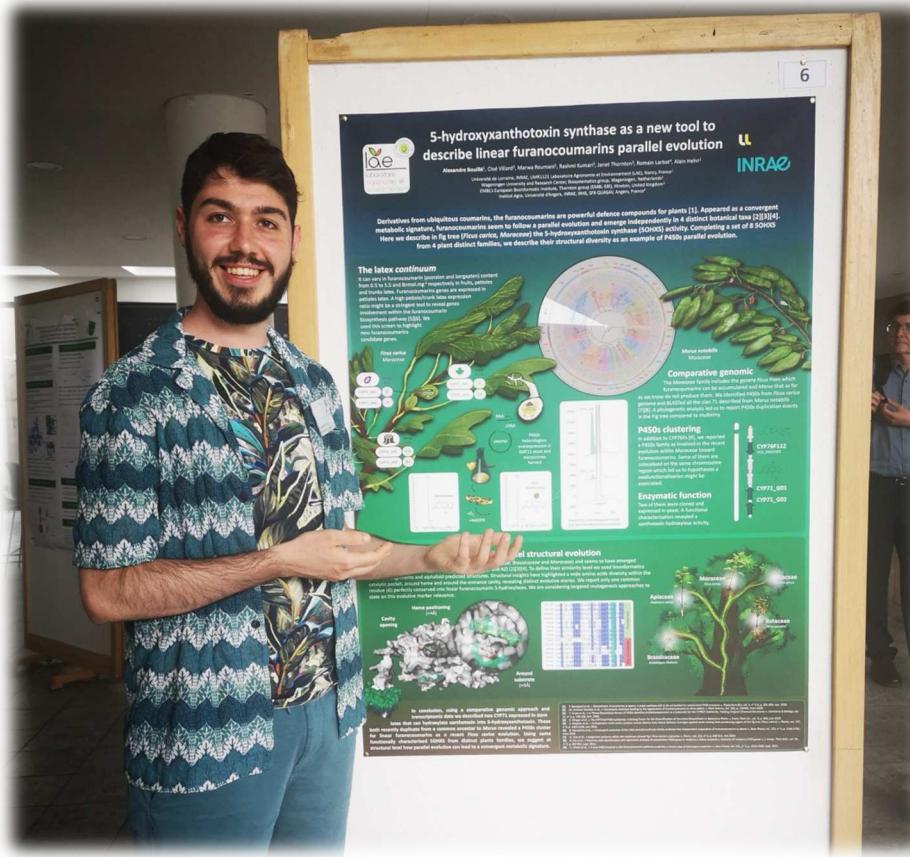
**Finale régionale du concours MT180s (en bas à gauche) avec son illustration (en haut), et échange avec les collégiens de Vittel dans le cadre du dispositif Experimentarium (en bas à droite) en 2022**

La culture scientifique c'est aussi tisser les liens entre chercheur.se.s. Je suis très reconnaissant à la directrice du réseau national MétaSP (réseau de spécialistes français du métabolisme spécialisé des plantes notamment), **Nathalie Guivarc'h** de m'avoir permis de cocréer avec **Julie Buges** (Institut Charles Violette, Université de Lille) le réseau des MétaSP en herbe dès 2021. Dans le cadre d'une expérience très enrichissante d'un an, j'ai pu apprendre à diffuser le savoir scientifique dans une communauté bienveillante, mais également de développer les bases d'un réseau dont les contributions perdurent au travers notamment de la gestion de réseaux sociaux et de webinaires mensuels (<https://sites.google.com/view/metasp/metasp-en-herbe>). Un grand merci pour la confiance et le respect de nos engagements. Merci également aux autres doctorants avec qui j'ai pu partager cette aventure, dont **Léo Bellenger, Jennifer Perrin, Manon Ferrier et Anne-Sophie Paguet**.

Je souhaite également remercier les organisateurs des colloques qui ont accepté que je présente mes résultats plus largement à la communauté scientifique, dont notamment : La 11<sup>e</sup> édition des journées MétaSP à Saint-Étienne les 12-13 mai 2022 où j'ai présenté à l'oral la capacité du tabac à produire du psoralène de façon hétérologue ; La 15<sup>e</sup> édition du colloque international « The cytochrome P450 biodiversity and biotechnology » à Copenhague les 27-30 juin 2023, où j'ai présenté mes travaux sur l'évolution des CYP71Bs et est remporté la deuxième place du meilleur Poster ; Le colloque Phytofactories au Luxembourg les 7-9 juin 2023 où j'ai présenté à l'oral mes premiers résultats sur la capacité de la tomate à produire du psoralène de façon hétérologue et les impacts sur sa croissance et son métabolome.



Présentation des travaux au colloque Phytofactories au Luxembourg en 2023



Présentation de mon poster au colloque P450 à Copenhague en 2023

La recherche c'est aussi le maillon qui accompagne les changements de la Société. En ce sens, avec les jeunes chercheur.se.s du LAE et dans le cadre initialement d'une journée scientifique en interne, nous avons développé le premier Symposium National du Houblon qui a regroupé plus de 200 personnes. Je remercie l'ensemble du groupe qui m'a laissé prendre les rênes du projet durant un an et qui m'a aidé à coconstruire ce projet très médiatique qui restera dans les mémoires. Je remercie **Séverine Piutti** et **Alain Hehn** pour votre soutien dans cette aventure hors norme. Je remercie **Florent Ducrocq** pour s'être prêté au jeu de la mascotte. Je remercie également **Claire Bergerot** et **Catherine Flauder** pour tous vos conseils dans les aspects communication sur le plan régional comme national. Je remercie la filière houblon et notamment **Matthieu Luthier** (directeur général de la filière InterHoublon) et **Antoine Wuchner** (directeur commercial chez Association Générale des Producteurs de Houblon (AGPH) - HOP France) pour les interactions qui ont permis d'aboutir à cet évènement tout en définissant les contours de la recherche pour la filière houblon en France à l'horizon 2035. Je remercie également plusieurs membres du gouvernement pour leurs soutiens symboliques dans cette initiative bien qu'ils n'aient pas pu se joindre en définitive à nous, dont **Marc Fesneau**, **Élisabeth Borne** et **Emmanuel Macron**. Je remercie également, le pôle Scientifique Agronomie, Agroalimentaire, Forêt (A2F) et le LAE pour avoir contribués au financement de la partie scientifique de ce beau projet. Enfin, je tiens à remercier particulièrement **Edwige Elmer-Laurent** d'avoir répondu favorablement à mon invitation lors de cet évènement. Une rencontre qui m'a beaucoup inspirée.



**Introduction du Symposium National du Houblon en 2022 à Nancy**



**La Dream Team du Symposium National du Houblon en 2022 à Nancy**

La recherche c'est aussi la formation et la reconnaissance des engagements. En ce sens, je remercie l'Université de Lorraine pour les formations doctorales qu'elle propose. Je remercie l'École International d'Agreenium (EIR-A) parmi laquelle je fais partie de la promotion 2021-2023 et grâce à laquelle je suis sanctionné du label d'excellence de la recherche en Agronomie pour avoir notamment suivis deux semaines de séminaires sur « Les enjeux de la bioéconomie » à Reims les 19-22 avril 2021, et sur "les transitions numériques en agriculture et en agroalimentaire" à Nantes les 25-29 avril 2022, ainsi que pour avoir investi 3 mois de recherches à l'international. Je remercie également l'Université de Lorraine de m'accorder le Label de la recherche doctorale européenne pour avoir conduit 3,5 mois de recherche en Europe.



**Atelier de formation entrepreneurial dans le cadre de l'EIR-A à Nantes en 2022**

Je remercie ensuite le groupe communication du LAE et notamment **Christophe Robin** et **Andréïna Nuttens** avec qui j'ai pu cocréer la page LinkedIn du LAE et que je gère jusqu'à maintenant. En 3 ans d'existence, la page a plus de 800 abonnés, ce qui permet au LAE de faire rayonner ses activités vers la Société. Merci à la direction du LAE de m'avoir permis d'élaborer cette vitrine extérieure.

Je tiens à remercier également les membres du pôle A2F qui m'ont permis de m'engager dans la vie politique et scientifique de l'Université de Lorraine et notamment à **Alain Hehn** pour m'avoir incité à le faire. Merci également à **Omar Hafidi, Anaïs Thomas** et **Oyut Dagva** d'avoir été colistier.ère.s en tant qu'élus doctorant.e.s de cette aventure durant 2 ans.

La thèse c'est également le maintien de liens qui permettent de tenir le cap. Je remercie infiniment **Nicolas Barbault** et **Benoit Monnereau** pour ces liens qu'on entretient depuis la licence. Nous voilà maintenant tous 3 docteurs. Je remercie également **Léo Bellenger, Cassandre Maillard** et **Maxime Aubourg**, qui chacun pour des raisons différentes ont été des soutiens et des amis de tout instant.

Enfin, je garde ces derniers mots pour ma famille sans qui cette thèse n'aurait pas la même saveur. Vous vous demandez sans doute toujours dans quoi je m'engage, pourtant votre soutien à permis que j'en sois là. Je pense bien sûr à ma sœur, **Marie Bouillé**, et à ma mère, **Sylvie Bouillé**. Tu n'as sans doute pas compris à quoi pouvaient bien servir mes recherches, mais la plus spécialiste de mon aventure de thèse ce n'est que toi maman. Un grand merci pour ton oreille attentive de chaque instant. Cette thèse je te la dédie pour grande partie. Enfin, je dédie à titre posthume cette thèse à mon père, **Richard Bouillé**, à mon grand-père, **Fernand Ray**, et à ma grand-mère, **Simone Ray**, qui n'ont pas pu voir ou même entrevoir cette partie de ma vie qui leur aurait permis de comprendre quel été le monde qui est le mien.

À vous tous, merci infiniment pour les espoirs partagés !



## Table des matières

<b>Remerciements .....</b>	<b>2</b>
<b>Liste des Figures .....</b>	<b>19</b>
<b>Liste des tableaux .....</b>	<b>21</b>
<b>Liste des abréviations.....</b>	<b>21</b>
<b>Chapitre 1 : Synthèse bibliographique .....</b>	<b>24</b>
<b>I. Les multiples dimensions du métabolisme spécialisé : Mieux comprendre la physiologie des plantes et l'évolution de leurs interactions avec l'environnement..... 24</b>	
1-1 Du concept de métabolisme spécialisé à la diversité phytochimique connue.....	24
2-Émergence et évolution fonctionnelle du MS des plantes .....	37
3- Théorisation de la gestion des ressources par la plante pour mieux percevoir le rôle physiologique des MS.....	50
3- L'exemple des furocoumarines chez les Moracées pour décrire l'évolution du métabolisme spécialisé des angiospermes .....	58
<b>II. Les enzymes de la famille des cytochromes P450s .....</b> <b>73</b>	
1- Acteurs majeurs de la biosynthèse de MS chez les plantes .....	73
2- Structure et fonction d'un P450 .....	77
<b>Enjeux et objectifs du projet de thèse.....</b>	<b>82</b>
<b>Chapitre 2 : Comprendre l'impact de l'insertion de la voie de biosynthèse hétérologue « p-coumaroyl-CoA-psoralène » sur la physiologie de la tomate..... 84</b>	
1- Contexte scientifique et stratégie de recherche .....	84
2- Résultats .....	85
2-1- Création de la construction « p-coumaroyl CoA - psoralène » (historique) .....	85
2-2- Validation expérimentale et fonctionnelle de la construction PP chez le tabac .....	87
2-3- L'intégration de la voie du psoralène chez la tomate conduit à une accumulation de coumarines et limite la croissance (Article 1) .....	91
2-4- Impact de la voie du psoralène sur la composition de la lignine (données complémentaires de l'article 1) .....	133
2-5- Impact de la « surproduction » de coumarines dans la lignée PP sur la résilience à la carence ou à l'absence de fer .....	135
3- Discussion et perspectives .....	138
4- Conclusions du chapitre 2 .....	145
<b>Chapitre 3 : Mieux comprendre l'évolution de la voie de biosynthèse des furocoumarines chez Ficus carica..... 146</b>	
3.1- Contexte scientifique et stratégie de recherche .....	146
3.2- Résultats.....	147
3.2.1- Les CYP71Bs : Outils pour mieux comprendre la voie de biosynthèse des FCs et son évolution .....	147

3.2.2 - Using lineage-specific patterns to understand convergence of enzymatic functions led to the identification of Moraceae-specific P450s involved in furanocoumarin biosynthesis (Article 2) ....	149
3.2.3- Étude complémentaire concernant CYP71B131a .....	180
3.2.4- Caractérisation fonctionnelle de CYP71AN90, CYP71AN94 et CYP71A164 chez <i>F. carica</i> .....	182
3.2.5- À la recherche d'autres P450s impliqués dans la synthèse des FCs .....	185
3.2.6- Organisation génomique : Présence d'un cluster de biosynthèse pour les FCs ? .....	186
<b>3.3- Discussion et perspectives .....</b>	<b>189</b>
3.3.1- Caractérisation de la voie de biosynthèse des FCs chez <i>F. carica</i> .....	189
3.3.2- Mieux comprendre l'évolution de la voie des FCs pour mieux la décrire.....	193
<b>4-Conclusion du chapitre 3 .....</b>	<b>196</b>
<b>Conclusions générales et perspectives .....</b>	<b>197</b>
<b>Matériel et Méthodes.....</b>	<b>201</b>
<b>Matériel .....</b>	<b>201</b>
1- Espèces modèles.....	201
2- Plasmides.....	202
<b>Méthodes .....</b>	<b>215</b>
1- Expression hétérologue dans <i>Nicotiana benthamiana</i> .....	215
2- Cultures <i>in vitro</i> de tomates .....	218
<b>Annexes.....</b>	<b>220</b>
<b>Bibliographie .....</b>	<b>227</b>
<b>Analyse bibliométrique .....</b>	<b>253</b>
<b>Résumé .....</b>	<b>255</b>
<b>Summary.....</b>	<b>257</b>

# Liste des Figures

Figure 1: Exemple de structures chimiques de métabolites « primaires » (vert) et « secondaires » (violet) retrouvés chez certaines plantes. ....	25
Figure 2: Schéma simplifié de l'évolution de la perception par la communauté scientifique du métabolisme spécialisé (MS) des plantes et de ses disciplines d'étude depuis le 19e siècle à nos jours en parallèle de quelques avancées scientifiques majeures. ....	28
Figure 3 : Diversité phytochimique au sein du règne du vivant d'après une analyse de la base de données « Dictionary of Natural Products ». ....	29
Figure 4: Aperçu des 7 voies métaboliques et 70 superclasses du jeu de données NPClassifier. ....	31
Figure 5: Schéma d'un itinéraire technique pour le traitement des données non ciblées par spectrométrie de masse à haute résolution et l'exploration statistique simple. ....	34
Figure 6: Nombre d'espèces végétales avec un génome séquencé et publiquement disponible en janvier 2021 (n = 798) par rapport au nombre attendu pour chaque ordre. ....	39
Figure 7: Mécanismes de duplication génique.....	42
Figure 8: Schéma de l'histoire évolutive globale du métabolisme des plantes, avec les estimations des principales périodes et événements géologiques (Ga, Ma ou Ka).. ....	43
Figure 9: Évolution du métabolisme spécifique aux plantes, essentielle pour l'évolution des plantes terrestres.....	46
Figure 10: Schéma simplifié de l'évolution de la perception de la gestion des ressources par les plantes au travers de théories descendante et ascendante de 1932 à nos jours. ....	51
Figure 11: Signaux en amont et régulateurs de la signalisation TOR chez les plantes... ..	57
Figure 12: Schéma de la position des carbones sur le squelette des coumarines (en haut) et des furocoumarines (FCs) linéaires (à gauche) et angulaires (à droite).....	59
Figure 13: Représentation simplifiée des étapes décrites de la voie de biosynthèse des furocoumarines (FCs) et des pyranocoumarines (PCs) chez certaines angiospermes..	63
Figure 14: Phylogénie des Moracées. ....	67
Figure 15: État de l'art des ramifications possibles du squelette des furanocoumarines (a,b) linéaires et (c) angulaires chez les Moracées. ....	71
Figure 16: Phylogénie des familles de cytochromes P450 (CYP) chez les algues vertes et les plantes terrestres. ....	77
Figure 17: Représentation simplifiée de la structure tridimensionnelle d'un P450 (CYP71B131a) et de sa Cytochrome P450 Réductase (CPR) (Atr1) à la surface d'un réticulum endoplasmique.....	78
Figure 18: Cycle catalytique d'une hydroxylation typique médiée par P450. ....	81
Figure 19: Vérification de l'expression hétérologue transitoire en tabac des gènes de la construction pDGB1Ω1_PP après 5 jours d'infiltration.....	88
Figure 20: Suivi cinétique de l'expression hétérologue dans les feuilles de la GFP entre 4 et 6 jours post-infiltration avec de l'eau stérile ou des agrobactéries contenant les constructions pDGB1Ω1_PP et pEAQ-HT_GFP.....	89

Figure 21: Vérification de la production de psoralène après 1 jour de supplémentation en p-coumarate dans des feuilles de tabac, à 5 jours post-infiltration avec de l'eau stérile ou des agrobactéries contenant la construction pDGB1 $\Omega$ 1_PP.....	90
Figure 22: Analyse par Résonnance Magnétique Nucléaire (RMN) de la composition de la lignine de tige des lignées Wild-Type (WT) et “Psoralen Pathway” (PP) de <i>Solanum lycopersicum</i> comparativement à la lignée accumulatrice de scopolétine SCOP1 chez <i>Arabidopsis thaliana</i> (Spectres RMN HSQC (Heteronuclear Single Quantum Coherence) 1H-13C à courte portée, pour un temps d'acquisition de 6h).....	134
Figure 23: Exemple de cultures en condition <i>in vitro</i> d'individus des génotypes Wild-Type (WT) et « Psoralen Pathway » (PP) en milieu complet (bleu), carencé en fer (rose) ou sans fer (vert). .....	136
Figure 24: Schéma simplifié de la voie de biosynthèse des coumarines décrivant les étapes à partir des dérivés de l'acide <i>t</i> -cinnamique.....	140
Figure 25: Prédiction de la structure tridimensionnelle des canaux d'entrée (T3C1) et de sortie (T4C1) de CYP71B131a et du cheminement de la xanthotoxine dans le cadre de sa fonction 5-hydroxyxanthotoxine synthase (5OHXS) (à droite). .....	180
Figure 26: Effet inhibiteur du métabolome soluble foliaire sur CYP71B130 et CYP71B131a dans le cadre de la fonction 5-hydroxyxanthotoxine synthase (5OHXS).....	182
Figure 27: Produits de métabolisation des enzymes CYP71AN94, CYP71AN90 et CYP71A164 en présence du cofacteur NADPH et du substrat psoralène.. .....	184
Figure 28: Produits de métabolisation des enzymes CYP71AN94, CYP71AN90 et CYP71A164 en présence du cofacteur NADPH et du substrat xanthotoxol.....	185
Figure 29: Schéma de l'organisation chromosomique et du profil d'expression dans le latex des cytochromes P450s (P450s) des familles CYP71s et CYP76s dans le génome de <i>Ficus carica</i> (Moracées).....	189
Figure 30: Schéma simplifié des étapes de biosynthèses identifiées dans la voie des coumarines et des furocoumarines linéaires chez <i>Ficus carica</i> (Moracées).....	190
Figure 31: Substrats métabolisables par CYP71AN24 en condition <i>in vitro</i> .....	193
Figure 32: Schéma de synthèse de l'évolution potentielle des fonctions de quelques P450s dans la voie des FCs linéaires chez <i>F. carica</i> . .....	194
Figure 33: Carte du plasmide linéaire pCR®8/GW/TOPO™. .....	203
Figure 34: Carte plasmidique de pYeDP60. .....	204
Figure 35: Carte plasmidique de pYeDP60_GW®.....	205
Figure 36: Cartes plasmidiques des vecteurs pDGB1 $\alpha$ 1, pDGB1 $\alpha$ 2, pDGB1 $\Omega$ 1 et pDGB1 $\Omega$ 2.....	210
Figure 37: Schéma simplifié de l'assemblage du plasmide pDGB1 $\Omega$ 1_PP.....	211
Figure 38: Carte plasmidique du vecteur pDGB1 $\Omega$ 1_PP.....	212
Figure 39: Carte plasmidiques du vecteur pSOUP.....	213
Figure 40: Carte plasmidiques du vecteur pEAQ-HT_GFP. .....	214
Figure 41: Photographies de la préparation des plants (a), des dispositifs d'agroinfiltration pour l'expression transitoire de jeunes plants de tabac (b), et de supplémentation (c). .....	217
Figure 42: Séquence du plasmide pDGB1 $\Omega$ 1_PP.....	220
Figure 43: Analyse bibliométrique en réseau du co-auteurat dans le cadre de ce manuscrit pour des articles publiés entre 2005 et 2024. .....	253
Figure 44: Analyse bibliométrique en réseau de mots clés dans le cadre de ce manuscrit pour des articles publiés entre 2005 et 2024. .....	254

# Liste des tableaux

Tableau 1: Liste de FCs linéaires et angulaires identifiées chez les Moracées. ....	221
Tableau 2: Cinétique de l'effet du génotype "Psoralen Pathway" (PP) comparativement au génotype Wild-Type (WT) sur la résilience à la carence ou l'absence de fer en conditions de culture in vitro.....	138
Tableau 3: Expression des P450s du clan 71 dans le latex de figuier et classification en niveau d'intérêt en tant que candidats pour la biosynthèse de FCs.....	187
Tableau 4: Liste des éléments de composition des milieux de culture in vitro de tomate dans le cadre d'une induction de la carence en fer. ....	219
Tableau 5: Liste de FCs linéaires et angulaires identifiées chez les Moracées.....	224

# Liste des abréviations

3D : tridimensionnelle	BBE : FAD-binding Berberine family protein
3OHPGSL : 3-hydroxypropyl-glucosinolate	bHLH : Basic Helix-Loop-Helix
4CL : 4-coumarate CoA-ligase	BRs : Brassinosteroids
AA : Acide Aminé	CaMV/TAV : virus de la mosaïque du chou-fleur/transactivateur-viroplasmin
AAO3 : Abscisic Aldehyde Oxidase 3	CB : Cycle Calvin Benson
ABA : Acide Abscissique	C3H : 4-Coumarate 3-Hydroxylase
AD : Acide DésoxyRiboNucléique	C4H : Cinnamate 4-Hydroxylase
AKG : $\alpha$ -ketoglutarate	CCM : Chromatographie sur Couche Mince
ANR : Agence Nationale de la Recherche	CG : Chromatographie Gazeuse
AP2 : APETALA2	CHI : Chalcone Isomérase
ARN : Acide RiboNucléique	Chl : Chloroplast
AUX : Auxin	CL : Chromatographie Liquide
AOS : Allène oxide synthase	CoA : Coenzyme A
ARF : Auxine Response Factor	COI1 : Coronatine Insensitive Protein 1

CoRAH : Coordinated Resource Allocation Hypothesis	GEA : GO enrichment analysis
COSY : Coumarin Synthase	GFP : Green Fluorescent Protein
COP 1 : constitutivement	GO : Gene Ontology
photomorphogénique 1	GW : GateWay
CPR : Cytochrome P450 Réductase	HCT : Hydroxycinnamoyl-CoA
CRY1/2 : cryptochrome 1/2	shikimate/quinate
CKs : Cytokinins	hydroxycinnamoyltransférase
CNRS : Centre National de la Recherche Scientifique	HR : Haute résolution
DAMPs : Damage Associated Molecular Patterns	IAA : Acide indole-3-acétique
DB : Database	IR : Infrarouge
DEGs : Differentially Expressed Genes	JA : Jasmonic acid
DET2 : Deetiolated 2	JAZ : Jasmonate Zim Domain
DIMBOA : 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one	Ka : Millier d'années
DMSO : Diméthylsulfoxyde	Km : Michaelis Menten
DS : Diméthylsubérosine Synthase	LACS : Long Acyl-CoA Chain Synthase
EDTA : acide éthylène-diamine-tétracétique	LED : Light-Emitting Diode
ER : Endoplasmic Reticulum	LINE : Long Interspersed Nuclear Element
ERF : Ethylene Response Factor	LOTUS : naturaL prOducTs occUrrence database
ET : Ethylene	LST8 : Lethal with SEC13 protein 8
FCs : Furocoumarines	Ma : Million d'années
FD : Ferredoxin	MAPK : Mitogen-Activated Protein Kinase
FNR : Ferredoxin-NADP+ reductase	MAPK-R : MAPK Receptor
GAs : Gibberellins	MC : Métabolisme Central
Ga : Milliard d'années	MES : acide 2-(N-morpholino)éthanosulfonique
GB : GoldenBraid	MEV : Mevalonate Pathway
GDBH : Growth/Defense Balance Hypothesis	MEP : Methylerythritol Phosphate
	MS : Métabolites Spécialisés ou Métabolisme Spécialisé

NAM : vitamine B3 (nicotinamide)	SA : Salicylic acid
NGS : New Generation Sequencing	SM : Spectrométrie de Masse
NCED : 9-cis-Epoxyxcarotenoid Dioxygenase	SnRK 1/2 : kinase 1/2 liée à la protéine kinase liée au sucrose non fermentant-1 (SNF1)
OPDA : 12-OxoPhytodienoic Acid	SAMPs : Signals Associated Molecular Patterns
OPPP : Oxidative Pentose Phosphate Pathway	SRSs : Substrate Recognition Sites
OPR3 : OPDA Reductase 3	TAA1 : Tryptophane Aminotransférase of <i>Arabidopsis</i> 1
PAL : Phenylalanine Ammonia-Lyase	TCA : Tricarboxylic Acid Cycle
PC : Plastocyanin	TCA-ETC : Cycle de l'acide tricarboxylique-chaîne de transport d'électrons mitochondriale
PCs : Pyranocoumarines	TE : Transposable Elements
PG : Plante Génétiquement Modifiée	TOR : Target of Rapamycin
PhyA/B : phytochrome A/B	TORC : TOR Complex
PIN : auxin efflux carriers	Trp : Tryptophane
PM : Plasma Membrane	UK : United Kingdom
P450s : cytochromes P450s	UHPLC-MS : Ultra high-Performance Liquid Chromatography - Mass Spectrometry
PPP : Pentose Phosphate Pathway	UM : Umbelliférone Synthase
PQ : Plastoquinone	USA : United State of America
PS : Psoralène Synthase	UT : Unité de Transcription
PSI/PSII : Photosystems I or II	UV : Ultraviolet
PT : Prenyltransferase	WD : Whole Genome Duplication
PUFA : acide gras polyinsaturé	WT : Wild Type
RAPTOR : Regulatory-associated protein of mTOR	YUC : YUCCA flavin monooxygénase
RMN : Résonance Magnétique Nucléaire	
ROS : Reactive Oxygen Species	
R2R3MYB : R2R3 Myeloblastosis	
ROP2/4/6 : protéine liée à rho des plantes 2/4/6	
RT : Retention Time	
RT-PCR : Reverse Transcribed - PolyChain Reaction	





# Chapitre 1 : Synthèse bibliographique

## I. Les multiples dimensions du métabolisme spécialisé : Mieux comprendre la physiologie des plantes et l'évolution de leurs interactions avec l'environnement

### 1-1 Du concept de métabolisme spécialisé à la diversité phytochimique connue

#### A- Le concept de métabolisme spécialisé des plantes

Les plantes sont des éléments centraux pour de nombreux services écosystémiques<sup>1</sup> et constituent une aménité environnementale<sup>2</sup> essentielle aux sociétés humaines (Forman, 2014; Molina-Venegas *et al.*, 2021). La phytochimie<sup>3</sup> est à l'origine de certains services rendus par les plantes (Romeo, 1999). L'étude de la production des métabolites<sup>4</sup> à l'origine de cette phytochimie fait l'objet de nombreuses recherches depuis plus de deux siècles (Dixon & Dickinson, 2024). La caractérisation et la quantification de plusieurs centaines de métabolites végétaux a conduit très tôt la communauté scientifique à partitionner le métabolisme<sup>5</sup> en deux catégories principales : (i) Les métabolites présents chez toutes les espèces, généralement fortement accumulés et essentiels à la survie de la plante (i.e. les métabolites primaires) ; (ii) les métabolites spécifiques à certaines espèces, généralement peu accumulés et non essentiels à la survie, dès lors considérés comme des déchets (Sachs, 1873), et qui seront nommés « métabolites

---

<sup>1</sup> Service écosystémique : Bénéfices que les humains tirent des écosystèmes, comme la pollinisation ou la purification de l'eau.

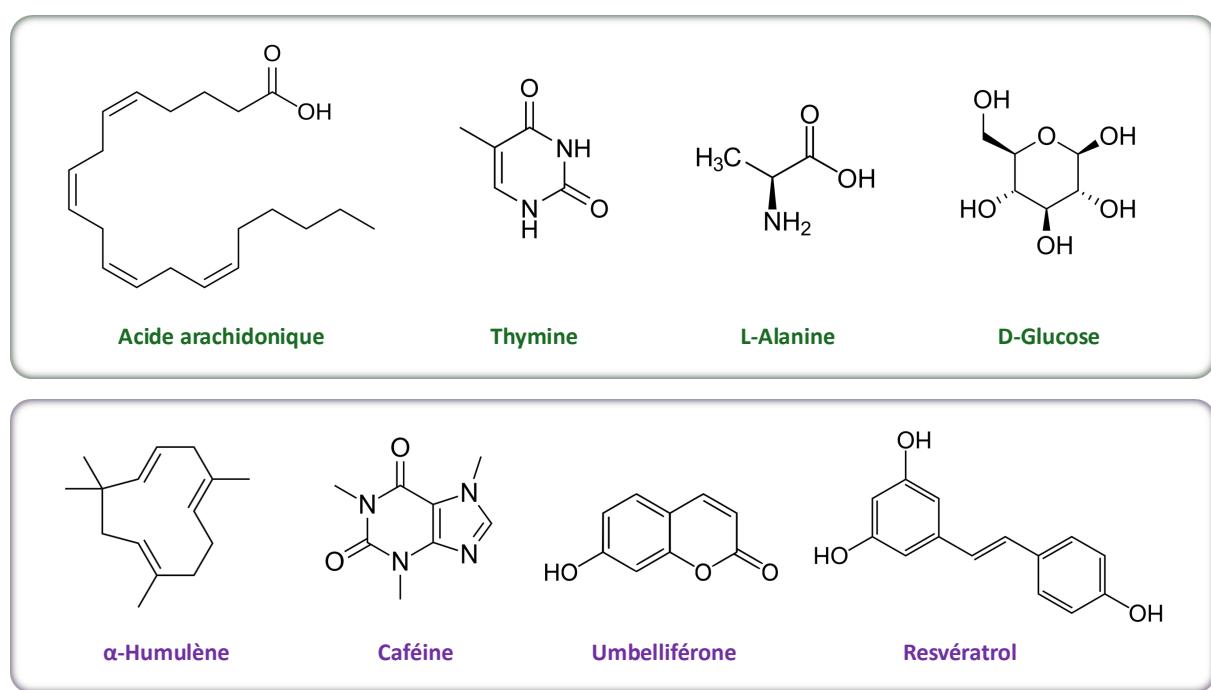
<sup>2</sup> Aménité environnementale : Avantage ou bénéfice fourni par les écosystèmes naturels qui contribue au bien-être humain, comme la purification de l'air, la régulation du climat ou la beauté des paysages.

<sup>3</sup> Phytochimie : Étude des composés chimiques produits par les plantes.

<sup>4</sup> Métabolite : Composé produit par le métabolisme d'un organisme, aussi nommé produit naturel ou biomolécule.

<sup>5</sup> Métabolisme : Ensemble des réactions biochimiques permettant à un organisme de produire de l'énergie et de la matière.

secondaires » (Kossel, 1891). Cette définition considère l'abondance et la fonction d'un métabolite plutôt que la façon dont il est produit ou son degré de complexité. À titre d'exemple, le D-glucose (glucide), la thymine (acide nucléique), l'acide arachidonique (lipide) ou encore l'alanine (acide aminé) peuvent être considérés comme des métabolites primaires. Le resvératrol (stilbène), l'umbelliférone (coumarine), l'α-humulène (terpène) ou la caféine (alcaloïde) sont associés à des métabolites secondaires (**Figure 1**).



**Figure 1: Exemple de structures chimiques de métabolites « primaires » (vert) et « secondaires » (violet) retrouvés chez certaines plantes.**

Parallèlement à l'identification de ces composés, la question de leurs rôles a émergé (Hartmann, 2007) (**Figure 2**). En 1888, l'allemand C. E. Stahl formalise l'hypothèse d'une corrélation entre la compartimentation<sup>6</sup> de certains composés et l'interaction des plantes avec le règne animal (Stahl, 1888). Il faudra attendre les années 1970s pour que ces travaux très critiqués aboutissent à un consensus à l'origine de la discipline de

<sup>6</sup> Compartimentation/Distribution (de métabolites) : Répartition des métabolites dans les différentes parties d'une cellule ou d'un organisme.

l'écologie chimique<sup>7</sup> (Hartmann, 2007). Durant trois décennies, cette discipline a été portée par deux visions divergentes : (i) celle d'une fonction écologique importante des métabolites secondaires et (ii) celle opposée qui réfutait cette vision finaliste en faveur de leurs existences en tant que métabolites déchets. En 1978, la capacité du soja à accumuler des flavonoïdes en réponse à l'application d'un extrait de bioagresseur<sup>8</sup> prouve la capacité des plantes à éliciter<sup>9</sup> la production de métabolites secondaires en réponse à un stress (Zähringer *et al.*, 1978). Ce précédent tranche le débat en faveur d'un rôle plus conséquent que celui imaginé jusqu'alors (Kutchan, 2001).

Peu de temps après, en 1983, la première plante génétiquement modifiée (PGM)<sup>10</sup> a été développée à partir du tabac (Zambryski *et al.*, 1983). Ce précédent a ouvert la porte à l'étude du rôle physiologique<sup>11</sup> potentiel des métabolites secondaires dont les étapes de biosynthèse étaient connues (Hartmann, 2007). Grâce à l'utilisation de mutants, les premières fonctions liées à la biosynthèse de certains flavonoïdes et brassinostéroïdes ont été découvertes. À titre d'exemple, ont été décrites : la capacité de la lutéoline à faciliter le dialogue chimique des plantes avec leurs symbiontes<sup>12</sup> dans le cadre d'établissement d'une nodulation<sup>13</sup> chez la luzerne (Peters *et al.*, 1986), le rôle de certains flavonols dans la fertilité du pollen chez le pétunia (Mo *et al.*, 1992; Van Der Meer *et al.*, 1992) ou encore le rôle des brassinostéroïdes dans le développement de l'Arabette en présence de lumière (J. Li *et al.*, 1996; Szekeres *et al.*, 1996). Depuis, de nombreuses fonctions associées à la croissance<sup>14</sup> et au développement<sup>15</sup> ont été décrites pour

<sup>7</sup> Écologie chimique : Branche de l'écologie qui étudie les interactions entre les organismes vivants et leur environnement à travers des signaux chimiques, incluant la communication entre espèces, les défenses chimiques contre les prédateurs et les relations symbiotiques.

<sup>8</sup> Bioagresseur : Organisme nuisible, comme un parasite, qui affecte la santé des plantes.

<sup>9</sup> Élicitation : Stimulation d'une réponse défensive chez une plante par des signaux extérieurs, comme la production de biomolécules de défense.

<sup>10</sup> Plante génétiquement modifiée : Plante dont le matériel génétique a été modifié par des techniques de biotechnologie pour introduire ou supprimer des traits spécifiques.

<sup>11</sup> Physiologie/Écophysiologie : Étude des processus biologiques des plantes et de ceux impliqués dans leur interaction avec l'environnement.

<sup>12</sup> Symbiose : Relation étroite entre deux organismes, bénéfique pour l'un ou les deux partenaires (symbionte).

<sup>13</sup> Nodulation : Processus par lequel certaines plantes, notamment les légumineuses, forment des nodules sur leurs racines en symbiose avec des bactéries fixatrices d'azote, permettant ainsi la conversion de l'azote atmosphérique en formes utilisables par la plante.

<sup>14</sup> Croissance : Augmentation de la taille ou du volume d'une plante par division et expansion cellulaire.

<sup>15</sup> Développement : Processus par lequel une plante atteint sa maturité en passant par divers stades morphologiques et fonctionnels.

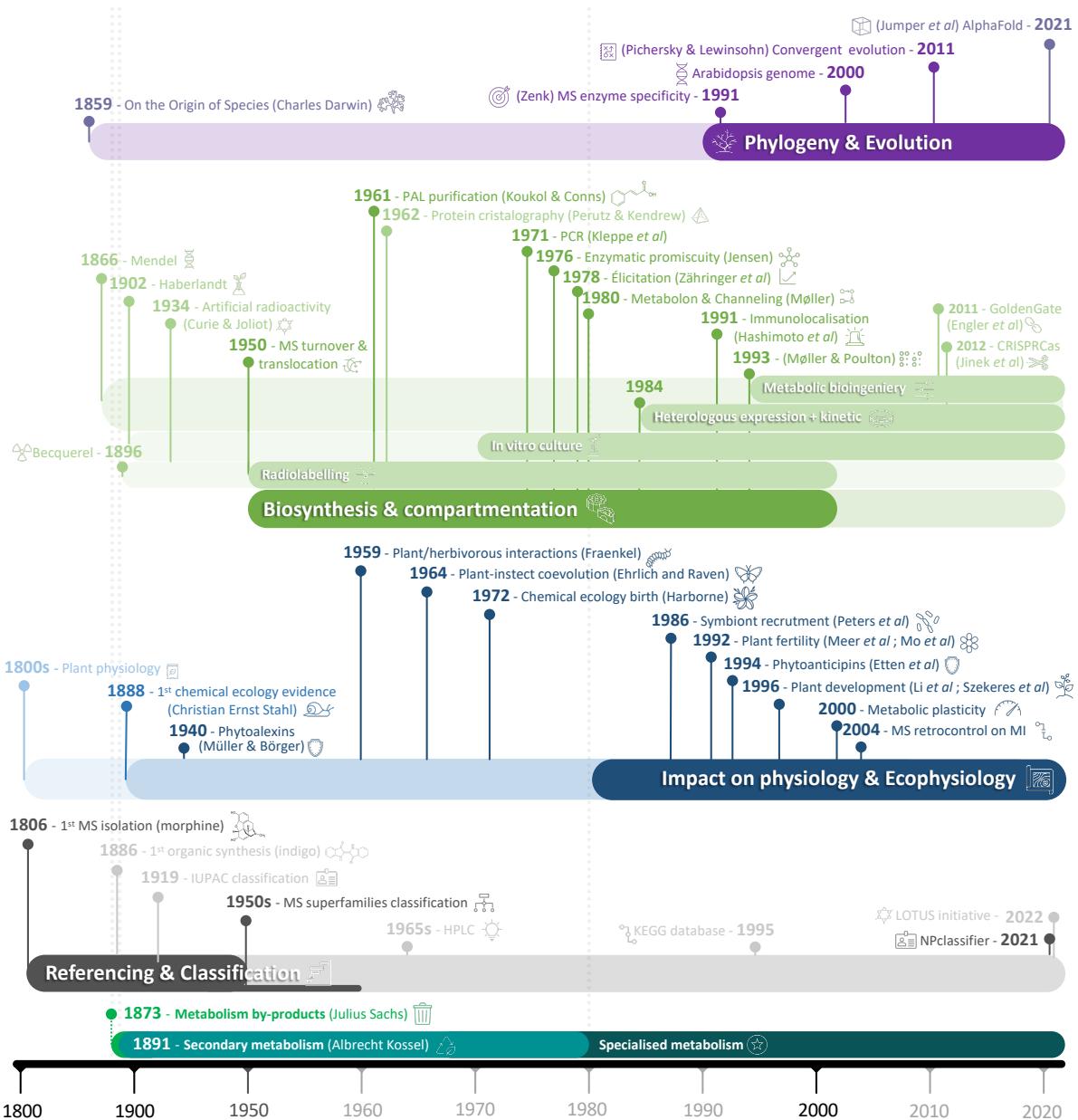
plusieurs familles de métabolites secondaires (e.g. apocaroténoïdes, diterpénoïdes, flavonoïdes, glucosinolates, etc.) (Voir Durán-Medina *et al.*, 2021 pour review). Ces fonctions spécialisées et bien souvent taxons<sup>16</sup>-spécifiques ont amené certains auteurs à modifier leur dénomination. Ce qui était considéré comme « métabolites secondaires » est devenu « métabolites spécialisés »<sup>17</sup> (MS) (Pichersky & Lewinsohn, 2011) (**Figure 2**). Parallèlement, l'utilisation du terme « métabolites primaires » a également évolué en « métabolites centraux »<sup>18</sup>. Une nomenclature qui place les métabolites centraux non pas seulement comme indispensables à la croissance et au développement des plantes, mais également comme des éléments centraux à l'origine du métabolisme spécialisé (Sweetlove *et al.*, 2017).

---

<sup>16</sup> Taxon botanique : Groupe de plantes classées ensemble dans la taxonomie selon des caractéristiques communes.

<sup>17</sup> Métabolisme spécialisé : Ensemble des réactions produisant des composés spécifiques à certaines fonctions comme la défense.

<sup>18</sup> Métabolisme central : Ensemble des réactions biochimiques nécessaires à la survie de base des cellules et de l'individu.

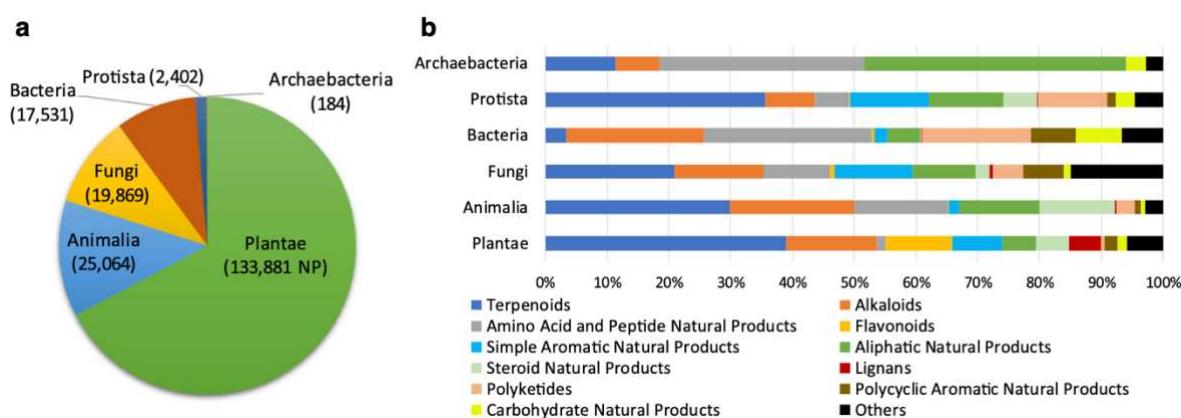


**Figure 2: Schéma simplifié de l'évolution de la perception par la communauté scientifique du métabolisme spécialisé (MS) des plantes et de ses disciplines d'étude depuis le 19<sup>e</sup> siècle à nos jours en parallèle de quelques avancées scientifiques majeures.** La chronologie de quatre grandes thématiques de recherches est décrite en fonctions de publications et de faits historiques de références pour l'étude du MS des plantes : Référencement et classification (gris), impact sur la physiologie et l'écophysiologie (bleu), biosynthèse et compartimentation (vert), phylogénie et évolution (violet). L'évolution de la perception et de la nomenclature du MS est représentée en turquoise. Les niveaux de contrastes de couleurs représentent par thématique : Période avec des évolutions majeures dans l'étude du MS notamment renforcée par des avancées techniques et la création de disciplines fondatrices (contraste sombre) (e.g. bioinformatique (violet), étude de voies de biosynthèse (vert), transgénèse (bleu), chromatographie (gris)) ; Période de courants de pensées ou de disciplines importantes à l'étude du MS (contraste intermédiaire) ; période peu active après la naissance de disciplines ou la publication de travaux pionniers (contraste faible). Adaptés des travaux de (Egerton, 2012; Ellis & Henderson, 1934; Gayon, 2016; Hartmann, 2007; Mo *et al.*, 1992; Moghe & Last, 2015; Nakanishi,

1998; Paxton, 1981; Peters *et al.*, 1986; Pichersky & Lewinsohn, 2011; Tawfik, 2010; Thorpe, 2007; Van Der Meer *et al.*, 1992; Zenk, 1991).

## B- Les plantes comme créatrices de diversité phytochimique

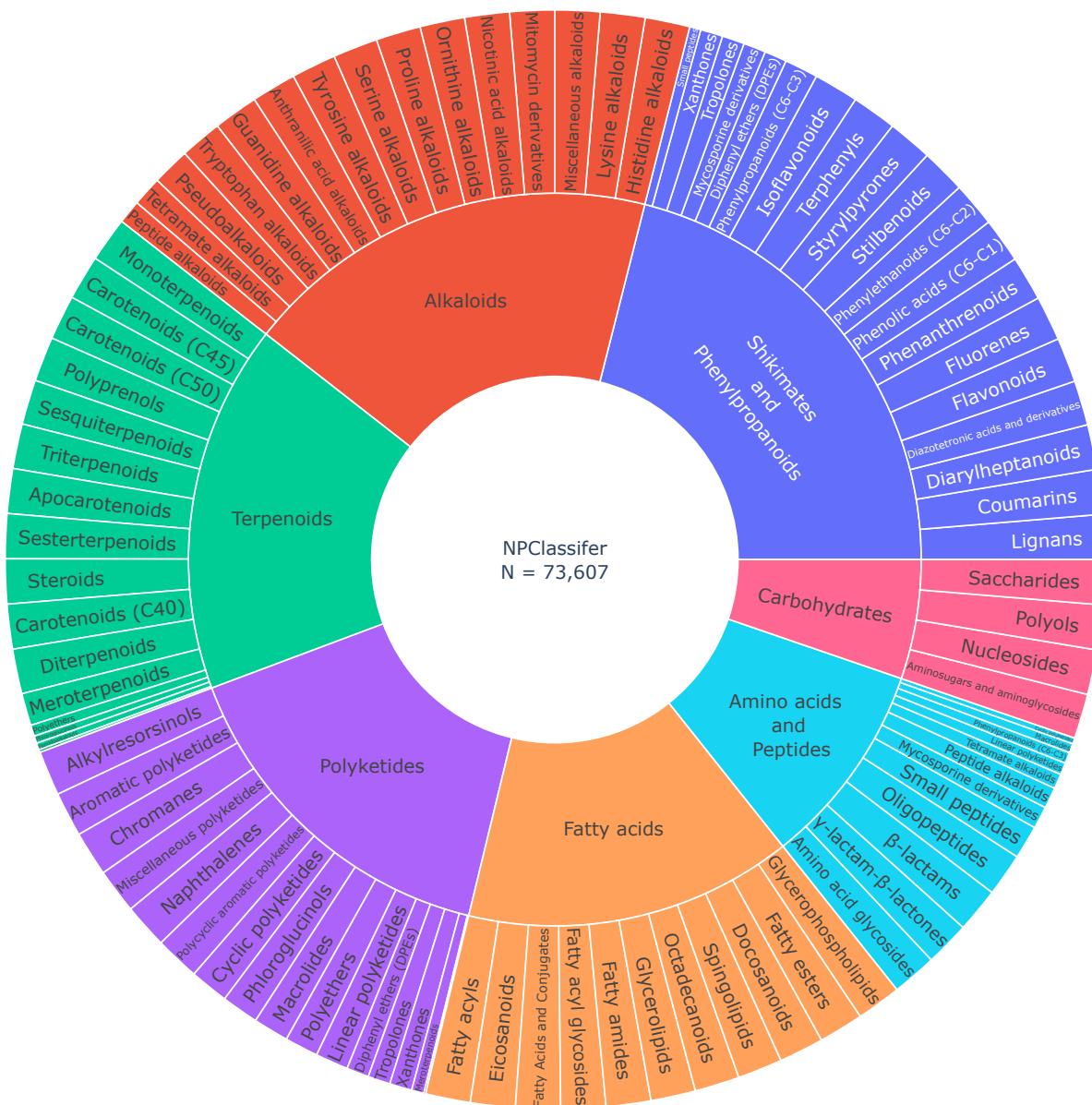
Si les plantes représentent entre 2,8 à 3,6% des espèces terrestres, 67,3% de la diversité des métabolites connus provient d'espèces végétales (Chassagne *et al.*, 2019; Christenhusz & Byng, 2016; Louca *et al.*, 2019; Mora *et al.*, 2011) (Figure 3a). Depuis la purification de la morphine au début du 19<sup>e</sup> siècle (Hartmann, 2007), pas moins de 300 000 MS ont pu être référencés chez les plantes (Chassagne *et al.*, 2019). Ce nombre est estimé à plus d'un million (Huang & Dudareva, 2023). Chez les plantes, toutes espèces confondues, le MS représente entre 96-99,2% de la diversité phytochimique (Chassagne *et al.*, 2019 ; Huang & Dudareva, 2023; Li & Gaquerel, 2021). D'un autre côté, le MC représente seulement 0,8-4% (i.e.  $\simeq$  8 000 métabolites) (D. Li & Gaquerel, 2021). À l'échelle d'une espèce végétale, le nombre de métabolites (MC + MS) est estimé à entre 5 000 et 15 000 métabolites (D. Li & Gaquerel, 2021).



**Figure 3 : Diversité phytochimique au sein du règne du vivant d'après une analyse de la base de données « Dictionary of Natural Products ».** (a) répartition des produits naturels par règne du vivant. (b) répartition des principales classes chimiques de produits naturels dans chaque règne du vivant. D'après Chassagne *et al.*, 2019.

L'importante diversité des métabolites identifiés depuis le 19<sup>e</sup> siècle a apporté dans les années 1950s le besoin de les classer (Hartmann, 2007). Aujourd'hui, les métabolites en

provenance de toutes espèces confondues sont décrits sous les traits de 7 voies principales, 70 superclasses (e.g. coumarines) et 672 classes (e. g. furocoumarines) (**Figure 4**). Les 7 voies principales correspondent aux voies du shikimate et des phénylpropanoïdes<sup>19</sup>, des alcaloïdes<sup>20</sup>, des terpénoïdes<sup>21</sup>, des polycétides, des acides gras, des acides aminés et peptides, et des carbohydrates (H. W. Kim *et al.*, 2021).



<sup>19</sup> Phénylpropanoïdes (polyphénols) : Famille de composés issus principalement de la phénylalanine, jouant notamment un rôle dans la défense et la structure des plantes.

<sup>20</sup> Alcaloïdes : Composés organiques azotés souvent toxiques, servant notamment de défense chimique chez les plantes.

<sup>21</sup> Terpénoïdes : Classe de métabolites dérivés des isoprènes, impliqués notamment dans les interactions biotiques et abiotiques.

**Figure 4: Aperçu des 7 voies métaboliques et 70 superclasses du jeu de données NPClassifier.** Les 73,607 métabolites référencés proviennent des banques de données : Pubchem (n=39 549), LIPID MAPS (n=23 506), UNPD (n=5 337), ChEBI (n=3 219), Chemspider (n=1 453) et la curation manuelle (n=534). D'après (H. W. Kim *et al.*, 2021).

Certaines voies de biosynthèse se sont préférentiellement développées chez certains *phylum*<sup>22</sup>. La vaste diversité de structures chimiques identifiées jusque-là témoigne d'une évolution<sup>23</sup> différente entre les *phylum* (Figure 3b). À noter que 96% des structures chimiques (MC + MS) identifiées chez les plantes ne sont retrouvées que chez les plantes. Les 4% restantes sont également décrites au sein d'autres *phylum* (Sorokina *et al.*, 2021). La plus grande diversité de structures chimiques (MC + MS) chez les plantes est retrouvée principalement au sein des voies des terpénoïdes (~40%), des alcaloïdes (~20-38%) et des phénylpropanoïdes (~12-21%) (Chassagne *et al.*, 2019; Wink, 2008). Plus précisément, parmi les plantes, la plupart de la diversité phytochimique provient des angiospermes<sup>24</sup> et notamment des dicotylédones<sup>25</sup> (~83%) (Chassagne *et al.*, 2019).

## C- De la phytochimie à la métabolomique

Décrire la diversité du MS nécessite une identification formelle et reproductible des structures moléculaires *via* des techniques analytiques. La séparation de composés dans des extraits<sup>26</sup> complexes est basée sur les propriétés physico-chimiques des MS (i.e. polarité, masse, etc.). La chromatographie<sup>27</sup> a été la première approche qui a permis d'isoler des MS *via* une élution par un solvant<sup>28</sup> sur une colonne de séparation (Tswett, 1906). Cette technique sera perfectionnée dès 1958 par l'utilisation d'un support de silice pour rendre plus résolutive la séparation (voir Stahl, 1969 pour review).

<sup>22</sup> Phylum : Rang taxonomique majeur dans la classification biologique, situé entre le règne et la classe, qui regroupe des organismes partageant des caractéristiques fondamentales communes.

<sup>23</sup> Évolution : Processus de transformation des espèces au fil du temps sous l'effet de mutations et de la sélection naturelle.

<sup>24</sup> Angiospermes : Groupe de plantes à fleurs produisant des graines enfermées dans des fruits.

<sup>25</sup> Dicotylédones/monocotylédones : Groupes de plantes définis par le nombre de cotylédons, ou premières feuilles, dans l'embryon.

<sup>26</sup> Extrait de plante : Substances extraites d'une plante par des méthodes physiques ou chimiques, utilisées pour l'étude ou la production de composés actifs.

<sup>27</sup> Chromatographie : Méthode analytique pour séparer et analyser les composants d'un mélange.

<sup>28</sup> Solvant : Substance capable de dissoudre d'autres substances, formant une solution.

Parallèlement, d'autres techniques ont été développées pour faciliter la séparation de molécules de petites tailles, dont initialement des nucléotides. Ces dernières permettent la détection de métabolites séparés par une colonne du fait de leurs affinités avec un solvant et de leurs détections par une lampe UV. L'apport de solutions logiciels pour faciliter la quantification des métabolites a favorisé dès les années 1990s le développement de la chromatographie préparative. Cette approche consiste à permettre la purification d'une quantité importante de métabolites à partir d'un extrait biologique, initialement pour le domaine pharmaceutique (Guiochon, 2002). Parallèlement, d'autres approches, telles que la chromatographie en phase gazeuse (CG), ont été développées pour permettre la séparation et la détection d'autres métabolites moins solubles comme les acides gras par exemple (James & Martin, 1951). Dans ce cas, les métabolites volatiles (ou rendus volatiles par la chaleur) sont séparés dans une colonne grâce à un gaz inerte. Les approches de CL et CG ont par la suite été améliorées par un couplage avec des systèmes de spectrométrie de masse<sup>29</sup> d'abord avec la CG-SM (Mc Lafferty, 1959) et plus tardivement la CL-SM (Baldwin & McLafferty, 1973). Ces approches ont permis la détection avec précision de la masse unitaire de certains métabolites ainsi que leurs fragmentations par des approches de SM en tandem<sup>30</sup> (Fenn *et al.*, 1989). Enfin, le développement de systèmes de spectrométrie de masse plus précis (dont le système à trappe orbitale) a permis jusqu'à la détection de la masse exacte de certains métabolites (Makarov, 2000). Les approches de CL-SM et CG-SM ont permis l'avènement de la phytochimie moderne puis de la métabolomique, et sont actuellement utilisées en routine dans les laboratoires de biologie (D. Li & Gaquerel, 2021). D'un autre côté, le développement de systèmes d'analyse à résonance magnétique nucléaire (RMN)<sup>31</sup> dès 1946 (Bloch, 1946), a joué un grand rôle dans l'identification formelle de structures chimiques complexes. La connaissance de ces structures a alors facilité l'identification

---

<sup>29</sup> Spectrométrie de masse : Technique analytique permettant de mesurer la masse des molécules pour identifier et quantifier des composés.

<sup>30</sup> \*\*Fragmentation en tandem : Technique de spectrométrie de masse où un ion sélectionné est fragmenté en ions plus petits, suivie d'une analyse de ces fragments. Elle permet d'obtenir des informations structurales détaillées sur les molécules, facilitant l'identification et la caractérisation des composés complexes.

<sup>31</sup> RMN : Technique d'analyse qui utilise les propriétés magnétiques des noyaux atomiques pour obtenir des informations sur la structure, la dynamique et l'environnement chimique des molécules.

formelle des molécules par des approches de CL-SM et CG-SM au sein de divers organismes, dont les plantes.

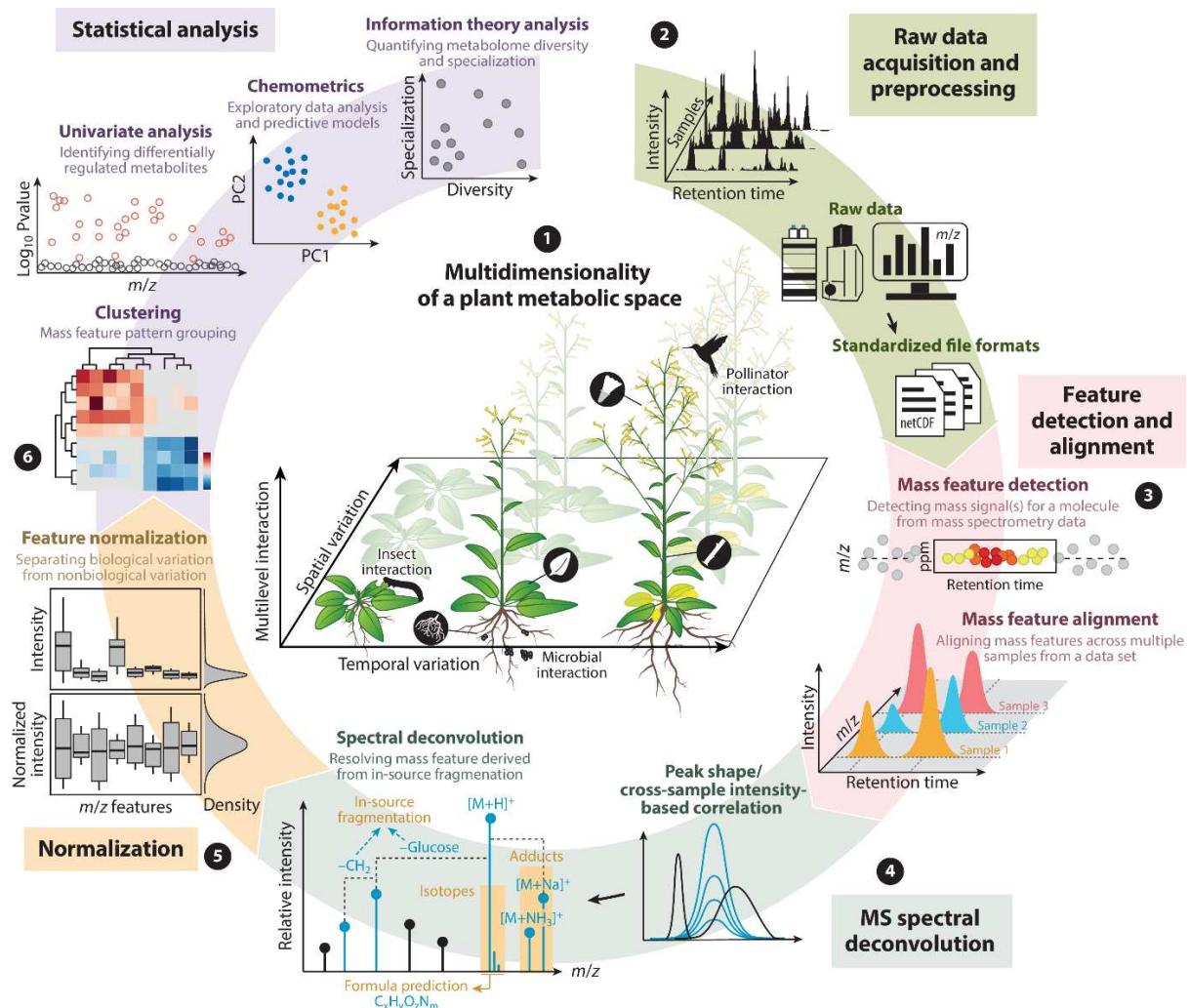
L'essor de la métabolomique<sup>32</sup> nommée ainsi en 1998 (Oliver *et al.*, 1998), ou l'étude à large échelle des métabolites extractibles a permis au cours de cette dernière décennie, une description beaucoup plus fine de la vaste diversité des MS (D. Li & Gaquerel, 2021). La métabolomique se différencie de la phytochimie par le regard plus exhaustif qu'elle permet (**Figure 5**). Couplée avec des approches de phylogénie<sup>33</sup>, la métabolomique permet entre autres d'imaginer les chemins évolutifs qui ont conduit à la vaste complexité du MS en réalisant des comparaisons à l'échelle de populations, d'espèces, de genres, voire de familles, comme illustré au sein du genre *Nicotiana* (Solanacées) (Elser *et al.*, 2023). De telles comparaisons peuvent permettre de caractériser des signatures métaboliques<sup>34</sup> sous réserve qu'elles existent (D. Li & Gaquerel, 2021).

---

<sup>32</sup> Métabolomique : Étude globale des métabolites (ou métabolome) présents dans un organisme pour comprendre son état physiologique.

<sup>33</sup> Phylogénie : Étude des relations évolutives entre les espèces à partir notamment de séquences nucléotidiques.

<sup>34</sup> Signature métabolique (ou caractères métaboliques idiosyncratiques) : Profil unique de métabolites révélant l'état physiologique ou environnemental d'un organisme.



**Figure 5: Schéma d'un itinéraire technique pour le traitement des données non ciblées par spectrométrie de masse à haute résolution et l'exploration statistique simple.** (1) Les hétérogénéités spatiotemporelles dans l'espace métabolique d'une plante sont détectables à différentes échelles hiérarchiques et pour différentes interactions écologiques. La conception expérimentale et la sélection des spécimens végétaux doivent être considérées de manière critique en fonction des niveaux explorés de diversité métabolique. (2) Après l'extraction des métabolites et la mesure par chromatographie liquide couplée à la spectrométrie de masse à haute résolution (HR-MS), les données brutes spécifiques au fournisseur sont converties en formats de fichiers universels pour un traitement automatisé ultérieur par une large gamme de logiciels open source. (3) Le traitement automatisé vise à convertir les données multidimensionnelles de la HR-MS en une matrice de masse pour l'ensemble des données et utilisée comme entrée pour l'exploration statistique. Les étapes de traitement automatisé comprennent de manière critique la détection des signaux métaboliques et l'alignement entre les échantillons ainsi que (4) la déconvolution des spectres MS par le regroupement des signaux métaboliques appartenant à un seul métabolite et générée lors de la fragmentation en source. (5) La normalisation des signaux et l'imputation des valeurs manquantes sont importantes pour les étapes suivantes de l'analyse statistique. (6) Des analyses statistiques multiéchelles sont généralement effectuées pour l'interprétation des tendances métaboliques globales et l'identification des signaux métaboliques marqueurs associés à ces tendances. Plus récemment, des descripteurs statistiques issus du cadre théorique de l'information ont été

transposés pour évaluer des indices de diversité et de spécialisation à partir des profils de métabolome, permettant ainsi de quantifier la reprogrammation de la diversité métabolomique en fonction des interactions écologiques. Abréviations : HR, haute résolution ; LC, chromatographie liquide ; MS, spectrométrie de masse ; m/z, rapport masse/charge. D'après Li et Gaquerel (2023).

L'acquisition de données de structures moléculaires est un point essentiel au référencement de MS. Elle est réalisée par la déréplication c'est-à-dire l'annotation unitaire des signaux acquis soit en métabolites, soit en bruits de fond (D. Li & Gaquerel, 2021). En effet, au sein du métabolome d'un organisme, il est estimé que seulement 5% des métabolites sont identifiés. Cette estimation suggère une grande part d'ombre sur la réalité du métabolome. Les autres métabolites peuvent être par exemple non extraits, transformés ou dégradés au moment de l'extraction, non détectés par les méthodes analytiques standards, etc. (Tsugawa *et al.*, 2019). Cette part d'inconnue persiste malgré une très nette amélioration de l'acquisition et de la standardisation de l'analyse des données au cours de ces 20 dernières années (D. Li & Gaquerel, 2021). Très récemment, il a été démontré que les analyses métabolomiques étaient soumises à un biais analytique qui démultipliait les signaux acquis par spectrométrie de masse. Ce biais représenterait jusqu'à 70% des signaux détectés (Giera *et al.*, 2024).

Avant 2018, les méthodes les plus avancées d'analyses métabolomiques permettaient l'identification d'un maximum de 3,5% des métabolites totaux de plantes modèles<sup>35</sup> comme *A. thaliana*. Une identification qui pouvait aller jusqu'à 7% chez *Homo sapiens*, 18% chez *Saccharomyces cerevisiae* et 19% chez *Escherichia coli* (Alseekh & Fernie, 2018). L'identification précise de certaines familles de métabolites permet d'atteindre parfois la finesse de certains processus physiologiques chez certaines espèces modèles (Lindon *et al.*, 2004). À noter que 80% des composés référencés proviennent de seulement 10% des espèces étudiées dont principalement des espèces modèles comme *A. thaliana* (Rutz *et al.*, 2022). La plupart des métabolites connus sont listés au sein de banques de données métabolomiques contenant comme informations leur temps de rétention dans des conditions standardisées, leur masse exacte, leur formule

---

<sup>35</sup> Organisme modèle : Espèce biologique largement utilisée pour la recherche en raison de ses caractéristiques génétiques, biologiques ou physiologiques bien comprises.

brute et leur profil de fragmentation en masse. MassBank (Horai *et al.*, 2010), PRIMe (Sakurai *et al.*, 2013), ReSpect (Sawada *et al.*, 2012), WEIZMASS (Shahaf *et al.*, 2016), METLIN (Tautenhahn *et al.*, 2012), GNPS (M. Wang *et al.*, 2016), constituent autant d'exemples de banques de données publiques. En questionnant tout ou partie de ces banques de données, il est possible d'établir des annotations putatives de métabolites à partir de données acquises au laboratoire. Néanmoins, ces banques de données ne sont pas exhaustives et incluent des données de qualités variables (D. Li & Gaquerel, 2021). Depuis moins d'une décennie, certaines initiatives tendent à rassembler les banques de données pour faciliter leur utilisation. C'est le cas par exemple de l'initiative « naturaL prOducTs occUrrence databaSe » (LOTUS), qui est l'une des banques de données publiques les plus fournies (Rutz *et al.*, 2022). Cette base de données a tout de même des limites comme ne pas renseigner les profils de fragmentation de ces métabolites. Cette dernière rassemble des données structurales de 290 000 structures chimiques uniques couplées à leurs origines taxonomiques<sup>36</sup> (40 000 organismes distincts). Ce recensement a permis de référencer plus de 750 000 paires structures-organismes<sup>37</sup> pour plus de 75 000 publications (Rutz *et al.*, 2022). L'annotation de métabolites et de leurs organismes d'origines est un outil précieux pour mieux comprendre l'évolution du métabolisme des plantes. Une tâche qui est complexe tant la diversité métabolique se recoupe au sein de taxons qui sont parfois évolutivement très éloignés chez les plantes (Pichersky & Lewinsohn, 2011).

Dès les premières purifications de MS, est apparue la volonté de retracer les liens entre taxons botaniques et diversité phytochimique. Les données acquises et mises à disposition dans les bases de données y contribuent très largement. Cette volonté a permis l'émergence de la discipline de la chemophylogénie<sup>38</sup> (Candolle, 1816; Hegnauer, 1986). Ces deux dernières décennies, l'étude de voies de biosynthèse<sup>39</sup> au sein de taxons éloignés a permis de démontrer la richesse des processus évolutifs à l'origine de leurs

---

<sup>36</sup> Taxonomie : Science de la classification des organismes vivants.

<sup>37</sup> Paires structures-organismes : Relation entre des structures chimiques spécifiques de métabolites et les organismes qui les produisent.

<sup>38</sup> Chemiophylogénie : Étude de l'évolution du profil en biomolécules en fonction des liens phylogénétiques entre les organismes.

<sup>39</sup> Voie de biosynthèse : Suite de réactions enzymatiques produisant des molécules spécifiques à partir de précurseurs simples.

acquisitions, comme synthétisé par (Negin & Jander, 2023). Une acquisition qui est bien souvent considérée comme convergente<sup>40</sup> et propre à chaque taxon, comme synthétisé par (Ono & Murata, 2023).

## 2-Émergence et évolution fonctionnelle du MS des plantes

### A- Évolution des génomes et fonctions liées au MS

Le séquençage<sup>41</sup> des génomes<sup>42</sup> est un outil qui a favorisé l'identification de gènes impliqués dans la biosynthèse de MS. L'identification de ces gènes et la comparaison de leurs évolutions chez des espèces apparentées permettent de retracer l'émergence et l'évolution du MS des plantes. L'essor des nouvelles techniques de séquençage (NGS) ces 20 dernières années a permis un bond en avant considérable dans l'annotation des génomes (Aksoy *et al.*, 2022; Marks *et al.*, 2021). En 2021, les génomes de 798 embryophytes<sup>43</sup> étaient répertoriés (i.e. 0,32% des plantes terrestres connues) (Marks *et al.*, 2021) (**Figure 6**). Ce référencement est continuellement mis à jour, notamment sur la plateforme NCBI (<https://www.ncbi.nlm.nih.gov/datasets/genome/>). À titre d'exemple, le génome de 0,74% des plantes terrestres connues était annoté en août 2024.

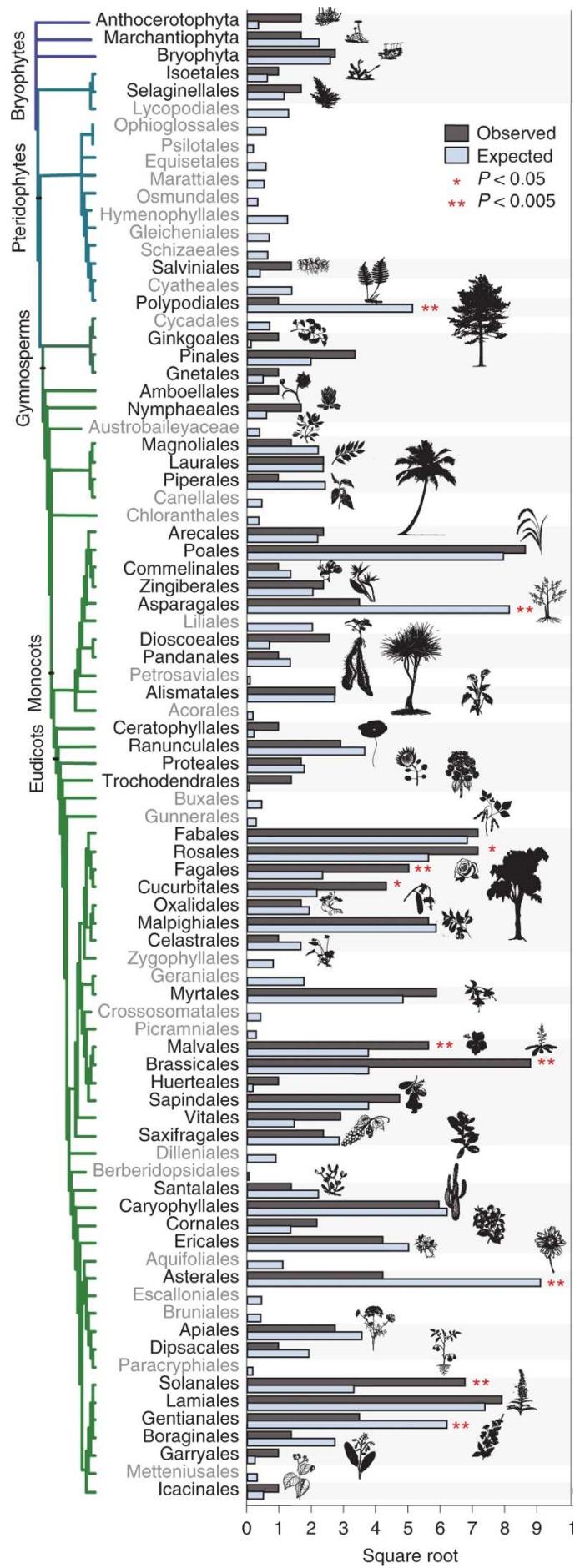
---

<sup>40</sup> Évolution convergente : Apparition de traits similaires chez des espèces non apparentées en réponse à des pressions environnementales similaires.

<sup>41</sup> Séquençage : Détermination de l'ordre des nucléotides dans une molécule d'ADN ou d'ARN.

<sup>42</sup> Génome : Ensemble du matériel génétique d'un organisme.

<sup>43</sup> Embryophytes : Groupe de plantes caractérisées par la présence d'un embryon développé à l'intérieur de structures protectrices, ce qui inclut toutes les plantes terrestres.



0 1 2 3 4 5 6 7 8 9 10  
Square root

**Figure 6: Nombre d'espèces végétales avec un génome séquencé et publiquement disponible en janvier 2021 (n = 798) par rapport au nombre attendu pour chaque ordre.** Les ordres montrant une surreprésentation ou une sous-représentation significative sont marqués par un astérisque. Les valeurs de significativité ont été calculées à l'aide du test exact de Fisher. D'après Marks *et al.*, 2021.

La capacité des plantes à produire une vaste diversité phytochimique passe par une évolution continue de leur génome (Morris *et al.*, 2018). Au sein du règne du vivant, les génomes de plantes sont considérés comme les plus évolutifs et les plus denses avec entre 8 000 et 95 000 gènes (Dimitrov *et al.*, 2023; Murat *et al.*, 2012; Panchy *et al.*, 2016). Cette évolutivité est notamment due à la duplication<sup>44</sup> couplée à la diversification<sup>45</sup> répétée de tout ou partie du génome (Flagel & Wendel, 2009; Panchy *et al.*, 2016). L'évolution d'un génome peut impacter des séquences géniques ou intergéniques<sup>46</sup> (e.g. promoteur<sup>47</sup>, séquences répétées, transposons<sup>48</sup>, etc.) (Schlötterer, 2015; Tautz & Domazet-Lošo, 2011; Wray, 2003). Ainsi, des séquences peuvent être cooptées<sup>49</sup> au profits de fonctions nouvelles (True & Carroll, 2002). Dans ce chapitre, je décris uniquement les phénomènes de duplications qui peuvent impacter la biosynthèse de MS.

Parmi les processus de duplication, celui du génome entier est relativement commun chez les plantes (Clark & Donoghue, 2018) (**Figure 7a**). Il peut favoriser l'évolution rapide du métabolisme (Van Hoek & Hogeweg, 2009). Les duplications de séquences respectivement d'ADN ou d'ARNm médiées par des transposons (**Figure 7c**) ou des rétrotransposons<sup>50</sup> (**Figure 7e**) sont également importantes dans l'évolution du génome des plantes. Ainsi, une séquence peut être répliquée et intégrée ailleurs dans le génome

---

<sup>44</sup> Duplication : Copie d'une portion du génome nommé dupliquon.

<sup>45</sup> Diversification : Processus d'augmentation des variations génétiques (ou polymorphisme) au sein d'une population, favorisant l'adaptation.

<sup>46</sup> Séquence génique/intergénique : Partie du génome correspondant à un gène ou située entre deux gènes.

<sup>47</sup> Promoteur : Région d'ADN située en amont d'un gène qui régule son expression en servant de site de liaison pour l'ARN polymérase et d'autres facteurs de transcription.

<sup>48</sup> Transposon : Séquence d'ADN capable de se déplacer dans le génome.

<sup>49</sup> Cooptation/Exaption : Processus par lequel un trait ou une fonction acquiert une nouvelle utilité au cours de l'évolution.

<sup>50</sup> Rétrotransposons : Éléments génétiques mobiles qui se déplacent dans le génome en convertissant leur ARN en ADN, qui s'insère ensuite à un nouvel emplacement

via un processus de transposition<sup>51</sup>. Ce type de duplication peut impacter très largement l'évolution du génome comme synthétisé par (Bennetzen & Wang, 2014; Kaessmann et al., 2009; Lisch, 2013).

D'autres processus comme la recombinaison homologue<sup>52</sup><sup>53</sup> peuvent aboutir à la duplication de gènes (Ray & Langer, 2002). La redondance des génomes implique parfois que des séquences très similaires, même si elles sont éloignées, puissent être interchangées. Ce type de réarrangement appelé recombinaison ectopique<sup>54</sup> peut aboutir à des transferts inégaux de fragments nucléotidiques (e.g. gènes) (Knoll et al., 2014). Des séquences qui peuvent parfois être de grandes tailles comme dans le cadre de la duplication segmentaire (e.g. portion de chromosome) comme synthétisé par Leister (2004) (**Figure 7d**).

La recombinaison inégale peut également impliquer des séquences plus proches dans le génome comme dans le cadre de la duplication en tandem<sup>55</sup> (Yu et al., 2015) (**Figure 7b**). Les séquences dupliquées en tandem sont abondantes dans le génome des plantes terrestres et peuvent représenter jusqu'à 26,1% du génome de *Malus domestica* (Rosacées) (Yu et al., 2015). À ce titre, elles jouent un rôle majeur dans l'évolution des espèces végétales, mais également dans la diversification du MS (Lichman, 2020).

Soumises à la pression de sélection<sup>56</sup>, ces séquences dupliquées peuvent évoluer pour former des paralogues<sup>57</sup>. La duplication de paralogues induit également la duplication des effets de ce gène. Dans le cadre de fonction de biosynthèse de métabolites toxiques à forte dose, l'évolution peut favoriser la perte de ces gènes. Ainsi, certains paralogues disparaissent au fil des générations par pseudogénéisation<sup>58</sup>. Parfois, l'apparition de

---

<sup>51</sup> Transposition : Processus par lequel des transposons se déplacent d'un endroit à un autre dans le génome, ce qui peut entraîner des modifications dans l'expression des gènes et la structure du génome.

<sup>52</sup> Homologie : Ressemblance entre deux gènes (dit homologues) ou structures en raison de leur origine évolutive commune.

<sup>53</sup> Recombinaison homologue : Échange de segments d'ADN entre deux molécules d'ADN similaires, essentiel à la réparation et à la diversité génétique.

<sup>54</sup> Recombinaison ectopique : Recombinaison entre des séquences d'ADN situées à des endroits éloignés du génome.

<sup>55</sup> Duplication en tandem : Type de duplication génétique où des copies de séquences d'ADN sont répétées de manière consécutive l'une à côté de l'autre dans le génome.

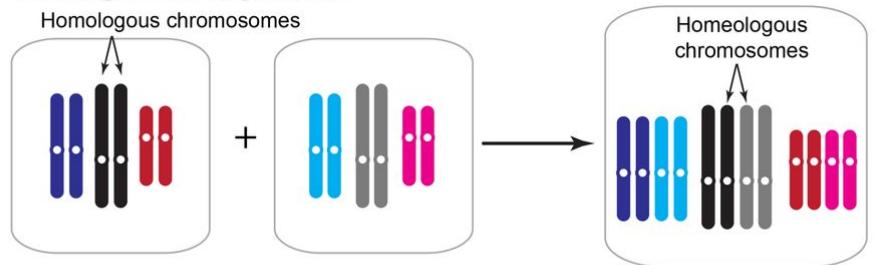
<sup>56</sup> Pression de sélection : Facteur environnemental influençant la survie et la reproduction des individus, favorisant certains traits.

<sup>57</sup> Paralogie : Relation entre deux gènes résultant d'une duplication dans un même génome aussi nommés paralogues.

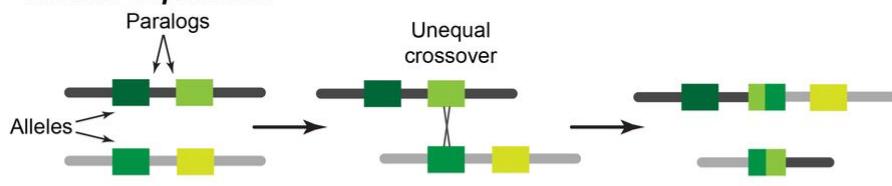
<sup>58</sup> Pseudogénéisation : Processus par lequel un gène devient non fonctionnel et est conservé comme pseudogène.

mutations<sup>59</sup> sur l'un des paralogues peut favoriser l'acquisition d'une nouvelle fonction par un processus nommé néofonctionnalisation<sup>60</sup>. Dans le cas où le gène possède plusieurs fonctions et que les mutations induites favorisent la spécialisation du gène vers l'une de ces fonctions, il est question de subfonctionnalisation<sup>61</sup>. L'évolution de la fonction des gènes dupliqués comme les paralogues est synthétisée par (Panchy *et al.*, 2016).

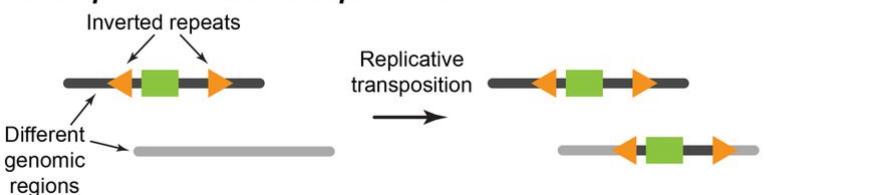
#### A Whole genome duplication



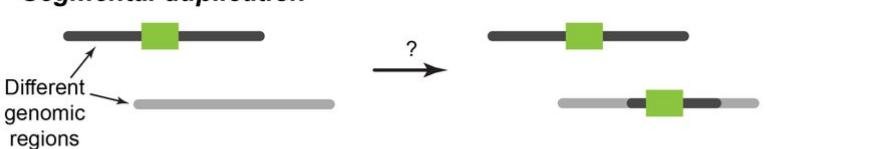
#### B Tandem duplication



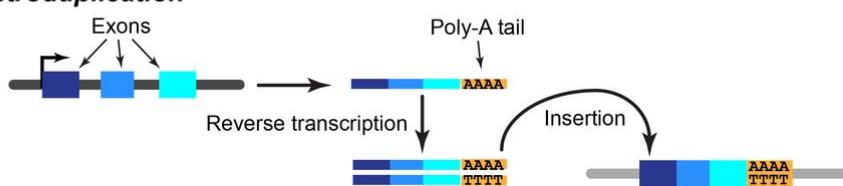
#### C Transposon-mediated duplication



#### D Segmental duplication



#### E Retroposition



<sup>59</sup> Mutation : Changement dans la séquence d'ADN, pouvant affecter la fonction d'un gène.

<sup>60</sup> Néofonctionnalisation : Acquisition d'une nouvelle fonction par un gène dupliqué.

<sup>61</sup> Subfonctionnalisation : Division des fonctions d'un gène dupliqué, chaque copie conservant une partie des fonctions d'origine.

**Figure 7: Mécanismes de duplication génique.** A. La duplication génomique totale (WGD), ou duplication de gènes par augmentation de la ploïdie. B. La duplication en tandem, ou duplication d'un gène via un croisement inégal entre des allèles similaires. C. La duplication médiée par transposon, ou duplication d'un gène associé à un élément transposable (TE) par transposition réplicative. D. La duplication segmentaire, ou duplication de gènes via la transposition réplicative d'éléments LINE. E. La rétroduplication, ou duplication d'un gène par transcription inverse d'ARNm. D'après Panchy *et al.*, 2016.

Plus largement, l'existence de séquences homologues peut favoriser les réarrangements chromosomiques dans le génome (Ray & Langer, 2002). Ce constat a permis de suggérer l'implication potentielle des réarrangements génomiques dans l'assemblage de clusters de biosynthèse (Field *et al.*, 2011; H. Nützmann *et al.*, 2016). Ce type de structure définit un ensemble de gènes impliqués dans la biosynthèse d'une même famille de composés qui sont regroupés à la suite sur une portion du génome. Le premier cluster de biosynthèse décrit chez les plantes intervient dans la biosynthèse du DIMBOA (i.e. benzoxazinoïde) chez le maïs (Frey *et al.*, 1997). En 2016, plus de 13 clusters étaient mis au jour, parmi les quelque 50 voies de biosynthèse du MS décrites à cette époque. Ce type de structure est peu abondant chez les plantes (H. Nützmann *et al.*, 2016; H.-W. Nützmann & Osbourn, 2014). Néanmoins, il porte l'intérêt qu'une fois identifié, il permet de découvrir l'intégralité ou presque d'une voie de biosynthèse. À noter que la plupart des gènes de biosynthèses décrits jusqu'alors chez certaines espèces sont dispersés dans le génome (e. g. voie des flavonoïdes, caroténoïdes, glucosinolates, etc.) (Nisar *et al.*, 2015; K. Saito *et al.*, 2013; Sønderby *et al.*, 2010).

## B- Émergence des « grandes voies » de biosynthèse de MS

La description des « grandes voies » de biosynthèse<sup>62</sup> de MS a permis de mieux en comprendre les origines (Figure 8). Actuellement, il est proposé que l'origine de la voie des phénylpropanoïdes provienne d'un transfert horizontal de gènes bactériens et/ou fongiques (van Santen *et al.*, 2022; Wink, 2008). Elle a été datée *a minima* antérieurement

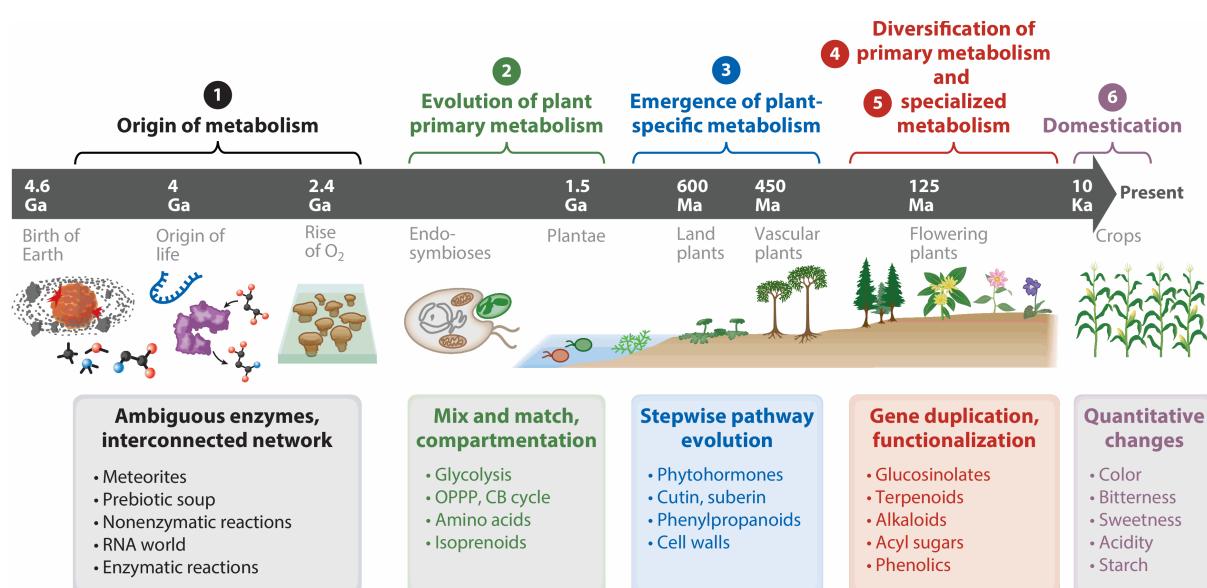
---

<sup>62</sup> Grandes voies de biosynthèse : Voies principales du MS comme les phénylpropanoïdes, terpénoïdes et alcaloïdes.

à l'émergence des plantes terrestres (taxon des Chlorophytes) (Vries *et al.*, 2021; Wink, 2008).

L'émergence des terpènes est, quant à elle, décrite comme au moins antérieure à la radiation<sup>63</sup> des premières formes de mousses (*phylum* des Marchantiophytes), c'est-à-dire acquise en parallèle du début de la colonisation terrestre (Wink, 2008). Enfin, l'émergence des alcaloïdes, moins bien comprise, semble être postérieure à l'émergence des mousses (*phylum* des Bryophytes) (Wink, 2008).

Ainsi, les gènes du début de la voie des phénylpropanoïdes, terpénoïdes et alcaloïdes ont définitivement été fixés dans le génome des plantes de la lignée verte dès les prémisses de la colonisation terrestre, il y a ~500 Ma. Ces voies se sont fortement diversifiées bien plus tard chez les angiospermes. La description de l'évolution du métabolisme des plantes permet de proposer qu'il a successivement été enrichi en phytohormones<sup>64</sup> chez les algues, puis en sucres et en phénylpropanoïdes chez les plantes terrestres, et enfin en terpénoïdes puis en alcaloïdes principalement chez les angiospermes (Chae *et al.*, 2014; Vries *et al.*, 2021; Wink, 2008).



<sup>63</sup> Radiation évolutive : Apparition rapide de nouvelles espèces à partir d'un ancêtre commun, souvent après une extinction ou un changement environnemental.

<sup>64</sup> Phytohormone : Hormone végétale régulant divers aspects de la croissance et du développement des plantes.

différentes étapes de l'évolution métabolique des plantes sont colorées et numérotées de (1) à (6). Les modes d'évolution métabolique ayant principalement contribué à chaque étape, ainsi que les métabolites clés ou les voies métaboliques qui ont émergé ou se sont développées, sont indiqués dans des encadrés. La chronologie n'est pas à l'échelle. Abréviations : cycle CB, cycle Calvin-Benson ; Ga, milliard d'années ; Ka, millier d'années ; Ma, million d'années ; OPPP, voie oxydative du pentose phosphate. D'après Maeda et Fernie, 2021.

## C- Émergence de nouvelles fonctions face aux enjeux de la vie fixée terrestre

Si l'évolution des génomes a permis la rétention et la diversification du MS chez les plantes terrestres, c'est notamment du fait de l'apport de nouvelles fonctions capables de faciliter la survie et la compétitivité des individus sessiles<sup>65</sup> (Maeda & Fernie, 2021; Weng *et al.*, 2021) (Figure 9). Une diversité de fonctions qui reste encore largement incomprise, bien que de plus en plus de voies de biosynthèse soient partiellement décrites (Weng *et al.*, 2021).

La première fonction reconnue des MS réside dans l'amélioration de la survie des individus *via* des propriétés attractives, répulsives, inhibitrices, voire léthales face aux stress biotiques<sup>66</sup> générés par l'environnement immédiat (Hartmann, 2007). À ce titre, on peut citer des propriétés anti- et pro- bactériennes, fongiques, herbivories, ou encore antivirales et allélochimiques<sup>67</sup> (Kumar *et al.*, 2023). Les plantes accumulent des MS sous forme de phytoanticipines<sup>68</sup> et/ou de phytoalexines<sup>69</sup> pour répondre aux compétiteurs avoisinants (Mérillon & Ramawat, 2020). Cette notion de défense a été, par la suite, étendue aux stress abiotiques également générés par l'environnement immédiat, mais non vivant, du fait de fonctions protectrices constitutives ou élicitées de

---

<sup>65</sup> Sessile : Désigne les organismes immobiles, comme les plantes, qui ne se déplacent pas.

<sup>66</sup> Stress Biotique/abiotique : Contraintes environnementales dues à des agents vivants (biotiques) ou non vivants (abiotiques).

<sup>67</sup> Composé allélochimique : Substance (dite allélopathique) libérée par une plante affectant la croissance d'autres plantes.

<sup>68</sup> Phytoanticipine : Molécule de défense produite de manière constitutive par les plantes, avant une attaque. Concept défini par Brian (1951).

<sup>69</sup> Phytoalexine : Molécule de défense produite par les plantes en réponse à une attaque. Concept défini par Muller, (1940).

certains MS (e. g. photoprotectrices<sup>70</sup>, osmoprotectrices<sup>71</sup>, chélatrices<sup>72</sup>) (Waadt *et al.*, 2022; H. Zhang *et al.*, 2022). Ces derniers protègent l'individu de stress générés par exemple *via* : des rayonnements ionisants<sup>73</sup> (e.g. UVs, IR, etc.), des changements de températures (e.g. dessiccation<sup>74</sup>, gel), des blessures mécaniques, une biodisponibilité<sup>75</sup> inégale en micro- et/ou macro-éléments<sup>76</sup> (e.g. carences, surabondances), etc. Les stress biotiques comme abiotiques sont souvent combinés dans l'environnement et déclenchent des régulations croisées auxquels les plantes savent généralement s'adapter (Suzuki *et al.*, 2014). Si l'évolution du MS chez les plantes est encore principalement considérée comme une course à l'armement<sup>77</sup> pour faire face aux stress biotiques et/ou abiotiques, de nombreux exemples permettent de leur attribuer une palette fonctionnelle bien plus variée (Durán-Medina *et al.*, 2021). Ceci notamment à travers la communication et la facilitation des interactions ou bien encore par la capacité de façonner (e.g. éléments polymérisables), voire de réguler la morphologie des plantes (Weng *et al.*, 2021).

---

<sup>70</sup> Photoprotection : Mécanismes permettant aux plantes de limiter les dégâts causés par la lumière excessive.

<sup>71</sup> Osmoprotection : Ensemble de mécanismes permettant aux plantes de survivre en conditions de stress hydrique.

<sup>72</sup> Chélation : Liaison d'une molécule organique à un ion métallique par exemple pour former un complexe stable.

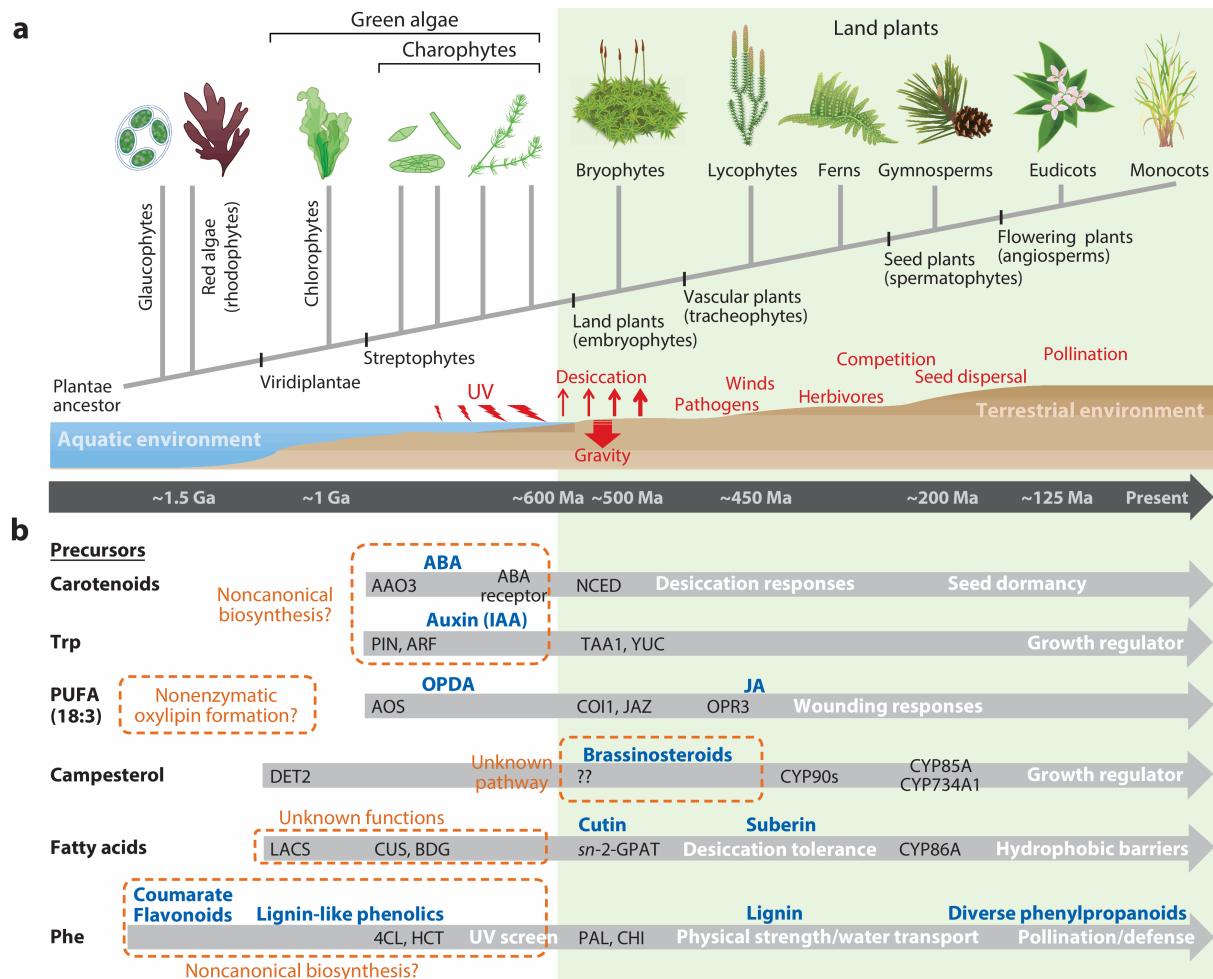
<sup>73</sup> Rayonnement ionisant : Rayonnement énergétique capable d'ioniser des molécules, utilisé en biologie pour induire des mutations.

<sup>74</sup> Dessiccation : Processus de déshydratation extrême des plantes, souvent lié à des mécanismes de tolérance à la sécheresse.

<sup>75</sup> Biodisponibilité : Fraction d'un nutriment ou d'un composé qui peut être absorbée et utilisée par un organisme.

<sup>76</sup> Macroéléments/microéléments : Nutriments essentiels pour les plantes, nécessaires en grandes (macro) ou petites quantités (micro).

<sup>77</sup> Course à l'armement : Évolution d'adaptations défensives dans une espèce en réponse à des adaptations d'espèces antagonistes.



**Figure 9: Évolution du métabolisme spécifique aux plantes, essentielle pour l'évolution des plantes terrestres.** (a) Les plantes terrestres (encadré vert) ont évolué à partir d'un ancêtre des algues charophytes il y a environ 600 millions d'années. Un certain nombre de voies métaboliques spécifiques aux plantes ont évolué au cours de la transition graduelle des environnements aquatiques aux environnements terrestres et ont probablement joué un rôle clé pour surmonter divers défis sur terre (rouge). La chronologie n'est pas à l'échelle. (b) Diverses innovations métaboliques au sein des plantes sont représentées, avec un accent particulier sur les décalages temporels observés entre les preuves chimiques (bleu) et biochimiques (noir). Ces contradictions suggèrent que les premières voies étaient peut-être catalysées par des enzymes analogues, mais non homologues ou médiées par des voies différentes (orange). Les principales fonctions chimiques sont indiquées en lettres blanches sur des flèches grises. Abréviations : 4CL, 4-coumarate CoA-ligase ; AAO3, abscisic aldehyde oxidase 3 ; ABA, acide abscissique ; AOS, allène oxide synthase ; ARF, facteur de réponse aux auxines ; BDG, BODYGUARD ; CHI, chalcone isomérase ; COI1, CORONATINE-INSENSITIVE PROTEIN 1 ; CUS, cutine synthase ; DET2, DEETIOLATED2 steroid 5-a-réductase ; Ga, milliard d'années ; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransférase ; IAA, acide indole-3-acétique ; JA, acide jasmonique ; JAZ, JASMONATE-ZIM DOMAIN ; LACS, long chain acyl-CoA synthase ; Ma, million d'années ; NCED, 9-cis-époxycarotenoid dioxygénase ; OPDA, 12-oxophytodienoic acid ; OPR3, OPDA reductase 3 ; PAL, phénylalanine ammonia-lyase ; Phe, phénylalanine ; PIN, PIN-FORMED auxin efflux carriers ; PUFA, acide gras polyinsaturé ; sn-2-GPAT, sn-2-specific glycerol 3-phosphate acyltransférase ; TAA1, tryptophane aminotransférase of *Arabidopsis* 1 ; Trp, tryptophane ; UV, ultraviolet ; YUC, YUCCA flavin monooxygénase. D'après Maeda et Fernie, 2021.

La colonisation terrestre par les plantes a notamment été permise par la biosynthèse de MS plus complexes, parfois polymérisés, voire complexés avec des MC (e. g. lignines, cutines, sporopollénines, caoutchouc, etc.) (Mizutani & Ohta, 2010). Ces structures chimiques plus complexes ont notamment permis aux plantes de diversifier la structure de leurs cellules et plus généralement de leurs morphologies (Bowman, 2022; Clark *et al.*, 2023). D'autre part, les voies du MS au fur et à mesure de leurs diversifications semblent s'être imbriquées à des réseaux de régulations transcriptionnels<sup>78</sup> et à certaines voies phytohormonales (Blázquez *et al.*, 2020; Lacchini & Goossens, 2020; Méteignier *et al.*, 2022). L'évolution des génomes qui a facilité cette régulation très contrôlée semble ainsi avoir permis aux plantes de s'adapter continuellement aux stress générés par l'environnement (Cannell *et al.*, 2020; Clark *et al.*, 2023; Maeda & Fernie, 2021) (**Figure 9**). Enfin, certains MS agissent dans la régulation physiologique de la croissance et du développement des tissus (Durán-Medina *et al.*, 2021).

Parmi les MS, les phénylpropanoïdes sont historiquement les plus étudiées. Au sein de cette famille, de nombreux rôles physiologiques ont été attribués principalement aux flavonoïdes, mais également aux coumarines, stilbènes, lignanes, tannins, etc. (Dong & Lin, 2021; Durán-Medina *et al.*, 2021). Ces rôles concernent la régulation de la croissance et du développement des plantes (germination, stade précoce), la fertilité<sup>79</sup> (attractivité des fleurs et des fruits, développement des anthères<sup>80</sup> et fertilité mâle), ainsi que la signalisation<sup>81</sup> et la protection face à l'environnement (soutien mécanique, détoxicification<sup>82</sup>) qu'il soit biotique (allélopathie, développement de symbioses pour la fixation d'azote) ou abiotique (filtre UVs, résistance au gel et à la dessiccation, captation de microéléments en cas de carences) (Barbehenn & Peter Constabel, 2011; Chong *et al.*, 2009; Dong & Lin, 2021; Ražná *et al.*, 2021; Samanta *et al.*, 2011; Stringlis *et al.*, 2019).

---

<sup>78</sup> Régulations transcriptionnelles : Mécanismes qui contrôlent l'expression des gènes en influençant la transcription de l'ADN en ARN.

<sup>79</sup> Fertilité : Capacité d'une plante à produire des gamètes viables et à se reproduire.

<sup>80</sup> Anthère : Partie de la fleur qui produit et contient le pollen.

<sup>81</sup> Voie de signalisation : Série de réactions moléculaires permettant à une cellule de répondre à des stimuli extérieurs.

<sup>82</sup> Détoxicification : Processus par lequel une plante ou un organisme élimine ou neutralise des substances toxiques.

## D- Les outils Omiques pour étudier et caractériser les voies de biosynthèse du MS

La biologie des systèmes<sup>83</sup> et les approches omiques sont des exemples contemporains d'outils précieux en fort développement (Jamil *et al.*, 2020; J. Yan & Wang, 2022). Ils sont basés sur des approches d'apprentissages automatiques ou profonds<sup>84</sup> et la corrélation de grands jeux de données (e.g. génomique, transcriptomique<sup>85</sup>, métabolomique, etc.). Ils facilitent la cartographie des voies de biosynthèses du MS peu ou non élucidées en donnant accès par exemple à la séquence de gènes, à leurs niveaux d'expression, à la structure et au niveau d'accumulation de certains MS, etc. Ces outils permettent d'intégrer plusieurs échelles de données pour identifier des différences entre des conditions biologiques contrastées. La diversité du polymorphisme génétique<sup>86</sup> peut faciliter l'identification de gènes candidats<sup>87</sup> codants pour des enzymes<sup>88</sup> impliquées dans la biosynthèse de MS (Abley & Locke, 2021; Delli-Ponti *et al.*, 2021; Dugé de Bernonville *et al.*, 2020; Mutwil, 2020; van der Hooft *et al.*, 2020). D'autres cibles complémentaires aux données omiques<sup>89</sup> peuvent faciliter l'identification de gènes de biosynthèse. À titre d'exemple, la position dans le génome, la phylogénèse<sup>90</sup> et le statut de duplication de familles de gènes, ou encore l'existence d'une signature métabolique comparativement chez plusieurs taxons peuvent faciliter la compréhension du cadre évolutif d'une voie de biosynthèse (Jacobowitz & Weng, 2020; Seligmann *et al.*, 2024; Shaw & Cheung, 2020). Enfin, certains outils peuvent permettre d'appuyer l'identification

---

<sup>83</sup> Biologie des systèmes : Approche intégrée visant à comprendre les interactions complexes entre les éléments biologiques d'un organisme.

<sup>84</sup> Apprentissage automatique et profond : Méthodes d'intelligence artificielle permettant aux ordinateurs d'apprendre des modèles à partir de données complexes.

<sup>85</sup> Transcriptomique : Étude de l'ensemble des ARN (transcriptome) transcrits dans un organisme à un moment donné.

<sup>86</sup> Polymorphisme : Variation génétique au sein d'une population, impliquant des différences dans les traits physiques ou fonctionnels.

<sup>87</sup> Gène candidat : Gène potentiellement impliqué dans un processus à vérifier par des études fonctionnelles.

<sup>88</sup> Fonction enzymatique : Activité catalytique spécifique d'une enzyme, facilitant une réaction biochimique.

<sup>89</sup> Sciences omiques : Ensemble des approches globales comme la génomique, transcriptomique, la métabolomique et protéomique pour l'étude des organismes.

<sup>90</sup> Phylogénèse : Étude des relations évolutives entre différentes espèces ou groupes d'organismes, généralement représentée sous forme d'un arbre phylogénétique.

de gènes candidats. À titre d'exemple, un profilage phylogénétique peut permettre de décrire l'histoire d'une famille d'enzymes et de plus facilement cibler certains candidats (e.g. gènes récemment dupliqués) (Weißenborn & Walther, 2017). Enfin, la prédition de fonctions enzymatiques putatives par une approche d'amarrage moléculaire<sup>91</sup> peut faciliter la sélection des gènes d'intérêt directement par un cible fonctionnel<sup>92</sup> (Gentile *et al.*, 2020).

---

<sup>91</sup> Amarrage moléculaire : Technique de simulation permettant de prédire comment une molécule se lie à une cible biologique comme une enzyme.

<sup>92</sup> Cible fonctionnel : Méthode expérimentale pour identifier des gènes ayant un effet spécifique.

### 3- Théorisation de la gestion des ressources par la plante pour mieux percevoir le rôle physiologique des MS

#### A- La gestion des ressources centrée sur la plante entre 1930 et 1970, aux prémisses de l'écologie chimique

L'origine des grandes voies de biosynthèse et la diversité du MS sont maintenant bien décrites chez les plantes (Chassagne *et al.*, 2019; Vries *et al.*, 2021; Wink, 2008). À l'inverse, l'impact des MS sur la physiologie et la gestion des ressources par la plante sont encore mal compris (Durán-Medina *et al.*, 2021; D. Li & Gaquerel, 2021). L'impact des MS sur la croissance et le développement des plantes, et inversement, questionne depuis longtemps la communauté scientifique et a permis l'essor de nombreuses théories (**Figure 10**). C'est en 1932 que les travaux pionniers de l'Américain W. E. Loomis ont posé les jalons d'une théorisation du partage des ressources entre MC et MS (**Figure 10a**). Il développe la théorie de la balance croissance/différenciation (GDBH) qui propose que le MS induise un coût métabolique dans le cadre d'une différenciation<sup>93</sup> compartimentée de la plante. Un coût qu'elle doit couvrir tout en priorisant l'allocation<sup>94</sup> de ses ressources à sa croissance pour survivre (Herms & Mattson, 1992). En 1964, les américains P. R. Ehrlich et P. H. Raven apportent un regard plus évolutif de ce compromis croissance/défense, en proposant la théorie de la coévolution<sup>95</sup>, ou hypothèse de l'évasion et de la radiation. Cette dernière prédit que les évolutions favorables à la diversification du MS sont également responsables de la diversification des lignées végétales. Autrement dit, que le MS joue un rôle prépondérant dans la sélection des plantes qui développent une phytochimie dont la régulation permet un compromis croissance/défense fonctionnel. Cela malgré ses potentielles contraintes ontogénétiques<sup>96</sup> (Ehrlich & Raven, 1964).

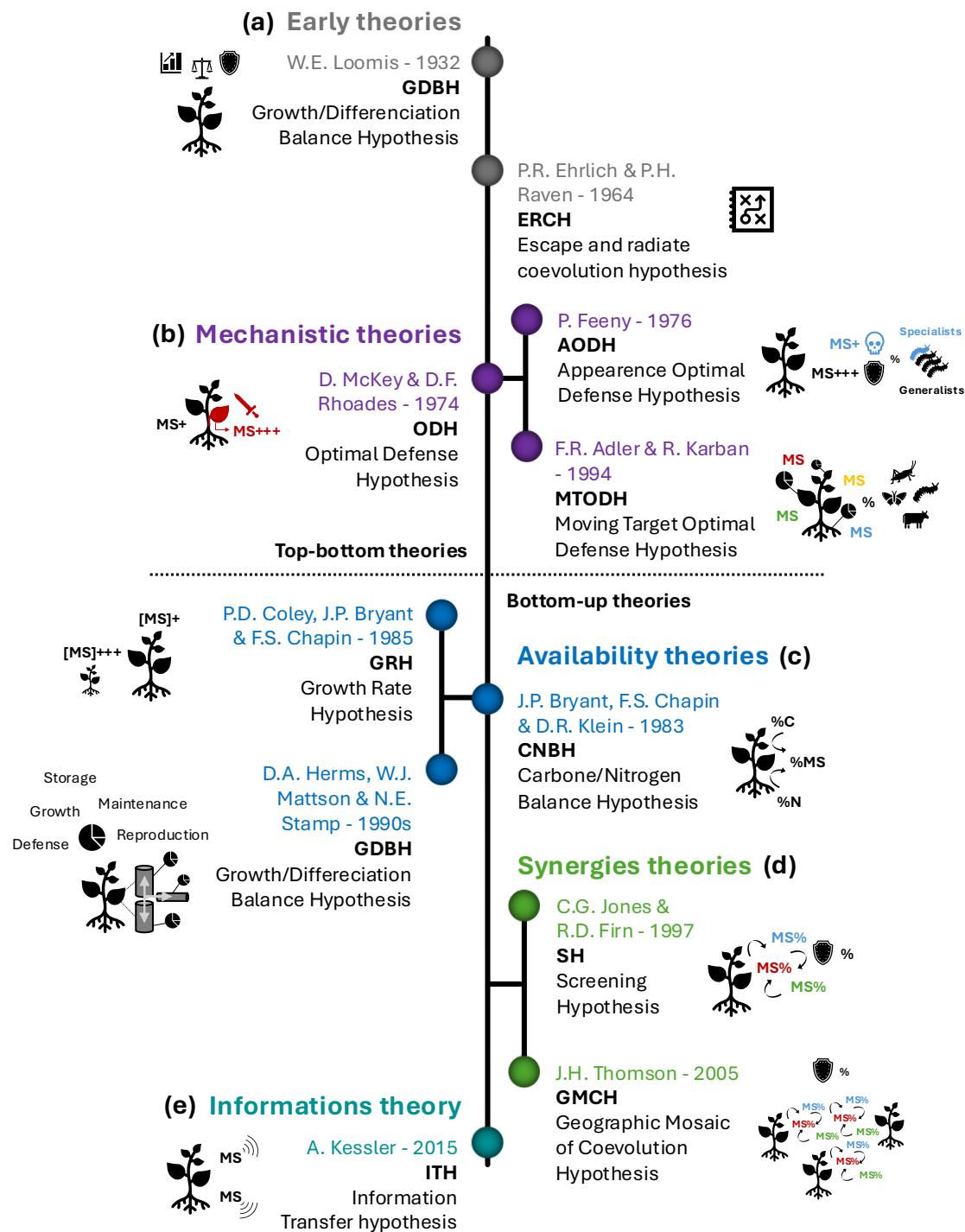
---

<sup>93</sup> Différenciation : Processus par lequel des cellules non spécialisées se transforment en cellules spécialisées dans un tissu.

<sup>94</sup> Allocation de ressources : Répartition des nutriments et de l'énergie entre les différentes parties et fonctions d'une plante.

<sup>95</sup> Coévolution : Évolution réciproque entre deux espèces interagissant étroitement.

<sup>96</sup> Contrainte ontogénétique : Limitation du développement d'un organisme en raison de contraintes inhérentes à son plan de développement.



**Figure 10: Schéma simplifié de l'évolution de la perception de la gestion des ressources par les plantes au travers de théories descendante et ascendante de 1932 à nos jours.** Les différents courants de pensés sont décrits selon les théories (a) premières, (b) mécanistes, (c) de disponibilité en ressources, (d) de synergies et (e) d'informations. Abréviation : MS (Métabolites Spécialisés). Les théories descendantes décrivent comment l'environnement façonne la gestion des ressources des plantes. À l'inverse les théories ascendantes décrivent comment la plante gère ces ressources pour faire face à l'environnement.

## B- Vision écocentré de la gestion des ressources pour la biosynthèse du MS chez les plantes : Entre compromis et équilibre

La description des processus physiologiques majeurs qui régissent la vie des plantes terrestres a permis de laisser émerger à partir des années 1970s de nouvelles théories davantage mécanistes pour tenter de déchiffrer les compromis croissance/défense (D. Li & Gaquerel, 2021) (**Figure 10b**). Les théories de défenses optimales constituent les premiers courants de pensée mécanistes qui décrivent le compromis croissance/défense (Hartmann, 2007; D. Li & Gaquerel, 2021). Ainsi, dès 1974, les Américains D. McKey et D. F. Rhoades apportent des éléments initiateurs à la théorie descendante (i.e. de l'environnement à la plante) de la défense optimale. Cette dernière propose une distribution spatio-temporelle orientée des MS avec un investissement préférentiel directement lié à la valeur défensive intrinsèque de chacun des MS. Ainsi, les plantes qui « prédisent » le plus efficacement quels tissus seront attaqués bénéficieront d'un meilleur succès reproducteur<sup>97</sup> (D. Li & Gaquerel, 2021; McKey, 1974). Cette théorie est relativisée par deux idées opposées. Premièrement, en 1976, l'américain P. Feeny développe la théorie de la défense optimale d'apparence. Cette théorie propose de distinguer les plantes « apparentes » (i.e. qui investissent leurs ressources dans une vaste diversité voire quantité de MS), des plantes « peu apparentes » (i.e. qui investissent leurs ressources dans une faible diversité et quantité de MS hautement toxiques). Cette distinction révèle deux stratégies de gestion opposées permettant aux plantes de faire face à une herbivorie<sup>98</sup> liée soit à des prédateurs spécifiques soit à un ensemble diversifié de prédateurs opportunistes (Feeny, 1976). En 1994, les Américains F. R. Adler et R. Karban détournent le concept de défense optimale en émettant une seconde hypothèse dite de « cibles mobiles ». Cette dernière propose que la régulation du MS soit la conséquence directe de la consommation de la plante par les herbivores et que la diversité spatio-temporelle en MS dépende de la diversité d'herbivores (i.e. cibles mobiles) ayant été en contact avec la plante. Sur des périodes longues, cela peut amener à la coévolution (Adler & Karban, 1994). Ces théories de défenses optimales basées

---

<sup>97</sup> Fitness (ou succès reproducteur) : Capacité d'un organisme à survivre et à transmettre ses gènes à la génération suivante.

<sup>98</sup> Herbivorie : Mode de nutrition où des organismes, appelés herbivores, se nourrissent principalement de matières végétales.

principalement sur le rôle de défense des MS ont permis de conceptualiser l'évolution de la gestion des ressources par les plantes.

Dès les années 1980s, un second courant de pensée vient relativiser la compréhension mécaniste du compromis croissance/défense et les travaux pionniers de W. E. Loomis, au travers de théories ascendantes (i.e. de la plante à l'environnement) se concentrant sur la disponibilité en ressources comme moteur de régulation (**Figure 10c**). En 1983, une équipe danoise propose que la nature des MS accumulés dépende de la biodisponibilité en carbone (C) et en azote (N) dans l'environnement immédiat de la plante en développant l'hypothèse de l'équilibre C/N. Ainsi, des conditions propices à l'accumulation de photoassimilats<sup>99</sup> faciliteront l'accumulation de MS carbonés de types terpénoïdes ou phénylpropanoïdes. À l'inverse, des conditions propices à l'accumulation de dérivés azotés faciliteront l'accumulation de MS azotés comme les alcaloïdes (Bryant *et al.*, 1983). En 1985, une équipe américaine développe l'hypothèse du taux de croissance. Cette dernière propose qu'au sein d'un environnement aux ressources limitées, les plantes à croissance lente favorisent une défense quantitative, tandis que les plantes à croissance rapide développent une défense qualitative (Coley *et al.*, 1985). Enfin, les Américains D. A. Herms, W. J. Mattson et N. E. Stamp complètent dans les années 1990-2000s la théorie GDBH de W. E. Loomis, qui est sous cette forme aujourd'hui encore considérée comme l'une des théories les plus intégrative et aboutie pour expliquer la gestion intégrée du MS par les plantes (Herms & Mattson, 1992; D. Li & Gaquerel, 2021; Stamp, 2003, 2004).

En guise de synthèse, l'allocation des ressources est visualisée comme un ensemble de transactions de matières entre des modules semi-autonomes, nommés sous-unités morphologiques spécialisées. Leurs fonctions varient du fait de leurs capacités à être graduellement et transitoirement sources et/ou puits. Ces différents niveaux d'organisations forment un réseau intégré de systèmes cellulaires soumis à une maturation séquentielle qui oscille entre croissance et différentiation en fonction de l'énergie assimilée et mobilisable. Les ressources acquises et limitées sont ainsi

---

<sup>99</sup> Photoassimilats : Produits de la photosynthèse, principalement des sucres, utilisés pour la croissance et le stockage dans la plante.

allouées pour soutenir les processus physiologiques inhérents à la croissance, l'entretien, le stockage, la reproduction et la défense. Ainsi, le génotype doit permettre la mise en place d'un phénotype qui facilite une gestion rapide et très localisée des ressources qui soit favorables à la survie des plantes fixées. Un phénotype qui prend la forme de modules qui doivent permettre notamment d'amortir les effets de variations spatiales et temporelles de la disponibilité en ressources (Herms & Mattson, 1992; Stamp, 2003, 2004).

## C- Du compromis à l'interaction

Depuis les années 1990s, le développement de la métabolomique a permis d'étoffer la perception de l'évolution et de la gestion du MS chez les plantes avec une vision de plus en plus holistique. Cette dernière a permis l'émergence de la théorie des synergies (**Figure 10d**). Cette théorie propose que la fonction d'un MS dépende de la fonction commune qu'il induit sous forme de synergies<sup>100</sup> au sein d'un paysage métabolique<sup>101</sup>. Cette théorie prend sa source dans les travaux anglo-américains de C. G. Jones et R. D. Firn qui développent dès 1991 l'hypothèse de criblage. Cette dernière propose que les MS qui apportent une nouvelle valeur adaptative<sup>102</sup> soient préférentiellement conservés dans le bouquet phytochimique<sup>103</sup> spécifiquement en réponse à des contraintes de l'environnement qui orientent le criblage évolutif (Firn & Jones, 2003; Jones & Firn, 1991). Cette théorie fait écho à l'hypothèse de la coévolution de P. R. Ehrlich et P. H. Raven développée en 1964. De façon contraire, l'américain J. N. Thompson développe en 2005 l'hypothèse de la mosaïque géographique de la coévolution. Cette dernière propose que la réponse des végétaux face à l'herbivorie soit l'accumulation non directionnelle de MS.

---

<sup>100</sup> Synergies métaboliques : Interactions entre différentes voies métaboliques ou métabolites qui produisent des effets cumulatifs ou complémentaires, améliorant l'efficacité des processus biologiques.

<sup>101</sup> Paysage métabolique : Ensemble des voies métaboliques d'un organisme ou d'une communauté, influençant sa physiologie et son environnement.

<sup>102</sup> Valeur adaptative : Aptitude d'un trait à améliorer la survie et la reproduction dans un environnement donné.

<sup>103</sup> Bouquet phytochimique : Ensemble de composés chimiques présents dans une plante, notamment des métabolites spécialisés, qui contribuent à ses caractéristiques biologiques, son goût, son odeur et ses propriétés médicinales.

Ceci créant ainsi des phénotypes<sup>104</sup> aléatoires qui s'incluent dans une mosaïque géographique dont il est difficile de s'accommoder tant la diversité d'armement phytochimique est importante dans le paysage immédiat (Thompson, 2005).

Enfin, l'émergence de la métagénomique et du concept d'*holobionte*<sup>105</sup> cette dernière décennie a permis de recentrer la théorisation de la gestion des ressources. Ceci non pas uniquement dans le cadre d'un face à face avec des herbivores, mais dans la perspective d'une facilitation des échanges d'informations des plantes avec l'environnement (**Figure 10e**). Ainsi, en 2015, l'américain A. Kessler développe l'hypothèse du transfert d'informations (Kessler, 2015; Kessler & Kalske, 2018). Cette hypothèse conceptualise les MS comme des informations que les plantes échangent avec leur environnement direct. Elle propose de redéfinir leurs rôles en étendant leurs valeurs adaptatives non pas uniquement vers la défense, mais également vers l'interaction et le fait de porter des messages (Kessler, 2015; Kessler & Kalske, 2018). À titre d'exemple, la régulation d'un cluster de biosynthèse de terpénoïdes chez l'Arabette lui permet de façonner spécifiquement ses interactions avec l'environnement rhizobactérien au travers d'échanges d'informations en continu avec la rhizosphère<sup>106</sup> (A. C. Huang *et al.*, 2019).

## D- Le système de contrôle central du métabolisme des plantes

Identifié pour la première fois chez la levure (Heitman *et al.*, 1991), le gène « Target of Rapamycin » (TOR) s'est imposé ces dernières années comme étant le commutateur<sup>107</sup> central de la gestion des ressources chez les plantes ainsi que chez tous les eucaryotes<sup>108</sup> (Burkart & Brandizzi, 2021). Fonctionnant de pair avec la « Regulatory-associated protein of mTOR » (RAPTOR) et « Lethal with SEC13 protein 8 » (LST8), ce

---

<sup>104</sup> Phénotype : Ensemble des caractéristiques observables d'un organisme, résultant de l'interaction entre son génotype (matériel génétique) et son environnement. Cela inclut des traits morphologiques, physiologiques et comportementaux.

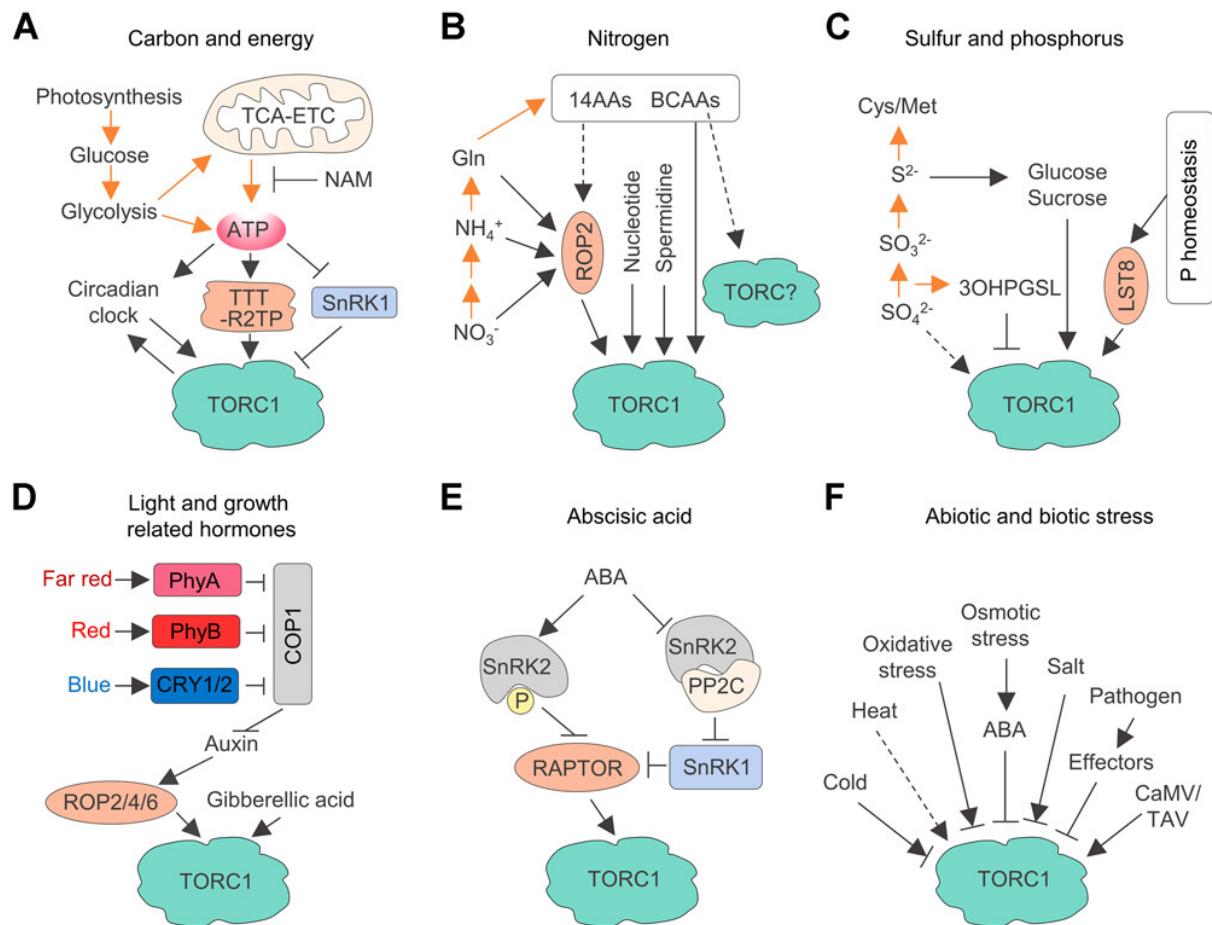
<sup>105</sup> Holobionte : Ensemble formé par un organisme et les microorganismes qu'il abrite.

<sup>106</sup> Rhizosphère : Zone du sol entourant les racines des plantes, où se développent des interactions entre la plante et les micro-organismes dont les rhizobactéries.

<sup>107</sup> Commutateur : Élément régulateur dans un système biologique qui permet de changer l'état d'expression d'un gène ou d'une voie métabolique, souvent en réponse à des stimuli spécifiques.

<sup>108</sup> Eucaryote : Organisme dont les cellules ont un noyau et des organites membranaires, incluant animaux, plantes et champignons.

régulateur central a un impact fondamental dans la gestion antagoniste du MC et du MS (**Figure 11**). Si beaucoup reste à découvrir quant à son fonctionnement (en amont et aval de sa fonction de régulation), des preuves convergentes indiquent qu'il intervient dans les interactions entre phytohormones et la gestion de la croissance et du développement (auxines, cytokinines, brassinostéroïdes, gibbérellines) ainsi que dans la réponse au stress (acide abscissique, acide jasmonique, acide salicylique, éthylène) (Choi *et al.*, 2022; De Vleesschauwer *et al.*, 2018; Fu *et al.*, 2020; Xiong *et al.*, 2023). Cette vision intégrative émergente et centralisée de la gestion des ressources par les plantes est en cours d'exploration et manque encore de données (De Vleesschauwer *et al.*, 2018; Rabeh *et al.*, 2024). Néanmoins, elle permet d'élaborer de nouvelles hypothèses plus intégratives comme celle de l'allocation coordonnée des ressources (CoRAH) (Monson *et al.*, 2022). Cette hypothèse propose que la gestion des ressources soit basée sur la perception de l'évolution des réserves. Une allocation médiée entre offre et demande par le système centralisé du « Complexe TOR » (TORC). Cette hypothèse suggère également qu'en réajustant les systèmes de coordination, il serait possible de découpler la croissance et la défense et ainsi d'améliorer simultanément la gestion de ces deux types d'allocations (Monson *et al.*, 2022).



**Figure 11: Signaux en amont et régulateurs de la signalisation TOR chez les plantes.** Les régulateurs positifs de la signalisation TOR sont montrés en orange, tandis que les régulateurs négatifs sont représentés en gris ou en violet clair. TORC1 et TORC2 sont indiqués en cyan. Les flèches orange représentent le chemin d'assimilation ou métabolique, tandis que les flèches noires représentent la voie de signalisation. TCA-ETC, cycle de l'acide tricarboxylique-chaîne de transport d'électrons mitochondriale ; NAM, vitamine B3 (nicotinamide) ; TTT-R2TP, complexe (Tel2-Tti1-Tti2)-R2TP ; AA, acide aminé ; BCAA, AAAs à chaîne ramifiée ; ROP2/4/6, protéine liée à rho des plantes 2/4/6 ; 3OHPGSL, 3-hydroxypropyl-glucosinolate ; PhyA/B, phytochrome A/B ; CRY1/2, cryptochrome 1/2 ; COP 1, constitutivement photomorphogénique 1 ; SnRK 1/2, kinase 1/2 liée à la protéine kinase liée au sucre non fermentant-1 (SNF1) ; PP2C, protéine phosphatase 2C ; CaMV/TAV, virus de la mosaïque du chou-fleur/transactivateur-viroplasmin. D'après Liu et Xiong, 2022.

### 3- L'exemple des furocoumarines chez les Moracées pour décrire l'évolution du métabolisme spécialisé des angiospermes

#### A- La voie de biosynthèse des coumarines et des furocoumarines chez les angiospermes

Au sein du MS des plantes, les furocoumarines (FCs) forment une classe de composés qui sont issues de la voie des phénylpropanoïdes et qui dérivent des coumarines (H. W. Kim *et al.*, 2021). Ce sont des composés aromatiques<sup>109</sup> hémpolaires<sup>110</sup> tricycliques planaires. Leur origine dépend de la furanylation<sup>111</sup> d'un squelette<sup>112</sup> coumarine (**Figure 12**), c'est-à-dire de la cyclisation<sup>113</sup> d'un groupement hydroxyle avec un groupement prényl sur un squelette  $\alpha$ -benzopyrone (Larbat *et al.*, 2007, 2009; Villard *et al.*, 2021). Deux zones de greffage du noyau furane sur le squelette coumarine sont principalement retrouvées dans la nature. Ainsi, la déshydratation du groupement hydroxyle (en C7) de l'umbelliférone avec un groupement prényl ajouté soit en position C6 (i.e. FCs linéaires), soit en C8 (i.e. FCs angulaires) permet la formation d'un lien éther à l'origine du noyau furane des FCs (**Figure 12**) (Fowlks, 1959).

---

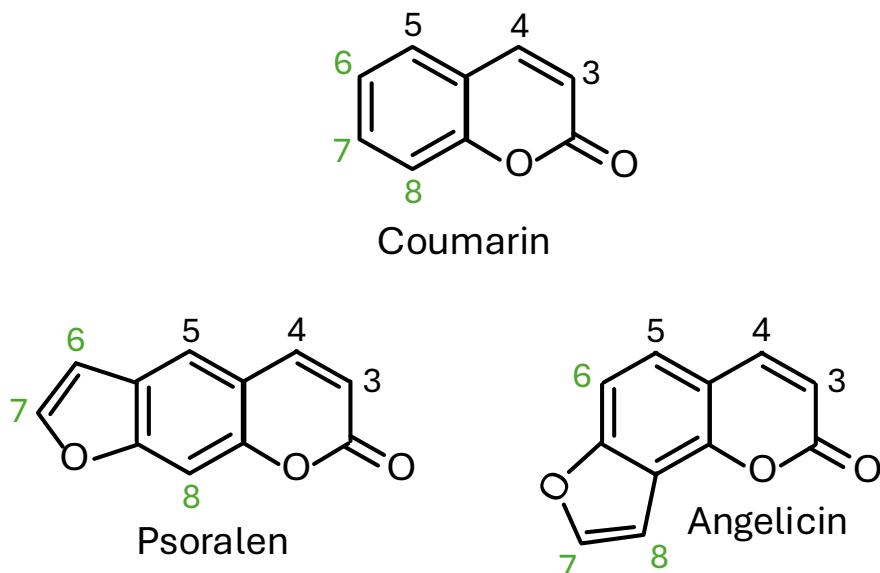
<sup>109</sup> Composés aromatiques : Molécules avec un ou plusieurs cycles constitués de plusieurs atomes de carbone, connus pour leur stabilité et leur arôme.

<sup>110</sup> Composé hémpolaire : Molécule ayant à la fois des parties polaires et non polaires, affectant sa solubilité.

<sup>111</sup> Furanylation : Modification chimique où un groupe furan, un cycle aromatique à cinq atomes contenant de l'oxygène, est ajouté à une molécule.

<sup>112</sup> Squelette moléculaire : Structure de base d'une molécule organique, déterminant ses propriétés chimiques.

<sup>113</sup> Cyclisation : Formation d'une structure cyclique à partir d'une chaîne linéaire, catalysée par exemple par une enzyme cyclase.



**Figure 12: Schéma de la position des carbones sur le squelette des coumarines (en haut) et des furocoumarines (FCs) linéaires (à gauche) et angulaires (à droite). Les carbones dont la numérotation change entre les coumarines et FCs sont en vert.**

Il est reconnu que les FCs constituent, pour les plantes, une défense efficace (Bourgaud *et al.*, 2014). Certaines FCs présentent notamment des propriétés antiherbivorie et antiappétantes<sup>114</sup> chez certaines larves de lépidoptères (Diawara *et al.*, 1993; Schuler, 2011; Stevenson *et al.*, 2003), tandis que d'autres sont capables de déclencher des photodermatites<sup>115</sup> chez certains mammifères et oiseaux (Collett, 2019). Certaines FCs disposent également de propriétés antifongiques (Razavi, 2010), nematicides (X.-B. Wang *et al.*, 2008) et antivirales (Hassan *et al.*, 2016). Les FCs sont des composés photoinductibles<sup>116</sup>. Cette propriété améliore l'impact des FCs sur la défense des plantes dans le cadre par exemple de fonctions antibactériennes, antifongiques et antibactériophages (Fowlks *et al.*, 1958; Hudson *et al.*, 1987; Lopes *et al.*, 2001). En effet, sous l'action des UVA (320-400nm), les FCs peuvent se lier aux bases pyrimidines de l'ADN par la création d'un pont cyclobutane avec les carbones C3-C4 et/ou C6-C7 (Dall'Acqua *et al.*, 1971; Musajo *et al.*, 1967). Cette propriété appelée

<sup>114</sup> Antiappétant : Composé chimique produit par une plante pour dissuader les herbivores de la consommer.

<sup>115</sup> Photodermatite : Inflammation de la peau causée par l'exposition à des substances photoactivables et à la lumière.

<sup>116</sup> Photoinductible : Caractéristique d'un processus activé par la lumière.

photocycloaddition<sup>117</sup> provient de la capacité des liaisons alcènes en C3-C4 et C6-C7 à maintenir un état élevé d'énergie après excitation aux UVs (Ben-Hur & Song, 1984; Scott *et al.*, 1976). D'autre part, certaines FCs sont capables d'affaiblir les systèmes de détoxification de certains organismes en inhibant certains cytochromes P450s (P450s)<sup>118</sup> capables de détoxifier les xénobiotiques<sup>119</sup>. Cette propriété est appelée « Effet Pomelo » du fait de sa découverte fortuite de l'impact du jus de Pomelo (riche en FCs) dans la diminution de la métabolisation de médicaments (D. Bailey *et al.*, 1991). Cette propriété est notamment due à la capacité de certaines FCs à se lier de façon covalente à la poche catalytique de P450s dont la fonction est notamment de dégrader les xénobiotiques, dont les médicaments (D. G. Bailey *et al.*, 1998; He *et al.*, 1998). Chez certains lépidoptères phytophages<sup>120</sup>, il a été démontré que cet effet inhibiteur<sup>121</sup> des FCs pouvait être contourné par l'apparition de nouveaux P450s (voir X. Li *et al.*, 2007 pour review). De même, l'effet inhibiteur pouvait être réaffirmé par les plantes *via* l'élaboration de nouvelles structures de FCs, en particulier des FC angulaires (**Figure 12**). Cette hypothèse a notamment été appuyée par les travaux pionniers de l'Américaine M. Berenbaum et son équipe qui ont suggéré le rôle de la course à l'armement et de la coévolution dans la capacité des plantes à produire de nouvelles FCs (M. Berenbaum & Feeny, 1981; M. R. Berenbaum *et al.*, 1986). Enfin, certaines FCs présentent des propriétés allélopathiques (e.g. inhibition/arrêt de la germination, limitation de la croissance cellulaire, racinaire ou plus largement de la plante) (Baskin *et al.*, 1967; Hale *et al.*, 2004; Razavi, 2010).

Face à cette toxicité, certaines plantes se sont adaptées pour compartimenter les FCs, par exemple, sous forme de cristaux foliaires (Apiacées), accumulées dans des

---

<sup>117</sup> Photocycloaddition : Réaction chimique induite par la lumière, où deux molécules se combinent pour former un cycle.

<sup>118</sup> Cytochrome P450s : Famille d'enzymes impliquées dans les processus de détoxification et de biosynthèse de nombreux composés.

<sup>119</sup> Agent xénobiotique : Composé étranger à un organisme, souvent toxique, nécessitant une détoxification.

<sup>120</sup> Phytophages : Organismes, généralement des insectes ou des animaux, qui se nourrissent exclusivement ou principalement de plantes.

<sup>121</sup> Inhibiteur : Molécule qui se lie à une enzyme et diminue ou arrête son activité catalytique.

laticifères<sup>122</sup> (*Ficus* chez les Moracées), sous formes glycosylées<sup>123</sup>, etc.) (Razavi & Zarrini, 2010; Weryszko-Chmielewska & Chwil, 2017). De plus, la gestion de l'accumulation des FCs dépend des espèces (e.g. phytoanticipines chez les Moracées, phytoalexines chez les Apiacées) (Bourgaud *et al.*, 2014; Munakata *et al.*, 2020). D'autre part, certaines FCs peuvent induire un rétrocontrôle de la biosynthèse de ses propres précurseurs<sup>124</sup> (Gravot *et al.*, 2004).

Ces deux dernières décennies, de nombreuses étapes de la voie des FCs ont été identifiées (**Figure 13**). Actuellement, la majorité des enzymes identifiées dans la voie des FCs sont décrites chez des Apiacées (Hehmann, Lukačin, *et al.*, 2004; Ji *et al.*, 2024; Karamat *et al.*, 2014; Krieger *et al.*, 2018; Larbat *et al.*, 2007, 2009; Munakata *et al.*, 2016) et plus récemment des Rutacées (Limones-Mendez *et al.*, 2020) et des Moracées (Munakata *et al.*, 2020; Villard *et al.*, 2021). Une description principalement de la partie linéaire de la voie qui aura permis de rendre accessible la production hétérologue<sup>125</sup> de ces composés à haute valeur ajoutée (Z. Wang *et al.*, 2023).

C'est à partir de l'umbelliférone que dérivent les FCs linéaires comme angulaires (Karamat *et al.*, 2014; Larbat *et al.*, 2007, 2009; Munakata *et al.*, 2016; Villard *et al.*, 2021). À noter que la biosynthèse de FCs est imbriquée à celle des pyranocoumarines linéaires et angulaires chez les Apiacées (Q. Li *et al.*, 2024; Y. Zhao *et al.*, 2024). Que ce soit pour la biosynthèse de FCs angulaires ou linéaires, l'umbelliférone est C-prénylée par une prényltransférase<sup>126</sup> pour produire respectivement de l'osthénol ou de la déméthylsubérosine (Karamat *et al.*, 2014; Munakata *et al.*, 2016, 2020). Ensuite, la cyclisation du noyau furane fait intervenir une enzyme de la famille des cytochromes P450 (P450) pour former respectivement soit de la columbianétine, soit de la marmésine (Villard *et al.*, 2021; K. Wang *et al.*, 2024; Y. Zhao *et al.*, 2024). Le squelette FCs est ensuite

---

<sup>122</sup> Laticifère : Cellule ou tissu végétal spécialisé dans la production et le transport du latex, un fluide souvent utilisé par les plantes pour la défense contre les herbivores.

<sup>123</sup> Glycosylation : Ajout d'un glucide à une molécule, souvent un composé non sucré (aglycone), modifiant sa solubilité et sa fonction.

<sup>124</sup> Précurseur de biosynthèse : Molécule initiale dans une voie métabolique, utilisée pour produire un composé plus complexe.

<sup>125</sup> Production hétérologue : Production d'une protéine (appelée recombinante) dans un organisme différent de celui d'origine grâce à des techniques de génie génétique.

<sup>126</sup> Prénylation : Ajout d'un groupe prénylique dérivé du diméthylallyl pyrophosphate (DMAPP) à une molécule, souvent dans les terpénoïdes.

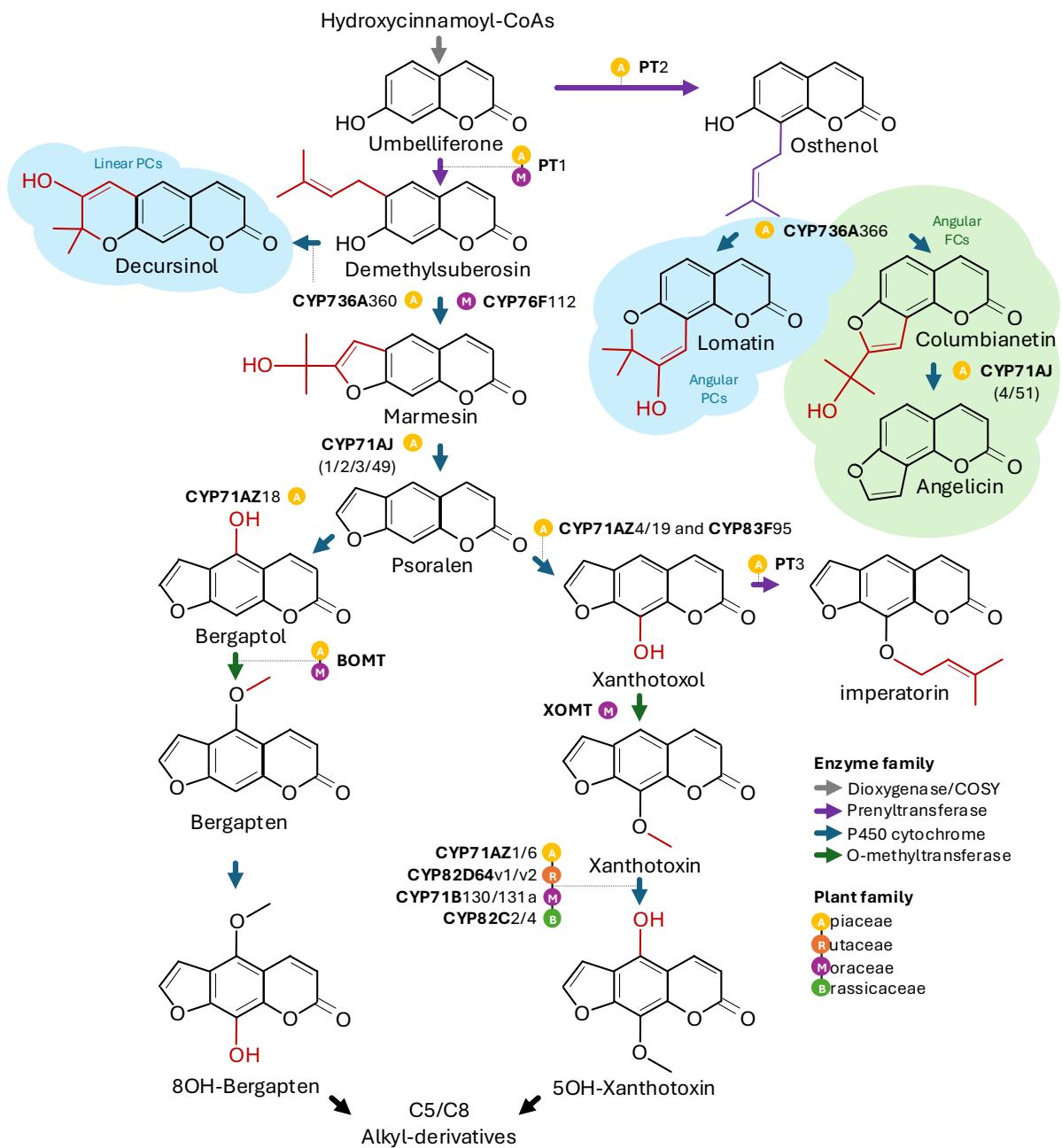
débarrassé de son groupement acétone en C7 par *syn*-élimination<sup>127</sup> (Jian *et al.*, 2020; Larbat *et al.*, 2007, 2009; Schelkunov *et al.*, 2024). Actuellement, aucune étape ultérieure de biosynthèse dans la voie des FCs angulaire n'a été décrite. Concernant la voie des FCs linéaires, le squelette psoralène peut être hydroxylé en C5 et/ou en C8 pour former respectivement du bergaptol ou du xanthotoxol (**Figure 13**) (Ji *et al.*, 2024; Krieger *et al.*, 2018). Ces produits peuvent être O-méthylés par une O-méthyltransférase<sup>128</sup> pour ensuite former respectivement du bergaptène ou de la xanthotoxine (Hehmann, Lukačin, *et al.*, 2004; Y. Zhao *et al.*, 2016). Le xanthotoxol peut également être O-prényle par une O-prényltransférase pour former de l'impératorine (Han *et al.*, 2023). Enfin, la xanthotoxine peut être hydroxylée<sup>129</sup> en C5 pour produire de la 5-hydroxyxanthotoxine (Krieger *et al.*, 2018; Kruse *et al.*, 2008; Limones-Mendez *et al.*, 2020) elle-même O-méthylée pour former de l'isopimpineline.

---

<sup>127</sup> Syn-élimination : Réaction d'élimination dans laquelle deux groupes partent simultanément de la même face d'une molécule, entraînant la formation d'une double liaison.

<sup>128</sup> Méthylation : Ajout d'un groupe méthyl (-CH<sub>3</sub>) à une molécule sur un groupe hydroxyl, catalysé par une O-méthyltransférase.

<sup>129</sup> Hydroxylation : Ajout d'un groupe hydroxyle (-OH) à une molécule, souvent catalysé par une hydroxylase, modifiant la solubilité.



**Figure 13: Représentation simplifiée des étapes décrites de la voie de biosynthèse des furucoumarines (FCs) et des pyranocoumarines (PCs) chez certaines angiospermes.** Les familles d'enzymes et leurs origines au sein des familles de plantes sont représentées par des flèches et des symboles colorés. Les étapes biocatalytiques sont détaillées dans : Hehmann et al., 2004 ; Larbat et al., 2007, 2009 ; Vialart et al., 2012 ; Karamat et al., 2014 ; Munakata et al., 2016, 2020 ; Roselli et al., 2017 ; Vanholme et al., 2019 ; Krieger et al., 2018 ; Zhao et al., 2016, 2024 ; Limones-Mendez et al., 2020 ; Villard et al., 2021 ; Han et al., 2023 ; Wang et al., 2024 ; Schelkunov et al., 2024 ; Li et al., 2024 ; Ji et al., 2024.

Jusqu'à présent, la capacité des plantes à accumuler des FCs a été décrite pour 4 familles botaniques, à savoir, les Apiacées, les Rutacées, les Fabacées et les Moracées

(Bourgaud *et al.*, 2014). Toutefois, la capacité des plantes à synthétiser des FCs a été démontrée chez au moins 15 familles botaniques (Munakata *et al.*, 2020). Elles pourraient être plus largement répandues au sein des angiospermes. En effet, les FCs sont retrouvées au sein de taxons primordiaux chez les angiospermes comme chez les Schisandracées (R.-T. Li *et al.*, 2005). Du point de vue de leur diversité, le référencement le plus abouti comptabilise un peu plus de 300 FCs, principalement des FCs linéaires. Ce référencement découle des travaux de synthèse de R. D. H. Murray (Murray, 2002) dans le cadre du plus grand rassemblement de produits naturels du siècle dernier comme synthétisé par (Soukup & Soukup, 2015). Peu de nouvelles structures ont été mises au jour depuis. Cependant, des FCs identiques ont été identifiées chez différents taxons (Sarker & Nahar, 2017).

La description de quelques étapes de la voie des FCs linéaires chez plusieurs familles botaniques a récemment permis d'apporter une vision évolutive de l'acquisition et de la diversification de la voie chez les angiospermes. C'est le cas par exemple des fonctions déméthylsubérosine synthase et 5-hydroxyxanthotoxine synthase (Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020; Villard *et al.*, 2021). Ainsi, la capacité de biosynthèse des FCs aurait été permise par le recrutement indépendant de sous-familles enzymatiques diverses chez plusieurs taxons botaniques. Une description de l'évolution de la fonction marmesine synthase chez les Moracées appuie ce constat (Villard *et al.*, 2021). Dans l'objectif de mieux comprendre comment la voie des FCs a évolué chez les Moracées, je présente dans la fin de ce chapitre introductif un état de l'art de la paléobiogéographie<sup>130</sup> des Moracées, de la chemophylogénie des FCs et de leur compartimentation au sein d'espèces décrites dans cette famille.

---

<sup>130</sup> Paléobiogéographie : Étude de la distribution des organismes fossiles et de leurs liens avec les changements géographiques passés.

## B- Émergence et évolution potentielle des FCs chez les Moracées

### a- Évolution paléobiogéographique et morphologique des Moracées du Crétacé au Miocène

Les angiospermes se sont fortement diversifiées entre la fin de l'ère jurassique<sup>131</sup> et durant le début du Crétacé<sup>132</sup> (Benton *et al.*, 2022; Dimitrov *et al.*, 2023). Parmi elles, les Moracées dérivent de l'ordre des Rosales et plus généralement des eudicotylédones dont les premiers taxons se sont diversifiés durant la fin du Crétacé. Les Moracées constituent une famille monophylétique<sup>133</sup> très diversifiée avec plus de 1100 espèces incluses dans 39 genres et 7 tribus (Gardner *et al.*, 2021) (**Figure 14**). Cette dernière a vraisemblablement émergé à la fin du Crétacé entre le Santonian<sup>134</sup> et le Maastrichtien<sup>135</sup> il y a ~73,2-84,7 Ma sur le supercontinent Laurasia<sup>136</sup> dans le contexte de la fragmentation de la Pangée<sup>137</sup> (Zerega *et al.*, 2005; Q. Zhang *et al.*, 2019). Ainsi, les 7 tribus<sup>138</sup> fondatrices des Moracées (**Figure 14**) ont pris forme durant cette période principalement au niveau des paléotropiques<sup>139</sup> et néotropiques<sup>140</sup>. Ce n'est que bien plus tard au début du Paléocène<sup>141</sup> impacté par un réchauffement atmosphérique que certains des premiers genres actuellement majoritaires se sont constitués (e.g. *Ficus*, *Morus*, *Fatoua*, *Maclura*) (Gardner *et al.*, 2017, 2023; Machado *et al.*, 2018; Pederneiras *et al.*, 2018; C.-X. Yang et

---

<sup>131</sup> Jurassique : Période géologique s'étendant d'environ 201 à 145 millions d'années, caractérisée par la dominance des dinosaures, l'épanouissement des plantes à fleurs et des changements significatifs dans les environnements marins et terrestres.

<sup>132</sup> Crétacé : Dernière période de l'ère mésozoïque, il y a environ 145 à 66 millions d'années, marquée par l'extinction des dinosaures et la diversification des angiospermes.

<sup>133</sup> Monophylétique : Groupe d'organismes descendant d'un ancêtre commun unique, incluant tous ses descendants.

<sup>134</sup> Santonian : Étape de la fin du Crétacé, il y a environ 86 à 83 millions d'années, marquée par l'apparition de nombreuses lignées d'angiospermes.

<sup>135</sup> Maastrichtien : Dernière étape du Crétacé, il y a environ 72 à 66 millions d'années, marquant la grande extinction de la fin de l'ère mésozoïque.

<sup>136</sup> Laurasia : Supercontinent du nord qui se sépara de la Pangée, comprenant l'Amérique du Nord, l'Europe et l'Asie.

<sup>137</sup> Pangée : Supercontinent unique qui regroupait toutes les masses terrestres il y a environ 300 millions d'années avant de se fragmenter.

<sup>138</sup> Tribus : Classification taxonomique regroupant plusieurs genres d'organismes végétaux.

<sup>139</sup> Paléotropiques : Région géographique tropicale comprenant l'Afrique, l'Asie du Sud et l'Océanie, ancienne de plusieurs millions d'années.

<sup>140</sup> Néotropiques : Région tropicale d'Amérique, incluant l'Amérique centrale, l'Amazonie et les Caraïbes.

<sup>141</sup> Paléocène : Première période de l'ère cénozoïque, entre 66 et 56 millions d'années, succédant à l'extinction des dinosaures.

*al.*, 2023). Plus spécifiquement, c'est durant l'Éocène<sup>142</sup> que la famille des Moracées s'est majoritairement diversifiée, avec le déploiement précoce des genres déjà constitués durant le Lutétien<sup>143</sup> (~45,8 Ma) dans les zones sèches Afroasiatiques et nord Eurasienne, et plus tardivement durant le Rupélien<sup>144</sup> (~29,3 Ma) dans le reste du monde (Dimitrov *et al.*, 2023). L'Éocène est également une période fondatrice pour les genres *Artocarpus* sur l'île de Bornéo et *Morus* en Amérique du Nord et dans la région Sino-himalayenne (Williams *et al.*, 2017; C.-X. Yang *et al.*, 2023). De la même façon, cette période a été faste pour la radiation de certains sous-genres *Ficus* en Amérique du Nord, en Asie du Sud-Est et en Inde. Cette radiation a notamment permis une grande diversification du genre *Ficus* durant le Miocène<sup>145</sup> et l'Oligocène<sup>146</sup> dans toutes les régions du monde notamment tropicales (Gardner *et al.*, 2023; Machado *et al.*, 2018; Pederneiras *et al.*, 2018). Enfin, ce n'est que durant l'Oligocène que le genre *Brosimum* a émergé au sein des *Dorsteniae*. Actuellement, parmi l'ensemble des espèces référencées par genres chez les Moracées, les genres les plus diversifiés sont *Ficus* (64,9%), *Dorstenia* (8,5%), *Artocarpus* (7,4%) et dans un moindre mesure *Morus* (2,5%), *Sorocea* (1,9%), *Brosimum* (1,8%), *Naucleopsis* (1,7%), *Maclura* (1,1%), etc. (**Figure 14**).

---

<sup>142</sup> Éocène : Période géologique de 56 à 34 millions d'années, marquée par une diversification importante des plantes à fleurs.

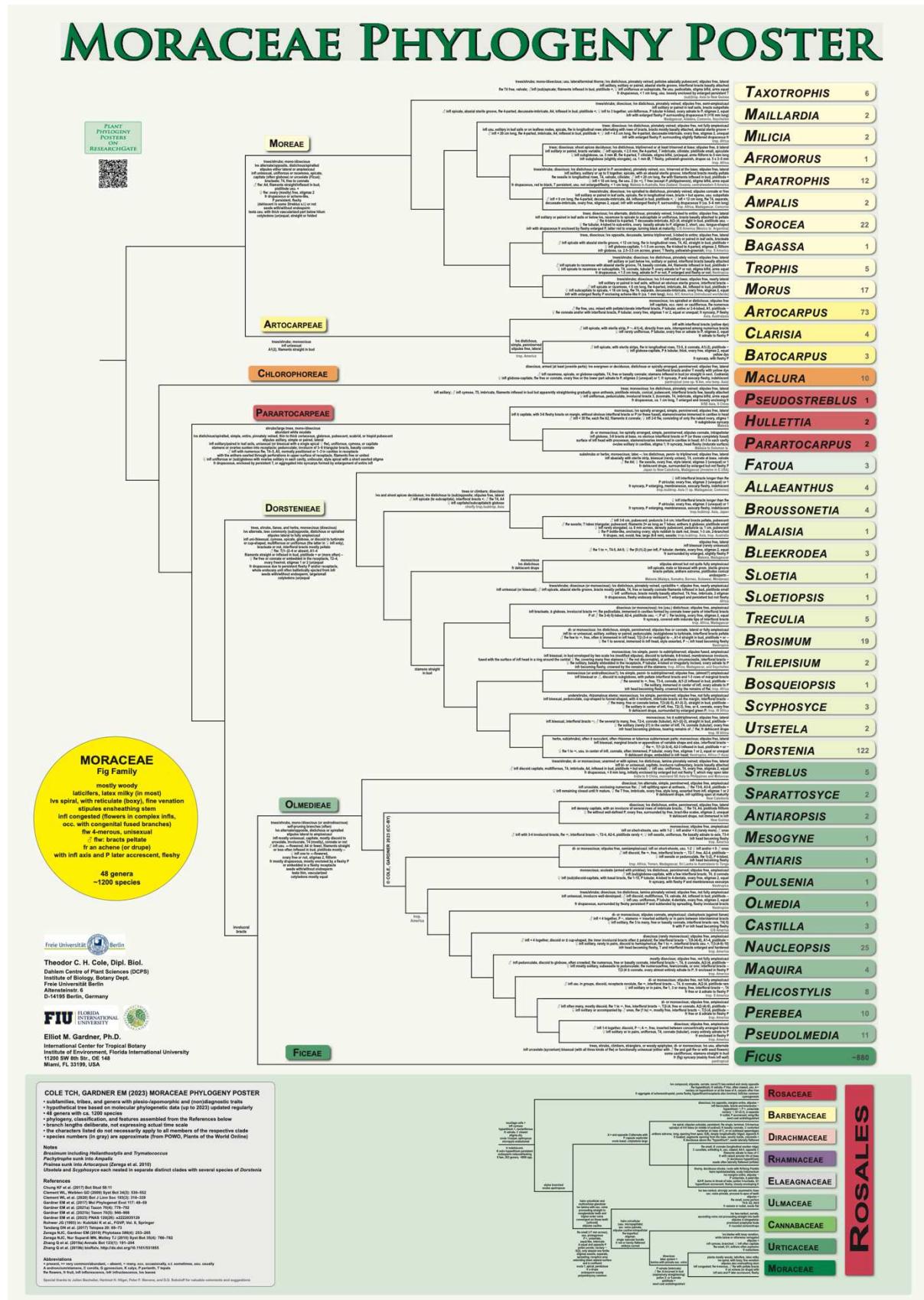
<sup>143</sup> Lutétien : Étage géologique de l'Éocène, datant d'environ 48 à 41 millions d'années.

<sup>144</sup> Rupélien : Premier étage de l'Oligocène, il s'étend de 33,9 à 28,1 millions d'années, marqué par des changements climatiques majeurs.

<sup>145</sup> Miocène : Période géologique allant de 23 à 5,3 millions d'années, durant laquelle de nombreuses familles de plantes modernes ont émergé.

<sup>146</sup> Oligocène : Période géologique de 34 à 23 millions d'années, caractérisée par un refroidissement global et l'extension des prairies.

# MORACEAE PHYLOGENY POSTER



**Figure 14: Phylogénie des Moracées.** D'après Gardner *et al.*, 2023.

## b- Diversité phytochimique des FCs chez les Moracées

Parmi les Moracées, je référence la présence de 54 FCs (Figure 15 ;Tableau 5). Globalement, les FCs identifiées le plus souvent chez les Moracées correspondent à des formes simples, comme le psoralène, ou le bergaptène qui est sa forme méthoxylée<sup>147</sup> en C5, bien qu'il existe une importante diversité de structures (Figure 13, 15a). La distribution en FCs varie en fonction des genres (Figure 15). Ainsi, les genres *Maclura* et *Broussonetia* produisent des formes simples et aucune autre forme plus complexe que le bergaptène n'a été répertoriée parmi ces genres. Seules les genres *Maquira*, *Ficus* et *Dorstenia* semblent produire une diversité plus importante de FCs ramifiées en C5 (e.g. pranferol, cnidiline, dorsténine, etc.). Ces ramifications<sup>148</sup> portent sur la O-prénylation (*Ficus*, *Dorstenia*), la O-glycosylation (*Ficus*), ainsi que sur la C-prénylation et la O-géranylation<sup>149</sup> (*Dorstenia*). D'autre part, les genres *Fatoua*, *Ficus* et *Dorstenia* sont capables de ramifications en C8 (e.g. rutarétine méthyl-éther, swietenocoumarines, etc.). Ces ramifications portent sur la méthoxylation (de la marmesine pour *Fatoua*), la O-glycosylation (*Ficus*), ainsi que sur la O-prénylation, la C-prénylation et la O-géranylation (*Dorstenia*). À noter que la ramification géanyl chez les *Dorstenia* est généralement retrouvée sous formes hydroxylées, époxydées<sup>150</sup> voire  $\gamma$ - ou  $\delta$ -lactonisées<sup>151</sup>. Chez les Moracées, certaines espèces produisent également des ramifications en C6 à partir de la marmesine (e.g. xanthoarnol, etc.), ce qui est très peu décrit par ailleurs dans la voie des FCs chez les angiospermes. Ces ramifications portent sur l'hydroxylation en C6 (*Brosimum*, *Dorstenia*), la mono- et di- O-glycosylation (*Dorstenia*), ainsi que sur la cétonisation<sup>152</sup>/déhydratation (*Brosimum*). Enfin, seul le genre *Dorstenia* semble capable de réaliser la ramification C-prenyl en C3 à partir de

<sup>147</sup> Méthoxylation : Ajout d'un groupe méthoxy (-OCH<sub>3</sub>) à une molécule, souvent impliqué dans la modification de composés organiques.

<sup>148</sup> Ramification chimique : Modification d'une structure moléculaire en ajoutant des groupes fonctionnels qui altèrent sa réactivité.

<sup>149</sup> Géranylation : Ajout d'un groupement géanyl à une molécule, souvent dans la biosynthèse des terpènes.

<sup>150</sup> Epoxydation : Formation d'un époxyde, un cycle à trois atomes, sur une molécule par l'action d'une époxydase.

<sup>151</sup> Lactonisation : Formation d'une lactone, un cycle fermé par un ester interne de 5 ( $\gamma$ ) ou 6 ( $\delta$ ) carbone.

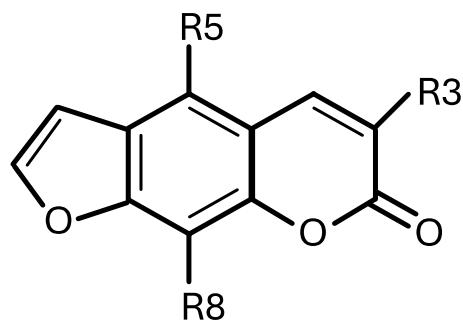
<sup>152</sup> Cétonisation : Réaction chimique par laquelle un alcool est converti en cétone, généralement via l'oxydation.

diverses structures de FCs linéaires (e.g. psoralène, bergaptène, oxypeucédanine, etc.) (**Figure 15b**).

Par ailleurs, parmi les Moracées j'ai référencé 10 FCs angulaires uniquement au sein des genres *Ficus* et *Dorstenia* (e.g. isobergaptène, pimpinelline, etc.) (**Figure 15c**). Ces ramifications portent sur l'hydroxylation en C5 et/ou C6, de la glycosylation (*Ficus*, *Dorstenia*), ainsi que sur la O-prénylation, la C-prénylation et la O-méthylation (*Dorstenia*).

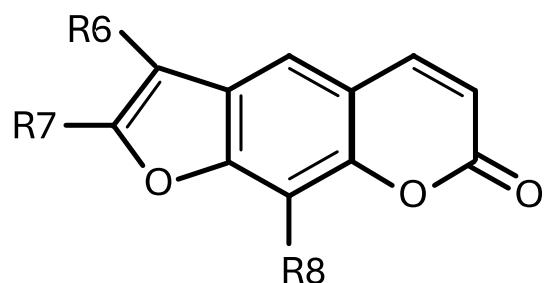
La caractérisation phytochimique de certaines Moracées ces 30 dernières années a permis d'identifier des FCs au sein d'espèces appartenant aux tribus sœurs des Dorsteniae (**Figure 14**). Ainsi, des FCs ont été retrouvées dans tous les organes des espèces étudiées et appartenant au genre *Dorstenia* et *Ficus*. À l'inverse, la présence de FCs n'a été que partiellement mise en évidence pour d'autres genres (*Brosimum* : racines et tiges, *Fatoua* : plante entière sans distinction, *Maquira* : tiges uniquement). La caractérisation phytochimique des FCs est principalement basée sur des approches de RMN. Ces approches considèrent généralement les composés fortement accumulés. Ainsi, ce référencement propose un aperçu partiel des compartiments capables d'accumuler des FCs, pour quelques tissus investigués parmi quelques espèces de Moracées.

a



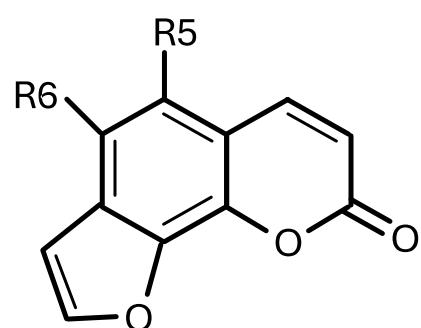
n°	Mc	Fa	Bo	Br	Do	Ma	Fi	R5	R8	R3
1										
2										
3								OH		
4								O-me		
5								O-me		
6								O-Pr		
7								O-Pr-OH		
8								O-Pr-diOH		
9								O-ge-γ-lactone		
10								O-ge-desat-γ-lactone		
11								OH	OH	
12								OH	O-β-D-G	
13								O-β-D-G	OH	
14								O-me	O-me	
15								O-Pr	O-me	
16								O-me	O-Pr	
17								O-me	C-Pr*	
18								O-me	C-Pr	
19								O-me	O-Pr-DiOH	
20								O-me	C-Pr-DiOH	
21								O-me	C-Pr*-DiOH	
22								O-me	O-ge	
23								O-me		C-Pr
24								O-me		C-Pr-DiOH
25								O-me		C-Pr-DiOH-Ac
26								O-Pr-diOH		C-Pr-DiOH-O-β-D-G
27								O-me	O-me	C-Pr
28								O-me	O-me	C-Pr-DiOH
29								O-me		C-Pr*
30								O-me		C-Pr*-DiOH
31								O-me		C-Pr*-O-β-D-G
32								O-Pr-epo		C-Pr*
33								O-Pr-diOH		C-Pr*
34								O-Pr-OH-O-me		C-Pr*
35								O-Pr-OH-Cl		C-Pr*
36								O-me	O-me	C-Pr*

b



n°	Mc	Fa	Bo	Br	Do	Ma	Fi	C7	C8	C6
37								Ter-But-OH		
38								Ter-But-DiOH-Ac		
39								Ter-But-OH	OH	
40								Ter-But-OH	O-me	
41								Ter-But-OH		OH
42								Ter-But-DiOH-O- $\beta$ -D-G		OH
43								Ter-But-DiOH-Di-O- $\beta$ -D-G		OH
44								IsoB		=O

c



n°	Mc	Fa	Bo	Br	Do	Ma	Fi	R5	R6
45									
46								O-me	
47								O-me	O-me
48								C-Pr-DiOH	O-me
49								O-Pr-DiOH-O- $\beta$ -D-G	O-me
50								O-Pr-OH=O	O-me
51								O-Pr	O-me
52								O- $\beta$ -D-G	OH
53								OH	O- $\beta$ -D-G
54								O- $\beta$ -D-G	O- $\beta$ -D-G

**Figure 15: État de l'art des ramifications possibles du squelette des furanocoumarines (a,b) linéaires et (c) angulaires chez les Moracées.** Descriptions des ramifications (R) sur les carbones (a) C5/C8/C3 et (b) C7/C8/C6 du psoralène et (c) C5/C6 de l'angélicine. Groupements chimiques : Pr\*=groupe prényl neohexene ; Pr=groupe prenyl diméthylallyle ; O-me=groupe O-méthyl ; OH=groupe hydroxyle ; O-ge=groupe O-géranyl ; O- $\beta$ -D-G=O- $\beta$ -D-Glucopyranoside ; epo=époxyde ; Ac=groupe acétyl ; Cl=groupe Chlore ; Ter-But=Tert-butane ; =O=formaldéhyde ; IsoB=Isobutyléne. Les dimères sont fusionnés en C3/C4. Pour plus d'informations sur les noms voir **Tableau 5** en annexes. Espèces de Moracées appartenant aux genres : Maclura (Mc), Fatoua (Fa), Broussonetia (Bo), Brossimum (Br), Dorstenia (Do), Maquira (Ma) et Ficus (Fi). D'après (Abegaz *et al.*, 2002; Caceres *et al.*, 2001; Chang *et al.*, 2005; Chiang *et al.*, 2010; Damjanić & Akačić, 1974; Gibernau *et al.*, 1997; Gottlieb *et al.*, 1972; Heinke *et al.*, 2011; Hou *et al.*, 2001; Kitajima *et al.*, 1998; Kuster *et al.*, 1994; Lansky *et al.*, 2008; Lee *et al.*, 2004; Llabres *et al.*, 1992; Lopes *et al.*, 2001; Nana *et al.*, 2012; Ngadjui *et al.*, 1998; Ngameni *et al.*, 2006; Pistelli *et al.*, 2000; Rojas-Lima *et al.*, 1999; Rovinski *et al.*, 1987; Terreaux *et al.*, 1995; Uchiyama *et al.*, 2002; Vieira *et al.*, 1999; J. H. Y. Vilegas *et al.*, 1997; W. Vilegas *et al.*, 1993; Yarosh & Nikonov, 1973).

## II. Les enzymes de la famille des cytochromes P450s

### 1- Acteurs majeurs de la biosynthèse de MS chez les plantes

Les gènes codants pour le MS prennent une place importante dans le génome des plantes supérieures et représentent ~10-20% des gènes codants (Tissier *et al.*, 2014). Parmi les enzymes qui produisent des MS sont principalement retrouvées des cytochromes P450s (P450s), dioxygénases<sup>153</sup> dépendantes de l'a-cétoglutarates, glycosyltransférases, méthyltransférases, prényltransférases et acyltransférases<sup>154</sup> (Modolo *et al.*, 2009). Chez les plantes supérieures, les P450s forment une superfamille qui compte parmi les plus diversifiées dans le génome. Elle regroupe jusqu'à 1% du génome codant (Mizutani & Ohta, 2010; D. Nelson & Werck-Reichhart, 2011). Sa démultiplication semble être une spécificité végétale puisque la majorité des P450s identifiés le sont chez les plantes (53%) et dans une moindre mesure chez les champignons (24%), les bactéries (18%), les animaux (4%), les protozoaires (0,2%), les archées (<0,1%) et même les virus (<0,01%) (D. R. Nelson, 2018).

Le premier P450 a été identifié dans le foie humain comme étant une hémoprotéine<sup>155</sup> capable de fixer le monoxyde de carbone et d'absorber la lumière à 450nm, ce qui lui a valu sa dénomination (Omura & Sato, 1962). La vaste diversité de P450s acquise dès lors a rapidement amené à classer les séquences en fonction de leurs homologies de séquences protéiques (i.e. sous-familles >55%, familles >40% et clans) (D. R. Nelson *et al.*, 1996). Cette classification regroupait en 2018, ~350 000 P450s pour 2252 familles (D. R. Nelson, 2018). À titre d'exemple, CYP71B130 représente un P450 appartenant à la famille 71 et à la sous-famille B au sein de laquelle il est référencé comme l'isoenzyme<sup>156</sup> 130.

---

<sup>153</sup> Dioxygénase : Enzyme qui catalyse l'incorporation de deux atomes de dioxygène (O<sub>2</sub>) dans un substrat.

<sup>154</sup> Acétylation : Ajout d'un groupe acétyle à une molécule, catalysé par une acyltransférase.

<sup>155</sup> Hémoprotéine : Protéine contenant un groupe prosthétique hémique, jouant un rôle clé dans les processus redox.

<sup>156</sup> Isoenzyme : Variantes d'une enzyme qui catalysent la même réaction mais avec des structures ou des régulations différentes.

Parmi les P450s, 22 activités catalytiques ont été décrites jusqu'alors chez les plantes comme synthétisé par (Hansen *et al.*, 2021). Néanmoins, l'activité principale reste la monooxygénéation<sup>157</sup>.

La diversification des P450s a permis aux plantes d'acquérir au cours de l'évolution de nouvelles fonctions catalytiques facilitant la biosynthèse de métabolites (MC + MS) (Mizutani, 2012). Ainsi, certains P450 des familles CYP51 et CYP710 ont apporté la capacité aux cyanobactéries de produire des éléments vitaux à l'homéostasie<sup>158</sup> cellulaire comme les phytostérols (Bak *et al.*, 1997; Morikawa *et al.*, 2006). Les CYP97 ont apporté la capacité de produire les xanthophylles qui sont impliquées dans la production de phytotpigments<sup>159</sup> photosynthétiques (J. Kim & DellaPenna, 2006; Tian *et al.*, 2004). Certains CYP86, CYP94 et CYP703 ont facilité la terrestrialisation<sup>160</sup> en permettant l'hydroxylation d'acides gras et la biosynthèse des précurseurs de la cutine et de la sporopollénine indispensables à l'homéohydrie<sup>161</sup> et la protection contre la dessiccation (Cannell *et al.*, 2020; Mizutani, 2012). La protection hydrique couplée avec la diversification de la voie des phénylpropanoïdes a permis l'acquisition d'un port érigé qui est à l'origine des trachéophytes (e.g. lignines, subérines). Cette diversification a également favorisé la diversification de MS permettant de faire face aux UVs (flavonoïdes) et aux bioagresseurs (isoflavonoïdes, coumarines, lignanes), tout en facilitant la communication avec l'environnement biotique (benzenoïdes et phénolpropènes) (Mizutani, 2012). Enfin, durant l'apparition des angiospermes, le CYPome<sup>162</sup> et notamment le clan 71 s'est fortement démultiplié. Ce dernier concentre jusqu'à la moitié du CYPome et la plupart des P450s fonctionnellement décrits (D. Nelson & Werck-Reichhart, 2011) (**Figure 16**). Cette radiation a notamment eu pour conséquences de permettre la biosynthèse de nombreux MS dont des phytohormones

---

<sup>157</sup> Monooxygénéation : Introduction d'un atome d'oxygène dans une molécule organique, catalysée par une enzyme monooxygénase.

<sup>158</sup> Homéostasie : Capacité d'un organisme à maintenir des conditions internes stables malgré les variations externes.

<sup>159</sup> Phytotpigment : Molécule pigmentaire produite par les plantes, souvent impliquée dans la capture de la lumière pour la photosynthèse.

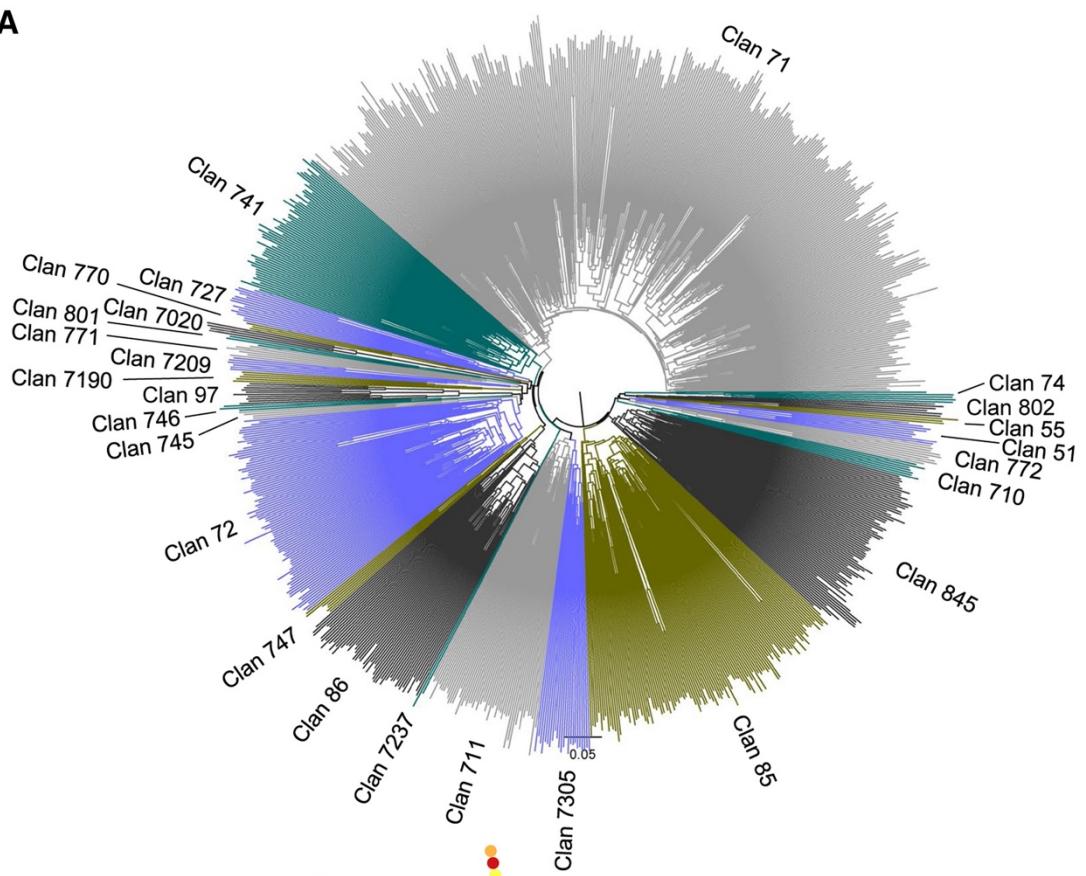
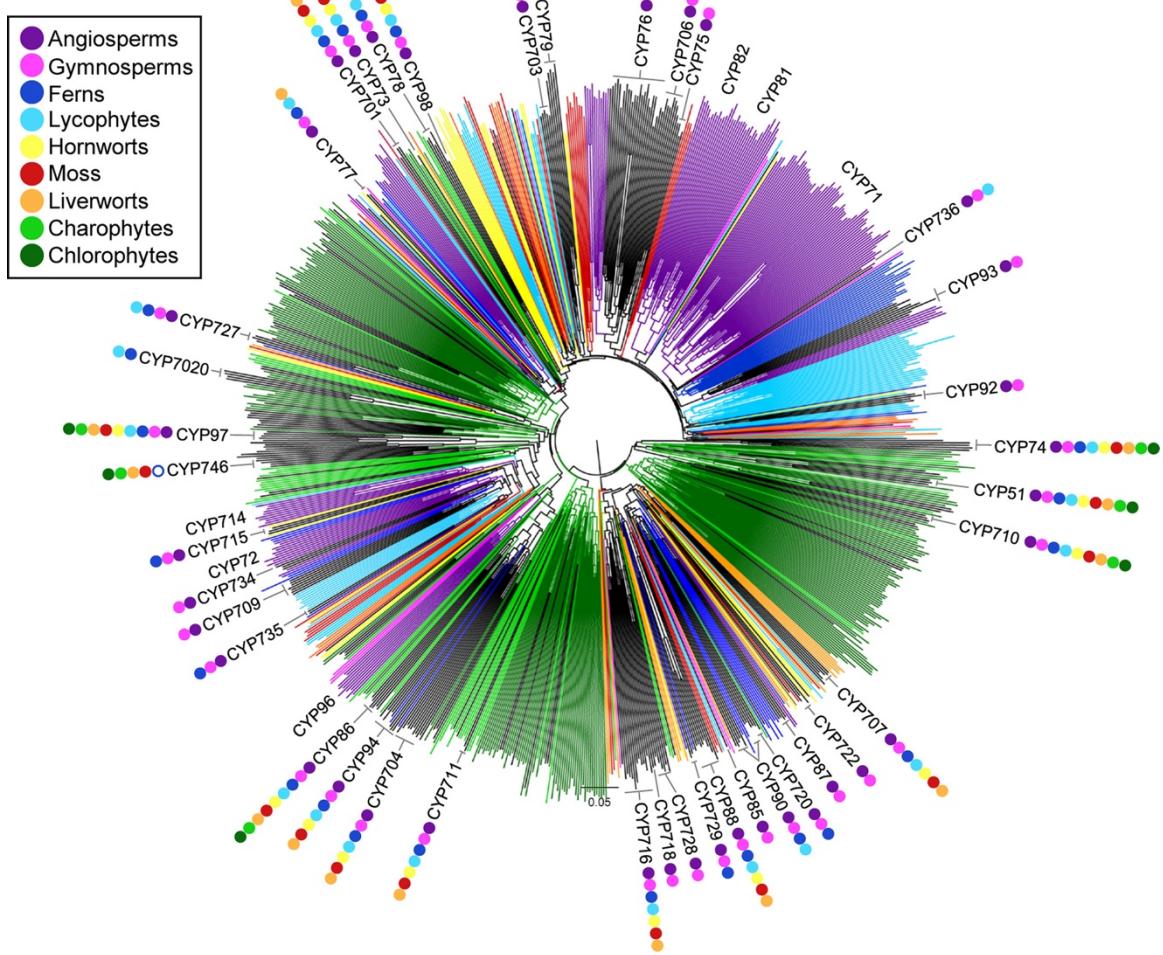
<sup>160</sup> Terrestrialisation : Transition évolutive des organismes vivant dans l'eau vers une vie sur terre ferme.

<sup>161</sup> Homéohydrie : Capacité d'un organisme à maintenir une hydratation constante malgré des conditions environnementales changeantes.

<sup>162</sup> CYPome : Ensemble des gènes codant pour les enzymes cytochrome P450 d'un organisme.

et de faciliter la régulation du phénotype et les interactions des plantes à fleurs avec l'environnement (Mizutani, 2012). Les gènes codants pour des P450s et impliqués dans la biosynthèse de phytohormones sont maintenant bien décrits (Blázquez *et al.*, 2020). À l'inverse, bien que la fonction de nombreux P450s impliqués dans le MS soit décrite (H. Wang *et al.*, 2021), celle de la plupart reste encore à être élucidées (X.-Q. Huang & Dudareva, 2023). En 2021, 578 P450s étaient fonctionnellement décrits (Hansen *et al.*, 2021), parmi les plus de 184 000 référencés chez les plantes (D. R. Nelson, 2018).

D'un point de vue évolutif, il apparaît que des activités enzymatiques identiques peuvent être portées par des P450s recrutés au sein de mêmes sous-familles ou bien au sein de familles totalement différentes pour des espèces évolutivement éloignées (Pichersky & Lewinsohn, 2011). C'est le cas par exemple pour la voie des FCs et plus particulièrement pour la fonction 5-hydroxyxanthotoxine synthase (Krieger *et al.*, 2018; Limones-Mendez *et al.*, 2020). Ainsi, il est reconnu que pour de nombreuses voies de biosynthèse, des phénomènes d'évolution convergente impliquant des P450s ont permis l'acquisition de signatures phytochimiques convergentes (e.g. voies des furocoumarines, caroténoïdes, bétalaïnes, etc.) (Hansen *et al.*, 2021; Pichersky & Lewinsohn, 2011).

**A****B**

**Figure 16: Phylogénie des familles de cytochromes P450 (CYP) chez les algues vertes et les plantes terrestres.** (A) Les branches sont colorées en fonction de la distribution des clans de CYP (A) ou des groupes taxonomiques (B). Les branches noires représentent les familles de CYP partagées entre plusieurs groupes taxonomiques. Certaines des familles conservées de CYP sont marquées avec des points colorés pour indiquer leur présence dans certains taxons. Les branches colorées illustrent les familles de CYP spécifiques à certains taxons. D'après Hansen *et al.*, 2021.

## 2- Structure et fonction d'un P450

Chez les eucaryotes, les P450s sont des enzymes membranaires constituées d'une queue d'ancrage N-terminale et d'une tête plongée à l'interface membrane/cytosol<sup>163</sup> (**Figure 17**) (Šrejber *et al.*, 2018). Ils sont retrouvés chez les eucaryotes principalement à la surface du réticulum endoplasmique<sup>164</sup>, et plus rarement dans d'autres organites<sup>165</sup> (Omura & Gotoh, 2017). La catalyse<sup>166</sup> des P450s est notamment permise par un cofacteur<sup>167</sup> prosthétique<sup>168</sup> de type protoporphyrine IX (i.e. hème b) qui contient un noyau central ferreux ( $Fe^{2+}$ ). Ce dernier est attaché *via* un groupement hème-thiolate (i.e. S-Fe) à un ensemble de cystéines très conservées dans la poche catalytique<sup>169</sup> (Green, 2009; Yosca *et al.*, 2017). Les P450s fonctionnent généralement de pair avec un partenaire redox<sup>170</sup> comme la NADPH Cytochrome P450 Réductase (CPR), ou le cytochrome b5 par exemple, qui transmettent les électrons nécessaires à la catalyse (Urban *et al.*, 1997; X. Zhao *et al.*, 2023). Il existe 4 classes de P450s conditionnées par la

<sup>163</sup> Cytosol : Composant liquide du cytoplasme d'une cellule, dans lequel se trouvent les organites et d'autres structures. Il est riche en enzymes, nutriments et ions, et joue un rôle clé dans les processus métaboliques.

<sup>164</sup> Réticulum endoplasmique : Organite impliqué dans la synthèse des protéines et des lipides, connecté au noyau cellulaire.

<sup>165</sup> Organite : Structure spécialisée au sein d'une cellule, comme le noyau ou la mitochondrie, ayant une fonction spécifique.

<sup>166</sup> Biocatalyse : Utilisation d'enzymes pour accélérer des réactions chimiques spécifiques, souvent dans des contextes industriels ou biologiques.

<sup>167</sup> Cofacteur : Molécule non protéique nécessaire à l'activité d'une enzyme, souvent un ion métallique ou une petite molécule organique.

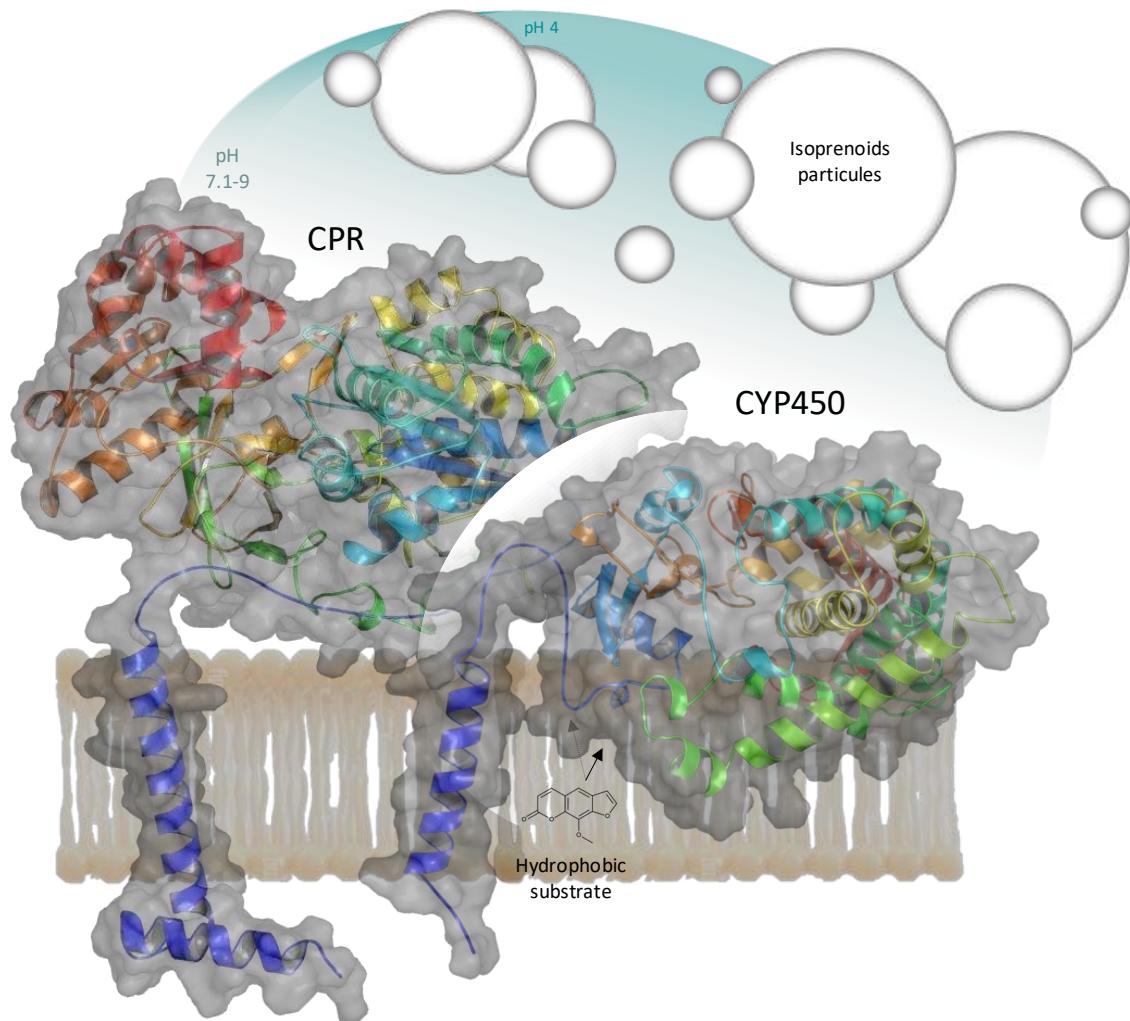
<sup>168</sup> Groupement prosthétique : Composant non protéique fixé de manière permanente à une enzyme, jouant un rôle dans la fonction enzymatique, comme l'hème chez les P450s.

<sup>169</sup> Poche catalytique : Site au sein de l'enzyme où se déroule la réaction catalytique, interagissant avec le substrat.

<sup>170</sup> Partenaire redox : Molécule impliquée dans des réactions d'oxydoréduction avec une enzyme, souvent un cofacteur.

nature de l'acquisition des électrons comme synthétisé par (Ortiz de Montellano, 2015).

Les P450s de la classe II concernent une catalyse spécifique aux eucaryotes.



**Figure 17: Représentation simplifiée de la structure tridimensionnelle d'un P450 (CYP71B131a) et de sa Cytochrome P450 Réductase (CPR) (Atr1) à la surface d'un réticulum endoplasmique.** La structure des deux protéines est réalisée via l'outil Pymol.

Si la superfamille des P450s est très diversifiée en termes de séquences protéiques, elle tend à conserver très fortement sa structure tertiaire<sup>171</sup> (Fujiyama et al., 2019; Lee et al.,

<sup>171</sup> Structure tertiaire : Arrangement tridimensionnel d'une chaîne polypeptidique, déterminant la fonction d'une protéine.

2008; Mestres, 2004; Niu *et al.*, 2020). Structuralement, le domaine globulaire<sup>172</sup> d'un P450 est constitué d'un ensemble de canaux hydrophobes<sup>173</sup> pour une longueur moyenne de 28Å dont, certains proches de la partie membranaire, permettent l'accueil de substrats<sup>174</sup> hémipolaires et le rejet des produits de catalyse (Navrátilová *et al.*, 2016; Pravda *et al.*, 2014; Urban *et al.*, 2018). Ces canaux sont continuellement en interactions avec les phospholipides et les stérols membranaires dont la composition détermine l'immersion du domaine globulaire dans la membrane, l'ouverture des canaux et plus largement l'efficacité de l'enzyme (Brignac-Huber *et al.*, 2016; Jeřábek *et al.*, 2016; K. Liu *et al.*, 2017). Plus en interne, les canaux s'ouvrent sur une poche catalytique constituée de 6 chaînes latérales. Cette région est également nommée Sites de Reconnaissance du Substrat (SRSs)<sup>175</sup> et est très évolutive chez les P450s (Gotoh, 1992).

Sur le plan de la catalyse, les P450s sont des enzymes peu spécifiques et généralement peu processives<sup>176</sup> mais capables dans certains cas d'une forte stéréo-<sup>177</sup> et régio-<sup>178</sup> sélectivité. Que ce soit au sein de familles très conservées ou très diversifiées, la structure des P450s peut leur permettre d'être multisubstrats et/ou multiproduits (Werck-Reichhart, 2023). Les P450s ont une durée de vie variable en fonction de comment ils sont étudiés. Par exemple le P450 humains CYP3A4 a une durée de demi-vie variable entre 26 et 140h en conditions *in vivo* et *in vitro* (J. Yang *et al.*, 2008). D'autre part, les P450s sont des enzymes dont la fonction peut être inhibée par de nombreux inhibiteurs naturels dans le cadre de régulations allostériques<sup>179</sup> (e.g. furocoumarines) (Denisov & Sligar, 2012; Francis & Delgoda, 2014; Gravot *et al.*, 2004; Halpert *et al.*, 1994; Wen *et al.*, 2006).

---

<sup>172</sup> Domaine globulaire : Partie d'une protéine enzymatique ayant une structure tridimensionnelle compacte et globulaire, souvent responsable de l'activité catalytique.

<sup>173</sup> Canaux hydrophobes : Voies au sein d'une enzyme qui permettent le passage de molécules hydrophobes, souvent vers le site actif.

<sup>174</sup> Substrat/Produit d'une enzyme : Le substrat est la molécule sur laquelle une enzyme agit, et le produit est la molécule résultante après la réaction.

<sup>175</sup> Sites de reconnaissance du substrat (SRSs) : Régions spécifiques d'une enzyme qui interagissent directement avec le substrat, permettant la spécificité de la réaction.

<sup>176</sup> Processivité enzymatique : Capacité d'une enzyme à catalyser de manière répétée plusieurs réactions sans se dissocier du substrat.

<sup>177</sup> Stéréosélectivité : Capacité d'une enzyme ou d'un processus chimique à produire préférentiellement un stéréoisomère plutôt qu'un autre composé.

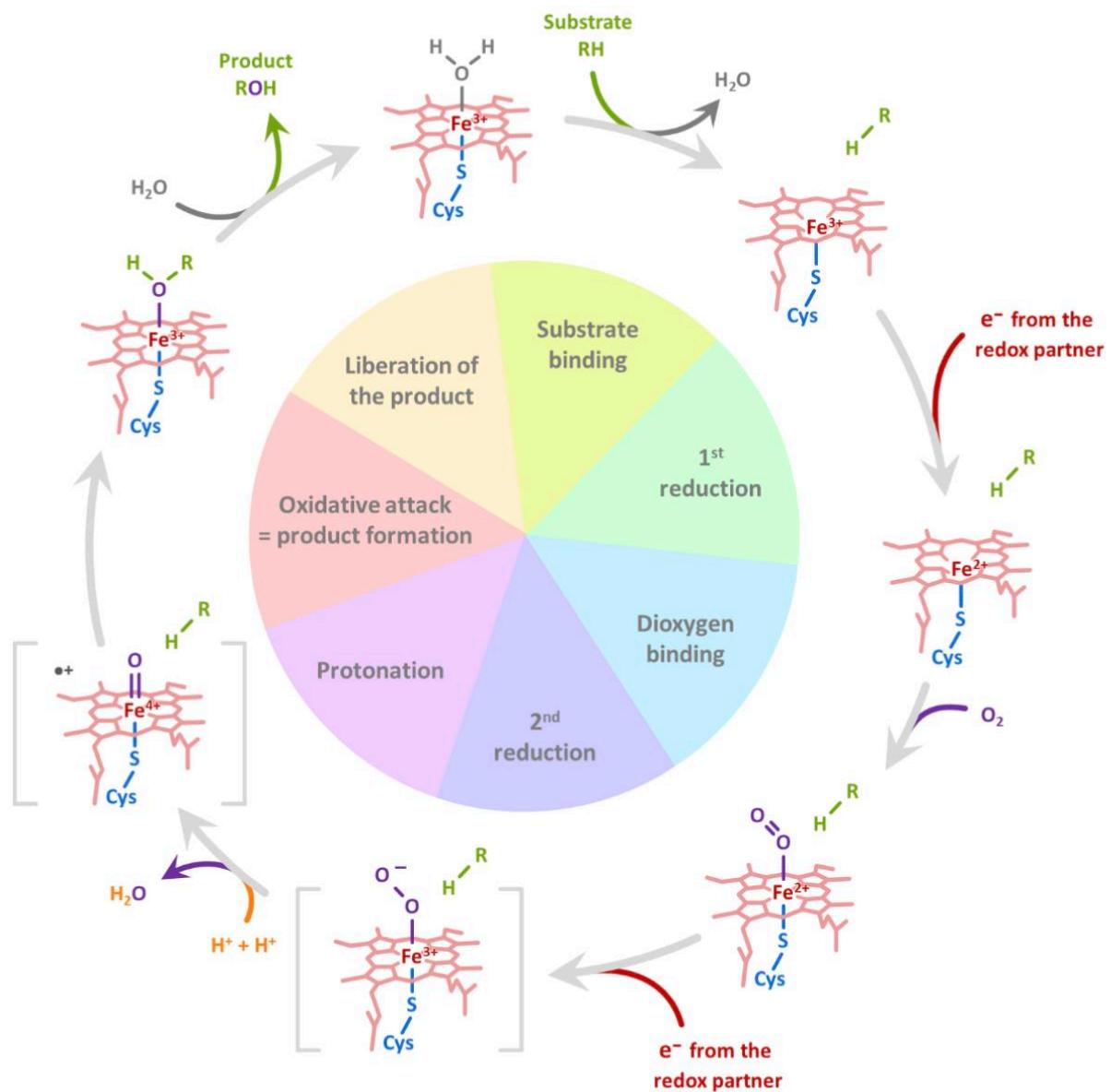
<sup>178</sup> Régiosélectivité : Préférence d'une enzyme ou d'un réactif chimique à interagir avec une région spécifique d'une molécule au cours d'une réaction.

<sup>179</sup> Allostéries : Modulation de l'activité d'une enzyme par la liaison d'une molécule (le composé allostérique) à un site autre que le site actif, modifiant sa conformation.

La description mécaniste du fonctionnement d'un P450s est plutôt bien comprise et notamment dans le cadre de sa fonction monooxygénase comme synthétisé par (Hrycay & Bandiera, 2015). Ainsi, lorsque le substrat se positionne sur le site actif<sup>180</sup>, une molécule d'eau stabilisatrice de l'hème est déplacée, ce qui a pour conséquence de modifier le potentiel redox du noyau ferreux sur l'hème. Par jeu de conséquences, une cascade de transfert de deux électrons et de deux protons permet alors la libération d'un radical réactif qui va s'intégrer au substrat et induire la libération d'un oxygène résiduel sous forme d'eau (**Figure 18**) (Fasan *et al.*, 2008; Ortiz De Montellano & De Voss, 2002). Le cycle est finement réglé et peu revenir à la phase initiale en cas de manque de partenaire redox ou d'un pH non adéquat en relarguant de l'H<sub>2</sub>O, de l'H<sub>2</sub>O<sub>2</sub> ou de l'O<sub>2</sub>.

---

<sup>180</sup> Site actif : Région d'une enzyme où le substrat se lie et où se déroule la réaction catalytique.



**Figure 18: Cycle catalytique d'une hydroxylation typique médiée par P450.** D'après (Mclean et al., 2011; Werck-Reichhart & Feyereisen, 2000; Zangar, 2004) et illustration par (Villard, 2020). La cystéine conservée est en bleu, son thiolate est désigné par "S", l'hème est en rose et le substrat en vert.



# Enjeux et objectifs du projet de thèse

Les plantes régulent leurs ressources pour s'adapter à leurs environnements en constante évolution. L'un des moyens de résilience des plantes consiste à produire des molécules actives (i.e. MS) en réponse à ces fluctuations (Weng *et al.*, 2021). Les plantes ont acquis ces molécules au cours de l'évolution, ce qui a eu pour conséquence de modifier leur physiologie et leurs interactions avec l'environnement (Durán-Medina *et al.*, 2021; Weng *et al.*, 2021).

Une partie des recherches initiées dans mon établissement d'accueil, le Laboratoire Agronomie et Environnement (LAE), s'intéresse à décrire comment certains de ces MS comme les furocoumarines pouvaient être produits chez les plantes. Ces travaux ont notamment permis de décrire la quasi-totalité des étapes de biosynthèse de la partie centrale de la voie des FCs linéaires. Les travaux de thèse de Cloé Villard (Villard, 2020) ont ainsi achevé la description de la voie par la découverte de la fonction marmésine synthase juste avant mon arrivée au LAE (Villard *et al.*, 2021). Contrairement à d'autres familles de MS, la compréhension du rôle des FCs sur la physiologie et les interactions avec l'environnement reste inexplorée. Dans le cadre de ma thèse, l'une de mes missions a consisté à développer un modèle végétal (la tomate) par bioingénierie métabolique pour tenter de mieux comprendre l'impact de la production de FCs linéaires sur la physiologie des plantes. Dans ce contexte, je me suis intéressé à répondre à la question : **Quels peuvent être les impacts d'une production constitutive de psoralène sur la physiologie et la croissance de la tomate ?**

La description de plusieurs activités enzymatiques chez différentes espèces de plantes a permis d'initier une meilleure compréhension de l'évolution de la voie des FCs linéaires (Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020). Ainsi, la capacité des plantes à produire des FCs a été acquise indépendamment au sein de différentes familles. La découverte de la fonction marmésine synthase chez le figuier a notamment permis de décrire comment cette fonction a récemment pu émerger au sein de la famille des Moracées (Villard *et al.*, 2021). Dans le cadre de ma thèse, l'un de mes objectifs a

consisté à rechercher de nouveaux gènes impliqués dans la voie des FCs chez le figuier et plus spécifiquement des P450s. Dans ce contexte, je me suis intéressé à répondre à la question : **Quels P450s peuvent être impliqués dans la voie des FCs chez le figuier et que nous apprennent-ils sur l'évolution de la voie des FCs chez les Moracées ?**

Pour répondre à ces questions, mon travail de recherche s'est appuyé sur des approches multidisciplinaires en travaillant sur deux modèles végétaux, la tomate (*Solanum lycopersicum*, Solanacées) et le figuier (*Ficus carica*, Moracées).



# Chapitre 2 : Comprendre l'impact de l'insertion de la voie de biosynthèse hétérologue « *p*-coumaroyl-CoA-psoralène » sur la physiologie de la tomate

## 1- Contexte scientifique et stratégie de recherche

Mon projet de recherche s'inscrit dans la continuité de travaux initiés au Laboratoire Agronomie et Environnement (LAE) et vise à mieux comprendre le métabolisme spécialisé (MS) des plantes. Cette dernière décennie, plusieurs projets ont permis au LAE de mieux comprendre comment la tomate gérait ses ressources dans le cadre d'un compromis croissance/défense pour faire face aux stress de l'environnement (Larbat *et al.*, 2014, 2016; Royer *et al.*, 2013). Dans ce même axe de recherches, plusieurs projets ont permis de décrire plus spécifiquement les voies de biosynthèse des coumarines, des furocoumarines (FCs) et des phénolamides (Karamat *et al.*, 2014; Krieger *et al.*, 2018; Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020; Roumani *et al.*, 2024; Villard *et al.*, 2021). Récemment, les travaux de thèse de Cloé Villard (2020) ont permis de décrire l'activité enzymatique de la marmésine synthase qui intervient au début de la voie des FCs linéaires (**Figure 13**) et qui complète les informations concernant la voie de biosynthèse menant au psoralène. L'ensemble des informations concernant la voie de biosynthèse des FCs a rendu accessible l'étude de leurs rôles physiologiques et notamment de ceux du psoralène.

L'étude du compromis croissance/défense se base sur la description de la gestion des ressources des plantes en réponse à des stress externes ciblés. Les approches mises en œuvre portent généralement la limite intrinsèque d'induire une vaste régulation de la physiologie du stress, ce qui complexifie fortement l'étude du compromis croissance/défense. Pour simplifier la perception des signaux importants dans la

régulation de ce compromis, nous avons choisi d'utiliser la voie de biosynthèse du psoralène comme outil. L'idée était de décrire les répercussions d'un stress endogène (e.g. biosynthèse d'un MS de défense comme le psoralène) sur la physiologie, la croissance et le développement d'une plante (e.g. la tomate) en limitant au maximum les contraintes environnementales (e.g. milieu en culture contrôlée et non limitante). Dans ce contexte et au travers de ce chapitre, je tente de répondre à la question : **Quels peuvent être les impacts d'une production constitutive de psoralène sur la physiologie et la croissance de la tomate ?**

Dans ce chapitre je décris d'une part (i) la création d'une construction génique permettant la biosynthèse de psoralène *in planta* à partir de *p*-coumaroyl CoA, (ii) la validation de la fonctionnalité de cette construction par expression hétérologue transitoire<sup>181</sup> dans le tabac, (iii) son intégration stable dans le génome de la tomate, (iv) et son impact sur la physiologie, la croissance et le développement de la tomate. Ces aspects sont décrits principalement sous la forme d'un article en cours de préparation. Des résultats expérimentaux complémentaires révélant d'autres aspects de l'impact de la voie du psoralène sur la physiologie de la tomate notamment en conditions de stress abiotiques sont présentés par la suite. L'ensemble de ces résultats est discuté en fin de chapitre.

## 2-Résultats

### 2-1- Création de la construction « *p*-coumaroyl CoA - psoralène » (historique)

La description de la majorité des étapes connues de la voie des FCs a été réalisée au LAE, ce qui m'a donné accès à l'ensemble des séquences déjà clonées et fonctionnellement testées. Les travaux visant à produire *in planta* et de façon hétérologue du psoralène ont été initiés dans le cadre de la thèse de Gianni Galati (Galati,

---

<sup>181</sup> Expression hétérologue transitoire : Production temporaire d'une protéine dans un organisme ou un système cellulaire différent de celui d'origine.

2019) au cours de laquelle il a tenté de produire par bioingénierie métabolique de la diméthylsubérosine, un précurseur de la voie des FCs linéaires chez le tabac (*Nicotiana benthamiana*, Solanacées). La construction génique a été réalisée en utilisant l'outil de clonage multigénique Goldenbraid 2.0 (GB). Cet outil permet l'assemblage séquentiel de gènes codant pour des enzymes de biosynthèse adjoints d'un promoteur et d'un terminator formant ainsi des unités de transcriptions (UTs). Grâce à cet outil, les UTs formées peuvent directement être intégrées dans le génome d'une plante par transfection médiée par *Agrobacterium tumefaciens*. Dans ce projet, il a été choisi d'utiliser le promoteur 35S pour permettre une expression constitutive<sup>182</sup> *in planta* et ainsi faciliter l'accumulation de FCs. L'outil GB à l'avantage d'être itératif ce qui permet de faciliter l'addition de nouvelles séquences codantes au fur et à mesure de leurs découvertes à des constructions déjà existantes et incluant par exemple les séquences codantes du début de la voie de biosynthèse des FCs linéaires. Un des objectifs derrière l'utilisation de cet outil est de permettre de mieux comprendre le rôle de différents intermédiaires de la voie des FCs linéaires *in planta*. Par la suite et dans le cadre de sa thèse, Cloé Villard, après avoir découvert une séquence codant pour une enzyme ayant la fonction marmesine synthase chez le figuier, a complété l'assemblage GB de la voie des FCs linéaires pour tenter de produire du psoralène dans la tomate (*Solanum lycopersicum*, Solanacées) (Villard, 2020). Ce travail n'avait pas pu aboutir à des lignées transformantes dans le temps imparti.

Dans le cadre de ma thèse, j'ai repris l'assemblage de cette construction génique pour produire un transgène pouvant être transféré dans le génome de la tomate. Pour la suite de ce document, ce transgène sera nommé PP (pour Psoralen Pathway). Le détail de la construction et de la vérification de l'assemblage de la construction PP est décrit dans la **section dédiée à l'article**.

---

<sup>182</sup> Expression constitutive : Expression d'un gène qui se produit de manière constante, indépendamment des conditions externes ou internes, assurant ainsi une production continue de la protéine.

## 2-2- Validation expérimentale et fonctionnelle de la construction PP chez le tabac

### Infiltration de plantules de *Nicotiana benthamiana* : approche classique

Avant de développer des lignées de tomates intégrant le transgène PP, nous avons souhaité vérifier sa fonctionnalité par expression hétérologue transitoire dans *Nicotiana benthamiana*. Afin d'avoir une expression importante dans les plantes, je suis passé d'une infiltration locale à la seringue à une infiltration globale et sous-vide des parties aériennes. Cette augmentation de surface infiltrée a été accompagnée par une augmentation de la quantité d'agrobactéries<sup>183</sup> utilisées permettant ainsi de limiter l'effet de dilution dans les tissus (Norkunas *et al.*, 2018). Enfin, l'infiltration a été réalisée sur de jeunes plants pour limiter les fuites de biosynthèse<sup>184</sup> en faveur de l'utilisation du *p*-coumaroyl-CoA vers les voies endogènes du tabac. En effet, au cours de son développement, le tabac oriente graduellement ces circuits de biosynthèse vers la production de plus en plus diversifiée de dérivés du *p*-coumaroyl-CoA (Drapal *et al.*, 2021). Ces adaptations nous semblaient nécessaires pour maximiser l'expression transitoire et simultanée des quatre enzymes hétérologues (Umbelliférone Synthase : UM, Diméthylsubérosine Synthase : DS, Marmesine Synthase : MS et Psoralène Synthase : PS) et pour faciliter la biosynthèse de psoralène dans un environnement cellulaire non natif.

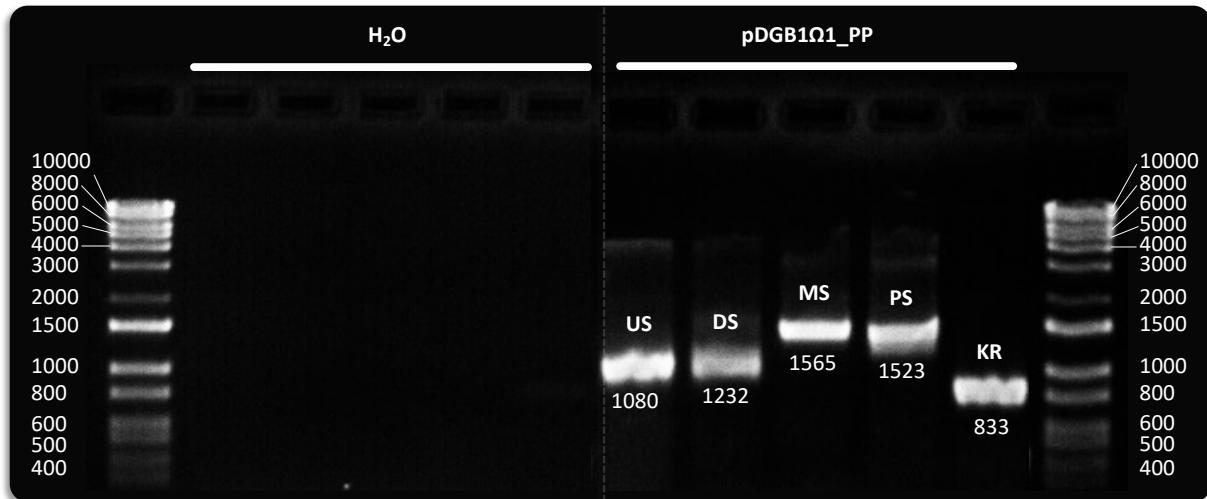
L'agroinfiltration a été réalisée avec un mélange de cultures d'agrobactéries transformées, soit par le plasmide pDGB1Ω1\_PP, soit par le plasmide pEAQ-HT\_GFP. Les feuilles de tabac ont été récoltées 4, 5 et 6 jours après l'infiltration. La production de la GFP mise en évidence par microscopie à épifluorescence nous a permis d'avoir un témoin indiquant que l'infiltration a été faite dans de bonnes conditions (**Figure 20**).

---

<sup>183</sup> Agrobactéries : Genre de bactéries, principalement *Agrobacterium tumefaciens*, capables de transférer une partie de leur ADN dans les cellules des plantes, provoquant des tumeurs ou « galles » et utilisées en biotechnologie pour introduire des gènes d'intérêt dans les plantes.

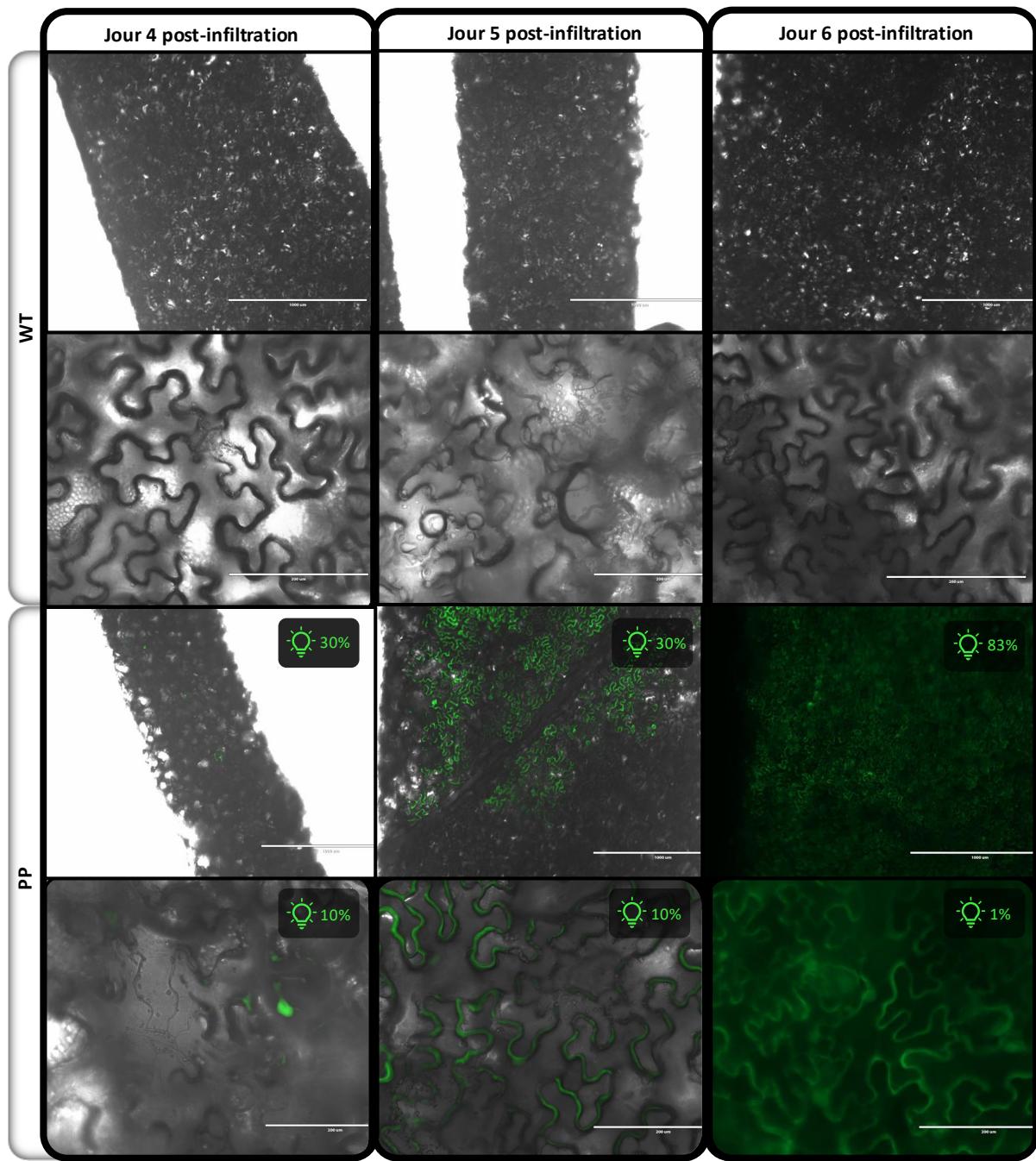
<sup>184</sup> Fuites de biosynthèse : Production non désirée de métabolites lors de la biosynthèse, souvent observée en expression hétérologue, où une régulation imparfaite ou une faible spécificité enzymatique dans l'organisme hôte conduit à des dérivés non ciblés.

L'extraction des ARNs à partir des plantes et une analyse réalisée par RT-PCR<sup>185</sup> nous a permis de montrer que les différents gènes étaient exprimés de manière efficace (**Figure 19**). Néanmoins, cette approche n'a pas permis de détecter de psoralène ni les intermédiaires de la voie par analyse métabolique ciblée des feuilles de tabac que ce soit après 4, 5 ou 6 jours d'expression transitoire.



**Figure 19: Vérification de l'expression hétérologue transitoire en tabac des gènes de la construction pDGB1Ω1\_PP après 5 jours d'infiltration.** Gel d'électrophorèse de produit PCR sur ADNc de 2 feuilles de tabac infiltrées avec de l'eau stérile (à gauche) ou agroinfiltrées avec la construction pDGB1Ω1\_PP (à droite). Pour l'agroinfection, la construction pDGB1Ω1\_PP est co-transformée avec des *Agrobacterium tumefaciens* (EHA105) contenant le plasmide d'aide à la réplication pSOUP. Cette première souche est co-infiltrée avec une seconde contenant le plasmide pEAQ-HT\_GFP. À noter que ce plasmide contient le gène de la séquence codant pour la protéine p19 qui a pour fonction de limiter la répression de l'expression de gènes hétérologues. Les deux cultures d'agrobactéries sont mélangées avant agroinfiltration en parallèle d'une infiltration à l'eau stérile comme témoin. Pour plus de détails, voir la section dédiée dans le matériel et méthodes. Abréviations : US : Umbelliférone Synthase ; DS : Diméthylsubérosine Synthase ; MS : Marmesine Synthase ; PS : Psoralène Synthase ; KR : Résistance à la Kanamycine.

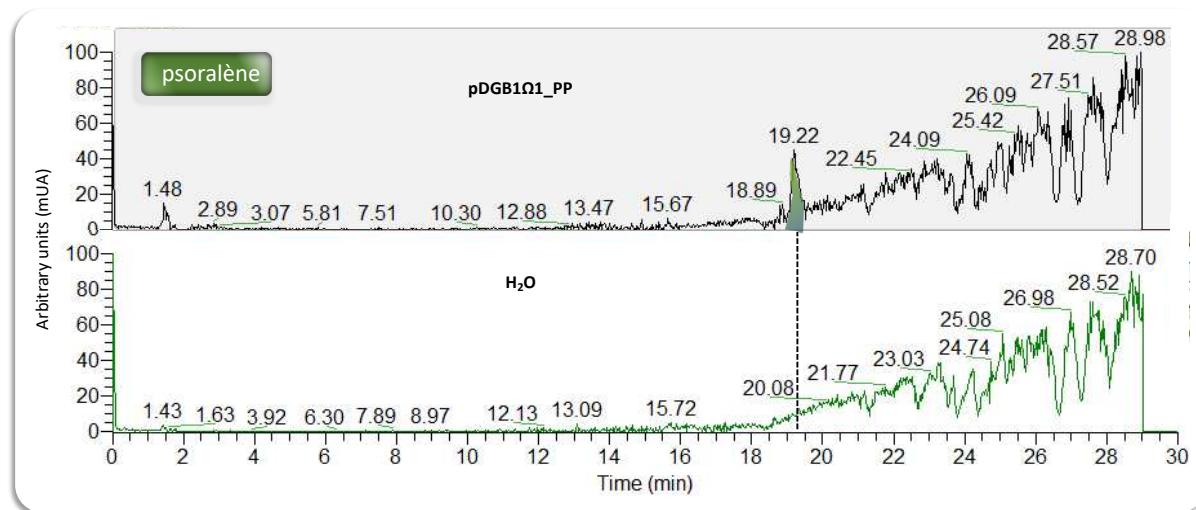
<sup>185</sup> RT-PCR (Reverse Transcription Polymerase Chain Reaction) : Technique permettant de convertir l'ARN en ADN complémentaire (ADNc) à l'aide d'une transcriptase inverse, puis d'amplifier cet ADN par PCR. Elle est utilisée pour étudier l'expression génique à partir de l'ARN messager.



**Figure 20: Suivi cinétique de l'expression hétérologue dans les feuilles de la GFP entre 4 et 6 jours post-infiltration avec de l'eau stérile ou des agrobactéries contenant les constructions pDGB1Ω1\_PP et pEAQ\_HT\_GFP.** Analyse par microscopie à épifluorescence de l'expression de la GFP dans l'épiderme de feuilles de tabac à un grossissement de 4x et 20x. Les luminosités blanche (72-100%) et UV (1-83%) sont adaptées pour permettre la meilleure visibilité possible de la GFP lorsqu'elle est détectée. Les images proviennent d'une superposition des spectres d'émission centrés sur la GFP (construction pDGB1Ω1\_PP) et de la lumière blanche pour visualiser les tissus ( $\text{H}_2\text{O}$  et construction pDGB1Ω1\_PP) via l'outil Image J. Aucune émission de la GFP n'a été détectée après infiltration d'eau.

### Infiltration de plantules de *Nicotiana benthamiana* : approche par supplémentation

L'absence de synthèse de psoralène peut être liée à un déficit de substrat dans les plantules de tabacs. Pour pallier à cette limite, nous avons imbibé pendant 24h les pétioles de feuilles 4 jours après infiltration dans une solution d'acide *p*-coumarique à 10mM (précurseur du *p*-coumaroyl-CoA). Les plantes sont ensuite broyées et les MS en sont extraits avant d'être analysés par CL-SM. Les résultats que nous avons obtenus démontrent que la supplémentation en *p*-coumarate favorise la production de psoralène dans les plantes agroinfectées (Figure 21). Ces résultats, obtenus en tripliquats, démontrent que la construction pDGB1Ω1\_PP permet bien *in planta* la production hétérologue des enzymes de la PP, ainsi que la production de psoralène. La même approche de supplémentation a été initiée avec de l'umbelliférone et a eu pour conséquence de rendre systématique la production de psoralène, ce qui a été légèrement moins le cas en présence de *p*-coumarate (données non présentées).



**Figure 21: Vérification de la production de psoralène après 1 jour de supplémentation en *p*-coumarate dans des feuilles de tabac, à 5 jours post-infiltration avec de l'eau stérile ou des agrobactéries contenant la construction pDGB1Ω1\_PP.** Spectrogramme de masse en ionisation ESI positive ciblée sur la masse du psoralène [M+H]<sup>+</sup> (186,50-187,50). Le *m/z* attendu du psoralène en ionisation positive est de 187,03. Le temps de rétention du psoralène attendu sur ce gradient UHPLC-MS de 30 minutes est de 19,21 minutes, par comparaison avec un standard pur.

Ainsi, par une approche d'expression transitoire en tabac supplémentée en *p*-coumarate ou en umbelliférone, j'ai démontré que (i) la biosynthèse hétérologue de psoralène était possible *in planta* grâce à la construction pDGB1Ω1\_PP, et (ii) que cette biosynthèse était dépendante de la disponibilité des substrats et notamment de l'umbelliférone dans le tabac.

### **2-3- L'intégration de la voie du psoralène chez la tomate conduit à une accumulation de coumarines et limite la croissance (Article 1)**

La fonctionnalité du transgène étant établie, nous l'avons intégré dans le génome de la tomate. Ce travail a permis de régénérer une lignée transgénique pour laquelle nous avons analysé finement la croissance, le développement et la physiologie en utilisant des approches croisées de mesures de traits morphométriques, de biochimie analytique, de métabolomique et de transcriptomique.

# **Bioengineering the furanocoumarin pathway in tomato led to coumarins accumulation and impaired growth**

Alexandre Bouillé<sup>1</sup>, Marwa Roumani<sup>1</sup>, Cloé Villard<sup>2</sup>, Aude Fauvet<sup>1</sup>, Jérémy Grosjean<sup>1</sup>, Lennart Hoengenaert<sup>3</sup>, Wout Boerjan<sup>3</sup>, Frédérique Hilliou<sup>4</sup>, Gianni Galati<sup>1</sup>, Christophe Robin<sup>1</sup>, Alain Hehn<sup>1</sup>, Romain Larbat<sup>5</sup>

1- Université de Lorraine, INRAE, LAE, 54000 Nancy, FR

2- Biosystematics group, Wageningen University and Research Center, Wageningen 6708 PB, NL

3- Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology, Ghent University, K.L. Ledeganckstraat 35, B-9000 Gent

4- INRAe, Institut Sophia Agrobiotech, Université Côte D'Azur, CNRS, Sophia Antipolis, France

5- Institut Agro, University of Angers, INRAE, IRHS, SFR QUASAV, F-49000 Angers, FR

#scopoletin #coumarins #furanocoumarins #tomato #growth #xenobiotic

## **Introduction**

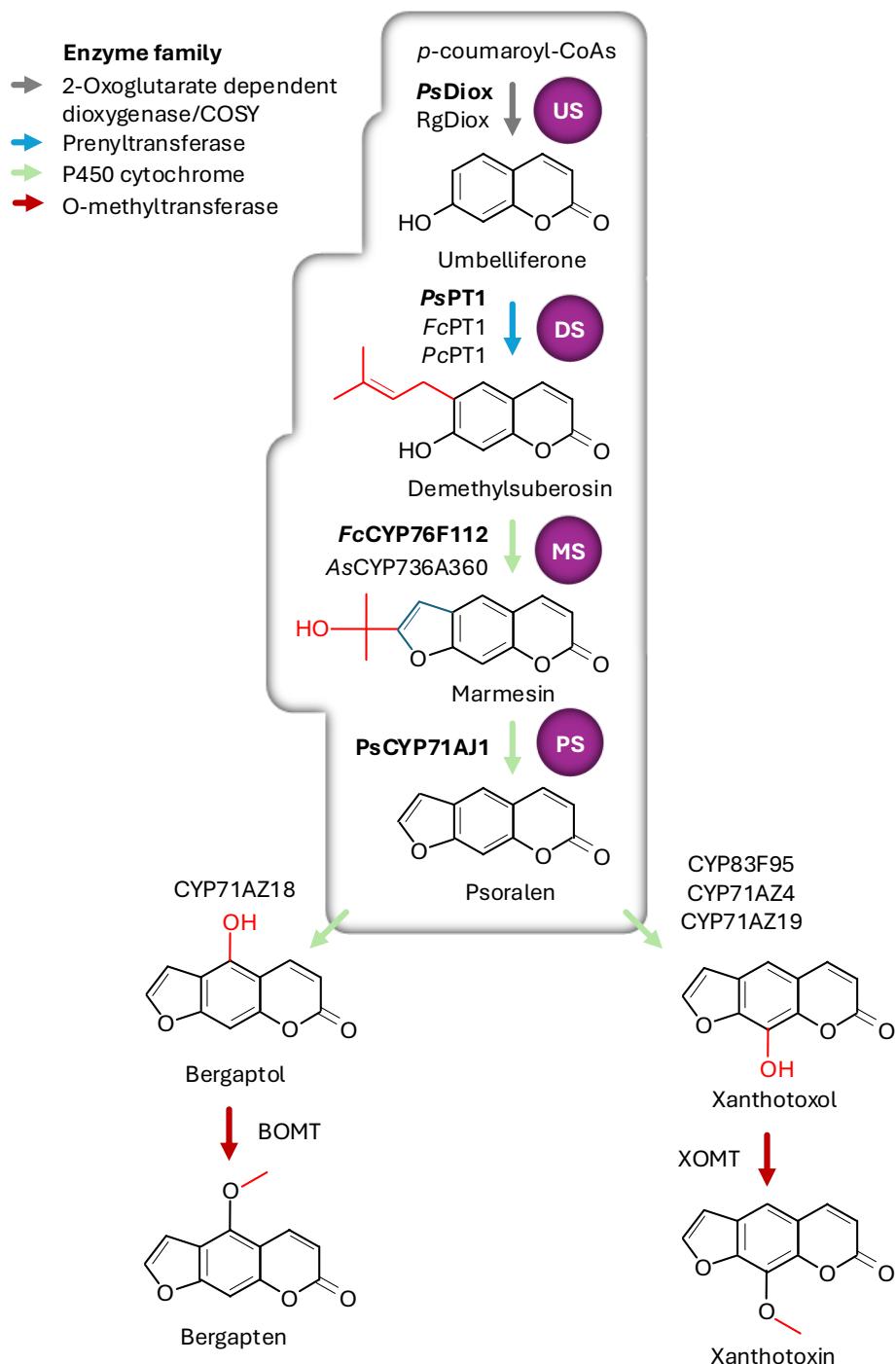
The sessile nature of plants necessitates adaptations to a constantly changing environment. The production of specialized metabolites (SM) exemplifies such an adaptation (Maeda & Fernie, 2021). Indeed, certain SMs play vital roles in enabling plants to cope with biotic and abiotic stresses (Weng *et al.*, 2021). Investment of resources and energy in processes favoring plant defense (and among them the biosynthesis of SM) may be at the expense of processes favoring plant growth. Then, plants must navigate a trade-off between the costly biosynthesis of SMs and the proper allocation of resources for growth (Neilson *et al.*, 2013). The management of these resources has been explored through various theories (D. Li & Gaquerel, 2021). Among them, the Growth-Differentiation Balance Hypothesis (GDBH) is one of the most widely accepted. It proposes that plants coordinate their resources in a trade-off that balances growth, maintenance, storage, reproduction, and defense, the trade-off being mainly conditioned by the resources availability (Herms & Mattson, 1992; Stamp, 2003, 2004). More recently, studies demonstrated that plant balance between growth and defenses is actively

regulated and orchestrated through a complex cross-talk involving phytohormone pathways together with the interaction between the two kinases SnRK1 (Sucrose non-fermenting-1-related protein kinase) and TOR (Target of Rapamycin) (Guo *et al.*, 2018; Margalha *et al.*, 2019; Monson *et al.*, 2022).

Terrestrial plants produce a wide range of SMs (Chassagne *et al.*, 2019). Among them, furanocoumarins are taxa restricted SMs (Sarker & Nahar, 2017). The biosynthesis of linear furanocoumarins has been well documented over the past two decades and notably with the recent discovery of the marmesin synthase function, that ends the characterization of the psoralen pathway (Villard *et al.*, 2021) (**Fig. 1**). They derive from ubiquitous coumarins, particularly umbelliferone that is produced from *p*-coumaroyl-CoA (Roselli *et al.*, 2017; Vialart *et al.*, 2012). For the biosynthesis of linear furanocoumarins, umbelliferone undergoes a C6-prenylation leading to the synthesis of demethylsuberosin by a prenyltransferase (Munakata *et al.*, 2016). This product is then cyclized into marmesin by a first cytochrome P450 (P450) (Villard *et al.*, 2021), followed by a decarboxylation *via* *syn*-elimination of acetone to form psoralen by a second P450 (Larbat *et al.*, 2007, 2009). The psoralen backbone can then be hydroxylated at C5 or C8 by additional P450s to yield bergaptol or xanthotoxol respectively (Ji *et al.*, 2024; Krieger *et al.*, 2018). These structures are typically processed by O-methyltransferases to form bergapten or xanthotoxin (Hehmann, Lukačin, *et al.*, 2004; Y. Zhao *et al.*, 2016). They can also undergo further modification by other enzymes, such as O-prenyltransferases or other P450s, to generate more complex alkylated structures (Bouillé *et al.*, 2024; Han *et al.*, 2023).

In this study, we aim to elucidate the potential costs associated with a newly SM production on the growth and physiology of a plant. To investigate this, we constitutively introduced 4 steps of the furanocoumarin biosynthesis pathway that converts *p*-coumaroyl-CoA into psoralen (referred to as the "Psoralen Pathway," or PP), within the tomato genome, a plant that do not naturally produce them (**Fig. 1**). Our findings indicate that tomato exhibits limited capacity for psoralen biosynthesis, favoring instead the accumulation of coumarins, particularly scopoletin. We present a detailed analysis of

the PP line's impact on growth and physiology, supported by morphophysiological, transcriptomic, and metabolomic data.



**Figure 1: Simplified diagram of the central part of the linear furanocoumarin biosynthetic pathway in angiosperms.** The pathway from *p*-coumaroyl-CoA to psoralen is framed, and the enzymes inserted into the tomato genome are in bold. US (Umbelliferone synthase), DS (Demethylsuberosin synthase), MS (Marmesin synthase), PS (Psoralen synthase). From

Hehmann *et al.*, 2007, Larbat *et al.*, 2007, 2009 ; Vialart *et al.*, 2012, Karamat *et al.*, 2014 ; Munakata *et al.*, 2016 ; Zhao *et al.*, 2016, Roselli *et al.*, 2017, Villard *et al.*, 2021, Wang *et al.*, 2024.

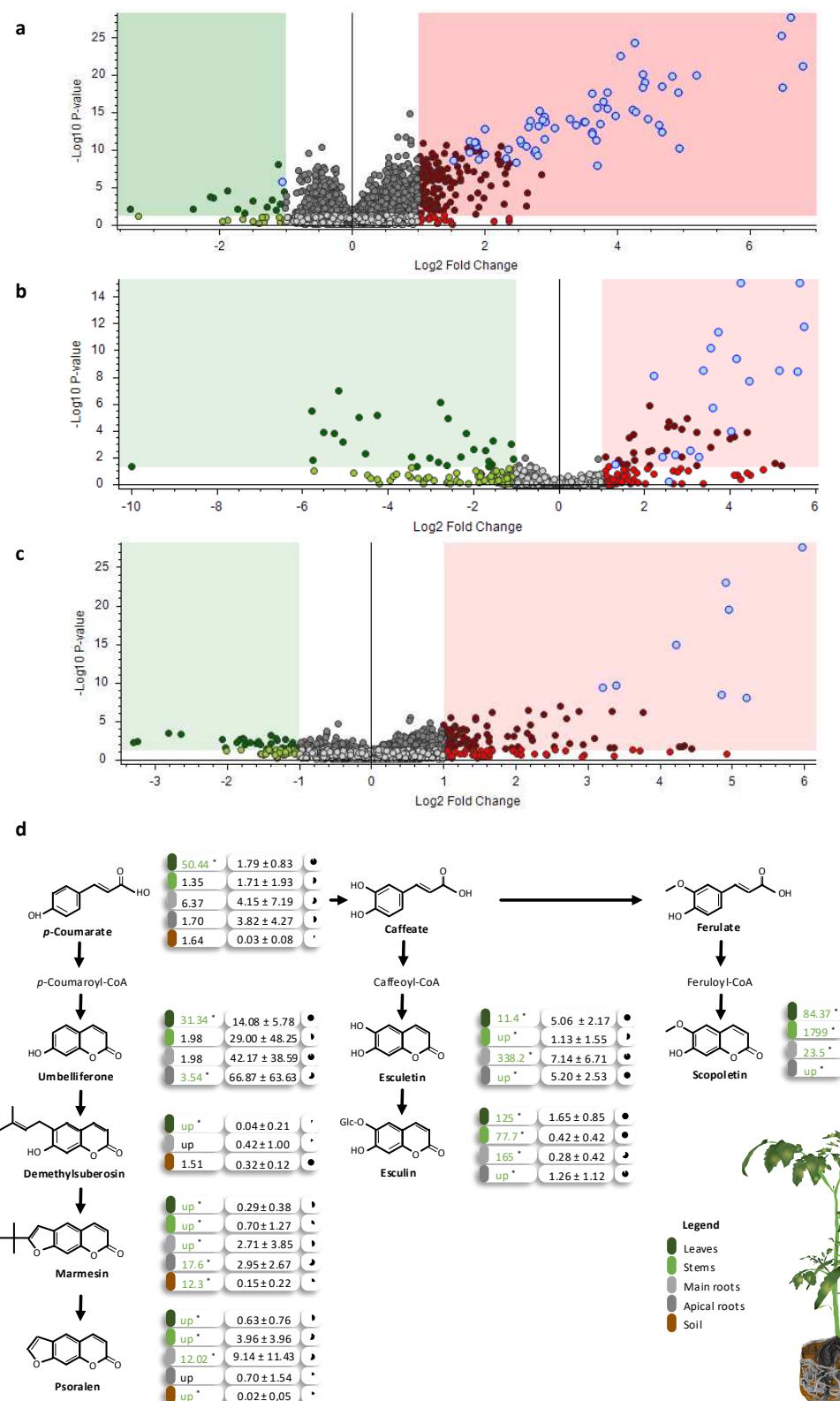
## Results

The tomato line designated "Psoralen Pathway" (PP) was constructed by integrating four genes encoding the enzymes required for the four-step conversion of *p*-coumaroyl-CoA into psoralen. The transgene was designed to make in one block of 4 transcription units (TUs) with the genes under the control of the constitutive CaMV35S promoter (**Fig. 1**), namely Umbelliferone synthase (*PsDiox*), Demethylsuberosin synthase (*PsPT1*), Marmesin synthase (*FcCYP76F112*), Psoralen synthase (*PsCYP71AJ3*) (Larbat *et al.*, 2009; Munakata *et al.*, 2016; Roselli *et al.*, 2017; Villard *et al.*, 2021). Plant selection was allowed by a fifth TU including the neomycin phosphotransferase (*EcNPTII*) also under the CaMV35S promoter control. Out of approximately 600 young tomato leaf fragments, we successfully regenerated two transformant lines. However, only one of these lines is effectively able to express the transgenes (**Fig. S1**). Our study is focused on this particular line.

### Untargeted and targeted metabolomics highlighted coumarin overaccumulation in the PP line

The consequences of the integration of the psoralen pathway on the tomato metabolic composition were investigated by comparing PP and WT metabolomes in leaves, stems and roots through untargeted LC-MS analysis of hydro-methanolic extracts. In order to strengthen the analysis and considering that only one PP line was available, three series of comparisons (including 6-8 plants per genotype in each comparison) were realized. Based on the mean of the three series, the PP line was shown to differentially accumulate 209, 74 and 128 metabolic features in comparison to WT in leaves, stems and roots respectively, which represent between 3.5 and 5% of the total metabolic features detected (**Fig. 2a-c**). The differentially accumulated features were mainly in favor of an overaccumulation in the PP line. The list of differentially accumulated metabolomic feature was shortened by considering only those being significant in each of the three series (Benjamini-Hochberg  $p<0.05$ ), which led to 80 metabolic features across leaves, stems, and roots (**Table I**). Putative annotation of these features allowed the

categorization of 36 of them as lipids, 8 as coumarins and 5 as phenolamides. In addition, one of them, present as trace amount in the PP line only, was annotated as psoralen, the final product of the metabolic pathway introduced in the PP line (**Table I**). Interestingly, only a few of the metabolic features were significantly affected in at least two organs. Among them, only coumarins were identified and quantified (**Fig. 2d** ; **Fig. S4**).



**Figure 2: Impact of the psoralen pathway (PP) line on the tomato metabolome across different compartments compared to the wild-type (WT) line. Volcano plot of differentially**

accumulated metabolites for the PP line compared to the WT line within (a) leaves, (b) stems and (c) roots. (d) Targeted analysis of the heterologous PP and the coumarin pathway. From left to right in boxes: X-fold difference (black) and significantly higher (green) or lower (orange) amounts of coumarins and furanocoumarins in the PP line compared to the WT line; relative quantification normalized by an internal standard per mg of DW; percentage of replicates containing the metabolite in the PP line, highlighted by pie charts. Statistical analysis was performed using the Kruskal-Wallis test ( $p<0.05^*$ ). “up” mean newly detected metabolite in the PP line. For the Volcano plots, green framed dots correspond to more WT specific and abundant metabolites. Oppositely, red framed dots correspond to more PP specific and abundant metabolites. The blue dots represent the differentially accumulated (adjusted  $p$ -value  $< 0.05$ ). Number of replicates: leaves, stems, main roots, and main roots soil ( $n=19-23$ ); apical roots and apical roots soil ( $n=14-16$ ). Based on 3 independent cultures of tomato growth in same controlled conditions.

**Table I: Subset of metabolic features differentially accumulated in the PP line in comparison to the WT.** Putative annotation and confidence followed the MSI identification notation (*Fiehn et al.*, 2007). The organ where the feature is differentially accumulated is indicated (Benjamini-Hochberg  $p<0.05$ ). Metabolic features differentially accumulated in two and three organs are in italics and bold respectively.

Name	Family	Confidence	Formula	Calc. MW	m/z	RT [min]	Leaf	Stem	Root
Esculin	Coumarin	L1	C15 H16 O9	340,07875	339,07148	10,213	PP>WT		
	Coumarin	L3	C18 H17 N2 P S	324,08463	325,09191	10,449	PP>WT	PP>WT	
CNP0278899	Coumarin	L3	C16 H18 O10	370,08959	369,08232	10,449	PP>WT	PP>WT	
CNP0306732/CNP0383992/CNP0306859	Coumarin	L3	C20 H24 O13	472,12209	471,11481	11,069	PP>WT		PP>WT
CNP0272825	Coumarin	L3	C21 H26 O14	502,13278	501,1255	11,418	PP>WT		
<b>CNP0271898</b>	<b>Coumarin</b>	<b>L3</b>	<b>C22 H28 O15</b>	<b>532,14387</b>	<b>531,1366</b>	<b>12,598</b>	<b>PP&gt;WT</b>	<b>PP&gt;WT</b>	<b>PP&gt;WT</b>
Scopoletin	Coumarin	L1	C10 H8 O4	192,04226	193,04954	16,314	PP>WT	PP>WT	
	Coumarin	L3	C17 H20 O11	400,10043	399,09315	12,192		PP>WT	
Psoralen	Furanocoumarin	L1	C11 H6 O3	186,03171	187,03899	19,332		PP>WT	
	Lipid	L3	C18 H43 N10 O5 P	510,3169	511,32418	23,836	PP<WT		
	Lipid	L3	C20 H44 N2 O4 P2 S	470,25135	469,24408	20,414	PP>WT		
	Lipid	L3	C21 H36 O7	400,24603	399,23875	20,448	PP>WT		
11 $\beta$ -Prostaglandin F2 $\alpha$	Lipid	L3	C22 H32 O5	376,22277	377,23004	20,712	PP>WT		
Prostaglandin F1 $\beta$	Lipid	L3	C20 H36 O5	402,26177	401,25449	20,858	PP>WT		
	Lipid	L3	C26 H40 N4 O9 S	584,25099	585,25827	21,228	PP>WT		
	Lipid	L3	C28 H51 O9 P3	624,27597	625,28324	21,286	PP>WT		
8-iso Prostaglandin F1	Lipid	L3	C20 H36 O5	378,23837	379,24565	21,602	PP>WT		
	Lipid	L3	C18 H37 N4 O7 P	452,23916	451,23189	21,777	PP>WT		
	Lipid	L3	C21 H36 O6	384,25098	383,24371	21,779	PP>WT		
Retinal	Lipid	L3	C20 H28 O	284,21412	285,22139	21,792	PP>WT		
	Lipid	L3	C16 H24 O	232,18276	233,19004	21,806	PP>WT		
	Lipid	L3	C18 H28 N6 O3	376,22264	377,22992	21,824	PP>WT		
Prostaglandin F2 $\alpha$	Lipid	L3	C20 H34 O5	400,24605	399,23878	21,824	PP>WT		
	Lipid	L3	C20 H42 N2 O4 P2 S	468,2341	467,22682	21,832	PP>WT		
	Lipid	L3	C16 H33 N8 O3 P	416,24093	415,23365	21,874	PP>WT		
	Lipid	L3	C18 H28 N6 O4	392,21765	393,22493	21,896	PP>WT		
(+/-)11(12)-DiHET	Lipid	L3	C20 H34 O4	360,22771	361,23498	21,979	PP>WT		

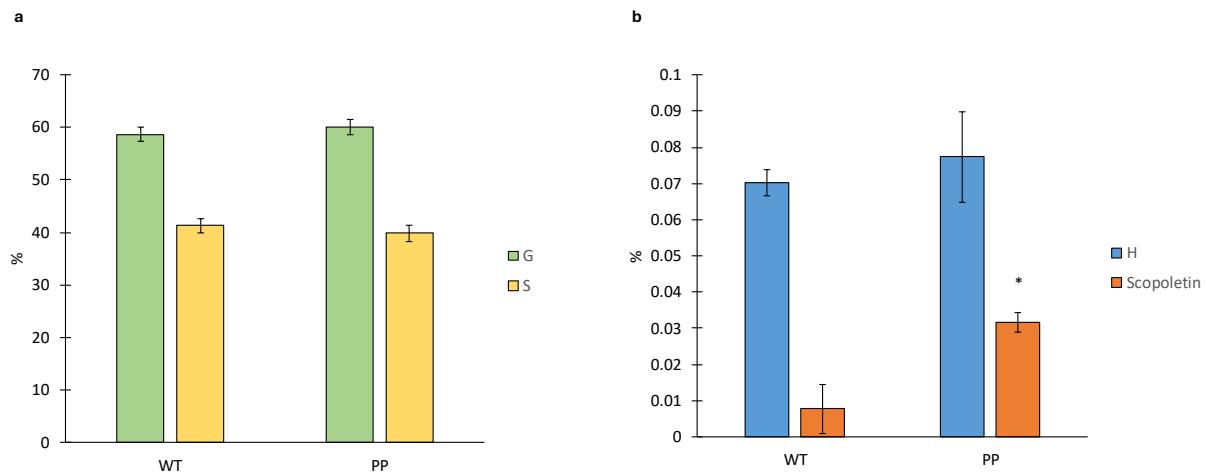
	Lipid	L3	C27 H46 O11 P2	608,25116	609,25844	22,454	PP>WT		
	Lipid	L3	C25 H33 N5 O3	451,25716	450,24988	22,576	PP>WT		
(5Z,9E,14Z)-(8xi,11xi,12S)-8,11,12,14-hydroxyicosanoic acid	Lipid	L3	C20 H34 O5	354,24014	353,23286	22,619	PP>WT		
11-Deoxy prostaglandin F1 $\alpha$	Lipid	L2	C20 H36 O4	362,24336	363,25063	22,752	PP>WT		
	Lipid	L3	C26 H45 Cl O8	520,28224	519,27496	22,963	PP>WT		
11-deoxy-PGE1	Lipid	L3	C20 H34 O4	338,24575	361,23497	23,676	PP>WT		
(9cis)-Retinal	Lipid	L2	C20 H28 O	284,2142	285,22147	23,697	PP>WT		
	Lipid	L3	C18 H28 N6 O2	360,22772	361,23499	23,698	PP>WT		
7Z,10Z,13Z,16Z,19Z-eicosapentaenoic acid	Lipid	L2	C20 H30 O2	302,22468	303,23196	23,71	PP>WT		
	Lipid	L3	C22 H25 N6 P	404,18776	405,19504	23,717	PP>WT		
5,12-Dihydroxy-6,8,10,14-icosatetraenoic acid	Lipid	L3	C20 H32 O4	336,22967	335,22239	23,719	PP>WT		
	Lipid	L3	C22 H27 N6 O P	422,19832	423,20559	23,723	PP>WT		
16,16-Dimethyl prostaglandin A2	Lipid	L3	C20 H34 O3	344,2328	345,24007	23,728	PP>WT		
	Lipid	L3	C16 H24	216,18787	217,19515	23,729	PP>WT		
	Lipid	L3	C22 H27 N6 P	406,20338	407,21065	23,738	PP>WT		
	Lipid	L3	C21 H36 O6	384,25098	383,24371	23,817	PP>WT		
	Lipid	L3	C20 H28	268,21911	269,22638	23,836	PP>WT		
Arachidonic acid	Lipid	L3	C20 H32 O2	286,22975	287,23702	23,838	PP>WT		
	Phenolamide	L3	C16 H24 N4 O4	336,17984	337,1871	13,525	PP>WT		
Feruloylagmatine	Phenolamide	L2	C15 H22 N4 O3	306,16928	307,17656	13,771	PP>WT		
	Phenolamide	L3	C14 H22 N2 O4	282,15802	283,1653	7,581			PP>WT
N-caffeoylelysine	Phenolamide	L2	C15 H20 N2 O5	308,13717	309,14456	10,139			PP>WT
p-coumaroylagmatine	Phenolamide	L3	C14 H20 N4 O2	276,15871	277,16598	13,055			PP>WT
	Unknown	L4	C14 H22 N2 O4	282,15813	283,16541	6,35	PP>WT		
	Unknown	L4	C17 H21 N4 O5 P3	454,07248	453,06521	10,202	PP>WT		
	Unknown	L4	C15 H17 Cl O8	360,06061	359,05334	10,441	PP>WT		
	Unknown	L4	C15 H14 N6 O10	438,07751	437,07023	10,442	PP>WT	PP>WT	
	Unknown	L4	C14 H18 O10 P2	408,03702	409,04429	10,451	PP>WT		
	Unknown	L4	C17 H18 O P2	300,08413	299,07685	10,953	PP>WT		
	Unknown	L4	C17 H31 O20 P	586,11599	585,10872	11,077	PP>WT		
	Unknown	L4	C19 H26 O14 P2	540,07929	541,08657	11,408	PP>WT		
	Unknown	L4	C24 H26 O16	570,12116	569,11388	11,415	PP>WT		
	Unknown	L4	C25 H24 O13	532,12278	531,11551	12,383	PP>WT		
	Unknown	L4	C11 H11 N4 P3	292,01927	291,012	12,514	PP>WT		
	Unknown	L4	C9 H21 N4 O9 P	360,10512	359,09784	13,546	PP>WT		
	Unknown	L4	C15 H26 N6 O11 S	498,13813	497,13085	14,817	PP>WT		PP>WT
	Unknown	L4	C11 H11 N O6	253,05829	252,05102	15,119	PP>WT		
	Unknown	L4	C11 H9 N2 P S2	263,99468	265,00196	15,828	PP>WT		
	Unknown	L4	C9 H12 O7 P2	294,00531	295,01259	16,287	PP>WT		
	Unknown	L4	C14 H20 N4 O	260,16374	261,17101	16,309	PP>WT		
	Unknown	L4	C19 H30 N6 O10 P2	564,14973	563,14245	17,685	PP>WT	PP>WT	PP>WT
	Unknown	L4	C10 H30 N8 O2 S	326,22062	327,2279	2,134			PP>WT
	Unknown	L4	C15 H20 O9	344,10997	343,10269	8,991			PP>WT
	Unknown	L4	C14 H20 N2 O4	280,14241	281,14969	10,361			PP>WT
	Unknown	L4	C18 H28 N2 O5	352,1999	353,20718	10,719			PP>WT
	Unknown	L4	C16 H22 O10	374,12083	373,11355	11,313			PP>WT
	Unknown	L4	C22 H35 O14 P	554,17476	555,18204	12,531			PP>WT

	Unknown	L4	C14 H27 O16 P	482,10397	481,09669	16,981		PP>WT
	Unknown	L4	C19 H16 O7	356,08966	357,09694	17,692		PP>WT
	Unknown	L4	C39 H59 N O3 P2	651,3983	652,40557	18,946		PP>WT
	Unknown	L4	C14 H14 O3	230,09406	229,08678	19,18		PP>WT
	Unknown	L4	C14 H24 O4	256,16723	255,15995	22,089		PP>WT
	Unknown	L4	C30 H48 S2	472,31881	473,32608	22,116		PP>WT

Psoralen and intermediates of the PP (umbelliferone, demethylsuberosin, marmesin) were detected only in trace amounts across PP line organs but were absent in the WT line (**Fig. 2b**). The PP line exhibited substantial increases in *p*-coumarate and umbelliferone levels (50- and 31-fold, respectively) in the leaves, with no significant changes in other plant parts. The root cortex is generally a hotspot for coumarin biosynthesis (Robe, Conejero, *et al.*, 2021). In this regard, we distinguished between apical roots, which contain cortical cells, and primary roots. Scopoletin concentrations were highly elevated in the PP line, especially in stems (1799-fold, up to 5 µg/g DW), leaves, and to a lesser extent, roots. Esculetin also showed higher accumulation, particularly in primary roots (338-fold) and leaves, whereas only traces were detected in stems and apical roots. Esculin levels increased in primary roots (165-fold), leaves, and stems. Interestingly, scopolin was not detected.

### Scopoletin is moderately accumulated into the PP line lignin

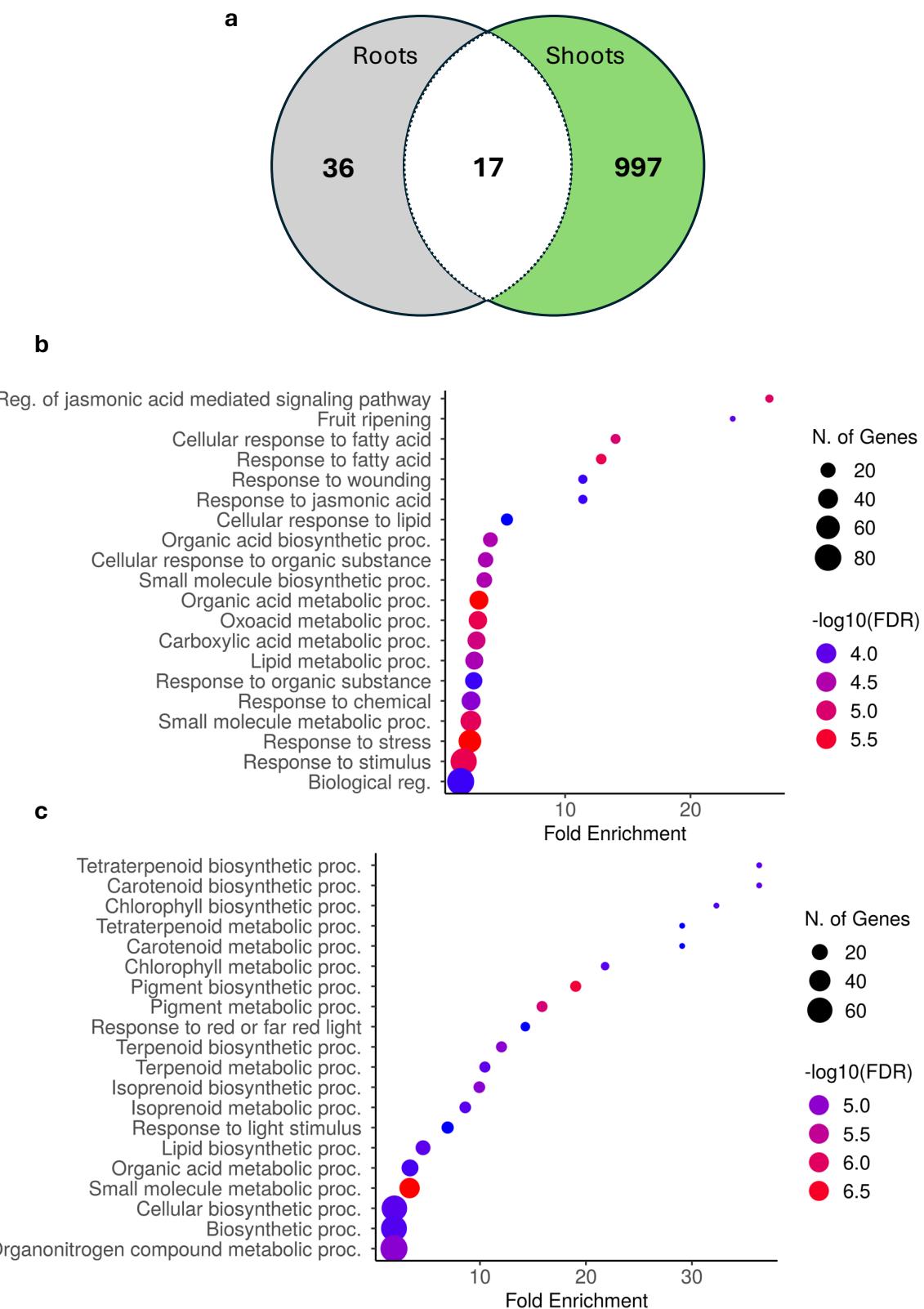
Scopoletin has recently been described to be able to incorporate lignin polymer in *Arabidopsis thaliana* transgenic lines overaccumulating this metabolite (Hoengenaert *et al.*, 2022). In accordance, the lignin composition was prospected in the PP line in comparison to WT. Lignin content and monomer composition (H, G, S) in the stems of the PP line were not significantly affected in the PP line (**Fig. 3a-b**). However, a moderate but significant accumulation of scopoletin was measured in the PP line (0.17±0.09% of cell wall residues) compared to the WT line (**Fig. 3b**).



**Figure 3: Impact of the Psoralen Pathway (PP) line on lignin composition for G, S (a), and H monomers, as well as scopoletin (b), compared to the Wild-Type (WT) line.** The percentage of monomers is calculated based on the content of HGS units and scopoletin in the lignin. Data were compared using the Kruskal-Wallis's test (p-value<0.05).

#### Specific gene expression in the PP line

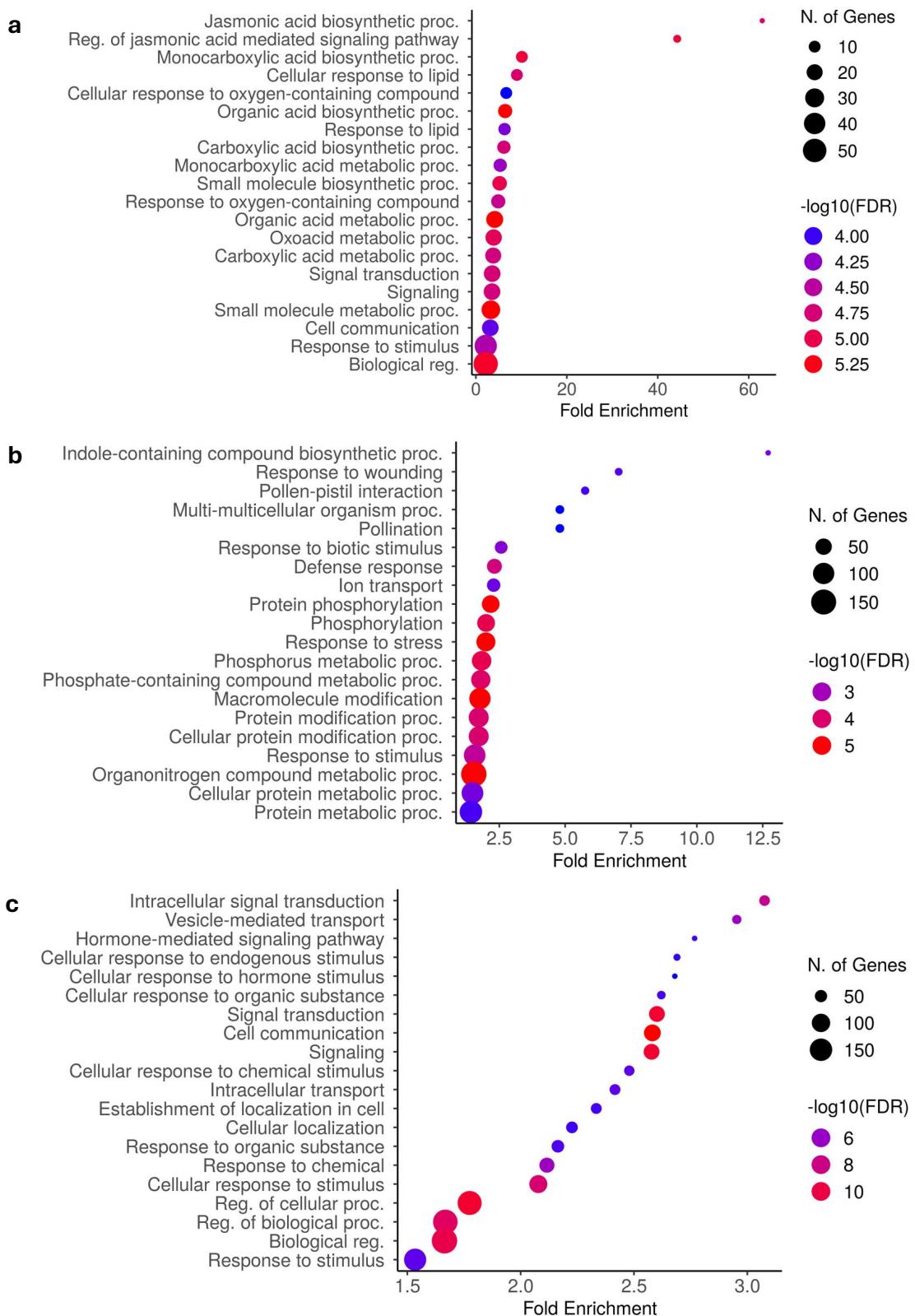
Transcriptomic analysis of the PP line and the WT highlighted 1014 and 53 differentially expressed genes (DEGs) in leaves and roots respectively (Fig. 4a). In leaves, the 1014 DEGs were shared in 659 up-regulated and 355 down-regulated genes in the PP line compared to WT. In roots, 41 were up-regulated and 12 down-regulated. Seventeen DEGs were common to leaves and roots. Gene Ontology enrichment analysis (GOEA) was conducted on the up- and down-DEGs identified in each organ. In leaves, this analysis highlighted an overall up-regulation of genes involved in processes related to response to stress and wounding (Fig. 4b) whereas the main trend was a down-regulation of genes involved associated to chlorophyll biosynthesis (Fig. 4c). In roots, the relative short list of DEGs did not lead to any significant information using GOEA.

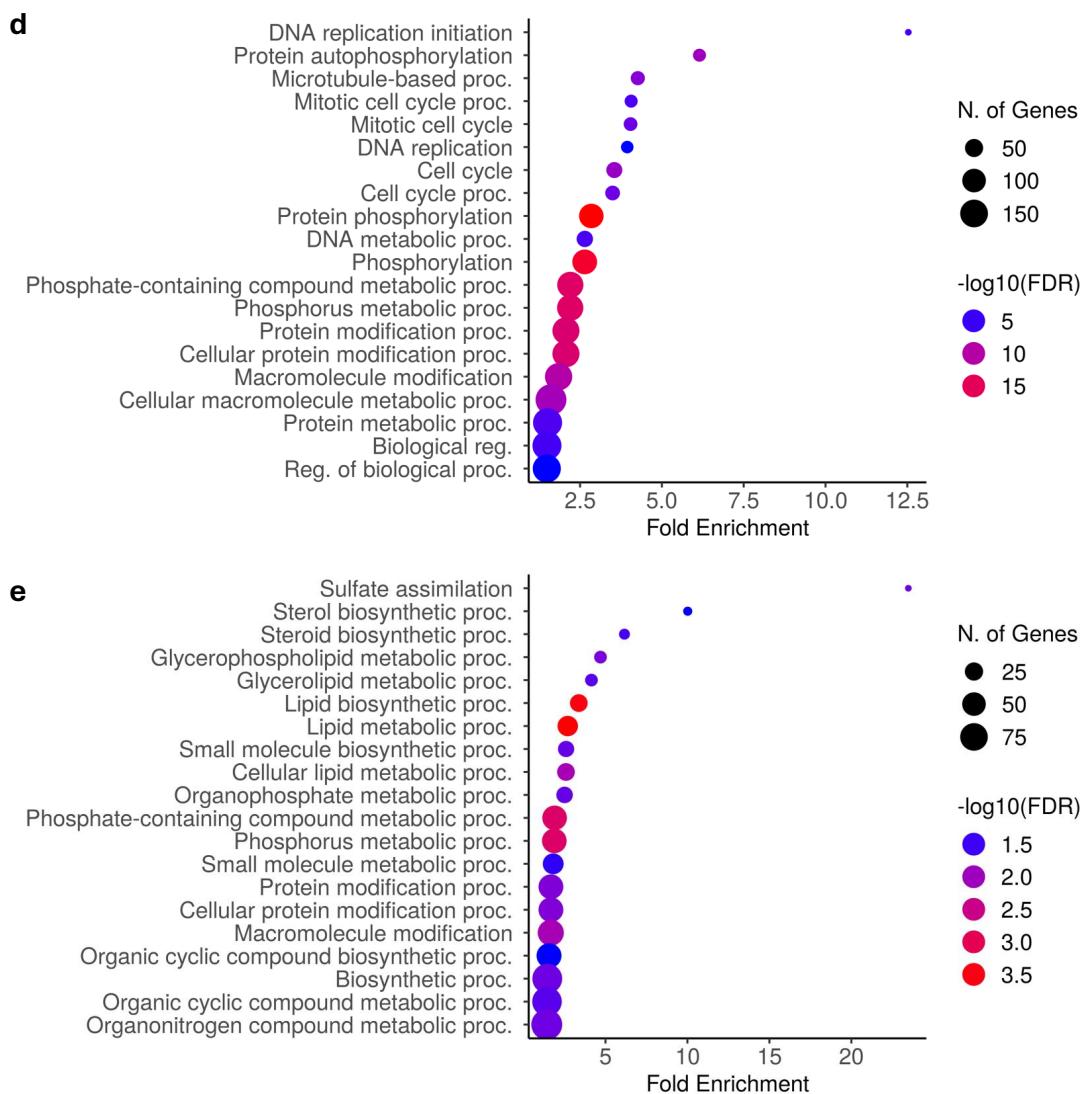


**Figure 4: Transcriptomic regulation in the Psoralen Pathway (PP) line compared to Wild-Type (WT).** (a) Differentially Expressed Genes (DEGs) in leaves and/or roots (FDR<0.05). Gene Ontology

Enrichment Analysis (GOEA) on biological processes for (b) up-regulated and (c) down-regulated genes in leaves (FDR<0.05). The position of the circles represents the Fold Change by considering the ratio PP/WT. The size of the circles represents the number of differentially expressed genes (DEGs) in each GO category. The color of the circles indicates the statistical significance. Cluster 8 contains 309 genes clustered using the k-means method and includes most of the genes specifically regulated in the leaves and roots of the PP line compared to the WT line (10/17 genes). The data consist of 3 replicates per genotype and tissue.

A main drawback of classical differential gene analysis, when applied on RNAseq data with limited replicates (only 3 replicates per organ and line here), is that only strong DEGs may be highlighted, impeding the identification of more subtle regulations. To limit this, we used a complementary analysis which consisted in clustering approach. Such approach allowed the clustering of genes in 11 clusters based on their expression profile among the four conditions (leaves and roots in PP and WT) (**Fig. S3**). Among them, 5 clusters showed patterns discriminating the PP line and the WT: The cluster 8 regrouped 309 genes with a higher expression in both the leaves and roots of the PP line in comparison to WT. In addition, the cluster 2 (1449 genes) contained genes specifically overaccumulated in the roots of the PP line and the cluster 7 (1113) genes overaccumulated in the leaves. Conversely, the clusters 1 (743 genes) and 5 (1466 genes) grouped genes downregulated in leaves and roots of the PP line compared to WT. Gene ontology enrichment analysis of cluster 8 indicates positive regulation of genes potentially linked to the PP line's response to stress, likely signaled by jasmonate and in response to lipids (**Fig. 5a**). The same trend was obtained regarding the genes from the cluster 7 (upregulated in leaves of the PP line) with additional processes highlighted like pollination and tryptophan biosynthesis (**Fig. 5b**). The cluster 2 (upregulated genes in roots of the PP line), processes related to hormonal regulation were enriched (**Fig. 5c**). Among the processes related to down regulated genes (**Fig. 5d,e**), DNA replication and cell cycle appeared affected in the roots of the PP line (cluster 5), whereas lipid and sterol biosynthesis were, in leaves (cluster 1).





**Figure 6: Gene Ontology Enrichment Analysis (GOEA) on biological processes in the Psoralen Pathway (PP) line compared to Wild-Type (WT) for clusters 8 (a), 7 (b), 2 (c), 5 (d) and 1 (e) (FDR<0.05).** The position of the circles represents the Fold Change by considering the ratio PP/WT. The size of the circles represents the number of differentially expressed genes (DEGs) in each GO category. The color of the circles indicates the statistical significance. Gene clusters was grouped using the k-means method. Cluster 8 includes 309 up-regulated genes and most of those specifically regulated in the leaves and roots of the PP line compared to the WT line (10/17 genes). Cluster 7 contains 1113 up-regulated genes in leaves. Clusters 1 and 5 contains 743 and 1465 down-regulated genes in roots and leaves. Cluster 2 contains 1449 up-regulated genes in roots. The data consist of 3 replicates per genotype and tissue.

### The PP line accumulate less biomass but more nitrogen

To identify potential impacts on the growth and development of the PP line compared to the WT line, we conducted a resource allocation and morphology comparative analysis.

The PP and WT plants exhibited equivalent morphology and size (**Fig. S1**). Comparing the number of leaves on the main axis and the stem length (between 13 and 28 days), we found no development modifications (**Table II**). However, the PP line accumulated significantly lesser dry biomass (-21.3%), with relative equivalent reductions in both shoot (-29.4%) and root (-20.8%) parts. Resource allocation patterns in the plant showed that the PP line accumulated proportionally lesser dry matter in leaf blades (-10%) and more in the internodes (+20%) (**Table II**). The specific leaf area (+39.4%) and specific stem length (+6.1%) were higher in PP line than in WT line.

**Table II: Mean alterations in growth and development parameters between WT and the PP line.** The variables of the two genotypes are compared via percentage change and supported by a non-parametric Kruskal-Wallis test. Significant variables are in bold. Measurements were taken at a vegetative growth stage (28 days after sowing) and compared for development with measurements at an early growth stage (13 days after sowing). Negative variations are highlighted in blue and positive's in green. Dry biomass allocation patterns are defined by the percentage of mean dry biomass in each compartment. The epigaea part refers to the main axis of the tomato, including the stem hypocotyl and the first six leaves. It excludes secondary axillary and apical branches and the first two cotyledons. Abbreviations: SLA (Specific Leaf Area); SRL (Specific Root Length); SSL (Specific Shoot Length).

	variable	WT mean	PP mean	modification (%)	p-value
dry weights	foliar blades (g)	2.46±0.36	1.74±0.24	-29.44	<b>0.0004</b>
	petioles (g)	0.49±0.07	0.4±0.05	-18.92	<b>0.0222</b>
	internodes (g)	1.11±0.28	1.05±0.17	-4.72	0.4142
	roots (g)	0.49±0.05	0.39±0.06	-20.87	<b>0.0025</b>
	epigea/hypogea ratio	8.20±1.07	8.15±0.70	-0.54	0.9349
lengths	epigeae (cm)	49.95±7.42	51.33±6.02	2.76	0.6532
	roots (cm)	44.21±15.1	44.22±9.38	0.02	0.7439
	epigea/hypogea ratio	1.24±0.43	1.2±0.28	-3.22	1
repartition	SLA (cm <sup>2</sup> .g <sup>-1</sup> )	236.48±33.67	329.73±35.8 <sub>1</sub>	39.43	<b>0.0002</b>
	SRL (cm.g <sup>-1</sup> )	90.95±39.82	115.21±31.8 <sub>5</sub>	26.67	0.0724
	SSL (cm.g <sup>-1</sup> )	45.06±14.62	47.84±6.17	6.16	<b>0.0274</b>
development	leaves number	8	8	0	na
	D13_D28_plant_length	0.1±0.02	0.1±0.02	0	0.8064

	variable	WT plant allocation (%)	PP plant allocation (%)	modification (%)	p-value
dry weights	foliar blades	54.06	48.6	-10.09	<b>0.0042</b>
	petioles	10.76	11.17	+3.81	0.2529
	internodes	24.39	29.32	+20.21	<b>0.0178</b>
	roots	10.76	10.89	+1.20	0.9349

To better understand the biomass loss in the PP line, we analyzed carbon (%C) and nitrogen (%N) accumulation across different tissues (**Table III**). We found that nitrogen content was significantly higher in the PP line compared to the WT, particularly in the shoots (+53%). Conversely, carbon content tended to decrease in the PP line, most notably in the stems (-13%). As a result, the carbon-to-nitrogen ratio (C/N) was significantly decreased in the PP line, especially in the shoots.

**Table III: Mean alterations in carbon (C), nitrogen (N) contents and C/N ratio between the Wild-Type (WT) and the Psoralen Pathway (PP) lines.** The variables of the two genotypes are compared via percentage change and supported by a non-parametric Kruskal-Wallis test. Measurements were taken at a vegetative growth stage (28 days after sowing).

	carbone (%)				nitrogen (%)			
	WT mean	PP mean	modification	p-value	WT mean	PP mean	modification	p-value
foliar blades	39.01±0.41	38.66±1.02	-0.89	0.1616	1.72±0.3	2.63±0.33	52.9	<b>0.0005</b>
petioles	34.57±0.62	33.89±0.66	-1.96	<b>0.0484</b>	1.34±0.22	1.96±0.3	46.26	<b>0.0005</b>
internodes	31.95±2.9	27.78±5.75	-13.05	0.0542	1.01±0.18	1.22±0.21	20.79	<b>0.0301</b>
roots	40.79±0.7	40.52±0.63	-0.66	0.885	2.02±0.12	2.27±0.17	12.37	<b>0.006</b>

	C/N ratio			
	WT mean	PP mean	modification	p-value
foliar blades	22.47±3.96	14.82±1.59	-34.02	<b>0.0002</b>

petioles	25.46±3.91	17.57±2.49	-30.96	<b>0.0002</b>
internodes	31.96±4.74	22.85±3.68	-28.49	<b>0.0004</b>
roots	20.33±0.95	17.89±1.29	-12	<b>0.0004</b>

## Discussion

The biosynthesis of SMs necessarily imposes a physiological cost on plants, either by diverting resources for their production or through the pleiotropic effects of the SM themselves (Neilson *et al.*, 2013). This study aimed to better understand the effects of constitutive expression of a part of the furanocoumarin biosynthesis pathway on tomato physiology and growth.

### Is furanocoumarin production toxic for tomato?

A first observation of this study was a low transformation success (0.33%) with only two lines successfully regenerated and demonstrated to have integrated the full transgene in their genome. This transformation success is very low compared to those observed in similar experiments using the same tomato variety and transfection methods in our lab (e.g., CRISPR-Cas9: 80,7-92,3%; Goldenbraid: 5,6%) (Roumani, 2021; Villard, 2020), however the size of the transgene, including 5 transcription units, may explain at least partly these differences compared to others that were around twice smaller. The second observation is that among the two transformed lines, the only one able to constitutively express all five genes (**Fig. S1e**) produce only trace amount of psoralen and all of the intermediates of the expected introduced pathway (**Fig. 2b**) indicating only a partial success in our initial strategy (**Fig. 2a,c**). Since independent hypotheses could be proposed to explain, in one side, this low psoralen accumulation in the transformed line (notably metabolic competition for substrate that will be discussed latter) and, in another side, the low transformation success (size and nature of the transgene), the co-occurrence of these two observations could also be interpreted as a potential lethal effect of the psoralen production for tomato plant. Indeed, furanocoumarins, among them psoralen, are well-known for their toxicity on a large range of organisms, including plants, occurring through a diversity of processes (photocycloaddition, lipid

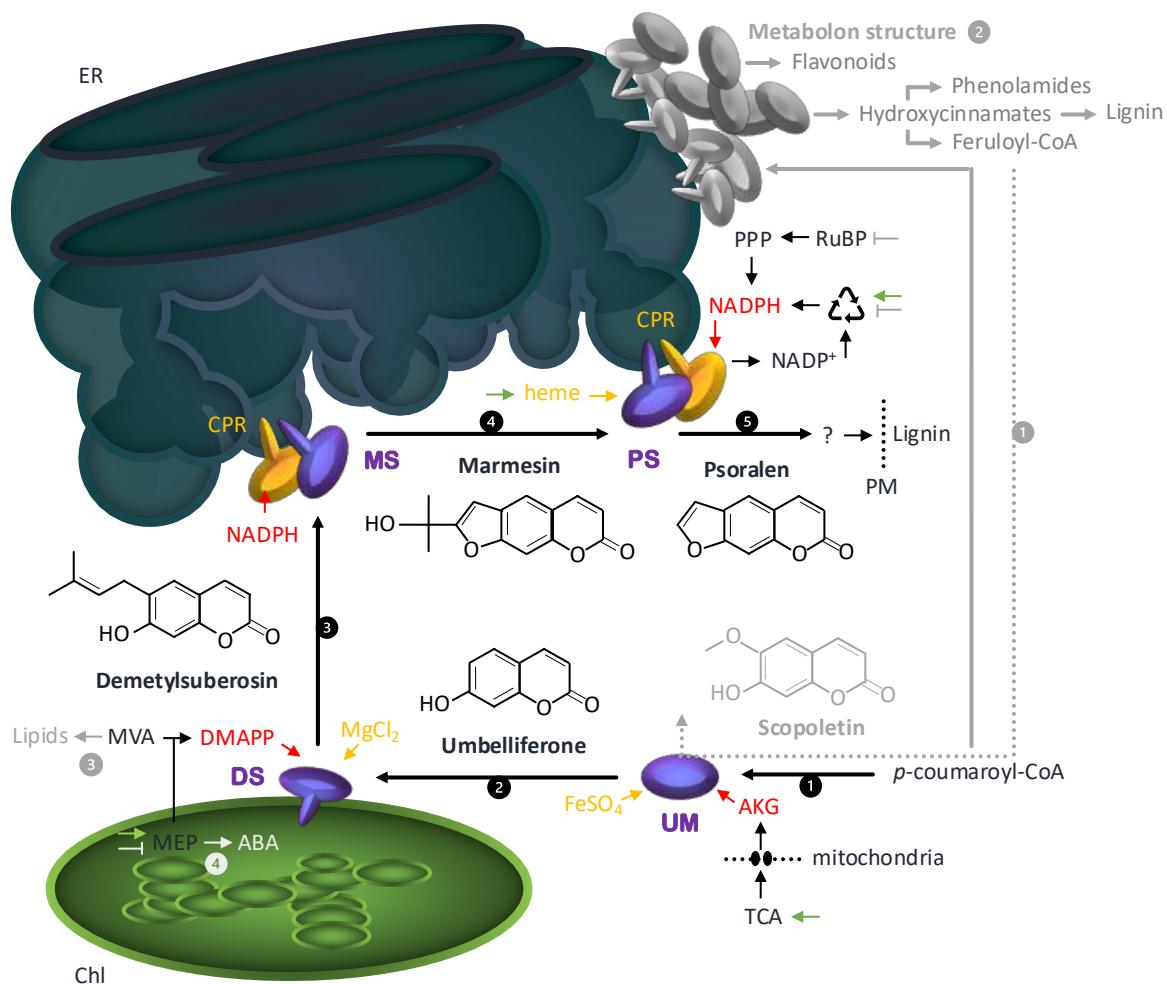
peroxidation, cytochrome P450 inhibition notably; (Baskin *et al.*, 1967; Bennet & Bonner, 1953; Diawara *et al.*, 1993; Frank *et al.*, 1998; Greenberg *et al.*, 2001; Hale *et al.*, 2004; Hassan *et al.*, 2016; Kitamura *et al.*, 2005; Razavi, 2010; X.-B. Wang *et al.*, 2008). It is tempting to hypothesize that an acute toxicity of psoralen accumulation in the tissue of tomato seedling could have contributed to select only individuals having integrated the transgene with the kanamycin resistance and producing no or very low level of psoralen. In accordance with a potential furanocoumarin's toxicity in the plant tissues, the PP line was shown to significantly over-expressed genes falling in the GO categories of “response to stress”, “response to chemicals” (**Fig. 4b**). A complementary analysis with a clear accumulation of psoralen could help to better appreciate if psoralen or an intermediate of the PP is, as we suggest, toxic for tomato.

### **Why is furanocoumarin accumulation limited in the PP line?**

Metabolic bioengineering approaches often fall short of expectations due to a limited understanding of the regulation of SM biosynthesis in plants (Lynch *et al.*, 2021). In the PP line, psoralen could be detected in almost all the plant parts but at a level just above the quantification limit (**Fig. 2d**). Since all the genes coding the enzymes necessary for its accumulation were correctly expressed (**Fig. S1e**), limiting factors should be investigated after the transcriptional step. Unfortunately, the protein accumulation, nor the enzymatic activities corresponding to the individual steps of the psoralen biosynthesis pathway were not measured in the plant tissues during this study impairing to conclude on these steps. However, metabolomics and transcriptomics data furnished elements enabling to elaborate hypothesis centered on a competition for the precursor *p*-coumaroyl-CoA involving both the catalytic properties of the integrated enzymes of the PP line and possible regulation in the PP line metabolism.

Firstly, instead of a clear and quantifiable accumulation of psoralen and intermediates in the PP line, the metabolomics approach revealed a marked and significant accumulation of several metabolites attributed to coumarins including the formally identified esculin and scopoletin (**Fig. 2d, Table I**). The biosynthesis of esculetin (the aglycone form of esculin) and scopoletin can be achieved respectively from caffeoyl-CoA and feruloyl-CoA through an *ortho*-hydroxylation mediated by a dioxygenase, followed by a cyclisation proved to happen spontaneously and/or catalyzed by the recently identified COSY

enzymes (Vanholme *et al.*, 2019) (**Fig. 1**). The *ortho*-hydroxylation of *p*-coumaroyl-CoA to produce umbelliferone constitutes the first metabolic step of the integrated psoralen pathway in the PP line. To catalyze this step we have chosen to express the C2'H  $\alpha$ -ketoglutarate dependent dioxygenase from *Pastinaca sativa* (*PsDiox*) which was characterized *in vitro* as very specific to *p*-coumaroyl-CoA (Roselli *et al.*, 2017a) by contrast to other C2'H orthologs from *Ruta graveolens*, *A. thaliana* and *Ipomea batatas* that accept also caffeoyl-CoA and feruloyl-CoA (Matsumoto *et al.*, 2012; Vialart *et al.*, 2012; S.-M. Yang *et al.*, 2015). Therefore, we hypothesize that *PsDiox* in tomato may exhibit a broader specificity than expected from *in vitro* characterization, increasing the carbon flux toward the biosynthesis of esculetin and/or scopoletin at the expense of umbelliferone and the desired final product psoralen. Non-exclusive hypotheses implies also an active regulation at the PP metabolism level in favor of competing metabolic pathway (**Fig. 6**). Indeed, metabolomics analysis highlighted the accumulation of phenolamides, in addition to coumarins, and transcriptomic pointed out the overexpression in the PP line of genes of the core phenylpropanoid pathway and putatively involved in the coumarin and phenolamide biosynthesis pathway (**Fig. 6-7**, **Table I**). Such metabolic competition for common substrate as the *p*-coumaroyl-CoA hub could also be enhanced by other factors like metabolic channeling. Such competition could imply an optimal organization of the endogenous enzymes from a same metabolic pathway (metabolon) to favor the use of metabolic intermediate and co-substrates (Dahmani *et al.*, 2023; Lallemand *et al.*, 2013; Laursen *et al.*, 2016).



**Figure 6: Schematic representation of the proposed subcellular pathway for psoralen biosynthesis from *p*-coumaroyl-CoA within tomato cell physiology.** The black arrows and numbers 1 to 5 indicate the key transport steps in the Psoralen Pathway (PP). Enzymes involved in the PP are shown in purple: US (Umbelliferone Synthase: *PsDiox*), DS (Demethylsuberosin Synthase: *PsPT1*), MS (Marmesin Synthase: *FcCYP76F112*), and PS (Psoralen Synthase: *PsCYP71AJ3*). Gray arrows and numbers 1 to 4 highlight potential bypass routes in the PP line. The cosubstrats (red) and potential impacts (down-regulation: gray inhibition arrow, up-regulation: green arrows, non-impacted: black arrows) on their production/recycling pathways are also illustrated. Enzymatic cofactors (cytochrome P450 reductase: CPR), prosthetic groups (heme), and ions ( $FeSO_4$  and  $MgCl_2$ ) are depicted in yellow. Endoplasmic Reticulum (ER), Plasma Membrane (PM), Chloroplast (Chl), Tricarboxylic Acid Cycle (TCA),  $\alpha$ -ketoglutarate (AKG), Mevalonate Pathway (MVA), Methylerythritol Phosphate (MEP), Pentose Phosphate Pathway (PPP), Abscisic Acid (ABA), Nicotinamide Adenine Dinucleotide Phosphate (NADP<sup>+</sup>/NADPH).

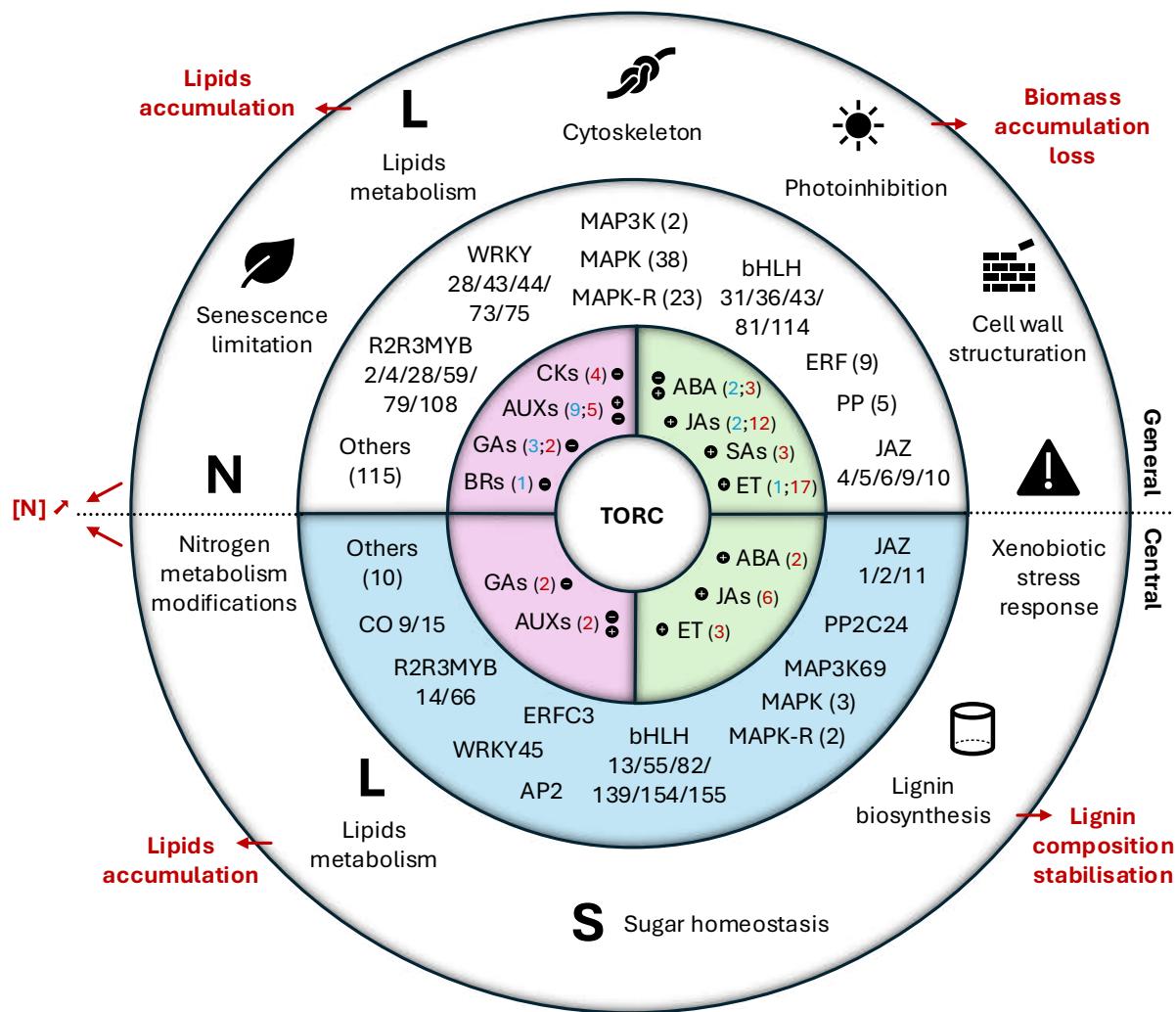
### Physiological traits of the PP line appear to be linked to the coumarin accumulation in plant tissues

The metabolomics analyses, repeated 3 times on the PP line and its WT control, highlighted a list of 80 metabolic features almost exclusively over-accumulated in the

transgenic line. Annotation outlined that at least 4 metabolic families (lipids, phenolamides, coumarins and furanocoumarins) were involved. Among them, coumarin accumulation appeared distinguishable from the other families for at least one of the following reasons: (i) the accumulation of all individual coumarins followed a presence/absence pattern in the PP/WT comparison, (ii) the accumulation of coumarin individuals was measured on the three tissues (leaf, stem, roots) in accordance to the systemic expression of the PP transgenes and (iii) based on the targeted analyses the accumulation of coumarins could be quantified, at least for scopoletin, to several  $\mu\text{g/g}$  DW whereas the furanocoumarin accumulation did not reach the quantification limit (**Fig. 2d; Table S3**). These observations made us consider that coumarin accumulation may be the main stimulus of the transcriptomic, physiological and developmental characteristics of the PP line. This assumption is coherent with the literature focused on plant response to external coumarin application (Araniti *et al.*, 2017; Graña *et al.*, 2017; Z. Yan *et al.*, 2016). Indeed, some coumarins, like umbelliferone, scopoletin or the simple coumarin, have been proved to exhibit allelopathic properties to plants, mediated notably by membrane peroxidation, ROS production, mitotic disruption or photoinhibition (Araniti *et al.*, 2017; Graña *et al.*, 2017; Z. Yan *et al.*, 2016). The transcriptomics analysis highlighted the up-regulation, in the PP line in comparison to WT, of a majority of genes associated to the responses to stress, to chemicals, to organic substances and wounding (**Fig. 4b + 5abc**). These categories comprise genes involved in ROS detoxification (e.g. glutathione-S-transferases, cytochrome P450s, peroxidases, detoxification proteins, chaperones, and heat shock proteins) (**Table S1**). On the opposite, the main categories of down regulated genes in the leaves of the PP line compared to WT were associated to the chlorophyll biosynthesis (**Fig. 4c+5d**). These transcriptomics data suggest an active regulation of the plant oxidative status, mediated by an increase in the ROS detoxification processes and a decrease in the ROS production linked to photosynthesis (**Fig. 7**). Indeed, photoinhibition usually limits photoinduced ROS accumulation (Didaran *et al.*, 2024). Additionally, the lipid over accumulation in the leaves of the PP line (**Table I**), together with the over-expression of genes associated to lipid metabolism (e.g. phospholipases, aldehyde dehydrogenases, lipid biosynthesis) (**Fig. 4a; Fig. 7; Table S1**) suggests membrane damage repair (e.g. lipid desaturases,

acyltransferase) (**Table S1**), likely due to coumarin-induced destabilization of chloroplast integrity and photochemistry as identified in *A. thaliana* (Araniti *et al.*, 2017).

By supplementation approaches, scopoletin and coumarins has been described as able to induce auxin-like hormetic effect in roots, likely presumed as due to its structural similarity to certain auxins (Bruno *et al.*, 2021; Graña *et al.*, 2017). This effects founds in auxinic herbicides is now pretty well understood as summarized in (Christoffoleti *et al.*, 2015; Grossmann, 2010). Most of the generally linked regulation are founds in the PP line regulation as for example: (i) the production of ABA and ethylene found as “fruit ripening” (**Fig. 4b**), (ii) a production of ROS, (iii) a cell wall reorganization, (iv) a proteins turnover induction via the ubiquitin-proteasome system and (v) a disturbance in root cytoskeleton (**Fig. 4b; Fig. 5c, Fig. 7; Table S1; Fig. S2a**). For all these reasons, we argue that the coumarins could have also induced an auxin-like hormetic effect in the PP line. Nevertheless, despite that we did not have deeply investigated tissue architecture, we have not seen any development impairments as dysmorphia or curved leaves as reported for tomato and more generally for plants (Grossmann, 2010; Warmund *et al.*, 2022) (**Fig. S1b**). Importantly, for instance, only high level of coumarins was shown to induce visible effects on plant development (Araniti *et al.*, 2017; Bruno *et al.*, 2021; Graña *et al.*, 2017). Then, this concentration appears far more important than the one measured in the tissues of the PP line.



**Figure 7: Conceptual representation of the changes in leaves of the Psoralen Pathway (PP) tomato line compared to the Wild Type (WT) line according to transcriptomic, metabolomic and morphophysiological data.** Central regulation corresponds to the DEGs from cluster 8. General regulation corresponds to the DEGs from all clusters except cluster 8. The diagram features concentric circles from the centre outward, representing: the theoretical central hub regulator (TORC, Target of Rapamycin Complex); Transcription of genes potentially linked to growth (pink) or resilience (green) via phytohormone metabolism; Transcription of genes associated with transcription factors and signaling-related proteins; Main potential physiological effects inferred from putatively annotated transcriptional signals; and measured outcomes based on metabolomic and morphophysiological data (red). Numbers in brackets indicate the number of transcripts associated with each gene family, while other numbers refer to specific gene names. Coloured numbers for phytohormones indicate up-regulation (red) or down-regulation (blue) and associated symbol correspond to the potential triggered phytohormone production. bHLH genes are named according to Wang *et al.* (2015), and R2R3MYB according to Li *et al.* (2016). Abbreviations: AUX (Auxin), CKs (Cytokinins), GAs (Gibberellins), BRs (Brassinosteroids), ABA (Abscisic acid), JA (Jasmonic acid), SA (Salicylic acid), ET (Ethylene), JAZ (Jasmonate-ZIM domain), PP (Phosphatase), MAPK (Mitogen-Activated Protein Kinase), MAPK-R (MAPK Receptor), bHLH (Basic Helix-Loop-Helix), AP2 (APETALA2), ERF (Ethylene Response Factor), R2R3MYB (R2R3 Myeloblastosis), CO (CONSTANS), N (Nitrogen).

### Coumarins accumulation could impact growth

To further understand the impact of coumarins on tomato physiology in the PP line, we investigated its effects on growth and resource allocation compared to the WT line. Morphological analysis of the PP line reveals a significant biomass reduction, particularly in the leaves (-30%) and, by dilution effect, in the rest of the plant (-21% overall). This biomass loss is supported by a clear photoinhibition identified in leaves with transcriptomic data. Such effect is consistent with the biomass loss observed following a coumarin or scopoletin supplementation in *A. thaliana* (Araniti *et al.*, 2017; Graña *et al.*, 2017). However, other effects, such as reduced leaf area (coumarin) (Araniti *et al.*, 2017) and simultaneous reductions in root and shoot elongation (umbelliferone) (Z. Yan *et al.*, 2016), were not observed in the PP line. These growth and development differences could depend on the plant models, the coumarins concentrations in tissues, and the physiological impacts they can triggered along times. Moreover, we identified that in the PP line, scopoletin is incorporated into lignin (**Fig. 3b**), albeit at concentrations approximately 10 times lower than in *A. thaliana* (Hoengenaert *et al.*, 2022). These data suggest that scopoletin is partly compartmentalized in the stem cell walls in the PP line, which could help mitigate the deleterious effects associated with its presence.

In the PP line, biomass and nitrogen are differentially accumulated, following a downward loss and an upward accumulation gradients, respectively (**Table II and III**). In parallel, coumarins diversity varies across compartments in line with these resource allocation patterns. Specifically, biomass is slightly less distributed in the leaves and more in the stems of the PP lineage compared to WT (**Table II**). This distribution aligns with the pattern of scopoletin accumulation in the PP line, which is ascending but higher in the stems and slightly lower in the leaves. The ascending distribution is consistent with scopoletin accumulation observed in *A. thaliana* leaves (Robe, Conejero, *et al.*, 2021) but diverges slightly by being less concentrated in tomato leaves. Assuming that scopoletin follows a similar pattern of accumulation in tomatoes as in *A. thaliana*, we propose that its accumulation may be driven by biomass fluxes.

Globally, apart for growth in the PP line, resource allocation seems to also supply investment in maintenance (e.g., potential ROS management, protein and lipid turnover), storage (e.g., resource allocation shift from leaves to stems), defense (e.g., constitutive

production of coumarins, furanocoumarins, and induced phenolamides) (**Fig. 4; Table II**; **Table S1**). All these identified regulations suggests a central management of resource allocation that is generally assumed to be linked to the TORC regulation system (Guo *et al.*, 2018; Margalha *et al.*, 2019; Monson *et al.*, 2022). One of the most prominently regulated genes in both leaves and roots is *SlPP2C24*, a putative protein phosphatase 2C (PP2C). The function of isoform 24 has not yet been described in tomato, and its expression profile has been reported as non-constitutive (Qiu *et al.*, 2022). Given the specific role of PP2Cs, typically involved in the regulation of the TORC complex, and the significant remobilization of resources, particularly nitrogen, observed in the PP line, we propose a potential role for *SlPP2C24* in resource management in the presence of coumarins (**Fig. 6**).

In conclusion, based on our transcriptomic, metabolomic, and morphophysiological data, combined with studies on the effects of coumarins on plant physiology (Araniti *et al.*, 2017; Bruno *et al.*, 2021; Graña *et al.*, 2017; Hoengenaert *et al.*, 2022; Z. Yan *et al.*, 2016), we suggest that the accumulation of coumarins at concentrations of 1–5 µg/g DW in tomatoes could induce several levels of regulation (**Fig. 6**): (i) xenobiotic stress leading to oxidative stress, which the plant can adapt to; (ii) a photoinhibition in response to oxidative stress; (iii) a potential auxin-like hormetic lowly induced stress globally in the plant; (iv) an increase of nitrogen uptake and storage, and (v) a potential TORC-mediated resource regulation driven by *SlPP2C24*.

## Material and methods

### Biological material and Standards

The tomato plants (*Solanum lycopersicum*, cv Wva106, Solanaceae) were used as control and for plant transformation. Analytical standards were sourced from Sigma (*p*-coumarate, scopoletin, esculin), Extrasynthese (umbelliferone, psoralen, esculetin), TransMIT (marmesin), Phytolab (scopolin), and R. Munakata from Kyoto University (demethylsuberosin).

### Creation of the PP transgene

The Psoralen Pathway (PP) transgene was created by combining the genes controlling the 4- steps metabolic pathway from *p*-coumaroyl-coA to psoralen (**Fig. 1**; **Fig S1c**). The four genes were Umbelliferone Synthase (UM or *PsDiox*; ncbi: KY081678) (Roselli *et al.*, 2017a), Demethylsuberosin Synthase (DS or *PsPT1* ; ncbi : KM017083) (Munakata *et al.*, 2016), Marmesin Synthase (MS or *FcCYP76F112* ; ncbi : MW348922) (Villard *et al.*, 2021a) and Psoralen Synthase (PS or *PsCYP71AJ3* (ncbi : C0SJS2) (Larbat *et al.*, 2009b). The selection marker used in the creation of the PP transgene is the gene conferring resistance to kanamycin (NPTII). This marker originates from the vector pICSL11024. The sequence of the five targeted genes were modified by removing internal restriction sites for *Bsal* and *BsmBI* (only necessary for *FcCYP76F112*). Following this, they were cloned into the vector pUPD. The genes were then flanked with extensions compatible with their insertion into the GoldenBraid 2.0 multigene cloning system (Sarrion-Perdigones *et al.*, 2013). Each domesticated gene was assembled between the Cam P35S promoter and the tNos terminator using the vectors pUPD-35S (Addgene plasmid #68163) and pUPD-tNOS (Addgene plasmid #68188), leading to individual Transcription Unit (TU). TUs build was done using a one-shot restriction/ligation of the donor plasmids (containing P35S, the gene of interest, tNOS) and the receptor GB2.0 adapted plasmid using the IIS restriction enzymes *Bsal* and *BsmBI* and T4 DNA ligase, following the procedure outlined in Villard, 2020. The resulting TUs were further fused two-by-two through restriction/ligation to assemble the PP transgene into the final vector pDGB1Ω1\_PP. The final vector was verified by PCR analysis, restriction digestion (**Fig. S1c**), and sequencing of the coding genes.

#### **Transformation and regeneration of the PP line**

The pDGB1Ω1\_PP vector was co-transformed into *Agrobacterium tumefaciens* EHA105 in the presence of the replication vector pSOUP. The transformed cells were selected on solid LB medium containing spectinomycin (100mg/L), tetracycline (5mg/L), and rifampicin (20mg/L) (Hellens *et al.*, 2000; Sarrion-Perdigones *et al.*, 2013). A pre-culture was performed overnight at 28°C, 200rpm, in 10mL of LB medium with appropriate antibiotics. A 50mL culture was then started with 5mL of the pre-culture to reach an OD<sub>600nm</sub> of 0.4-0.6. The culture was centrifuged for 10 minutes at 7000g, then resuspended in 40mL of sterile water and diluted to 30mL to achieve an OD<sub>600nm</sub> of 0.08. Meanwhile, axenic cotyledon fragments approximately 4mm in size were harvested from

7-day-old *Solanum lycopersicum* var. West Virginia 106 (Wva106) seedlings grown under controlled conditions (24°C, 16h/8h light/dark cycles; 150 $\mu$ mol.m<sup>-2</sup>.s<sup>-1</sup>, Fitoclima 1200, Aralab) (**Fig. S1a**). The cotyledon fragments were cultured for 24 hours at 24°C in darkness before being co-incubated with the Agrobacteria solution for 30 minutes at ambient temperature (gentle agitation every 5 minutes). The fragments were then dried on axenic paper and placed adaxial side up on co-cultivation medium for 48 hours at 24°C in darkness. Partially transformed tissues were selected on kanamycin-enriched callus medium (150mg/L) in Petri dishes (15 fragments per dish) at 24°C (16h/8h light/dark cycles; 150 $\mu$ mol.m<sup>-2</sup>.s<sup>-1</sup>) and subcultured weekly. After several months of subcultures, differentiated seedlings from calli were cut and transferred to rooting medium before being acclimated in soil in phytotrons (24°C; 16h/8h light/dark cycles; 400  $\mu$ mol.m<sup>-2</sup>.s<sup>-1</sup>; 70% humidity). Several T1 heterozygous lines were obtained by self-pollination. The insertion of the 5 genes into the genome was verified by PCR (**Fig. S1d**) using genomic DNA extracted with the kit following the supplier's recommendations and the primers described as follow: UM\_F: CTAGTCGAAATGGCTCCATCTCC; UM\_R: CGAAGATCCATTCCAAGCTCACATTG; DS\_F: CTAGTCGAAATGGCTCAAACAATTATGC; DS\_R: GAAGATCCATTCCAAGCTCACCTC; MS\_F: CTATTCTAGTCGAAATGGATATTCACCTC; MS\_R: CGAAGATCCATTCCAAGCCTAATGATG; PS\_F: CTATTCTAGTCGAAATGAAGATGCTTGAG; PS\_R: GATCGAAGATCCATTCCAAGCTCATC; KR\_F: CAATTACTATTCTAGTCGAAATGGTTGAACAAG; KR\_R: GAAGATCCATTCCAAGCTCAGAAGAAC. The expression of the 5 genes in each individual was confirmed by RT-PCR for each tomato line (**Fig. S1e**). Total RNA was extracted using the E.Z.N.A.® HP kit (Omega Bio-Tek) instructions, treated with DNase (AMPD1-1KT, Sigma) to remove any contaminating genomic DNA, and then reverse-transcribed into cDNA using the High-Capacity RNA-to-cDNA Kit (Applied Biosystems™), according to the supplier's.

#### **Plant cultivation and measurement of morphological and physiological traits**

Tomato plants were grown in 1.5L pots with 0.5L of potting soil (Stender D400, NPK 14/10/18, pH 6.2, sieved to 5 mm). The growth conditions in the phytotron were 24°C with 16h/8h light/dark cycles, LED illumination at 400 $\mu$ mol.m<sup>-2</sup>.s<sup>-1</sup>, and 70% humidity. Plants were watered to maintain 70% soil moisture three times a week with osmosed water.

Weekly, tomatoes were fertilised with horticultural fertiliser (50 mL at 2 g.L<sup>-1</sup>) (MultiTec™ 9.16.36 + 1.5 + trace elements). Tomatoes were harvested at 28 days after sowing, at the vegetative stage. Roots were gently washed with a brush and untangled for length measurements. Tissues were carefully separated to minimize stress during harvest. For metabolomic and transcriptomic analyses, tissues were rapidly snap-frozen in liquid nitrogen before to be finely grinded and stored at -80°C. Leaf areas (folioles and foliolules) were measured using a planimeter (Li-3100 C Area Meter, Li-Cor, Lincoln, Nebraska, USA). For the morphophysiology measurements, tissues were dried in an oven at 70°C for 1 week and then finely ground using an orbital mill (Pulverisette 6, Fritsch, France) before to be stored at room temperature in a dry-preserving package. Resulting powder were used for carbon, nitrogen and lignin measurements.

A total of four tomato plant cultivation series were conducted. The first three series were staggered but grown under identical conditions, and used for the metabolomics and transcriptomics analyses. The fourth series was dedicated to measuring morphometric traits (**Table S2c**), carbon and nitrogen allocation, and lignin composition and content in the stems. At harvest, stem and petiole tissues were separated from the leaves and roots. For metabolomics analyses, petioles and internodes were combined, while primary roots were separated from secondary roots.

### **Total Nitrogen and carbon analyses**

Total C and N contents in dry tissues were determined on ca. 3 mg dry powder weighed exactly, according to the Dumas method, with an elemental auto-analyser (Unicube, Elementar, Langenselbold, Germany).

### **Lignin extraction and analysis**

Lignin and its components were quantified according to Hoengenaert *et al.*, 2022. Briefly, cell wall residues (CWR) were extracted using successive boiling solvent baths. Lignin was extracted from CWR and quantified using the Cysteine-Assisted Sulfuric Acid (CASA) method. Lignin was then depolymerized by thioglycolic acid treatment, and its H, G, and S monomer composition was analyzed by GC-MS after derivatization.

### **Extraction and metabolomics analysis**

Frozen powders (100 mg) were extracted twice using maceration (16 hours in the dark at 22°C with slow vertical agitation) followed by sonication (37 kHz, 10 minutes at room temperature), each time with 800  $\mu$ L of methanol/deionized water (80/20). The macerates were then centrifuged (10 minutes, 16100 g, 22°C). The supernatants (1.6 mL) were collected, dried (5-7 hours under vacuum using a SpeedVac), and concentrated in 100 $\mu$ L of methanol/deionized water (80/20). The extracts were filtered (0.2  $\mu$ m, Minisart® RC4 membrane, Sartorius, REF17821) and stored at -20°C. Extracts were supplemented with 25  $\mu$ L of taxifoline as an internal quantification standard.

Chromatographic analyses coupled with High Resolution Mass Spectrometry were performed on a Thermo Scientific Vanquish UHPLC system equipped with a binary pump, an autosampler, a temperature-controlled column (40°C) and a Diode Array Detector. Extracted samples were separated on a XB-C18 Kinetex (150  $\times$  2.1 mm, 2.6  $\mu$ m) (Phenomenex Inc., Torrance, CA, USA) using a gradient of mobile phase composed of 0.1% formic acid in Water (A) and 0.1% formic acid in Methanol (B) at a flow rate of 200  $\mu$ L·min $^{-1}$ . The separation of the compounds was optimized using a 30 min mobile phase gradient (A: B; v/v). The elution method is as follows: (90:10) at 0 min, (90:10) up to 2 min, (70:30) up to 10 min, (5:95) up to 20 min remaining 5 min, (90:10) up to 26 min. The molecules are detected according to their absorbance at different wavelengths between 190 and 800 nm.

HRMS<sup>1</sup> detection was performed on an Orbitrap IDX™ (ThermoFisher Scientific, Bremen, Germany) mass spectrometer in positive and negative electrospray ionization (ESI) modes. The capillary voltages were set at 3.5 kV and 2.5 kV for positive and negative modes, respectively. The source gases were set (in arbitrary unit min $^{-1}$ ) to 40 (sheath gas), 8 (auxiliary gas) and 1 (sweep gas) and the vaporizer temperature and ion transfer tube was respectively 320°C and 275°C. Full scan MS<sup>1</sup> spectra were acquired from 120 to 1200 m/z at a resolution of 60,000. MS<sup>2</sup> analysis was performed on the pooled-samples vial using the data dependent acquisition (DDA) mode and fragments were analyzed by Orbitrap at a resolution of 15000. For this analysis, the AcquireX data acquisition workflow (ThermoFisher Scientific) was applied.

Identification of coumarins and furocoumarins, was done via comparisons with commercial standards. For compounds present as trace, name attribution was realized through their exact mass (MS<sup>1</sup>) and retention time (RT). For compounds associated to

stronger signal, MS<sup>1</sup> and MS<sup>2</sup>, together with TR were used to confirm their identity in comparison with the commercial standards. For scopoletin quantification, we use the commercial standard to establish standards curves. Areas of the samples peak were measured and normalized with an internal standard and according to the dry biomass of each tissues.

### **Construction and analysis of RNAseq libraries**

Total RNAs from leaves and roots were collected for 3 plants per genotype (WT and PP) as previously described (Roumani *et al.*, 2022). RNA quality was checked by electrophoresis on a 1% agarose gel. RNA samples were quantified using absorbance at 260 nm, their purity was evaluated by the 260/280 nm absorbance ratio and their integrity by determining and considering an RNA Integrity Number (RIN) > 7 (Agilent BioAnalyzer 2100). Transcripts data was obtained by RNA-Seq Illumina pair-end sequencing (2x150pb) using the NovaSeq 6000 methods for 30 million read pairs. Raw data were cleaned using Trimmomatic (Bolger *et al.*, 2014). The quality of the sequences was assessed using FastQC. The transcripts were mapped on the tomato genome (ITAG2.4) using the BWA-MEM software (H. Li & Durbin, 2009). Count tables were generated using the FeatureCounts software (Version 1.6.4) and Samtools merge workflows (Liao *et al.*, 2014). The differential gene expression was conducted with the glmfit option within the edgeR package for a FDR < 0.05 (Robinson *et al.*, 2010). Functional annotations, such as gene ontologies and interproscan features, were extracted from ITAG2.4\_protein\_functional.gff3 file. Differentially expressed genes (DEGs) between WT and PP lines DEGs were used to perform GO enrichment analysis (GEA) restricted to the Biological Processes on ShinyGO (version 0.80) using the “*Solanum lycopersicum* genes SL3.0” and a FDR cutoff < 0.05 (**Fig. 5**) (Ge *et al.*, 2020).

### **Statistical Analysis**

Measured and calculated morphometric and physiological variables, as well as relative and absolute quantified metabolites were statistically analyzed by comparing WT and PP lines using a non-parametric Kruskal-Wallis test.

### **Acknowledgments**

We would like to thank Alexandre Olry and Claude Gallois, members of the Experimental phytotronic platform of Lorraine PEPLor, Université de Lorraine, France for their technical assistance, as the Metabolomic and Structural Analytic Platform facility (PASM, Université de Lorraine, France). We thank Florent Ducrocq (LAE, Université de Lorraine) and Florent Magot (BBV, Université de Tours) for their valuable assistance in the critical analysis of the metabolomic data.

## Bibliography

- Araniti, F., Scognamiglio, M., Chambery, A., Russo, R., Esposito, A., D'Abrosca, B., Fiorentino, A., Lupini, A., Sunseri, F., & Abenavoli, M. R. (2017). Highlighting the effects of coumarin on adult plants of *Arabidopsis thaliana* (L.) Heynh. By an integrated -omic approach. *Journal of Plant Physiology*, 213, 30-41. <https://doi.org/10.1016/j.jplph.2017.02.013>
- Baskin, J. M., Ludlow, C. J., Harris, T. M., & Wolf, F. T. (1967). Psoralen, an inhibitor in the seeds of *Psoralea subacaulis* (Leguminosae). *Phytochemistry*, 6(9), 1209-1213. [https://doi.org/10.1016/S0031-9422\(00\)86083-5](https://doi.org/10.1016/S0031-9422(00)86083-5)
- Bennet, & Bonner. (1953). Isolation of Plant Growth Inhibitors from *Thamnosma montana*. 6.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bouillé, A., Rashmi Kumari, R., Olry, A., Charles, C., Nelson, D. R., Larbat, R., Thornton, J., Villard, C., & Hehn, A. (2024). Using lineage-specific patterns to understand convergence of enzymatic functions led to the identification of Moraceae-specific P450s involved in furanocoumarin biosynthesis. <https://doi.org/10.1101/2024.08.19.608558>
- Bruno, L., Talarico, E., Cabeiras-Freijanes, L., Madeo, M. L., Muto, A., Minervino, M., Lucini, L., Miras-Moreno, B., Sofo, A., & Araniti, F. (2021). Coumarin Interferes with Polar Auxin Transport Altering Microtubule Cortical Array Organization in *Arabidopsis thaliana* (L.) Heynh. Root Apical Meristem. *International Journal of Molecular Sciences*, 22(14), 7305. <https://doi.org/10.3390/ijms22147305>
- Chassagne, F., Cabanac, G., Hubert, G., David, B., & Marti, G. (2019). The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products®. *Phytochemistry Reviews*, 18(3), 601-622. <https://doi.org/10.1007/s11101-019-09606-2>
- Christoffoleti, P. J., Figueiredo, M. R. A. D., Peres, L. E. P., Nissen, S., & Gaines, T. (2015). Auxinic herbicides, mechanisms of action, and weed resistance : A look into recent plant science advances. *Scientia Agricola*, 72(4), 356-362. <https://doi.org/10.1590/0103-9016-2014-0360>
- Dahmani, I., Qin, K., Zhang, Y., & Fernie, A. R. (2023). The formation and function of plant metabolons. *The Plant Journal*, 114(5), 1080-1092. <https://doi.org/10.1111/tpj.16179>
- Diawara, M. M., Trumble, J. T., White, K. K., Carson, W. G., & Martinez, L. A. (1993). Toxicity of linear furanocoumarins to *Spodoptera exigua*: Evidence for antagonistic interactions. *Journal of Chemical Ecology*, 19(11), 2473-2484. <https://doi.org/10.1007/BF00980684>
- Didaran, F., Kordrostami, M., Ghasemi-Soloklui, A. A., Pashkovskiy, P., Kreslavski, V., Kuznetsov, V., & Allakhverdiev, S. I. (2024). The mechanisms of photoinhibition and repair in plants

- under high light conditions and interplay with abiotic stressors. *Journal of Photochemistry and Photobiology B: Biology*, 259, 113004. <https://doi.org/10.1016/j.jphotobiol.2024.113004>
- Fiehn, O., Robertson, D., Griffin, J., Van Der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., Van Ommen, B., Lindon, J. C., & Sansone, S.-A. (2007). The metabolomics standards initiative (MSI). *Metabolomics*, 3(3), 175-178. <https://doi.org/10.1007/s11306-007-0070-6>
- Frank, S., Caffieri, S., Raffaelli, A., Vedaldi, D., & Dall'Acqua, F. (1998). Characterization of psoralen-oleic acid cycloadducts and their possible involvement in membrane photodamage. *Journal of Photochemistry and Photobiology B: Biology*, 44(1), 39-44. [https://doi.org/10.1016/S1011-1344\(98\)00103-1](https://doi.org/10.1016/S1011-1344(98)00103-1)
- Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO : A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8), 2628-2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Graña, E., Costas-Gil, A., Longueira, S., Celeiro, M., Teijeira, M., Reigosa, M. J., & Sánchez-Moreiras, A. M. (2017). Auxin-like effects of the natural coumarin scopoletin on *Arabidopsis* cell structure and morphology. *Journal of Plant Physiology*, 218, 45-55. <https://doi.org/10.1016/j.jplph.2017.07.007>
- Greenberg, R. B., Alberti, M., Hearst, J. E., Chua, M. A., & Saffran, W. A. (2001). Recombinational and Mutagenic Repair of Psoralen Interstrand Cross-links in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 276(34), 31551-31560. <https://doi.org/10.1074/jbc.M103588200>
- Grossmann, K. (2010). Auxin herbicides : Current status of mechanism and mode of action. *Pest Management Science*, 66(2), 113-120. <https://doi.org/10.1002/ps.1860>
- Guo, Q., Major, I. T., & Howe, G. A. (2018). Resolution of growth-defense conflict : Mechanistic insights from jasmonate signaling. *Current Opinion in Plant Biology*, 44, 72-81. <https://doi.org/10.1016/j.pbi.2018.02.009>
- Hale, A. L., Meepagala, K. M., Oliva, A., Aliotta, G., & Duke, S. O. (2004). Phytotoxins from the Leaves of *Ruta graveolens*. *Journal of Agricultural and Food Chemistry*, 52(11), 3345-3349. <https://doi.org/10.1021/jf0497298>
- Han, L., Zhang, L., He, Y., Liao, L., Li, J., Xu, S., Zhao, Y., Bian, X., & Xia, Y. (2023). Three carbon-/oxygen-prenyltransferases responsible for furanocoumarin synthesis in *Angelica dahurica*. *Industrial Crops and Products*, 200, 116814. <https://doi.org/10.1016/j.indcrop.2023.116814>
- Hassan, Mohd. Z., Osman, H., Ali, M. A., & Ahsan, M. J. (2016). Therapeutic potential of coumarins as antiviral agents. *European Journal of Medicinal Chemistry*, 123, 236-255. <https://doi.org/10.1016/j.ejmech.2016.07.056>
- Hehmann, M., Lukačin, R., Ekiert, H., & Matern, U. (2004). Furanocoumarin biosynthesis in *Ammi majus* L. : Cloning of bergaptol O-methyltransferase. *European Journal of Biochemistry*, 271(5), 932-940. <https://doi.org/10.1111/j.1432-1033.2004.03995.x>
- Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., & Mullineaux, P. M. (2000). [No title found]. *Plant Molecular Biology*, 42(6), 819-832. <https://doi.org/10.1023/A:1006496308160>
- Herms, D. A., & Mattson, W. J. (1992). The Dilemma of Plants : To Grow or Defend. *The Quarterly Review of Biology*, 67(3), 283-335. <https://doi.org/10.1086/417659>
- Hoengenaert, L., Wouters, M., Kim, H., De Meester, B., Morreel, K., Vandersyppe, S., Pollier, J., Desmet, S., Goeminne, G., Ralph, J., Boerjan, W., & Vanholme, R. (2022). Overexpression of the scopoletin biosynthetic pathway enhances lignocellulosic biomass processing. *Science Advances*, 8(28), eab05738. <https://doi.org/10.1126/sciadv.ab05738>
- Ji, J., Han, X., Zang, L., Li, Y., Lin, L., Hu, D., Sun, S., Ren, Y., Maker, G., Lu, Z., & Wang, L.

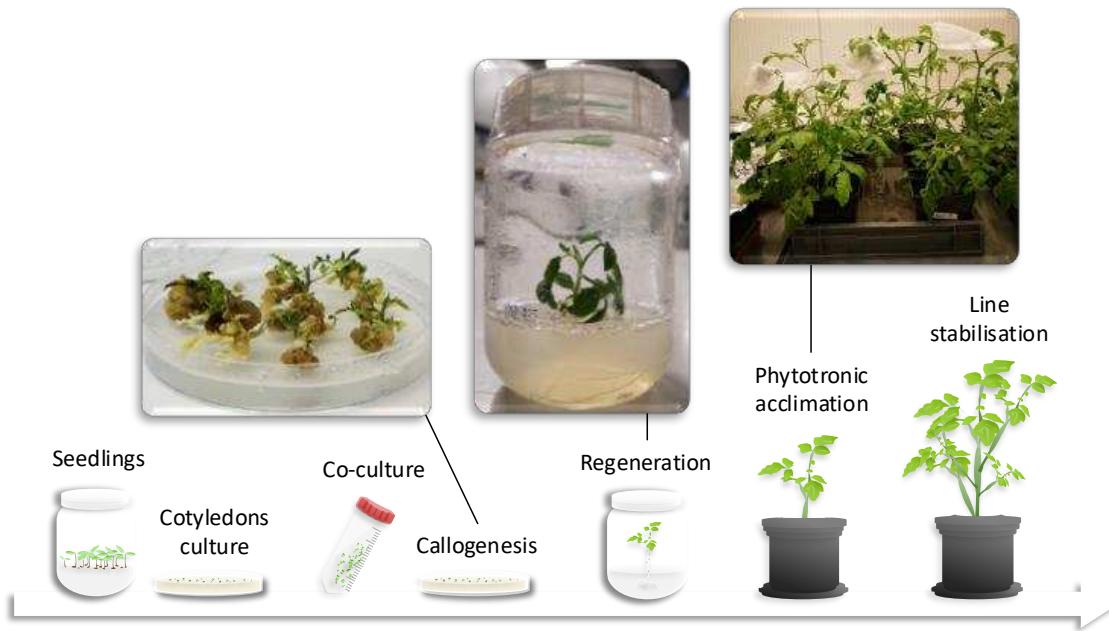
- (2024). Integrative multi-omics data elucidating the biosynthesis and regulatory mechanisms of furanocoumarins in *Angelica dahurica*. <https://doi.org/10.1101/2024.07.23.604792>
- Kitamura, N., Kohtani, S., & Nakagaki, R. (2005). Molecular aspects of furocoumarin reactions: Photophysics, photochemistry, photobiology, and structural analysis. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 6(2-3), 168-185. <https://doi.org/10.1016/j.jphotochemrev.2005.08.002>
- Krieger, C., Roselli, S., Kellner-Thielmann, S., Galati, G., Schneider, B., Grosjean, J., Olry, A., Ritchie, D., Matern, U., Bourgaud, F., & Hehn, A. (2018). The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants. *Frontiers in Plant Science*, 9, 820. <https://doi.org/10.3389/fpls.2018.00820>
- Lallemand, B., Erhardt, M., Heitz, T., & Legrand, M. (2013). Sporopollenin Biosynthetic Enzymes Interact and Constitute a Metabolon Localized to the Endoplasmic Reticulum of Tapetum Cells. *Plant Physiology*, 162(2), 616-625. <https://doi.org/10.1104/pp.112.213124>
- Larbat, R., Hehn, A., Hans, J., Schneider, S., Juggé, H., Schneider, B., Matern, U., & Bourgaud, F. (2009). Isolation and Functional Characterization of CYP71AJ4 Encoding for the First P450 Monooxygenase of Angular Furanocoumarin Biosynthesis. *Journal of Biological Chemistry*, 284(8), 4776-4785. <https://doi.org/10.1074/jbc.M807351200>
- Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., Bourgaud, F., & Matern, U. (2007). Molecular Cloning and Functional Characterization of Psoralen Synthase, the First Committed Monooxygenase of Furanocoumarin Biosynthesis. *Journal of Biological Chemistry*, 282(1), 542-554. <https://doi.org/10.1074/jbc.M604762200>
- Laursen, T., Borch, J., Knudsen, C., Bavishi, K., Torta, F., Martens, H. J., Silvestro, D., Hatzakis, N. S., Wenk, M. R., Dafforn, T. R., Olsen, C. E., Motawia, M. S., Hamberger, B., Møller, B. L., & Bassard, J.-E. (2016). Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science*, 354(6314), 890-893. <https://doi.org/10.1126/science.aag2347>
- Li, D., & Gaquerel, E. (2021). Next-Generation Mass Spectrometry Metabolomics Revives the Functional Analysis of Plant Metabolic Diversity. *Annual Review of Plant Biology*, 72(1), 867-891. <https://doi.org/10.1146/annurev-arplant-071720-114836>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923-930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lynch, J. H., Huang, X.-Q., & Dudareva, N. (2021). Silent constraints: The hidden challenges faced in plant metabolic engineering. *Current Opinion in Biotechnology*, 69, 112-117. <https://doi.org/10.1016/j.copbio.2020.12.014>
- Maeda, H. A., & Fernie, A. R. (2021). Evolutionary History of Plant Metabolism. *Annual Review of Plant Biology*, 72(1), 185-216. <https://doi.org/10.1146/annurev-arplant-080620-031054>
- Margalha, L., Confraria, A., & Baena-González, E. (2019). SnRK1 and TOR: Modulating growth-defense trade-offs in plant stress responses. *Journal of Experimental Botany*, 70(8), 2261-2274. <https://doi.org/10.1093/jxb/erz066>
- Matsumoto, S., Mizutani, M., Sakata, K., & Shimizu, B.-I. (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.

- <https://doi.org/10.1016/j.phytochem.2011.11.009>
- Monson, R. K., Trowbridge, A. M., Lindroth, R. L., & Lerdau, M. T. (2022). Coordinated resource allocation to plant growth-defense tradeoffs. *New Phytologist*, 233(3), 1051-1066. <https://doi.org/10.1111/nph.17773>
- Munakata, R., Olry, A., Karamat, F., Courdavault, V., Sugiyama, A., Date, Y., Krieger, C., Silie, P., Foureau, E., Papon, N., Grosjean, J., Yazaki, K., Bourgaud, F., & Hehn, A. (2016). Molecular evolution of parsnip (*Pastinaca sativa*) membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytologist*, 211(1), 332-344. <https://doi.org/10.1111/nph.13899>
- Neilson, E. H., Goodger, J. Q. D., Woodrow, I. E., & Møller, B. L. (2013). Plant chemical defense : At what cost? *Trends in Plant Science*, 18(5), 250-258. <https://doi.org/10.1016/j.tplants.2013.01.001>
- Qiu, J., Ni, L., Xia, X., Chen, S., Zhang, Y., Lang, M., Li, M., Liu, B., Pan, Y., Li, J., & Zhang, X. (2022). Genome-Wide Analysis of the Protein Phosphatase 2C Genes in Tomato. *Genes*, 13(4), 604. <https://doi.org/10.3390/genes13040604>
- Razavi, S. M. (2010). Plant Coumarins as Allelopathic Agents. *International Journal of Biological Chemistry*, 5(1), 86-90. <https://doi.org/10.3923/ijbc.2011.86.90>
- Robe, K., Conejero, G., Gao, F., Lefebvre-Legendre, L., Sylvestre-Gonon, E., Rofidal, V., Hem, S., Rouhier, N., Barberon, M., Hecker, A., Gaymard, F., Izquierdo, E., & Dubos, C. (2021). Coumarin accumulation and trafficking in *Arabidopsis thaliana* : A complex and dynamic process. *New Phytologist*, 229(4), 2062-2079. <https://doi.org/10.1111/nph.17090>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR : A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roselli, S., Olry, A., Vautrin, S., Coriton, O., Ritchie, D., Galati, G., Navrot, N., Krieger, C., Vialart, G., Bergès, H., Bourgaud, F., & Hehn, A. (2017). A bacterial artificial chromosome (BAC) genomic approach reveals partial clustering of the furanocoumarin pathway genes in parsnip. *The Plant Journal*, 89(6), 1119-1132. <https://doi.org/10.1111/tpj.13450>
- Roumani, M. (2021). Les phénolamides de la tomate : Développement d'une approche d'ingénierie métabolique pour l'étude de leurs fonctions in planta, et évaluation de leurs activités biologiques.
- Sarker, S. D., & Nahar, L. (2017). Progress in the Chemistry of Naturally Occurring Coumarins. In A. D. Kinghorn, H. Falk, S. Gibbons, & J. Kobayashi (Éds.), *Progress in the Chemistry of Organic Natural Products 106* (Vol. 106, p. 241-304). Springer International Publishing. [https://doi.org/10.1007/978-3-319-59542-9\\_3](https://doi.org/10.1007/978-3-319-59542-9_3)
- Sarrion-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarolo, P., Blanca, J., Granell, A., & Orzaez, D. (2013). GoldenBraid 2.0 : A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *PLANT PHYSIOLOGY*, 162(3), 1618-1631. <https://doi.org/10.1104/pp.113.217661>
- Stamp, N. (2003). Out Of The Quagmire Of Plant Defense Hypotheses. *The Quarterly Review of Biology*, 78(1), 23-55. <https://doi.org/10.1086/367580>
- Stamp, N. (2004). Can the growth-differentiation balance hypothesis be tested rigorously? *Oikos*, 107(2), 439-448. <https://doi.org/10.1111/j.0030-1299.2004.12039.x>
- Vanholme, R., Sundin, L., Seetso, K. C., Kim, H., Liu, X., Li, J., De Meester, B., Hoengenaert, L., Goeminne, G., Morreel, K., Haustraete, J., Tsai, H.-H., Schmidt, W., Vanholme, B., Ralph, J., & Boerjan, W. (2019). COSY catalyses trans-cis isomerization and lactonization in the biosynthesis of coumarins. *Nature Plants*, 5(10), 1066-1075. <https://doi.org/10.1038/s41477-019-0510-0>
- Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., & Bourgaud, F. (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: C2'H involved in umbelliferone synthesis. *The Plant Journal*, 70(3), 460-470.  
<https://doi.org/10.1111/j.1365-313X.2011.04879.x>
- Villard, C. (2020). HOW TO PRODUCE YOUR FURANOCOUMARINS: THE HIDDEN PATHWAY.
- Villard, C., Munakata, R., Kitajima, S., Velzen, R., Schranz, M. E., Larbat, R., & Hehn, A. (2021). A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution. *New Phytologist*, 231(5), 1923-1939.  
<https://doi.org/10.1111/nph.17458>
- Wang, X.-B., Li, G.-H., Li, L., Zheng, L.-J., Huang, R., & Zhang, K.-Q. (2008). Nematicidal coumarins from *Heracleum candidans* Wall. *Natural Product Research*, 22(8), 666-671.  
<https://doi.org/10.1080/14786410701766463>
- Warmund, M. R., Ellersieck, M. R., & Smeda, R. J. (2022). Sensitivity and Recovery of Tomato Cultivars Following Simulated Drift of Dicamba or 2,4-D. *Agriculture*, 12(9), 1489.  
<https://doi.org/10.3390/agriculture12091489>
- Weng, J.-K., Lynch, J. H., Matos, J. O., & Dudareva, N. (2021). Adaptive mechanisms of plant specialized metabolism connecting chemistry to function. *Nature Chemical Biology*, 17(10), 1037-1045.  
<https://doi.org/10.1038/s41589-021-00822-6>
- Yan, Z., Wang, D., Cui, H., Zhang, D., Sun, Y., Jin, H., Li, X., Yang, X., Guo, H., He, X., Pan, L., Ren, X., Guo, K., & Qin, B. (2016). Phytotoxicity mechanisms of two coumarin allelochemicals from *Stellera chamaejasme* in lettuce seedlings. *Acta Physiologiae Plantarum*, 38(10), 248.  
<https://doi.org/10.1007/s11738-016-2270-z>
- Yang, S.-M., Shim, G. Y., Kim, B.-G., & Ahn, J.-H. (2015). Biological synthesis of coumarins in *Escherichia coli*. *Microbial Cell Factories*, 14(1), 65.  
<https://doi.org/10.1186/s12934-015-0248-y>
- Zhao, Y., Wang, N., Zeng, Z., Xu, S., Huang, C., Wang, W., Liu, T., Luo, J., & Kong, L. (2016). Cloning, Functional Characterization, and Catalytic Mechanism of a Bergaptol O-Methyltransferase from *Peucedanum praeruptorum* Dunn. *Frontiers in Plant Science*, 7.  
<https://doi.org/10.3389/fpls.2016.00722>

## Supplemental data

a



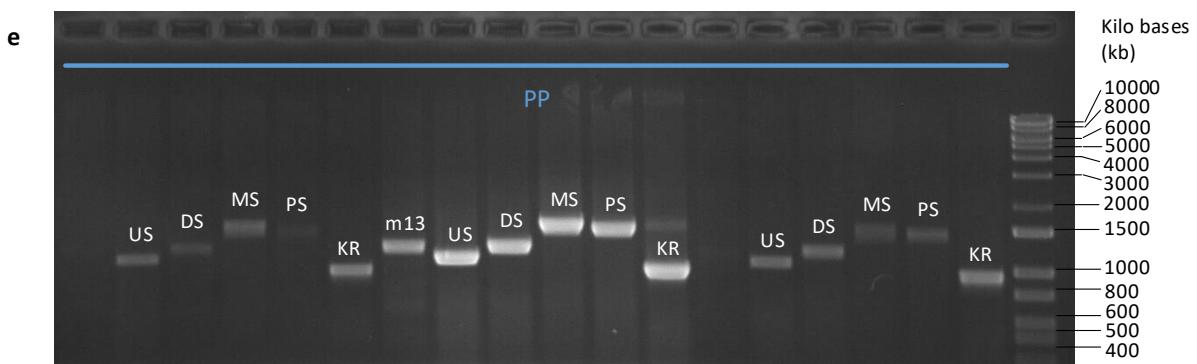
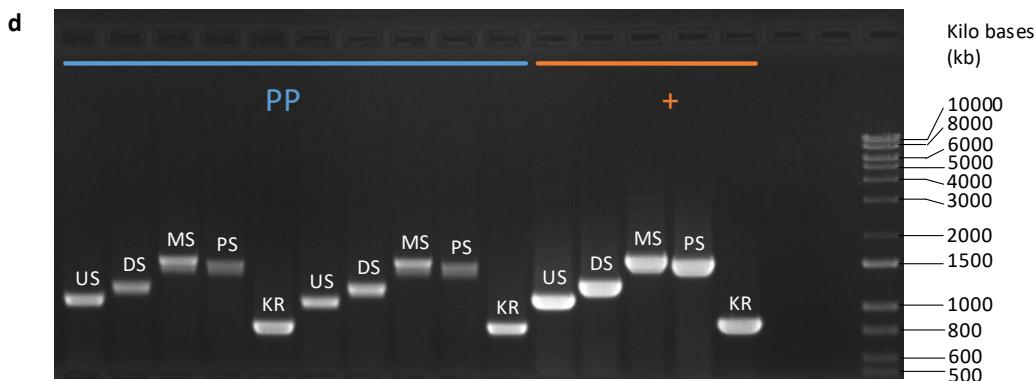
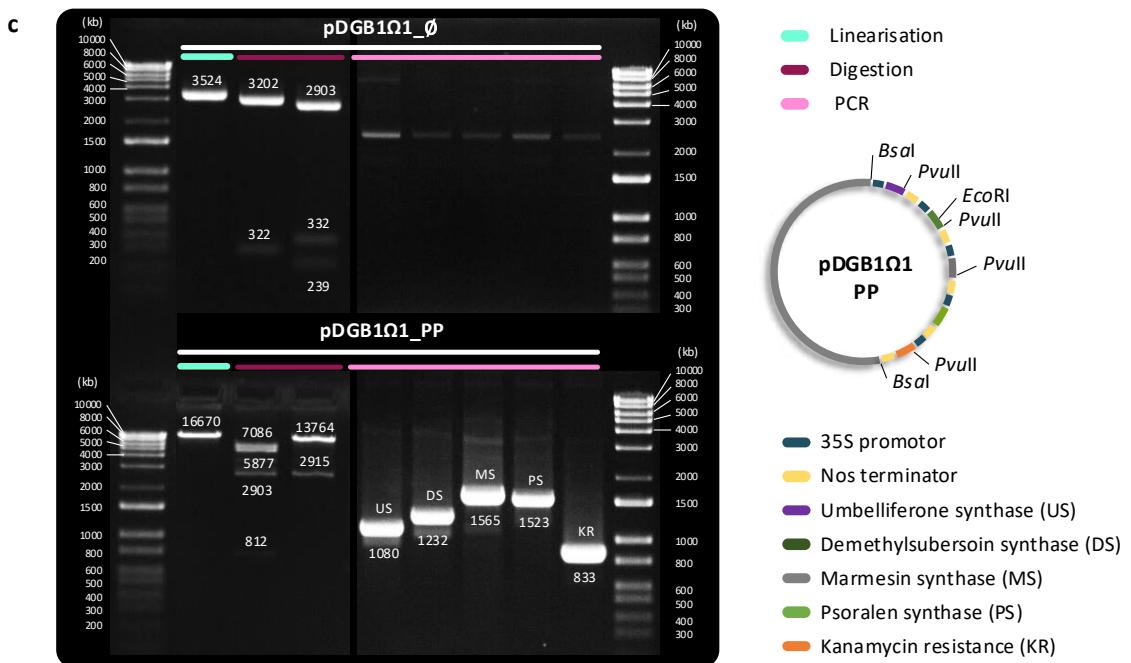
b



Wild-Type line

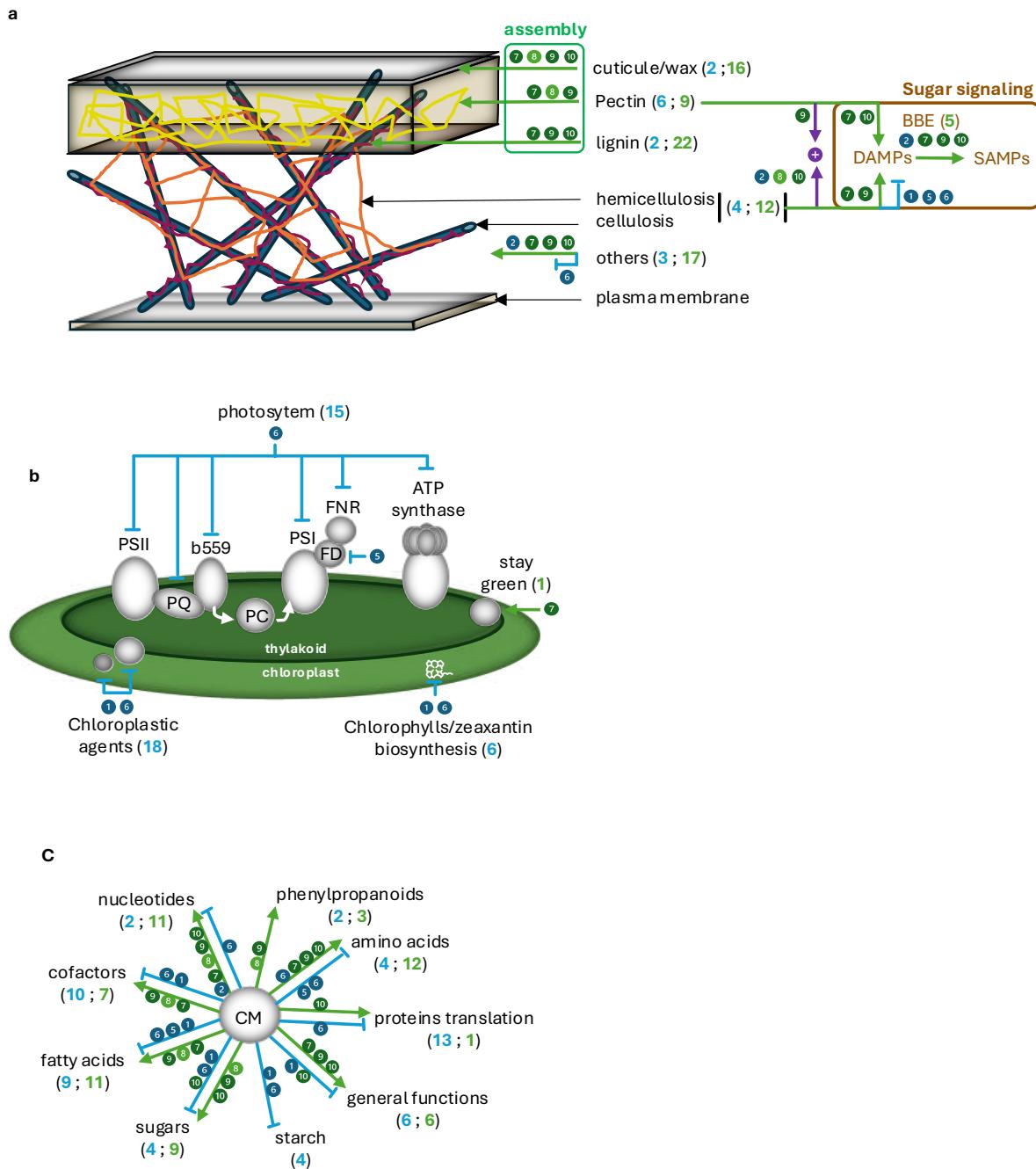


Psoralen Pathway line



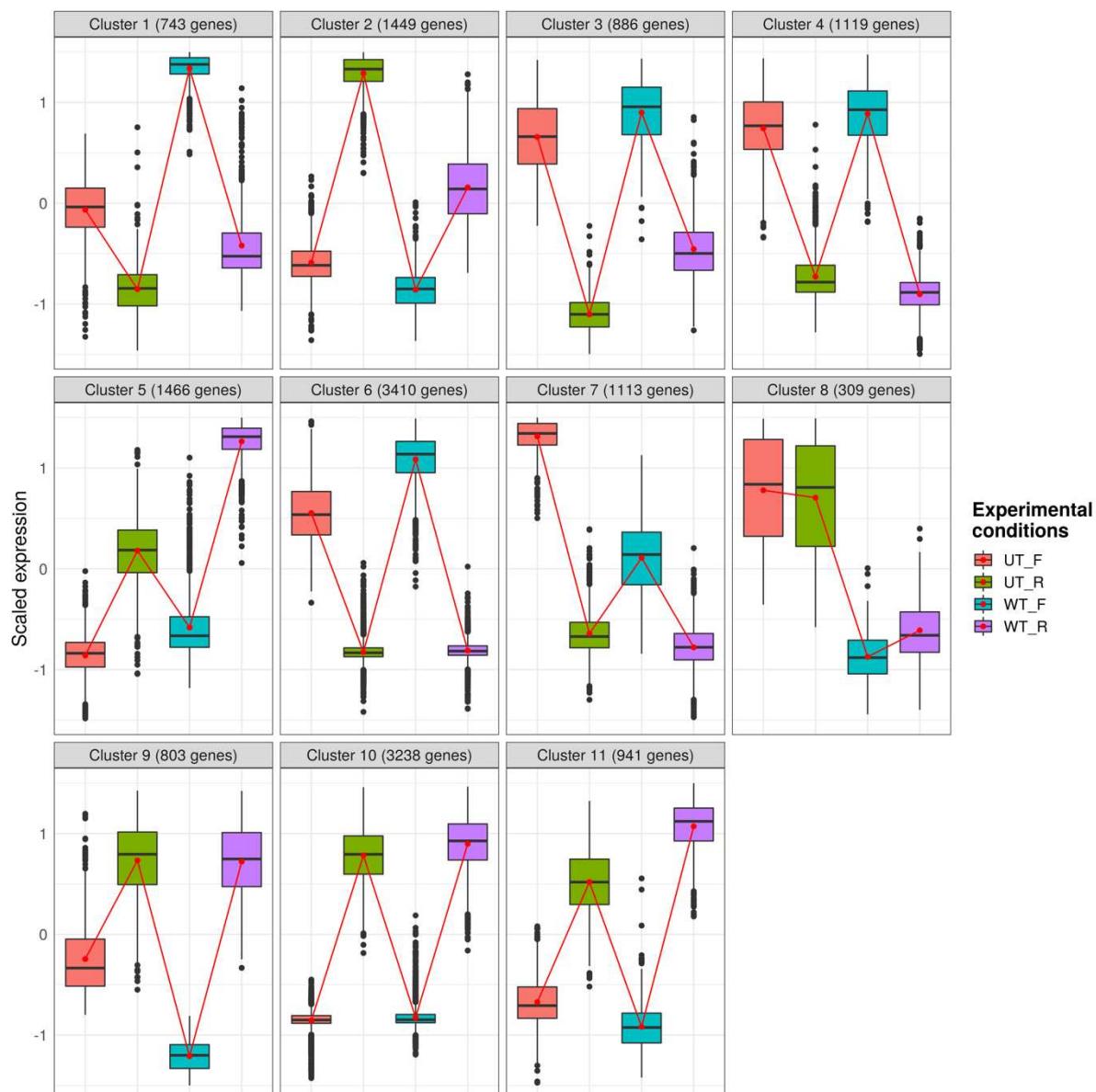
**Figure S1: The Psoralen Pathway (PP) establishment.** (a) Schematic overview of the technical steps involved in generating the PP line, alongside photographs highlighting key stages in the lineage development. (b) Photographs of representative tomato plants (*Solanum lycopersicum* var. Wva106, Solanaceae) from the Wild-Type (left) and PP (right) lines, 28 days post-sowing under controlled, non-limiting growth conditions. The plants exhibit a generally similar phenotype. (c)

Electrophoresis gel verification after linearization, restriction, and PCR of the pDGB1Ω1\_PP construct and the empty vector pDGB1Ω1\_Ø as a negative control. (d) PCR Electrophoresis gel verification of two T1 transformant lines using genomic DNA for the 5 genes of the PP transgene. Positive control (+) is a PCR amplification of the pure pDGB1Ω1\_PP plasmid under the same conditions. (e) Expression verification of the 5 PP transgene genes in T1 lines, with Myb13 (m13) housekeeping gene as the reference.

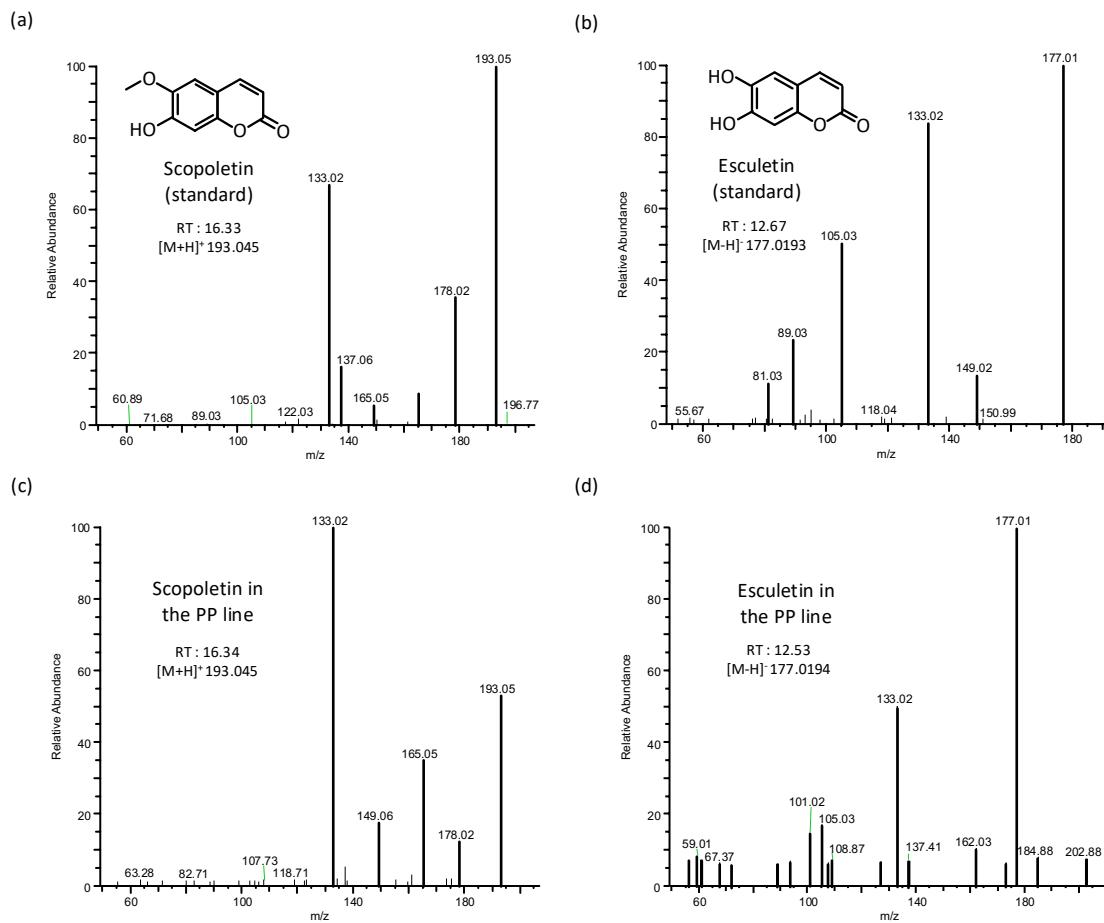


**Figure S2: Schematic overview of specific transcriptional regulations in tomato leaves of the 5UTs lineage (3n) compared to the WT lineage (3n) for genes putatively associated with cell wall remodeling (a) and Central Metabolism (CM) adjustment coupled with photoinhibition (b), by coregulation clusters.** The annotations represent specific (FDR<0.05 threshold) and convergent signals of putative functions annotated from a gene ontology analysis on the tomato genome (vs ITAG2.4) and consistent with manually annotated categories: cell wall related,

photoinhibition, and central metabolism. For more details refer to **Table S1**. The circled numbers represent the regulation clusters coloured according to their preferential regulations, positive (green: 7, 8, 9, 10) or negative (blue: 1, 2, 5, 6). The light green cluster refers to cluster 8, identified as having the highest specificity of regulation in response to the 5UTs genotype compared to the WT lineage. The arrows indicate positive (green) or negative (blue) transcriptional regulation. The assembly of cell wall portions is outlined in **emerald-green**, while the disassembly of sugar portions is outlined in **brown**. The consolidation of sugar polymers by methylations or acetylations is annotated in **purple**. "Others" refers to the regulations related to transcripts putatively annotated to expansins, extensins, Protein trichome birefringence, etc. For more details refer to **Table S1** DAMPs (Damage Associated Molecular Patterns); SAMPs (Signals Associated Molecular Patterns); BBE (FAD-binding Berberine family protein); PSI/PSII (Photosystems I or II); b559 cytochrome; PQ (plastoquinone); PC (Plastocyanin); FD (ferredoxin); FNR (ferredoxin-NADP<sup>+</sup> reductase); MC (Central Metabolism).



**Figure S3: Clustering analysis of the specific gene regulation between the Psoralen Pathway (PP) line and the Wild-Type (WT) line in roots (n=3) and leaves (n=3). Legend: UT: The PP line; WT: WT line; F: leaves; R: Roots.**



**Figure S4: Formal identification of scopoletin and esculetin in the PP line.** Tandem mass spectrometry fragmentation pattern of standards of (a) scopoletin and (b) esculetin and (c-d) their respective forms identified in the PP line. Analysis was performed after negative (for esculetin) or positive (for scopoletin) electrospray ionisation (ESI) and Targeted Single Ion Monitoring (SIM) mode between 50 and 210 m/z. The Higher Energy Collisional Dissociation (HCD) fragmentation mode was moth investigated.

**Table S3: Count table of the heterologous genes expression in roots and leaves according to the transcriptomic analysis.**

		WT		PP			
Tissu		Roots					
Sample		1	2	3	4	5	6
<b>Umbelliferone Synthase</b> (US; PsDiox)		0	0	0	5052	11307	5446
<b>Demethylsuberosin Synthase</b> (DS; PsPT1)		0	0	0	4539	10198	4447
<b>Marmesin Synthase</b> (MS ; FcCYP76F112)		0	0	0	1905	3585	2283
<b>Psoralen Synthase</b> (PS; PsCYP71AJ3)		0	0	0	4971	11100	5397

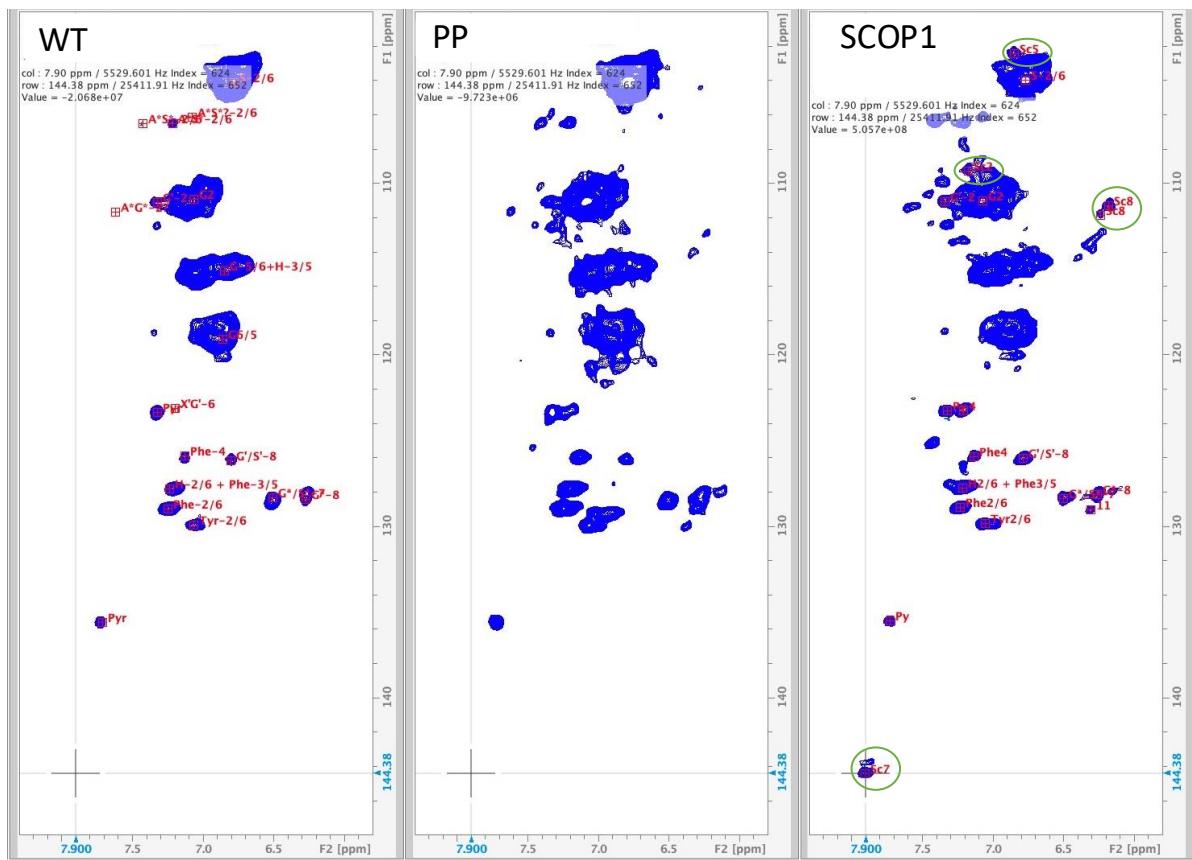
<b>Kanamycin Resistance</b> (KR; NPTII)	0	0	0	9081	16846	8335
--	---	---	---	------	-------	------

Tissu	WT		PP			
	Leaves					
Sample	1	2	3	4	5	6
<b>Umbelliferone Synthase</b> (US; PsDiox)	0	0	0	5747	9563	6039
<b>Demethylsuberosin Synthase</b> (DS; PsPT1)	0	0	0	5951	9591	5271
<b>Marmesin Synthase</b> (MS ; FcCYP76F112)	0	0	0	1996	2487	2572
<b>Psoralen Synthase</b> (PS; PsCYP71AJ3)	0	0	0	7703	12702	7923
<b>Kanamycin Resistance</b> (KR; NPTII)	5	3	0	7969	14413	10055

## 2-4- Impact de la voie du psoralène sur la composition de la lignine (données complémentaires de l'article 1)

L'analyse métabolomique de différents organes de la lignée PP a montré une accumulation de différentes molécules de la famille des coumarines, en particulier de la scopolétine, qui est peu présente et non quantifiable dans le contrôle WT. Au cours de ma deuxième année de thèse, un article sur la capacité de l'Arabette à intégrer la scopolétine dans la lignine a été publié (Hoengenaert *et al.*, 2022). Cet article démontre que la localisation subcellulaire et l'accumulation locale de scopolétine sont des facteurs importants pour permettre ou non son intégration dans la lignine. Afin d'explorer une éventuelle intégration de scopolétine dans la lignine dans la lignée PP, j'ai initié ainsi une collaboration avec Lennart Hoengenaert, Wout Boerjan (VIB, Ghent, Belgique) et John Ralph (Wisconsin-Madison University, USA), auteurs de l'article concernant *A. thaliana*.

Dans le cadre de cette collaboration, nous avons pu montrer que la scopolétine était bien accumulée dans la lignine de tiges de la lignée PP par une approche de GC-MS ciblée (**Figure 3 de l'article**). D'autre part, une analyse RMN sur des extraits de paroi cellulaire de tige de la lignée PP comparativement à la lignée WT a été menée par John Ralph aux USA. Par cette approche, la scopolétine n'a pas pu être détectée dans la lignine (**Figure 22**). Une seconde approche complémentaire d'analyse RMN plus poussée (54h d'acquisition au lieu de 6h) et centrée uniquement sur les cycles aromatiques de la lignine pour limiter les bruits de fond n'a pas non plus permis de visualiser la scopolétine (données non présentées). Ces résultats suggèrent que la scopolétine représente au maximum 0,1% de la composition de la lignine et par conséquent que sa quantité dans nos échantillons est en dessous du seuil de détection. Ce constat est en cohérence avec notre quantification en CG-SM de la scopolétine ( $0,017 \pm 0,009\%$  par unité de lignine). La RMN est une approche moins sensible que la GC-MS pour de faibles quantités de matrices à analyser. Par conséquent, nous n'avons pas pu décrire les différents types de liaisons que la scopolétine a pu établir avec la lignine dans les tiges de la lignée PP. Les données CG-SM suffisent néanmoins à prouver la capacité de la scopolétine à s'intégrer dans la lignine et à se retrouver dans le compartiment pariétal.



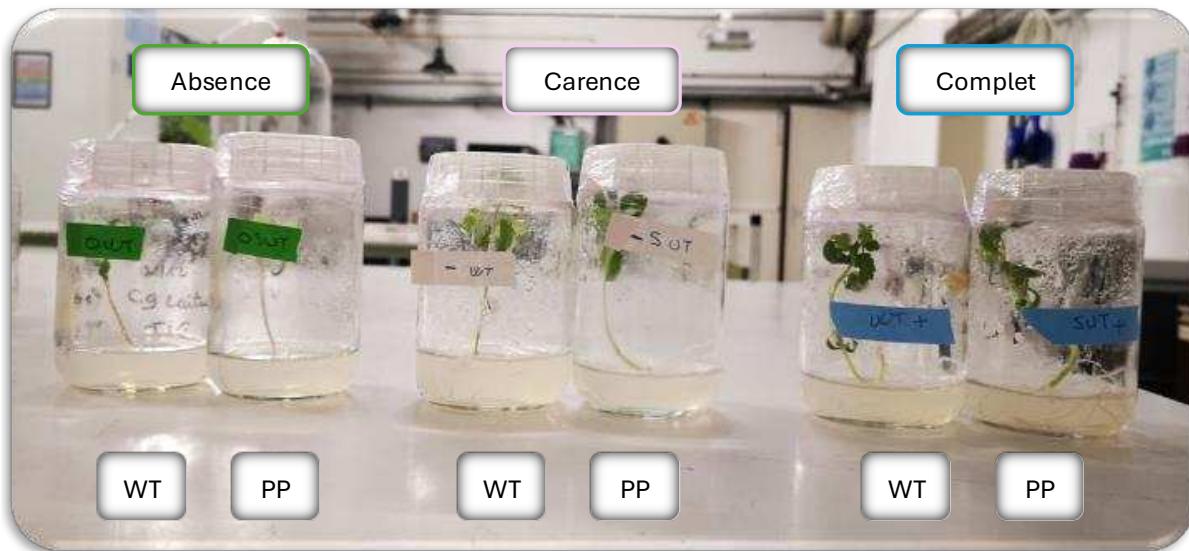
**Figure 22: Analyse par Résonnance Magnétique Nucléaire (RMN) de la composition de la lignine de tige des lignées Wild-Type (WT) et “Psoralen Pathway” (PP) de *Solanum lycopersicum* comparativement à la lignée accumulatrice de scopolétine SCOP1 chez *Arabidopsis thaliana* (Spectres RMN HSQC (Heteronuclear Single Quantum Coherence)  $^1\text{H}$ - $^{13}\text{C}$  à courte portée, pour un temps d’acquisition de 6h). La solution de lignine analysée provient d’une déstructuration par digestion aux cellulases de 20mg de poudre de parois cellulaires de tiges lavées de tout composé soluble et resuspendu dans du DMSO. La lignée SCOP1 est utilisée comme contrôle positif (d’après Hoengenaert et al., 2022). Les pics Sc retrouvés dans la lignée SCOP1 témoignent de l’établissement de liaisons entre la scopolétine (Sc) et la lignine (ronds verts). Abréviation : ppm, parties par million (Hz/MHz).**

En conclusion de cette partie, par des approches couplées de biochimie séparative et de chimie analytique de la lignine, j'ai pu identifier dans la lignée PP (i) l'accumulation de scopolétine dans les parois cellulaires et (ii) proposer son intégration même à faible concentration dans la structure de la lignine en utilisant une approche de GC-MS. Par contre, l'accumulation est insuffisante pour permettre de décrire comment la scopolétine s'associerait structuralement à la lignine dans la lignée PP.

## **2-5- Impact de la « surproduction » de coumarines dans la lignée PP sur la résilience à la carence ou à l'absence de fer (données complémentaires de l'article 1)**

Certaines coumarines possèdent un groupement catéchol qui peut, en cas de carence et sur substrat légèrement alcalin, favoriser la captation de fer dans le sol en agissant comme des sidérophores (Rajniak *et al.*, 2018; Siwinska *et al.*, 2018; Tsai *et al.*, 2018). L'esculétine est un exemple de coumarines catécholiques qui peut favoriser la résilience à une carence en fer (Robe, Izquierdo, *et al.*, 2021; Vanholme *et al.*, 2019). La lignée PP accumule spécifiquement de la scopolétine, et dans une moindre mesure de l'esculétine et de l'esculine principalement dans les parties supérieures (**Figure 2 de l'article**). Ce profil en coumarines est identique à celui des racines d'*A. thaliana* en cas de carence en fer (Schmid *et al.*, 2014).

Afin d'explorer un rôle physiologique de la scopolétine, de l'esculétine et de l'esculine et évaluer leur impact sur l'homéostasie du fer, j'ai cultivé les plantes en conditions *in vitro* sur des substrats différents (**Figure 23**).



**Figure 23: Exemple de cultures en condition *in vitro* d'individus des génotypes Wild-Type (WT) et « Psoralen Pathway » (PP) en milieu complet (bleu), carencé en fer (rose) ou sans fer (vert).**

La carence en fer chez les plantes induit généralement une chlorose qui démarre par un jaunissement des feuilles (Chaney, 1984). Dans le protocole mis en place, j'ai cultivé des plantes en condition *in vitro* jusqu'à ce qu'elles atteignent une taille d'environ 5 cm. L'ensemble des plantes utilisées, qu'elles soient WT ou PP, avait un phénotype identique au moment de leur transplantation sur un milieu carencé en fer. Après 6 jours de culture sur le milieu de culture modifié, on peut observer l'apparition d'une chlorose chez toutes les plantes placées sur les milieux carencés ou sans fer (**Tableau 2**), tandis qu'aucune plante n'a présenté un jaunissement sur milieu complet. Les premières « nouvelles » feuilles sont apparues entre 6 et 12 jours. On considère qu'à ce moment, les plantules passent d'un état de dépendance aux cotylédons à un état de dépendance au milieu de culture. Les résultats que nous avons obtenus indiquent qu'en condition de carence, les individus WT ont jauni dès les premiers moments de dépendance au milieu (**Tableau 2**). Les plantules sauvages placées sur le milieu carencé sont redevenues vertes environ 18 jours après la première détection des symptômes de chlorose. Concernant les individus PP, le jaunissement est apparu pour la plupart 6 jours après les plantes WT. Ces symptômes ont disparu en moyenne 6 jours avant les WT. À noter que les individus PP n'ont subi aucune chlorose et sont restés verts durant toute l'expérience. En absence

complète de fer dans le milieu, tous les individus (à l'exception d'un) sont devenus chlorosés sans reverdissement par la suite (**Tableau 2**).

**Tableau 1: Cinétique de l'effet du génotype "Psoralen Pathway" (PP) comparativement au génotype Wild-Type (WT) sur la résilience à la carence ou l'absence de fer en conditions de culture *in vitro*.** (a) Nombre d'individus présentant une coloration jaunie des tissus foliaires. La coloration des cases indique la tendance des tissus à rester vert, jaunir ou reverdir. Éléments clés du milieu de culture : pH 7.3 (légèrement alcalin) ; concentration en fer des milieux complet (40  $\mu\text{M Fe}^{2+}$ ), carencé (10  $\mu\text{M Fe}^{2+}$ ) et absent (0  $\mu\text{M Fe}^{2+}$ ) de fer d'après (Siwinska *et al.*, 2018). Échantillonnage initial: n=7-8 par génotype et par condition de culture.

		J0	J+6	J+12	J+18	J+24	J+30	J+36	Phénotype foliaire
Absence	WT	0/0	0/0	2/2	3/3	3/3	3/3	3/3	
	PP	0/0	0/0	6/6	7/8	7/8	7/7	7/7	
Carence	WT	0/0	0/0	2/2	2/3	3/3	0/2	0/2	
	PP	0/0	0/0	2/7	4/7	1/7	0/7	0/7	
Complet	WT	0/0	0/0	0/2	0/5	0/5	0/5	0/5	
	PP	0/0	0/0	0/7	0/7	0/7	0/7	0/7	

Jaunis
Vert

L'ensemble de ces résultats suggère que le génotype PP, au travers une accumulation supérieure en scopolétine que chez les WT, pourrait avoir une meilleure résilience à une carence en fer. Toutefois, l'apparition de symptômes de chlorose dans la lignée PP sur un milieu dépourvu de fer indique également que la présence de coumarines ne serait pas suffisante pour faire face à la carence en fer.

En conclusion de cette partie, par une approche de culture *in vitro*, j'ai pu identifier que la lignée PP était plus résiliente à une carence en fer. Une résilience qui prend la forme à la fois (i) d'une limitation des effets de la carence et (ii) d'une adaptation plus rapide des individus carencés. À la vue de la précédente caractérisation du métabolome de la lignée PP, j'émets l'hypothèse que cette résilience soit notamment due à l'accumulation de coumarines.

### 3-Discussion et perspectives

#### A- Quels rôles physiologiques et écophysiologiques pour les coumarines et les FCs chez la tomate ?

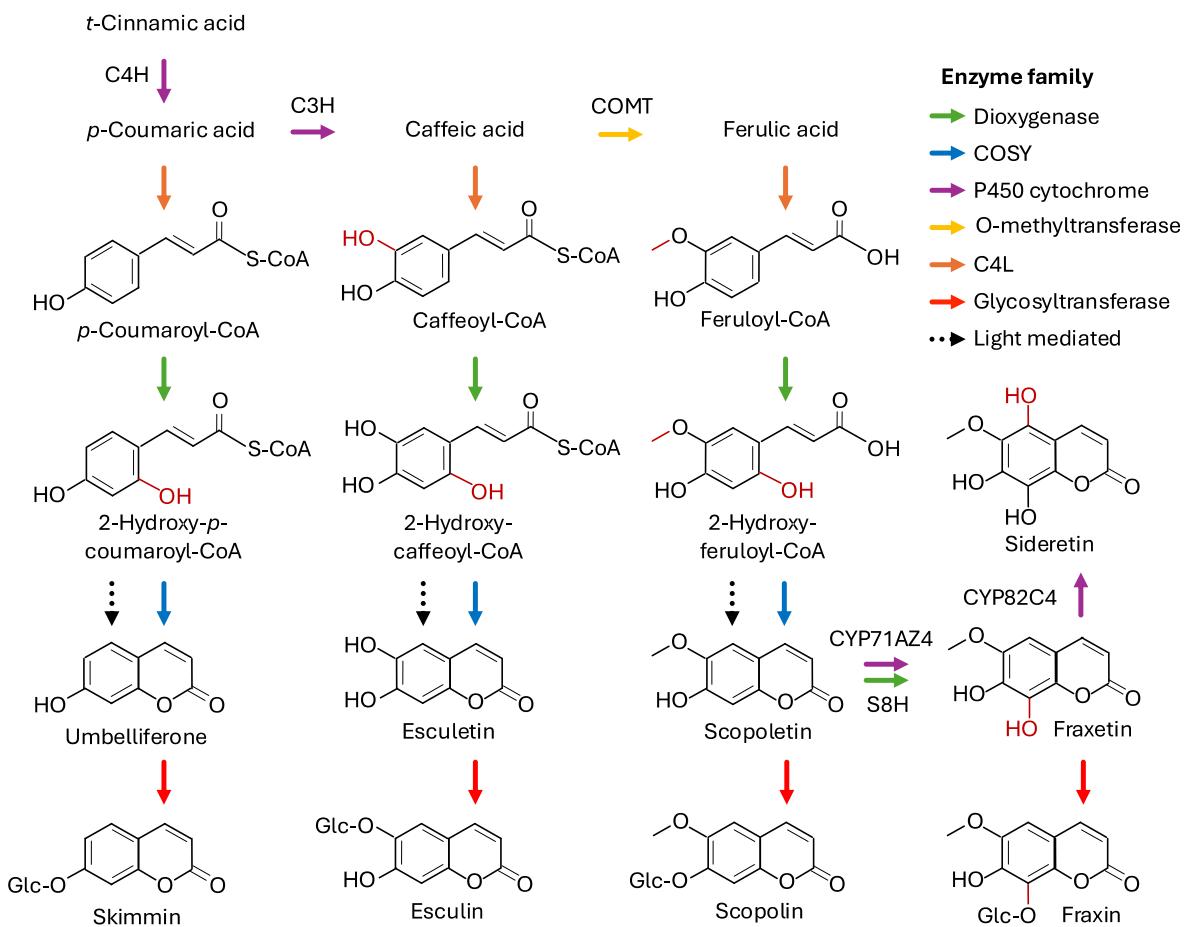
La physiologie est une science interdisciplinaire qui intègre l'organisation, le fonctionnement et le rôle des composants d'un organisme. Pour comprendre le rôle physiologique d'un MS, il est essentiel de prendre en compte toutes ces dimensions de manière simultanée (Weng *et al.*, 2021). L'étude de l'impact d'un MS sur la physiologie d'une plante doit pouvoir être découplé des effets écophysiologiques qu'il peut induire en réponse à l'environnement. Dans les travaux réalisés dans le cadre de cette thèse, j'ai utilisé la voie du psoralène pour approfondir la compréhension du rôle des coumarines sur la physiologie et l'écophysiologie de la tomate. Cette section discute principalement des aspects complémentaires à l'article (partie 4).

##### a- Distribution phytochimique et spatiale des coumarines dans la lignée de tomate transformée

Pour mieux comprendre l'impact potentiel de l'accumulation de coumarines dans les tomates transformées, il est nécessaire d'étudier leur distribution. La lignée PP accumule constitutivement de 1 à 5 µg de scopolétine par gramme de matière sèche, une quantité faible par rapport à d'autres MS majeurs comme la rutine (1200 µg/g de matière sèche) ou l'acide chlorogénique (1500 µg/g de matière sèche) dans les feuilles de tomates (Larbat *et al.*, 2014). Sur le plan phytochimique, la lignée PP synthétise la scopolétine, l'esculétine, l'esculine ainsi que d'autres coumarines non caractérisées (**Figure 2 de l'article**) (**Figure 24**). L'accumulation de coumarines varie entre les tissus de la tomate. Ainsi, la scopolétine est plutôt accumulée dans les tiges, tandis que l'esculétine et l'esculine le sont plutôt dans les feuilles et les extrémités racinaires (**Figure 2 de l'article**). Le suivi de l'expression des transgènes dans les différents tissus de la lignée PP a montré que l'ensemble des gènes était exprimé de façon stable dans la plante (**Annexe S3 de l'article**), ce qui suggère que le profil d'accumulation en coumarines dépend de facteurs physiologiques propres à l'environnement cellulaire des

tissus autres que l'expression des transgènes. Cette distribution pourrait ainsi dépendre d'une expression locale et constitutive de gènes dédiés à la biosynthèse et au stockage de coumarines comme c'est le cas dans les racines d'*A. thaliana* (Robe, Conejero, et al., 2021). À titre d'exemple, la distribution spatiale de coumarines pourrait dépendre de variations de disponibilité en précurseurs (e.g. *p*-coumaroyl-CoA, cafféoyl-CoA, feruloyl-CoA, umbelliférone), en cofacteurs (e.g.  $\alpha$ -cétoglutarate, DMAPP) ou en coenzymes (e.g. acyl-CoA) entre les tissus. Je note que la disponibilité en précurseurs est d'autant plus importante qu'elle impacte directement la capacité de la plante à produire du psoralène chez le tabac. Ainsi, la supplémentation en *p*-coumarate permet parfois la production de psoralène, tandis qu'une supplémentation en umbelliférone le permet systématiquement. Ces données permettent d'appuyer le fait que l'abondance de *p*-coumaroyl-CoA pourrait jouer un rôle important dans la distribution des coumarines dans la lignée PP, au même titre que l'abondance et la distribution d'acyl-CoA et de la 4CL.

Plus précisément, à l'échelle subcellulaire, chez l'arabette, la scopolétine peut être compartimentée dans la paroi des cellules racinaires corticales (Schmid et al., 2014) voire intégrées dans la lignine des tiges (Hoengenaert et al., 2022). De façon similaire dans la lignée PP, la scopolétine est compartimentée dans la paroi et intégrée à la lignine des tiges. D'autre part, l'esculétine est davantage présente sous sa forme glycosylée dans les racines apicales et dans les feuilles que dans les tiges. Ce profil d'accumulation est cohérent avec sa capacité à être accumulée dans les vacuoles des cellules corticales des racines de l'arabette (Robe, Conejero, et al., 2021), mais dénote quant à sa capacité à être accumulée dans les feuilles. Enfin, le profil d'accumulation similaire entre l'esculétine et son dérivé glycosylé l'esculine suggère que l'accumulation d'esculine soit notamment une conséquence locale de la production de l'esculétine. La localisation subcellulaire de l'esculétine n'est pas connue chez d'autres plantes, excepté sa capacité à être déglycosylée dans le cortex des racines pour être excrétée dans le sol (Robe, Conejero, et al., 2021). Une analyse cinétique des profils d'accumulation de coumarines entre les tissus pourrait éventuellement permettre de mieux comprendre leurs dynamiques chez la tomate.



**Figure 24: Schéma simplifié de la voie de biosynthèse des coumarines décrivant les étapes à partir des dérivés de l'acide *t*-cinnamique.** Les familles d'enzymes sont représentées par des flèches colorées. Les étapes biocatalytiques sont détaillées dans : Bourgaud *et al.*, 2014, Rajniak *et al.*, 2018, Krieger *et al.*, 2018, Vanholme *et al.*, 2019 et Robe, Izquierdo *et al.*, 2021. Abréviations : C4H, Cinnamate 4-Hydroxylase ; C3H, 4-Coumarate 3-Hydroxylase ; 4CL, 4-Coumarate:CoA Ligase ; COSY, Coumarin synthase ; S8H : Scopoletin 8-Hydroxylase.

### b- Régulations physiologiques médiées par l'accumulation de coumarines dans les tomates transformées

Les régulations physiologiques induites par les coumarines sont encore peu décrites. Actuellement, seule une analyse protéomique décrit la régulation de quelques protéines face au stress oxydatif induit par la présence de la coumarine chez *A. thaliana* (Araniti *et al.*, 2017). L'étude transcriptomique de la lignée PP montre que les coumarines pourraient induire : (i) un stress xénobiotique conduisant à un stress oxydatif auxquels les tomates peuvent faire face, (ii) une photoinhibition en réponse au stress oxydatif; (iii)

un potentiel stress hormétique auxinique léger globalement dans la plante; (iv) une augmentation de la captation et le stockage d'azote, et (v) une potentielle régulation de TORC pour la régulation des ressources via *SIP2C24*. Cette analyse ouvre ainsi la porte à une meilleure compréhension de l'impact de l'accumulation de coumarines sur la physiologie des plantes. Cette analyse transcriptomique et l'analyse protéomique de Araniti et al, (2017) mettent en lumière la régulation de processus biologiques similaires, dont notamment l'induction d'un stress oxydatif et d'une photoinhibition ce qui appuie une éventuelle fonction commune des coumarines chez les plantes. Cependant, je note que les voies de régulation qui ont amené au phénotype de la lignée PP sont pour grande partie non mentionnées dans la littérature. Ceci malgré que ces familles de gènes soient plutôt bien caractérisées chez la tomate (e.g. R2R3Myb, JAZ, bHLH, etc.) comme décrit pour quelques exemples précis par (Davis & Glover, 2024; R. Saito et al., 2021; Shuai et al., 2022). Ce constat suggère que la régulation transcriptionnelle dans la lignée PP pourrait révéler des réseaux de régulation d'une partie encore inconnue du génome de la tomate. Ces régulations nouvelles pourraient contribuer à compléter la compréhension du transcriptome au-delà notamment des processus de mûrissement des fruits et la mise en place de défenses qui sont les mieux référencés chez la tomate (Campos et al., 2022; H. Li et al., 2023). Plus précisément, la régulation spécifique à la lignée PP comparativement à la lignée WT pourrait aider dans la compréhension des processus physiologiques hormétique auxine-like et du stress xénobiotique. À noter que les effets hormétiques auxine-like correspondent à des effets similaires à ceux d'une surproduction d'auxine acide indole-3-acétique (Graña et al., 2017), et que le stress xénobiotique correspond à des effets délétères de composés non constitutifs à l'organisme hôte (X. Li et al., 2007). Toutefois, bien que les effets identifiés soient cohérents avec des fonctions connues pour les coumarines, plusieurs éléments viennent relativiser ces constats. L'obtention d'une seule lignée transgénique via une méthode d'insertion non ciblée dans le génome de la tomate, ne permet pas de s'affranchir des effets pléiotropes<sup>186</sup> potentiels associés à l'insertion du transgène à un ou plusieurs locus, comme synthétisé par (Chu & Agapito-Tenfen, 2022). La délimitation de la ou des zones d'insertions du transgène dans la lignée PP et le séquençage partiel

---

<sup>186</sup> Effets pléiotropiques : Effet d'un gène sur plusieurs traits différents d'un organisme.

voir total du génome pourrait permettre de mieux déceler l'impact de l'insertion du transgène sur son intégrité et les conséquences potentielles notamment sur les voies de régulation.

Actuellement, la description du rôle physiologique des coumarines chez les plantes s'appuie sur peu d'éléments bien qu'elle propose de nombreux rôles : (i) défense sélective face à l'herbivorie mais pas contre les polliniseurs, (ii) régulation de la rhizosphère bactérienne et élicitation du MS médié par le JA, (iii) facilitation de la mycorhization, (iv) amélioration de la résilience à une carence en fer, (v) modification des flux de fer des racines vers les feuilles, (vi) déstabilisation de l'organisation du cytosquelette, (vii) modification de la gestion des ressources pour la croissance et le développement, (viii) déstabilisation du statut redox, etc. (Bruno *et al.*, 2021; Cosme *et al.*, 2021; Graña *et al.*, 2017; Rajniak, Giehl, Chang, Murgia, Von Wirén, *et al.*, 2018; Robe, Izquierdo, *et al.*, 2021; Robe, Stassen, *et al.*, 2021; Siwinska *et al.*, 2018; Tsai *et al.*, 2018; Vanholme *et al.*, 2019; Vismans *et al.*, 2022; Z. Yan *et al.*, 2016; Zhou *et al.*, 2023). La tomate produit constitutivement peu de coumarines, et ces composés allélochimiques peuvent induire des effets délétères chez les plantes comme synthétisé par (Razavi, 2010). Ainsi, les coumarines pourraient avoir des effets importants sur la physiologie de la lignée PP. En ce sens, j'identifie que la concentration de coumarines dans la lignée PP suffit à impacter négativement la croissance (perte de 21% de la biomasse sèche dans la plante entière) tout en préservant un phénotype sain et en favorisant la résilience intrinsèque des tomates à une carence en fer. Ces éléments suggèrent que les coumarines, à la concentration mesurée dans la lignée PP, induisent un coût physiologique que la tomate peut supporter voire même qui peut lui être bénéfique dans certaines conditions. Les données transcriptomiques et de croissance suggèrent plusieurs voies potentiellement impactées par la présence de coumarines comme l'accumulation de ressources pour la croissance, la déstabilisation du statut redox, la déstabilisation de l'organisation du cytosquelette et l'amélioration de la résilience à une carence en fer.

La lignée PP accumule plus d'azote que la lignée WT ce qui a pour conséquence une forte augmentation du ratio C/N dans toute la plante (**Tableau II de l'article**). Jusqu'alors, peu

de travaux ont établi un lien entre coumarines et azote à l'exception de l'étude de (Abenavoli *et al.*, 2001) montrant une amélioration de l'acquisition d'azote chez le blé dur et le maïs en présence de faibles et de fortes concentrations en coumarine (Abenavoli *et al.*, 2001 ; Lupini *et al.*, 2018). Une analyse approfondie des processus pouvant conduire à cette plus forte accumulation d'azote pourrait être conduite en s'appuyant sur les compétences développées par le LAE et ses partenaires (INRAe PSH) en matière de suivi de l'acquisition d'azote chez la tomate. L'unité PSH a notamment développé un système automatisé permettant le suivi en temps réel de la consommation en azote dans une solution nutritive (système TOTOMATIX) (Adamowicz *et al.*, 2012). L'utilisation de ce système permettrait de quantifier l'acquisition de nitrate par la lignée PP en comparaison de WT, ce qui permettrait de savoir si les différences d'accumulation en azote sont plutôt liées à des différences d'acquisition ou bien de remobilisation. D'un point de vue mécaniste, même si peu de pistes sont ressorties de l'analyse transcriptomique sur le lien entre coumarine et accumulation d'azote, une hypothèse serait que ce lien passe par les propriétés auxine-like proposées pour les coumarines. L'auxine est, en effet, une phytohormone étroitement liée à la gestion des flux d'azote comme synthétisé par (Kiba *et al.*, 2011). En ce sens, mes données transcriptomiques permettent de référer l'expression positive et négative de gènes pouvant être liés aux transports d'azote (NRT1 : Nitrate Transporter 1; ammonium transporter; Amino acid transporter; Cyclic nucleotide gated channel 1, etc.) (**Table S1 de l'article**).

L'implication des coumarines dans les processus d'acquisition du fer a été rapportée chez *A. thaliana* il y a une dizaine d'années et suscite encore d'actives recherches dans la compréhension des mécanismes sous-jacents (Rajniak *et al.*, 2018; Siwinska *et al.*, 2018; Tsai *et al.*, 2018; Vanholme *et al.*, 2019). Les coumarines catécholiques comme l'esculétine agissent comme des phytosidérophores. La résilience observée de la lignée PP pour la carence en fer, suggère que l'accumulation en coumarines favorise la captation de fer chez la tomate. Je souhaite porter une attention particulière sur la captation potentielle du fer dans la lignée PP. Une première analyse consisterait à mesurer l'accumulation en Fer dans les tissus de la lignée PP comparativement au contrôle WT. Ces analyses n'ont pas pu être réalisées au cours de ma thèse faute de temps. L'accumulation importante de fer chez la tomate peut induire un stress oxydatif

sans porter atteinte à l'efficacité photosynthétique des feuilles (Chu & Agapito-Tenfen, 2022), ce qui est similaire aux effets retrouvés dans la lignée PP. Une analyse approfondie des niveaux de stockage du fer dans les tissus de la lignée PP pourrait permettre de se rendre compte de son éventuel impact sur la physiologie. De plus, la production de coumarines dans les racines pourrait éventuellement suggérer qu'elles puissent être excrétées dans le sol. Néanmoins, l'esculétine qui est la seule coumarine catécholique identifiée comme produite dans les racines de la lignée PP n'a pas été retrouvée dans le sol (**Figure 2 de l'article**). Il est cependant possible que la méthode d'extraction utilisée ne soit pas la plus adaptée pour l'extraction de coumarines dans du terreau. À titre d'exemple la scopolétine est fortement adsorbée dans le sol argileux (Galán-Pérez *et al.*, 2022), mais aucune littérature ne décrit sa capacité de sorption dans le terreau. Il est également possible que parmi le profil inconnu de métabolites dans les racines et le sol soient présentes d'autres coumarines catécholiques non identifiées. Une analyse plus précise du profil en coumarines dans le sol de la lignée PP pourrait permettre de mieux appréhender le rôle écophysiologique des coumarines dans la gestion des ressources de la tomate.

D'autre part, la production induite de la coumarine a été identifiée comme capable de modifier le microbiome racinaire chez l'Arabette en faveur d'une meilleure résistance de sa descendance après un stress biotique (Vismans *et al.*, 2022). On peut ainsi se poser la question des conséquences d'une accumulation accrue de coumarines sur le microbiome associé à la lignée PP. L'accumulation de scopolétine notamment dans les parties foliaires nous a amenés à considérer la lignée PP comme un modèle intéressant pour décrire le rôle potentiel de la scopolétine sur le microbiote épiphyte et notamment endoépiphyte de la tomate. L'analyse comparative, par métabarcoding, des populations microbiennes associées à la lignée PP et au WT a été conduite à la fin de ma thèse en collaboration avec la Pr. Sophie Slezack du LAE. Ces résultats sont en cours d'acquisition et constituent une perspective dans la meilleure compréhension du rôle écophysiologique des coumarines.



## 4-Conclusions du chapitre 2

En conclusion, dans ce chapitre j'ai présenté l'impact de l'introduction de 4 gènes de la voie du psoralène dans le génome de la tomate sur sa physiologie. Par des approches d'expression transitoire et de bioingénierie métabolique, je montre que cette construction est fonctionnelle et permet la production de psoralène dans *Nicotiana benthamiana* et *Solanum lycopersicum*. Je suggère néanmoins que pour que cette production soit efficace, la plante doit développer des adaptations physiologiques, soit pour accumuler du psoralène, soit pour limiter sa toxicité ou celle de ces intermédiaires. Je montre également que l'accumulation de coumarines est corrélée avec la régulation de l'expression de nombreux gènes. Les gènes dont l'expression est régulée semblent être impliqués dans l'adaptation de la tomate au déploiement d'un stress xénobiotique et notamment oxydant. Ils semblent également faciliter la mise en place d'un effet hormétique auxinique. Ces deux processus physiologiques sont largement incompris dans le cadre d'une réponse liée à la présence de coumarines. Par une analyse croisée d'approches de mesures morphophysiologiques<sup>187</sup>, d'analyses métabolomiques et transcriptomiques, j'apporte de nouveaux éléments quant aux rôles potentiels des coumarines sur la physiologie et la gestion des ressources des tomates. Enfin, via l'étude de la lignée PP, je soulève plusieurs questions qui pourraient orienter la meilleure compréhension des rôles physiologique des coumarines et des FCs.

---

<sup>187</sup> Morphophysiologie : Étude de la relation entre la structure et la fonction des organismes vivants.



# Chapitre 3 : Mieux comprendre l'évolution de la voie de biosynthèse des furocouumarines chez *Ficus carica*

## 3.1- Contexte scientifique et stratégie de recherche

Mon second axe de recherche s'inscrit dans la continuité des travaux menés au LAE sur la compréhension du métabolisme spécialisé des plantes et notamment sur le recrutement des gènes, et donc des enzymes, à l'origine des voies de biosynthèse. Divers travaux réalisés sur la caractérisation moléculaire de la voie de biosynthèse des FCs ont permis de montrer que des enzymes impliquées dans la voie de biosynthèse et ayant la même fonction catalytique pouvaient appartenir à des sous-familles d'enzymes très différentes chez des espèces évolutivement éloignées. Ces travaux suggèrent que la voie des FCs soit apparue indépendamment à plusieurs reprises chez les angiospermes *via* une évolution convergente (Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020; Villard *et al.*, 2021). L'hypothèse basée sur le principe « même famille, même fonction » est, de fait, mise à mal et, identifier les gènes impliqués dans la voie de biosynthèse de ce type de molécules peut donc se révéler être particulièrement compliquée.

Dans le cadre de sa thèse, Cloé Villard (Villard, 2020) a montré que la fonction marmésine synthase était apparue suite à une diversification récente du CYPome (ensemble des P450s) restreint au clan 71 chez les Moracées (Villard *et al.*, 2021). Ces travaux suggèrent que les FCs sont produites au sein des espèces du genre *Ficus*, mais pas du genre *Morus*. Ainsi, à partir d'une approche de phylogénie comparative entre une partie du CYPome de *F. carica* et de *Morus notabilis*, Cloé Villard a suggéré que plusieurs sous-familles de P450s pouvaient potentiellement être impliquées dans la biosynthèse de FCs chez *F. carica*.

Dans ce chapitre, nous avons mis en œuvre une stratégie basée sur la comparaison des CYPomes pour cibler certains P450s et explorer leur fonction potentielle dans la

synthèse des FCs. Les résultats obtenus ont confirmé que certains de ces P450s étaient impliqués dans la synthèse de FCs et que d'autres jouaient un rôle dans la production de coumarines. Ces résultats m'ont ainsi amené à m'interroger sur **les liens évolutifs entre la voie de biosynthèse des coumarines et des FCs ?**

Ainsi, au travers d'approches de génomique, transcriptomique et métabolomique comparatives, j'ai tenté de mieux comprendre comment la voie de biosynthèse des FCs s'est diversifiée chez les Moracées et plus largement chez les angiospermes. Les résultats obtenus m'ont amené à réaliser (i) la caractérisation fonctionnelle de P450s appartenant à la sous-famille des CYP71Bs chez le figuier et (ii) leur évolution structurale à l'origine de la sélectivité de substrats coumarines et FCs. Les premières données font l'objet d'un article soumis au journal *New Phytologist*. Ces travaux sont complétés par la caractérisation fonctionnelle de P450s des sous-familles CYP71ANs et CYP71As impliquées dans la voie des FCs. Les approches mises en œuvre me permettent de proposer de nouveaux gènes candidats à l'échelle du CYPome du figuier.

## 3.2- Résultats

### 3.2.1- Les CYP71Bs : Outils pour mieux comprendre la voie de biosynthèse des FCs et son évolution

#### A- Étude des CYP71Bs

La sous-famille CYP71Bs est la plus fortement dupliquée dans le clan 71 chez *F. carica* comparativement à *M. notabilis*. J'ai sélectionné et cloné 3 séquences codantes de cette sous-famille de P450s. Leur caractérisation fonctionnelle m'a amené à décrire les fonctions 5-hydroxyxanthotoxine synthase (5OHXS) et umbelliférone hydroxylase (UH). La fonction 5OHXS est l'étape la plus décrite de la voie des FCs parmi les angiospermes (Krieger *et al.*, 2018; Kruse *et al.*, 2008; Limones-Mendez *et al.*, 2020). J'ai utilisé ces différentes isoformes pour mieux comprendre les processus évolutifs qui ont pu conduire à l'apparition de cette fonction 5OHXS.

Le développement de l'outil Alphafold a révolutionné l'analyse structurale de protéines (Jumper *et al.*, 2021). J'ai utilisé cet algorithme pour mieux comprendre les origines structurales de la fonction 5OHXS. Des travaux réalisés en collaboration avec Janet Thornton et Rashmi Kumari (European Molecular Biology Laboratory - European Bioinformatic Institute) ont permis d'identifier le degré de similarité structurale des 5OHXS décrites dans la littérature et de les comparer avec leurs plus proches paralogues issus de 4 familles botaniques éloignées. Les résultats de ces travaux sont décrits dans un article prépublié BioRXIV (<https://doi.org/10.1101/2024.08.19.608558>) et soumis au journal *New Phytologist*.

### **3.2.2 - Using lineage-specific patterns to understand convergence of enzymatic functions led to the identification of Moraceae-specific P450s involved in furanocoumarin biosynthesis (Article 2)**

Alexandre Bouillé<sup>1</sup>, Rashmi Kumari<sup>2</sup>, Alexandre Olry<sup>1</sup>, Clément Charles<sup>1</sup>, David R Nelson<sup>3</sup>, Romain Larbat<sup>4</sup>, Janet Thornton<sup>2</sup>, Cloé Villard<sup>5</sup>, Alain Hehn<sup>1</sup>

1- Université de Lorraine, INRAE, LAE, 54000 Nancy, FR

2-European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, CB10 1SD Cambridge, UK

3- Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis TN, USA 38163

4- Institut Agro, University of Angers, INRAE, IRHS, SFR QUASAV, F-49000 Angers, FR

5- Biosystematics Group, Wageningen University and Research, Wageningen 6708 PB, the Netherlands

#### **Summary**

- Specialized metabolites are molecules involved in plants interaction with their environment. Elucidating their biosynthetic pathways is a challenging but rewarding task, leading to societal applications and ecological insights. Furanocoumarins emerged multiple times in Angiosperms, raising the question of how different enzymes evolved into catalyzing identical reactions.
- To identify enzymes producing lineage-specific metabolites, an evolutionary-based approach was developed and applied to furanocoumarin biosynthesis in *Ficus carica* (Moraceae). This led to the characterization of CYP71B129-131a, three P450 enzymes whose evolution of the function was investigated using phylogenetics, structural comparisons and site-directed mutagenesis.

- CYP71B129 and CYP71B130,131a were found to hydroxylate umbelliferone (coumarin) and xanthotoxin (furanocoumarin), respectively. Results suggest that CYP71Bs xanthotoxin hydroxylase activity results from duplications and functional divergence of umbelliferone hydroxylase genes. Structural comparisons highlighted an amino acid affecting CYP71Bs substrate specificity, which may play a key role in allowing xanthotoxin hydroxylation in several P450 subfamilies.
- CYP71B130-131a characterization validates the proposed enzyme-discovery approach, which can be applied to different pathways and help to avoid the classic bottlenecks of specialized metabolism elucidation. The CYP71Bs also exemplify how furanocoumarin-biosynthetic enzymes can stem from coumarin-biosynthetic ones and provides insights into the molecular mechanisms underlying the multiple emergences of xanthotoxin hydroxylation in distant P450 subfamilies.

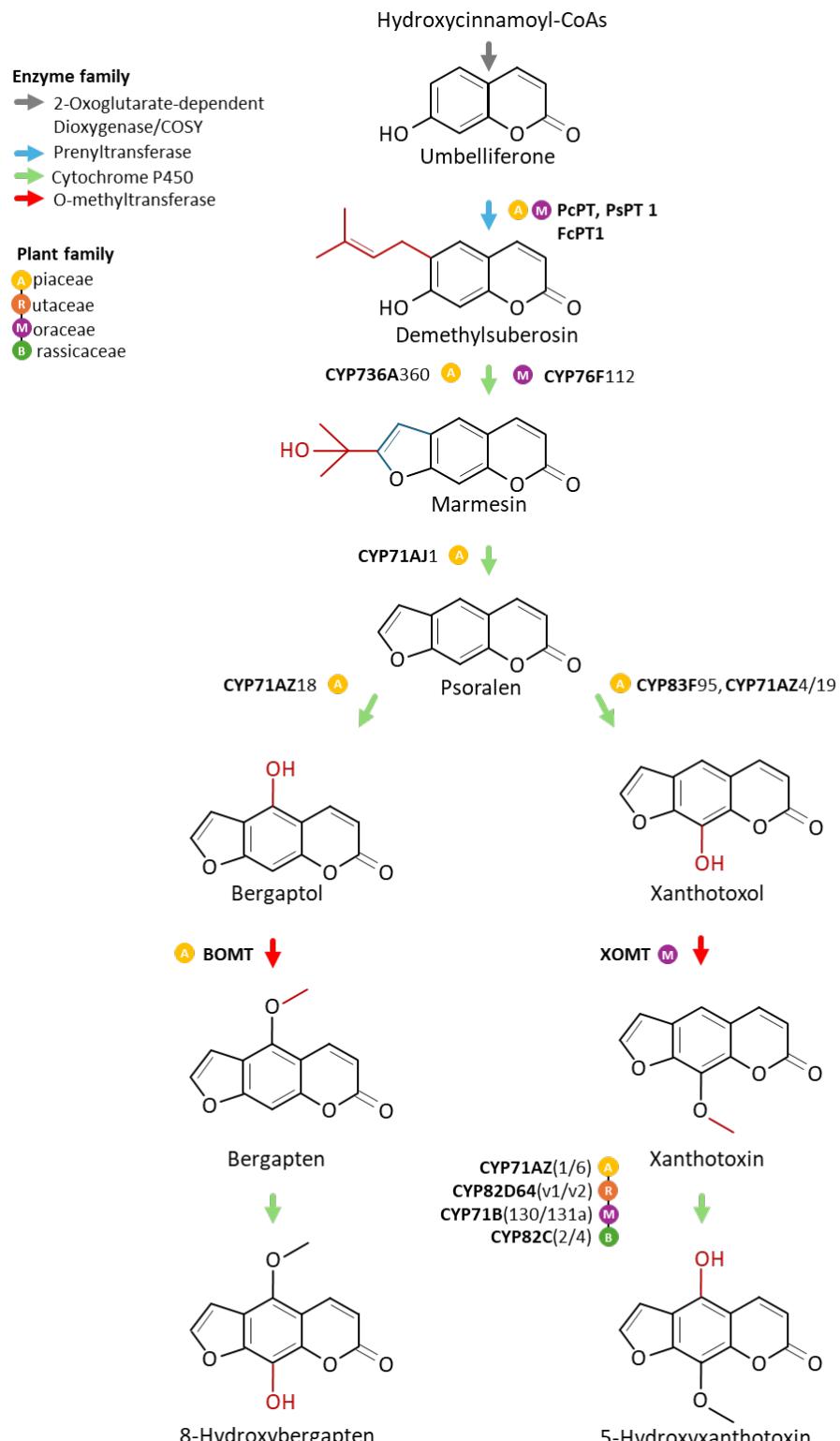
**Keywords:** Coumarins, cytochrome P450, enzyme functional convergence, furanocoumarins, Moraceae, specialized metabolism.

## Introduction

To adapt to their environment, plants have evolved various strategies including a large metabolic expansion which led to the production of myriads of molecules called plant specialized metabolites (PSMs) (Ono & Murata, 2023). Overall, the plant kingdom produces millions of structurally different PSMs, and each plant lineage possesses a unique metabolic profile. Remarkably, distant plants can produce similar PSMs, but rely on different enzymes to synthesize them. This phenomenon of convergent evolution towards an identical function is very common in specialized metabolism, making the biosynthesis of a given PSM potentially unique in each plant (Pichersky & Lewinsohn, 2011). Hence, PSM biosynthesis pathways are generally challenging to elucidate. To identify enzymes involved in PSM biosynthesis, various approaches were developed, relying on strategies such as sequence homology search, transcriptomic analyses, or gene mapping. For instance, it is common to analyze PSMs' accumulation patterns (e.g., tissues distribution, stress inducibility) to identify candidate enzymes displaying correlated gene expression (Ono & Murata, 2023). Such strategies are efficient and have already been successfully used, but they can be limited by factors such as low metabolite or transcription levels or recent divergent and convergent evolution (Pichersky & Lewinsohn, 2011). Promising advanced bioinformatic approaches such as deep learning and large-scale docking are also being developed, but still have limited efficiency (Daniel *et al.*, 2015; Moore *et al.*, 2019).

Furanocoumarins are a class of defensive PSM effective against various bioaggressors such as pathogens and phytophagous insects (Stevenson *et al.*, 2003). Although derived from ubiquitous coumarins, furanocoumarins have only been identified in a few phylogenetically distant plant lineages including the Apiaceae, Rutaceae, Fabaceae and Moraceae (Sarker & Nahar, 2017). Over the past decades, the furanocoumarin biosynthesis pathway has been significantly elucidated in Apiaceous, Rutaceous and Moraceous species (**Fig. 1**). This pathway involves enzymes from the 2-oxoglutarate-dependent dioxygenase (Roselli *et al.*, 2016), prenyltransferase (Munakata *et al.*, 2020), O-methyltransferase (Hehmann, Lukacin, *et al.*, 2004), and cytochrome P450 (P450) families (Villard *et al.*, 2021) (**Fig. 1**). Interestingly, identifying furanocoumarin-

biosynthetic enzymes from different species showed that furanocoumarins are a case of convergent evolution: they emerged independently in distant plant lineages, in which they are synthesized by different enzymes (Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020; Villard *et al.*, 2021).



**Figure 1: Simplified representation of the furanocoumarin biosynthetic pathway in Angiosperms.** All furanocoumarin-biosynthetic enzymes described so far are detailed, together with associated plant families (Hehmann, Lukacin, *et al.*, 2004; Ji *et al.*, 2024; Karamat *et al.*, 2014; Krieger *et al.*, 2018; Kruse *et al.*, 2008; Larbat *et al.*, 2007; Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020; Roselli *et al.*, 2016; Villard *et al.*, 2021; K. Wang *et al.*, 2024). Abbreviations: PcPT, *Petroselinum crispum* prenyltransferase; PsPT1, *Pastinaca sativa* prenyltransferase 1; FcPT1, *Ficus carica* prenyltransferase 1; BOMT, bergaptol O-methyltransferase; XOMT, xanthotoxin O-methyltransferase.

Currently, the most referenced step of the furanocoumarin pathway corresponds to xanthotoxin hydroxylation (Fig. 1), catalyzed by 5-hydroxyxanthotoxin synthases (5OHXSs). All 5OHXSs characterized so far are P450s but belong to different (sub)families. Indeed, 5OHXSs from the Rutaceae (*i.e.*, *Citrus hystrix* and *Citrus paradisi*) are CYP82Ds (Limones-Mendez *et al.*, 2020), while 5OHXS from the Apiaceae (*i.e.*, *Pastinaca sativa*) belong to the CYP71AZs (Krieger *et al.*, 2018). Curiously, CYP82Cs from *Arabidopsis thaliana* (Brassicaceae) also display *in vivo* 5OHXS activity, although *Arabidopsis* does not naturally accumulate furanocoumarins (Kruse *et al.*, 2008).

To pursue the elucidation of furanocoumarin biosynthesis and better understand the evolutionary process underlying the emergence of this pathway, we sought new furanocoumarin-biosynthetic enzymes in *Ficus carica* (Moraceae). To facilitate the identification of enzymes involved in the production of lineage-specific PSMs such as furanocoumarins, we established an evolutionary-based approach. Applying this approach, we compared the CYPomes of two Moraceae: *F. carica*, which produces furanocoumarins, and *Morus notabilis*, which as far as described does not. This led to identifying three *F. carica* CYP71Bs displaying 5OHXS and umbelliferone hydroxylase (UMBH) activity. Phylogenetics and comparative structural biology shed light on a lineage-specific diversification and highlighted an important residue which impacts enzyme activity and may have played a key role in the acquisition of the 5OHXS function.

## Materials and Methods

### Plant material and chemicals

*Ficus carica* plants were provided by INRAE Colmar (France). They were grown for four years at room temperature and under natural light. Standard specimens of phenolic compounds were purchased from various suppliers, as described in **Table S1**.

### **Comparison of the CYP71 clan in *F. carica* and *M. notabilis***

A P450 nucleotide dataset was assembled with *F. carica* and *M. notabilis* CYP71 clan sequences (**Table S2**). *M. notabilis* sequences were retrieved from Ma *et. al.* (2014). *F. carica* sequences were identified by screening the PRJNA565858 reference genome (Usai *et al.*, 2020) through BLASTN searches, using *M. notabilis* genes as queries and an e-value of 1000. Resulting hits were used to isolate full-length putative P450 genes and pseudogenes. Introns were predicted and removed using the NETGENE2 Server (<http://www.cbs.dtu.dk/services/NetGene2/>) (Hebsgaard, 1996). Putative pseudogenes containing frameshifts were altered by the addition of one or two “N” at the site of the indel(s) to recover a functional reading frame and allow subsequent analyses (**Table S2a**). CYP51G sequences were added to the dataset, to constitute the outgroup of the following tree. Sequences received a standardized name from the P450 nomenclature committee (<https://drnelson.uthsc.edu/CytochromeP450.html>, **Fig. S1**). Nucleotide sequences were then translated into proteins and aligned (**Fig. S2**) with MAAFT v.7 (<https://mafft.cbrc.jp/alignment/server/>) relying on default parameters (Katoh *et al.*, 2019; Kuraku *et al.*, 2013). Aligned sequences were used to build a Bayesian inference protein tree with MRBAYES v.3.2.7 (Ronquist *et al.*, 2012; Ronquist & Huelsenbeck, 2003) at CIPRES v.3.3 (Cipres Science Gateway). The Bayesian Markov Chain Monte Carlo analysis was run with the following parameters: nst = mixed, rates = gamma, shape = (all), 2 million generations, two independent runs, four chains, temperature heating 0.05. The tree was drawn with FIGTREE v.1.4.4. Posterior probabilities supporting tree-branches were considered weak below 0.7, strong above 0.9.

### **Cloning of CYP71B129, CYP71B130 and CYP71B131a coding sequences**

Total RNAs were extracted from young *F. carica* leaves and reverse-transcribed into complementary DNAs as described in Villard *et al.* (2021). CYP71B129, CYP71B130 and CYP71B131a coding sequences (CDSs) were PCR-amplified using the PrimeSTAR® Max polymerase (Takara), according to supplier’s recommendations. Associated primers

were 5'-GAAGAAAAATCATAGTAGAAATCTGAGTCATAAACTCCA-3' (forward) and 5'-CAAATTTATTCTCTGTTATTTCTACTAGAAGATTGTTCAATGCAA-3' (reverse) for *CYP71B129*, 5'-GGGAATCGAATTAGAAATCTAAGTGATAAGCTCCA-3' (forward) and 5'-GTTTTCTAGAGAAAATTAACAACTTCGCTAACATATGTC-3' (reverse) for *CYP71B130*, and 5'-GCGAATTGAACTAGAAATCAAACGTATAAGCTCCA-3' (forward) and 5'-GTTACTAGTTTATTGAGAAAACATATTCAAGTCTATTGAAGAAGTTT-3' (reverse) for *CYP71B131a*. CDSs were then inserted into the pCR™8/GW/TOPO® TA vector (Invitrogen™, Thermo Fisher Scientific). Each gene was independently cloned and sequenced twice. A 3'-end 6xHis-tag was added by PCR, using the following primers: 5'-CCAATGGCTATTGATGCCAAC-3' (forward) and 5'-CAACTAATGATGATGATGATGGAGGTGCTTGGGAC-3' (reverse) for *CYP71B129*, 5'-CCAATGGCTATGGATGCCAAC-3' (forward) and 5'-GTTCTAATGATGATGATGATGTGCAAATAGGTGCTCTTG-3' (reverse) for *CYP71B130*, and 5'-CCAATGGCTCTAGATGCCCG-3' (forward) and 5'-GTTTAATGATGATGATGATGTGCAAACAGGTGCTCTTG-3' (reverse) for *CYP71B131a*. Finally, His-tagged CDSs were subcloned into the pYeDP60\_GW® expression vector (Dueholm *et al.*, 2015) using the Gateway LR Clonase™ II Enzyme Mix (Invitrogen™, Thermo Fisher Scientific).

### Synthesis of *CYP71B* mutants

Nucleotide sequences of His-tagged *CYP71B* mutants were chemically synthesized and subcloned into the pYeDP60 vector by GenScript Biotech (Netherlands). Cloning was performed using the *BamH*I and *EcoR*I restriction enzymes.

### Heterologous expression of recombinant *CYP71Bs*

The pYeDP60 and pYedDP60-GW® vectors containing the various *CYP71Bs* genes, and the empty pYeDP60-GW® as control, were transformed into the WAT21 *Saccharomyces cerevisiae* strain according to Pompon *et al.* (1996) and Urban *et al.* (1997). P450 heterologous expression was performed as described in Larbat *et al.* (2007), with the following modifications: the yeasts were cultured at 18°C, and 30 mg. L<sup>-1</sup> adenine was added in the YPGE medium (Kokina *et al.*, 2014). Microsomal fractions containing heterologous P450s and P450 reductase AtCPR2 were collected as described in Larbat

et al. (2007). P450 expression was assessed by western blot, using 6xhistidine Epitope Tag antibodies (Acris, OriGene Technologies, USA, **Fig. S3**).

### ***In vitro* characterization of CYP71Bs**

A functional screening was performed on CYP71B129, CYP71B130, CYP71B131a and associated mutant enzymes. A total of 77 putative phenolic substrates were tested, including phenylpropenes, coumarins, furanocoumarins and pyranocoumarins (**Table S1**). For this, microsomal fractions were incubated at 28°C for 30 min in the presence of 200 µM substrate and 200 µM NADPH, in a final volume of 200 µL phosphate buffer, pH 7. Reactions were stopped by the addition of 65 µL acetonitrile/HCl 50:50 (1N). Reaction products were extracted twice with ethyl acetate, air dried, and resuspended in 100 µL methanol/water (80:20) for subsequent mass spectrometry analysis. Optimal pH and temperature were determined for CYP71B129, CYP71B130 and CYP71B131a;  $Km_{app}$  were determined for all enzymes and mutants using CYP71B130 and CYP71B131a optimal conditions, as described in Villard *et al.* (2021).

### **Identification and quantification of reaction products by mass spectrometry**

Reaction products were analyzed using a UHPLC-MS system (Shimadzu, Japan), following the procedure described in Villard *et al.* (2021). Compounds were identified by comparing their retention time and *m/z* ratio to those of standard molecules, and quantified based on 320 nm UV signal, using the LABSOLUTION software (Shimadzu, Japan). Product identity was further confirmed using a UHPLC-HRMS system composed by a Thermo Vanquish liquid chromatographic system (Thermo Scientific) coupled to a tandem mass spectrometer (Orbitrap-IDX; Thermo Scientific). Analyses were performed as described in Villard *et al.* (2021), except for the gradient phase which was modified as follows (A/B; v/v): 10:90 between 0 and 2min, 30:70 at 10 min, 95:5 at 20 min, 95:5 between 20 and 25 min, and 10:90 from 26 to 30 min. Data were recorded and analyzed on the XCALIBUR™ software (v.2.1.SP1. Build1160; Thermo Fisher Scientific).

### **Phylogenetic analysis of Moraceous CYP71Bs**

A CYP71B dataset (**Table S3a**) was constituted by performing a similarity search in the genomes and transcriptomic databases of 24 plant species from the Nitrogen Fixing

Clade, available at GenBank (Benson *et al.*, 2012). Genetic resources were screened through BLASTN search using *F. carica* CYP71B131a as query and an e-value of 1000. Full length putative P450 sequences were isolated, introns were removed, and putative pseudogenes were processed as explained earlier. Most sequences closer to *F. carica* CYP71AS33 than to *F. carica* CYP71B131a were removed. Putative genes received a standardized name from the P450 nomenclature committee (**Table S2a**). Previously characterized CYP71Bs (Hansen *et al.*, 2018; Irmisch *et al.*, 2014; Schuhegger *et al.*, 2006) and a few representative CYP71ASs were added to the dataset. *F. carica* CYP71AH65 was added to constitute the outgroup of the following tree. Nucleotide sequences were aligned with GENEIOUS (Geneious prime 2019, **Fig. S4**), partitioned on MESQUITE v.3.6 (<http://www.mesquiteproject.org>) and used to build a Bayesian inference gene-family tree with MRBAYES v.3.2.7 (Ronquist *et al.*, 2012; Ronquist & Huelsenbeck, 2003), as described in Villard *et al.* (2021). This first CYP71B gene-tree (**Fig. S5**) was used to reduce the initial CYP71B dataset to a smaller subset of Moraceae-specific sequences (**Table S3a**) that were realigned (**Fig. S5**) and used to generate a second tree, as described above. The trees were drawn with FIGTREE v.1.4.4. Posterior probabilities supporting tree-branches were considered weak below 0.7, strong above 0.9.

### Comparison of 5-hydroxyxanthotoxin synthases and closely related enzymes

A P450 dataset was assembled using the protein sequences of the eight 5OHXS enzymes described so far (i.e., CYP71B130; CYP71B131a; CYP71AZ1 (GenBank EF127863.1); CYP71AZ6 (GenBank MH000221.1); CYP82C2 (Genbank NM\_119348.2); CYP82C4 (GenBank: NM\_119345.3); CYP82D64v1 (C. clementina v0.9 genome clementine0.9\_007296m); CYP82D64v2 (C. sinensis DHSO v3.0 genome Cs\_ont\_1g009220.1) and seven closely related enzymes which do not display this activity (i.e., CYP71B129; CYP71AZ3 (GenBank MH000218.1); CYP71AZ4 (GenBank MH000219.1); CYP71AZ5 (GenBank MH000220.1); CYP82C3, (GenBank NM\_119346.3); CYP82D92 (Phytozome ID Ciclev10000846m); CYP82D179, (Phytozome ID Ciclev10000844m)). Protein sequences were aligned (**Fig. S6**) with MAAFT v.7 (<https://mafft.cbrc.jp/alignment/server/>) relying on default parameters (Katoh *et al.*,

2019; Kuraku *et al.*, 2013), and analyzed on MESQUITE v.3.6 (<http://www.mesquiteproject.org>).

### ***In silico* modeling and docking experiments**

CYP71B131a three-dimensional (3D) structures was folded using the ColabFold v1.5.2: AlphaFold2 (Mirdita *et al.*, 2022), using mmseqs2\_uniref\_env and unpaired-paired MSA options, with standard options for monomeric structure predictions. To model the enzyme-substrate interactions, docking experiments were performed. First, a heme was docked into CYP71B131a 3D-structure. For this, CYP76AH1/miltiradiene (PDB ID: 7cb9) from *Salvia miltiorrhiza* (Gu *et al.*, 2019), was used as a reference and overlapped with CYP71B131a 3D-structure to define a 18Å x 18Å x 18Å docking grid-box around the heme. Protein and heme files were prepared by adding polar hydrogens. A '+2' charge was also added to the heme iron. Rigid docking of the heme (*i.e.*, not allowing the protein to move) was performed using AutoDock Vina v1.2.3 (Trott & Olson, 2010). Subsequently, xanthotoxin and umbelliferone substrates were docked into the CYP71B131a-heme complex. The human microsomal P450 2A6 crystal structure (PDB ID: 1z11) was used as a reference (Yano *et al.*, 2005), and overlapped with the CYP71B131a-heme complex to define a grid-box of 28Å x 24Å x 24Å around the xanthotoxin. Protein file was prepared by adding polar hydrogens, and a '+2' charge was added to the heme iron. Residues within 5Å of xanthotoxin in the reference structure were identified as flexible docking residues. The ligand file was prepared by building bonds and adding polar hydrogen atoms. Flexible docking of the xanthotoxin or the umbelliferone was performed using AutoDock Vina v1.2.3 (Trott & Olson, 2010). CYP71B131a\_G373A and CYP71B131a-G373S mutant enzymes were folded, and substrates were docked as described above. Visualizations of the docking poses were prepared with PyMOL (DeLano, 2002).

## **Results**

### ***Ficus-Morus* CYPome comparison highlights a CYP71B expansion**

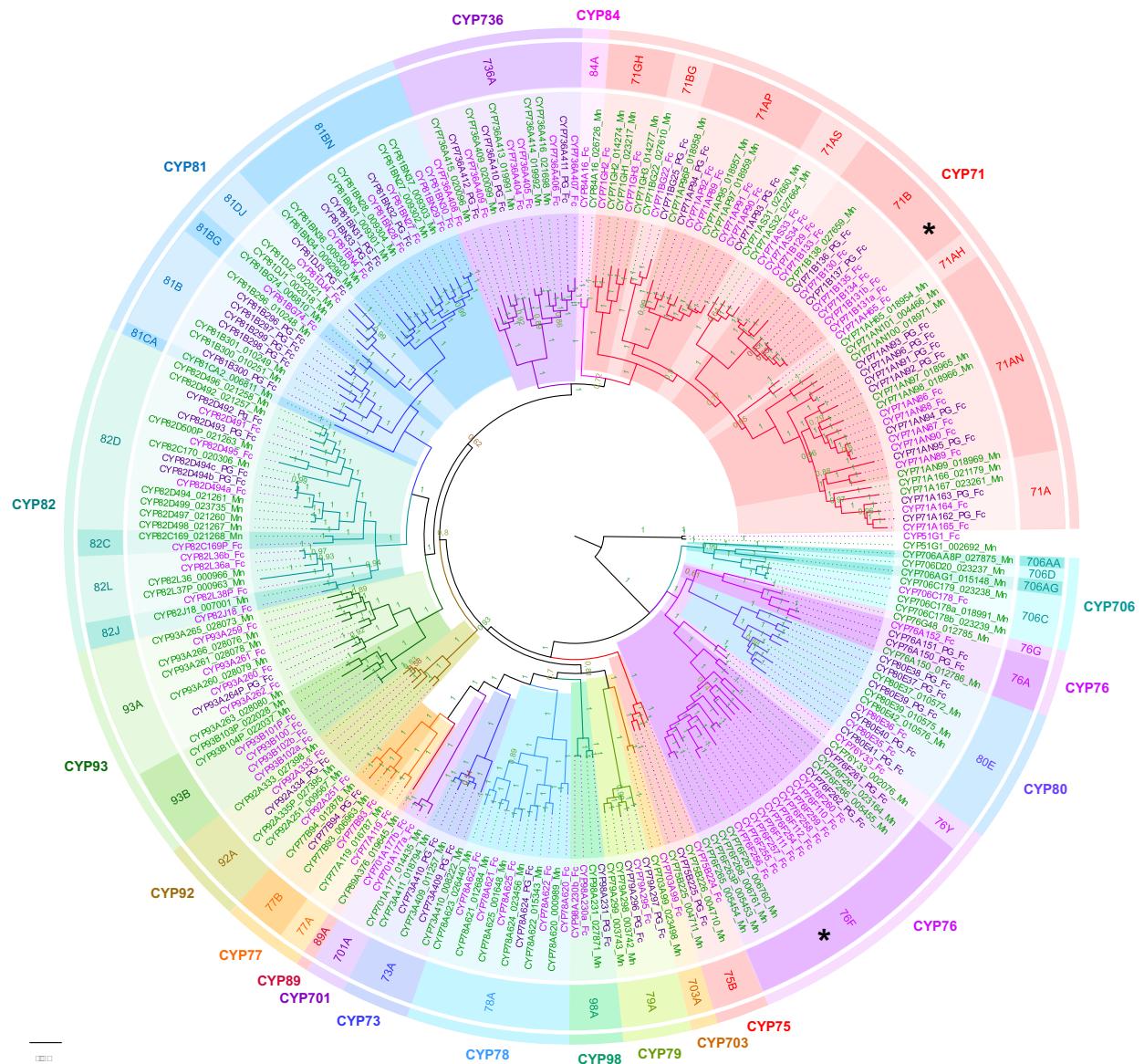
Since the furanocoumarin pathway appeared multiple times in higher plants, it is not advised to rely on classic similarity search to pursue its elucidation (Villard *et al.*, 2021). Instead, to identify new enzymes involved in furanocoumarin biosynthesis, we developed

a novel evolutionary-based approach. In previous work, we demonstrated that the marmesin synthase activity (**Fig. 1**) emerged relatively recently in the Moraceae, through a CYP76F expansion that occurred after the divergence of the *Ficus* and *Morus* ancestors (Villard *et al.*, 2021a). Therefore, we hypothesized that other *Ficus* enzymes catalyzing downstream furanocoumarin-biosynthetic reactions might have similar origins. This means other gene expansions that occurred after the *Ficus-Morus* divergence might enclose other furanocoumarin-biosynthetic enzymes. As target enzymes, we focused on P450s from the CYP71 clan, which play a central role in plant specialized metabolism (D. Nelson & Werck-Reichhart, 2011) and include all furanocoumarin-related P450s described so far (**Fig. 1**). Our approach therefore consisted in comparing the CYP71 clan in *F. carica* and *M. notabilis* to identify *Ficus*-specific expansions.

To identify *F. carica* CYP71 clan sequences, we screened the *F. carica* reference genome (Usai *et al.*, 2020). A total of 133 full-length P450 sequences were isolated, corresponding to 85 genes and 48 pseudogenes (**Table S2a**). These sequences were classified into 18 families and 35 subfamilies of various sizes (**Table S2b**). Twelve families comprised a single subfamily or single *F. carica* sequence (e.g., CYP84A). The six other families contained multiple subfamilies, with a maximum of 37 *F. carica* CYP71s distributed in eight subfamilies (**Table S2b**). Interestingly, most multi-sequence subfamilies result from tandem duplications and form physical gene clusters in *F. carica* genome. For instance, all nine *F. carica* CYP71Bs were clustered on a  $\approx$ 115 kb fragment of chromosome 11 (**Table S2a**). *F. carica* P450s were then compared to the 97 *M. notabilis* CYP71 clan sequences, previously reported by Ma *et. al.* (2014). Even though more sequences were identified in *F. carica* (133) than *M. notabilis* (97), all 35 subfamilies observed in *F. carica* were present in *M. notabilis*, and six additional single-sequence subfamilies were only observed in *M. notabilis* (**Table S2b**).

*F. carica* and *M. notabilis* CYP71 clan sequences were then used to generate a protein-family phylogenetic tree (**Fig. 2**). Across this phylogeny, some families and subfamilies were highly conserved. For example, the CYP78A subfamily comprises six *F. carica* and *M. notabilis* pairs of orthologs, meaning that the last common *Ficus-Morus* ancestor already possessed six CYP78As. This analysis also revealed significant lineage-specific

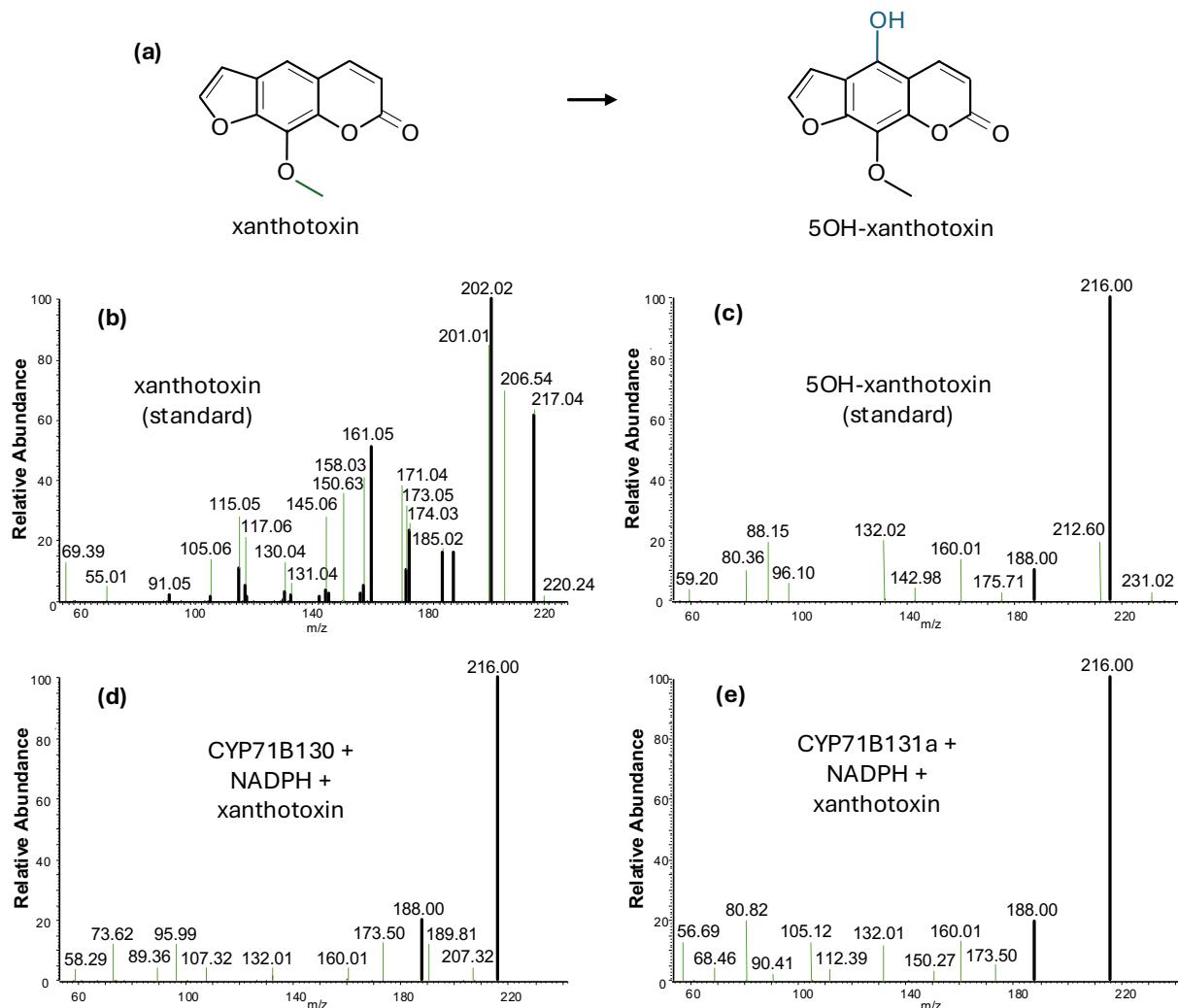
expansions of the CYP71B, CYP71AN, and CYP76F subfamilies in *F. carica*, and of the CYP82D subfamily in *M. notabilis* (**Fig. 2**). Since the CYP76F diversification gave rise to the marmesin synthase activity (Villard *et al.*, 2021), we assumed the CYP71B and CYP71AN expansions might have given rise to other furanocoumarin-biosynthetic enzymes. Among these two subfamilies, the most drastic expansion occurred in the CYP71Bs, with one single *M. notabilis* sequence compared to nine *F. carica* orthologs (CYP71B129, CYP71B130, CYP71B131a, CYP71B131b, CYP71B133, CYP71B134, CYP71B135, CYP71B136, CYP71B137, **Fig. 2**). We therefore decided to focus on *F. carica* CYP71Bs for further characterization.



**Figure 2: Comparison of the CYP71 clan in *F. carica* and *M. notabilis*.** Bayesian inference tree of *F. carica* and *M. notabilis* CYP71 clan, rooted on the CYP51 clan. *F. carica* genes are colored in light purple, *F. carica* pseudogenes in dark purple, *M. notabilis* genes in green. For an easier understanding of the tree, clades are colored by P450 (sub)families, P450 (sub)family names are detailed, and the CYP76F and CYP71B subfamilies are highlighted with a black asterisk. *M. notabilis* sequence names include the accession numbers given in Ma *et al.* (2014).

**CYP71Bs include two 5-hydroxyxanthotoxin synthases and one umbelliferone hydroxylase.**

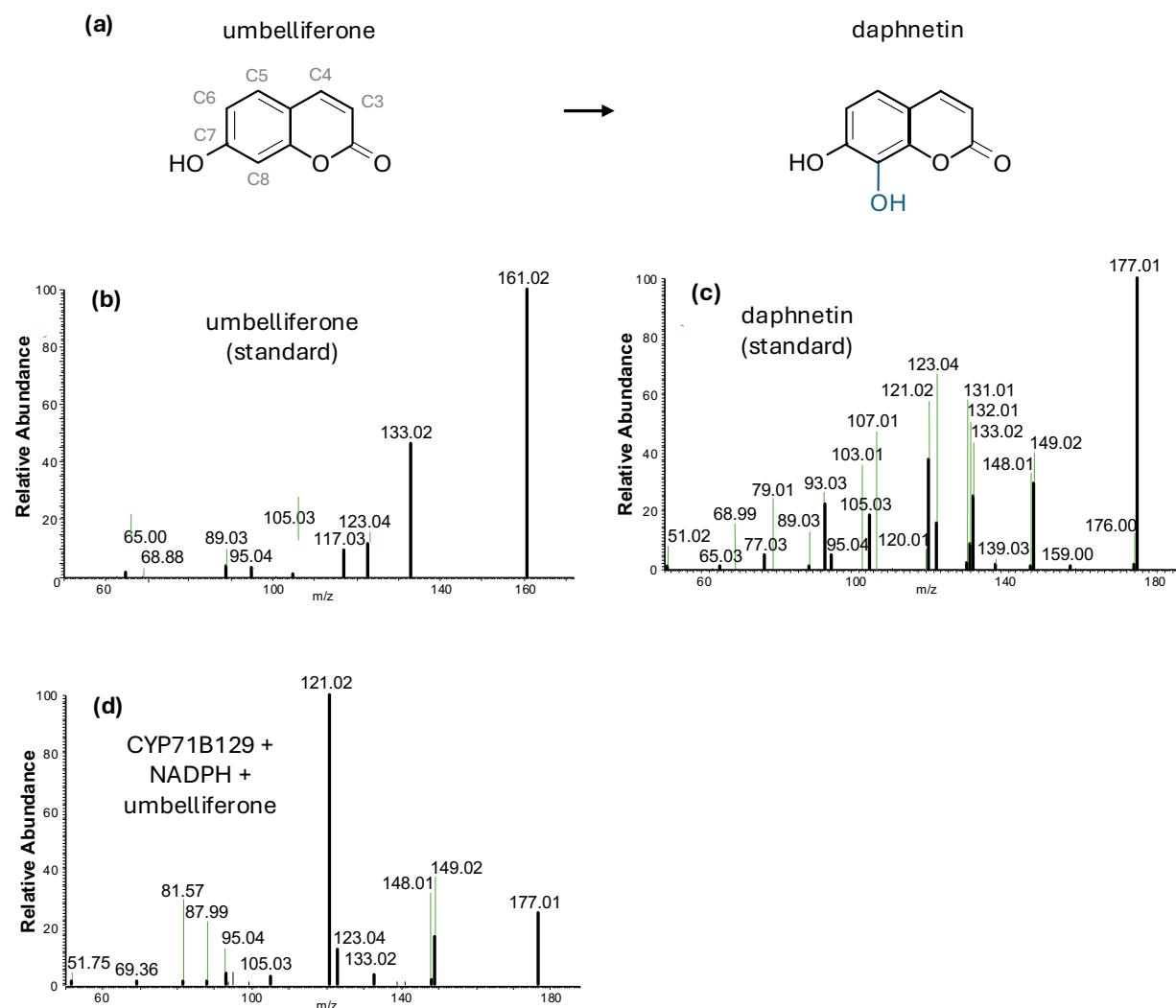
To investigate the function of the nine *F. carica* CYP71Bs, we cloned them from *F. carica* leaf RNA. Three genes were successfully amplified: *CYP71B129*, *CYP71B130* and *CYP71B131a*. CYP71B129-131a enzymes were heterologously produced in yeast (Fig. S3) and used to perform *in vitro* enzyme assays in the presence of NADPH and 77 putative substrates belonging to the coumarin, furanocoumarin, pyranocoumarin and phenylpropene families (Table S1). Results demonstrated that *CYP71B130* and *CYP71B131a* were both able to convert xanthotoxin into 5-hydroxyxanthotoxin, making them 5OHXS enzymes (Fig. 3; Fig. S7). The identity of the product was formally identified by comparing its mass spectrometry characteristics to a commercial 5-hydroxyxanthotoxin standard.



**Figure 3: Conversion of xanthotoxin into 5-hydroxyxanthotoxin by CYP71B130 and CYP71B131a.** (a) Molecular representation of the reaction. (b-e) Tandem mass spectrometry fragmentation patterns of the standards of (b) xanthotoxin and (c) 5-hydroxyxanthotoxin, and of the reaction products of (d) CYP71B130 and (e) CYP71B131a.

In contrast, CYP71B129 did not metabolize any of the tested furanocoumarins, but hydroxylated the coumarin umbelliferone (UMBH activity, **Fig. 4**). Theoretically, umbelliferone can be hydroxylated on five different carbons, leading to the production of 3,7-, 4,7-, 5,7-, 6,7- or 7,8- dihydroxycoumarin. Out of these five molecules, only four standards are commercially available, *i.e.*, 4,7-, 5,7-, 6,7- and 7,8- dihydroxycoumarin. Using mass spectrometry (**Fig. S8**), we compared these four standards to CYP71B129 reaction product, showing that only 7,8-dihydroxycoumarin (daphnetin) had the same retention time as the product. This indicates that CYP71B129 produces either daphnetin or the untested 3,7-dihydroxycoumarin. As the amount of product synthesized by

CYP71B129 was too low to consider an NMR investigation, we could not further confirm where the hydroxylation was performed.



**Figure 4: Metabolization of umbelliferone by CYP71B129.** (a) Molecular representation of the reaction. Carbon numbering is indicated on the molecule. (b-d) Tandem mass spectrometry fragmentation patterns of the standards of (b) umbelliferone and (c) daphnetin, and (d) of the reaction product of CYP71B129.

Next, further enzymatic characterization showed that optimal activities of the three enzymes CYP71B129-131a with their respective substrates were obtained at pH 7.3-7.9 and a temperature of 24-26°C (Fig. S10; Fig. S11). The apparent  $K_m$  values associated to these enzymes were also determined, showing that CYP71B130 ( $K_m$  of  $0.9 \pm 0.7 \mu\text{M}$ ) has a 10-times higher affinity for xanthotoxin than CYP71B131a ( $K_m$  of  $9.4 \pm 1.2 \mu\text{M}$ )

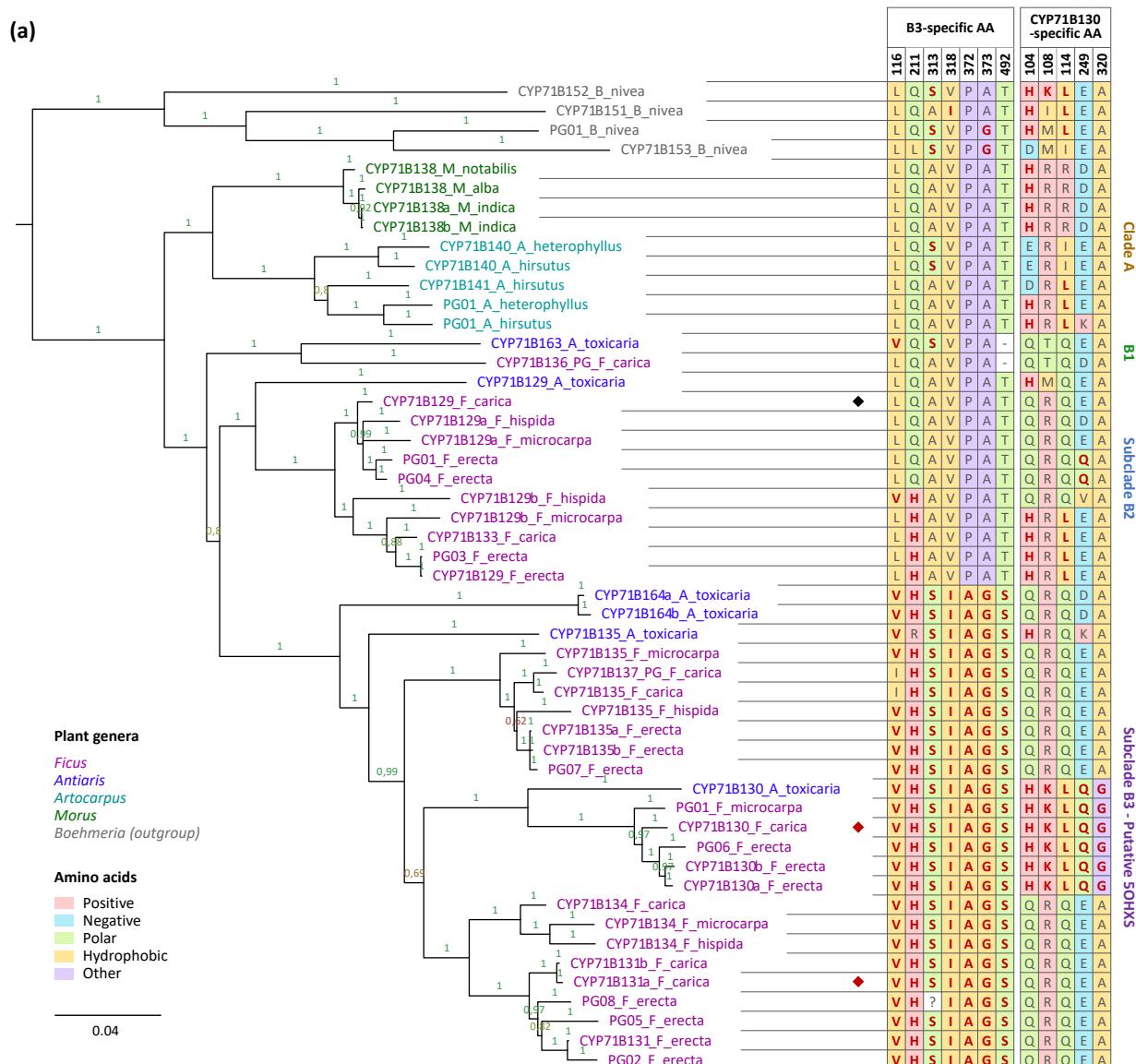
(Fig. S12a-b). The  $K_m$  associated to umbelliferone hydroxylation by CYP71B129 was  $28.5 \pm 9.3 \mu\text{M}$  (Fig. S12d).

### The 5-hydroxyxanthotoxin synthase activity emerged in subclade B3 of Moraceous CYP71Bs

CYP71B129-131a are three *F. carica* enzymes sharing 83-88% nucleotide identity. However, CYP71B129 metabolizes a coumarin while CYP71B130 and CYP71B131a metabolize a furanocoumarin. To better understand the evolution of these enzymes, we performed a gene phylogenetic analysis. Screening 24 plant genomes, we identified 82 CYP71B sequences corresponding to 58 genes and 24 pseudogenes (Table S3a). Sequences were used to generate a gene-family phylogenetic tree (Fig. S5). On this gene-tree, all Rosales CYP71Bs form a well-supported clade which topology perfectly reflects the Rosales phylogeny (Zuntini *et al.*, 2024). This indicates that the last common Rosales ancestor possessed a single CYP71B sequence, which either remained low-copy (e.g., Rosaceae family) or diversified in a lineage-specific manner (e.g., Moraceae family, Fig. S5).

Zooming on the gene-tree (Fig. 5a), we focused on Moraceae-specific CYP71Bs and classified them into two clades (A-B) comprising a total of four subclades (A, B1-B3). Clade A consists of all *Morus* and *Artocarpus* CYP71Bs. These CYP71Bs are well-conserved and present as single- or low-copy in the five screened *Morus* and *Artocarpus* genomes (Fig. 5a). Contrarily, clade B comprises all *Ficus* and *Antiaris* CYP71Bs. Each screened *Ficus* and *Antiaris* genome contained 4-14 CYP71B sequences, indicating significant expansion. Clade B is further subdivided into subclades B1-B3: each of these subclades contains both *Ficus* and *Antiaris* sequences, and therefore pre-dates the divergence of the *Ficus* and *Antiaris* ancestors (Fig. 5a-b). *F. carica* CYP71B129 belongs to subclade B2, CYP71B130 and CYP71B131a to subclade B3. The relatively low expansion and short branch-lengths observed in subclade B2 suggest that the CYP71B129 UMBH activity might not be a B2-specific novelty. Instead, it may be shared with clade A and subclade B1. Additional experiments are however required to support this hypothesis. Contrarily, sequences from subclade B3 diversified and accumulated

more mutations, suggesting the 5OHXS activity observed in *F. carica* CYP71B130 and CYP71B131a emerged in this subclade. This means all B3-enzymes might be 5OHXSs.



**Figure 5: Emergence of the 5OHXS activity in the Moraceae family.** (a) Gene-family tree of the CYP71Bs from the Moraceae family, rooted on CYP71Bs from *Boehmeria nivea* (Urticaceae). Characterized 5-hydroxyxanthotoxin synthase (5OHXS) are highlighted with a red diamond, characterized umbelliferone hydroxylase (UMBH) with a black diamond. Amino acids (AAs) of

interest from the Substrate Recognition Sites (SRSs) are reported through the phylogeny, numbered according to *F. carica* CYP71B131a equivalent. Amino acids corresponding to CYP71B130 derived character states are highlighted in bold red, others are in gray. (b) Phylogeny of the Moraceae family, adapted from (Gardner *et al.*, 2021). The tribes containing known furanocoumarin producing genera (*Ficus*, *Antiaris*, *Maquira*, *Brosimum* and *Dorstenia*) are highlighted in red (Abegaz *et al.*, 2004; Rovinski *et al.*, 1987; Shi *et al.*, 2014; Takahashi *et al.*, 2017; Vieira *et al.*, 1999). The putative emergence of the 5OHXS activity in Moraceous CYP71Bs is highlighted in purple.

To further investigate the emergence of the 5OHXS activity within the CYP71Bs, we used our sequence alignment (**Fig. S7**) to compare the substrate recognition sites (SRSs, **Table S3b**) throughout the CYP71B phylogeny (Gotoh, 1992). Results highlighted seven SRS-residues (**Fig. 5a**) that were well-conserved in subclades A-, B1, and B2 (L116, Q211, A313, V318, P312, A373, T492, numbered according to CYP71B131a equivalents) but substituted in all B3-sequences (V116, H211, S313, I318, A372, G373, S492). These B3-specific synapomorphies might have played a key role in the emergence of the 5OHXS activity. Remarkably, SRS-residues were relatively well conserved throughout the B3 subclade, suggesting that even early branching B3-genes likely encode 5OHXSs (*i.e.*, *CYP71B135s*, *CYP71B164s*). For instance, only nine SRS-residues were different between the two characterized 5OHXS, *i.e.*, H/Q104, K/R108, L/Q114, H/R248, Q/E249, I/V315, G/A320, L/I481 and A/T487 in CYP71B130/131a, respectively. These mutations probably explain the activity difference measured between the two enzymes. Among them, H104, K108, L115, Q248 and G320 are unique in subclade B3, as they are only observed in the six CYP71B130 sequences (**Fig. 5a**). These five recent mutations might thus be responsible for an increased 5OHXS activity in CYP71B130s.

Taken together, our results indicate that the 5OHXS activity emerged in subclade B3 of Moraceous CYP71Bs, probably from the UMBH activity. This evolution correlates with a few mutations affecting the SRSs. The 5OHXS activity therefore results from a lineage-specific CYP71B-diversification that happened after the divergence of the Ficeae and Moreae tribes, but before the divergence of the Ficeae and Olmedieae tribes. This pattern is consistent with the repartition of known furanocoumarin-producing species of the Moraceae family (**Fig. 5b**).

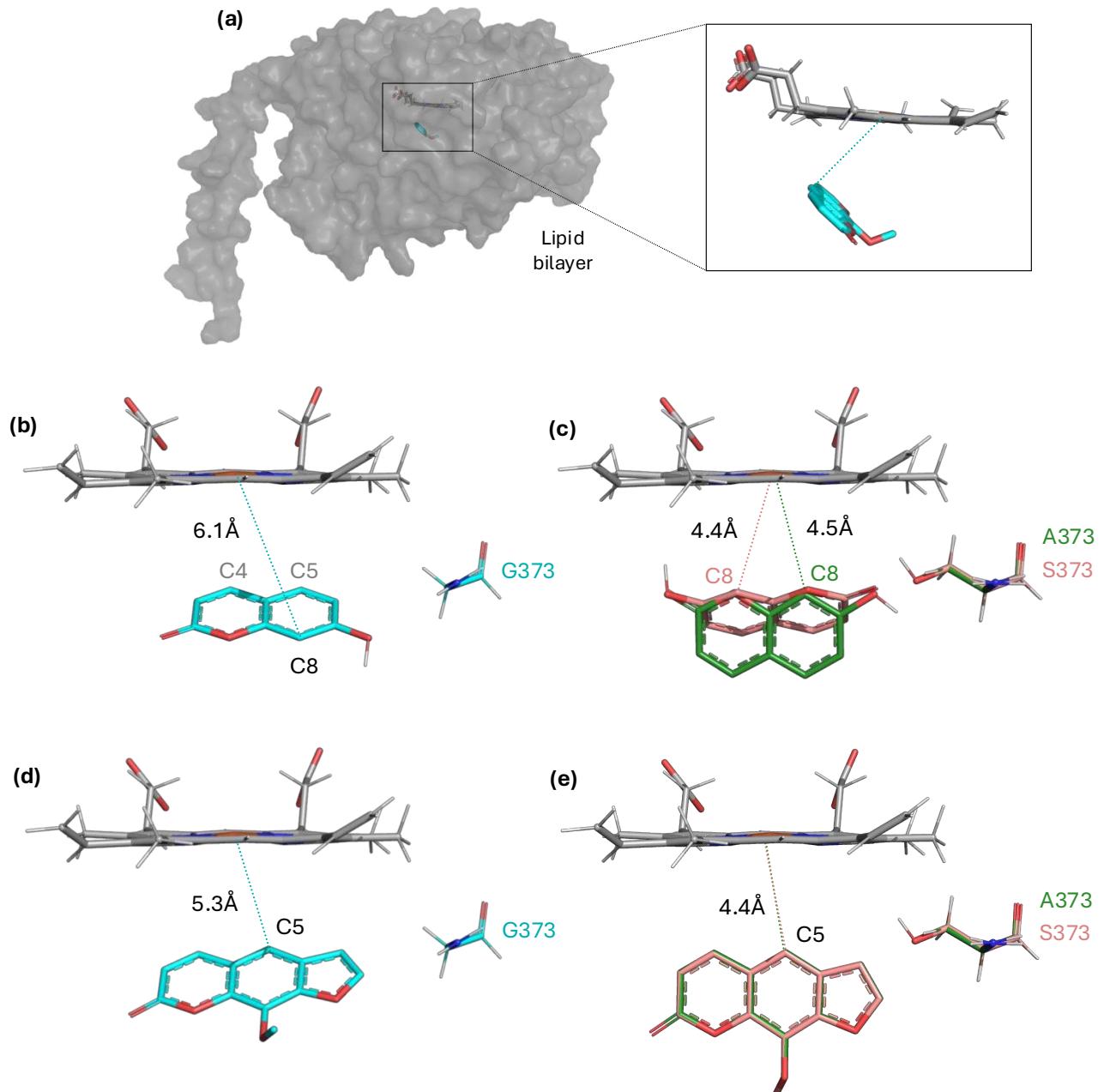
### 5-hydroxyxanthotoxin synthases possess a glycine in position 373

If the 5OHXS function emerged recently in the CYP71Bs, it has also been described in other P450 subfamilies, *i.e.*, CYP71AZ, CYP71B, CYP82C and CYP82D (Krieger *et al.*, 2018; Kruse *et al.*, 2008; Limones-Mendez *et al.*, 2020). To understand 5OHXS independent emergence in distant P450 subfamilies, we compared the eight characterized 5OHXS enzymes (*i.e.*, CYP71AZ1, CYP71AZ6, CYP71B130, CYP71B131a, CYP82C2, CYP82C4, CYP82D64v1, CYP82D64v2) with seven closely related enzymes which do not exhibit this activity (*i.e.*, CYP71AZ3, CYP71AZ4, CYP71AZ5, CYP71B129, CYP82C3, CYP82D92, CYP82D179). For this, we aligned the protein sequences of the different enzymes and analyzed the SRSs throughout the alignment (**Fig. S7**). Overall, SRS-residues were either well-conserved across all sequences, varied according to P450 families or subfamilies, or varied without a clear pattern (**Table I**, **Fig. S7**). However, out of the 96 SRS-residues, residue 373 displayed a unique pattern (**Table I**, **Fig. S7**). Indeed, all investigated 5OHXS enzymes possessed a glycine in position 373 while non-5OHXS enzymes possessed either an alanine (CYP71B129, CYP71AZ3, CYP71AZ5, CYP82C3), a valine (CYP82D92) or a serine (CYP82D179). The only non-5OHXS enzyme possessing a glycine at this position was CYP71AZ4, which catalyzes another furanocoumarin-biosynthetic reaction (*i.e.*, xanthotoxin synthase) (Krieger *et al.*, 2018). Additionally, G373 was also one of the seven B3-specific amino acids highlighted above (**Fig. 5a**). This unique and intriguing pattern hinted toward a critical role of residue 373 for the 5OHXS activity.

To explore this further, we docked the xanthotoxin substrate into CYP71B131a 3D-model (**Fig. 6a**) and defined the enzyme's catalytic pocket as the 27 residues located at less than 5Å from the substrate (**Table 1**). These 27 residues form a subgroup of the 96 SRS-residues. Interestingly, residue 373 was included in this catalytic pocket, in very close proximity ( $\simeq$ 3Å) to xanthotoxin's furan ring (**Fig. 6d**).

**Table 1: Residues belonging to the substrate recognition sites (SRSs) and positioned close to the substrate (5Å) in 5-hydroxyxanthotoxin synthases (5OHXSs) and non-5OHXSs P450 enzymes.** Residues conserved in most investigated sequences are highlighted in light blue, residues with a clear P450 family or subfamily patterns are highlighted in light yellow, residue varying according to P450 function is highlighted in light red, more variable residues are in white. Amino acids identical to the CYP71B131a equivalents are in bold red, others are in black. Residues are numbered according to the CYP71B131a equivalent.

SRS		1												2		3		4				5				6		
Residues		104	106	111	114	115	116	117	118	120	121	122	213	217	247	303	304	305	307	308	309	312	372	373	376	377	379	491
5OHXS	CYP71B130	H	A	Y	L	D	V	A	F	P	Y	G	A	M	F	L	N	M	L	G	G	T	A	G	L	L	R	L
	CYP71B131a	Q	A	Y	Q	D	V	A	F	P	Y	G	A	M	F	L	N	M	L	G	G	T	A	G	L	L	R	L
	CYP71AZ1	T	M	Y	L	D	V	A	F	P	Y	S	L	L	F	L	N	V	N	G	G	T	T	G	L	I	R	I
	CYP71AZ6	T	M	Y	L	D	V	A	F	T	Y	S	I	L	F	L	N	V	N	G	G	T	T	G	L	I	R	L
	CYP82D64v1	V	G	Y	S	L	L	G	F	P	Y	G	F	M	A	L	A	L	L	G	G	T	A	G	L	A	R	L
	CYP82D64v2	V	G	Y	S	L	L	G	F	P	Y	G	F	M	A	L	A	L	L	G	G	T	A	G	L	A	R	L
	CYP82C2	A	A	Y	A	V	F	G	F	P	Y	S	F	V	G	L	A	L	L	G	G	T	A	G	L	G	R	L
	CYP82C4	A	A	Y	A	V	F	G	F	P	Y	S	F	I	G	L	A	L	L	G	G	T	A	G	L	G	R	L
Non-5OHXS	CYP71B129	Q	A	Y	Q	D	L	A	F	P	Y	G	A	M	F	L	N	M	L	G	G	T	P	A	L	L	R	L
	CYP71AZ3	T	M	Y	L	D	V	A	F	P	Y	T	I	F	F	L	N	V	N	G	G	T	P	A	L	V	R	L
	CYP71AZ4	T	S	Y	L	E	M	A	F	P	Y	G	I	F	F	L	N	V	S	G	G	T	P	G	L	I	R	L
	CYP71AZ5	T	T	Y	L	D	M	A	F	P	Y	S	L	S	F	Q	N	I	N	G	G	T	P	A	L	L	R	-
	CYP82D92	L	M	Y	S	M	F	G	F	P	Y	G	S	S	A	L	G	L	L	A	A	T	A	V	L	V	H	L
	CYP82D179	L	M	Y	S	M	F	G	F	P	Y	G	F	S	A	M	G	L	L	A	A	N	A	S	S	I	H	L
	CYP82C3	A	A	Y	-	-	-	-	-	-	-	-	F	V	R	L	A	L	L	G	G	T	A	A	L	G	R	L



**Figure 6: Docking of xanthotoxin and umbelliferone into CYP71B131a and associated mutant enzymes.** (a) Three-dimensional (3D) model of CYP71B131a in complex with a heme (grey) and umbelliferone or xanthotoxin. (b-e) Most probable binding mode of umbelliferone (b,c) and xanthotoxin (d,e) into CYP71B131a (b,d) and associated mutants (c,e). Substrates are colored according to the enzyme they are docked in: blue for CYP71B131a, green for CYP71B131a\_G373A, pink for CYP71B131a-G373S. Distance between substrates' carbons and the heme iron are highlighted with dashed lines.

#### Residue 373 strongly impacts CYP71B131a enzymatic activity.

To experimentally investigate the impact of amino acid 373 on the 5OHXS enzyme activity, we designed a series of CYP71B131a mutants, in which G373 was substituted with an alanine (A), serine (S) or valine (V). These substitutions correspond to the amino

acids observed in the non-5OHXS P450 enzymes described above. Each substitution introduces a gradual steric hindrance (G<A<S<V) and/or variable polarity (S). CYP71B131a mutant genes were synthesized, and associated enzymes were expressed and characterized. Results showed that CYP71B131a\_G373A and CYP71B131a\_G373S could still catalyze the 5OHXS reaction (**Fig. S7**). The affinity of CYP71B131a\_G373A for xanthotoxin was 3-times lower than that of the wild-type enzyme ( $Km_{app} = 20.8 \pm 8\mu\text{M}$ ). The activity of CYP71B131a\_G373S was too low to accurately determine its  $Km_{app}$ . For CYP71B131a\_G373V, no 5OHXS activity was detected. The three mutant enzymes were also incubated in the presence of umbelliferone. Interestingly, they all exhibited the UMBH activity (**Fig. S8**), and showed a rather high affinity for umbelliferone ( $Km_{app} = 9.3 \pm 8.9 \mu\text{M}$  for CYP71B131a\_G373A;  $12.3 \pm 9.9 \mu\text{M}$  for CYP71B131a\_G373S, **Fig. S11e-f**). Next, to test if a glycine in position 373 would be enough to recover the 5OHXS activity, we designed a CYP71B129\_A373G mutant and characterized the associated enzyme. No 5OHXS activity was detected, and CYP71B129\_A373G still exhibited traces of UMBH activity (**Fig. S8h**).

To understand the impact of the G373A and G373S mutations on CYP71B131a 5OHXS activity, we built 3D-models of the associated mutant enzymes and performed docking experiments with xanthotoxin and umbelliferone. Xanthotoxin dockings showed that substituting G373 by an alanine or a serine slightly pushes the xanthotoxin and modifies its orientation toward the heme (**Fig. 6e**). Given the experimental results, this altered position might not be favorable for xanthotoxin binding. In the case of umbelliferone, amino acid 373 had an even stronger impact. Indeed, the best docking state obtained with CYP71B131a positioned umbelliferone with its carbons C4 and C5 oriented toward the heme iron. In contrast, docking performed with CYP71B131a\_G373A and CYP71B131a\_G373S resulted in a horizontal rotation of umbelliferone, positioning carbon C8 towards the heme iron (**Fig. 6c**). These results strongly suggest that CYP71Bs' UMBH activity corresponds to a C8 hydroxylation, and thus, to daphnetin production. This would also explain why the wild-type CYP71B131a does not exhibit this activity.

In summary, our results demonstrated that amino acid 373 plays an important role in shaping the CYP71B131a active site and directly impacts substrate specificity and

affinity. Indeed, increased steric hindrance at position 373 influences the positioning of the substrate in the active site, resulting in a decreased affinity for xanthotoxin and an increased affinity for umbelliferone. However important, amino acid 373 was not critical for the acquisition of the 5OHXS function in CYP71Bs, as CYP71B131a\_G373A still exhibited 5OHXS activity while CYP71B129\_A373G did not. This suggests additional residues must also play an essential role in the 5OHXS activity.

## Discussion

PSMs are molecules of ecological and societal interest with applications in medicine, agronomy, cosmetics, and in the agri-food industry (Weng *et al.*, 2021). As plants do not usually produce enough PSMs quantities to cover large-scale applications, a major challenge of the field consists in elucidating PSM biosynthetic pathways, to help to develop biotechnological tools for large-scale production (Ono & Murata, 2023). This also provides new insights into plant biochemistry and can help better understand PSM ecological roles.

In this study, we developed an evolutionary-based approach to identify enzymes involved in the biosynthesis of lineage-specific PSMs, and we applied it to furanocoumarin biosynthesis. For this, we compared the genes of two Moraceae with different PSM profiles: *F. carica* and *M. notabilis*. As far as we know, furanocoumarins are indeed present in *Ficus* but not in *Morus*, and likely appeared in Moraceae after the *Ficus-Morus* divergence (Villard *et al.*, 2021). To identify furanocoumarin biosynthetic specific genes, we thus looked for lineage-specific gene expansions which could correlate with furanocoumarin lineage-specific emergence. This led to the identification of a CYP71B diversification, amongst which two enzymes exhibited the 5OHXS activity. Our approach therefore proved successful and could be generalized to other families of PSMs, enzymes, and plants, provided that the target PSM is lineage-specific and its repartition in plants is known. If phylogenetics is widely used to analyze characterized-enzymes, it has so far rarely been used to select candidate-enzymes, and only as a minor tool complementing classic approaches (Hodgson *et al.*, 2019; Pluskal *et al.*, 2019). Our proposed approach is therefore original but supported by many examples of lineage-

specific enzymes producing lineage-specific PSMs, as reviewed by Pichersky and Lewinsohn (2011). Moreover, as it only requires genomic resources which are now plentiful and easy to generate, it can be applied to virtually any lineage-specific PSM, regardless of previously limiting factors such as convergent evolution and low PSM production. Our approach therefore has great potential to help address the classic bottlenecks of specialized enzyme discovery. Of course, it also has limits: for instance, it cannot highlight newly functionalized enzymes that evolved without prior major gene duplications. However evolving new functions without gene duplications is relatively rare (Yu *et al.*, 2015). Our approach thus constitutes an excellent tool to elucidate PSM biosynthesis pathways and can be used in combination with other existing approaches. Finally, it also means that the CYP71AN expansion observed in *F. carica* likely includes enzymes involved in the production of furanocoumarins or another lineage-specific PSM.

Experimental results showed that CYP71B129-131a are highly selective enzymes hydroxylating umbelliferone (CYP71B129) or converting xanthotoxin into 5-hydroxyxanthotoxin (CYP71B130,131a). The affinity of CYP71B130 ( $Km_{app} = 0.9 \pm 0.7 \mu\text{M}$ ) and CYP71B131a ( $Km_{app} = 9.4 \pm 1.2 \mu\text{M}$ ) for xanthotoxin falls within the classic affinity range determined for other furanocoumarin-biosynthetic P450s (*i.e.*,  $Km_{app} \approx 0.5\text{-}13 \mu\text{M}$  for CY71AJ1-4, CYP71AZ1,4,6, CYP82D64) (Krieger *et al.*, 2018; Limones-Mendez *et al.*, 2020). The only exception so far is CYP76F112, which displays a very high affinity ( $Km_{app} = 32.2 \pm 3.9 \text{ nM}$ ) for demethylsuberosin (Villard *et al.*, 2021a). On the contrary, P450s involved in coumarin biosynthesis often have lower affinity for their substrate. This is for instance, the case of CYP71AZ3, which displays a rather low affinity ( $Km_{app} = 248.6 \pm 51.9 \mu\text{M}$ ) for esculetin (Krieger *et al.*, 2018). This is also consistent with CYP71B129 slightly lower affinity for umbelliferone ( $Km_{app} = 28.5 \pm 9.3 \mu\text{M}$ ). In addition, according to the Kitajima *et al.* *F. carica* RNAseq libraries, CYP71B130 and CYP71B131a are overexpressed in petiole latex (Kitajima *et al.*, 2018), which has a high furanocoumarin content (Munakata *et al.*, 2020c). Taken together, the high selectivity and relatively good affinity of CYP71B129-131a, their expression patterns, and the presence of both umbelliferone and xanthotoxin substrates in *Ficus carica* (Mamoucha *et al.*, 2016), suggest that the UMBH and 5OHXS activities might be CYP71B129 and CYP71B130-131a physiological functions, respectively. However, to the best of our knowledge, associated

products were never reported in *Ficus*. This could be because these products are not often included in metabolomic studies, or because they do not accumulate in *Ficus* but are instead quickly converted into different compounds.

Furanocoumarins are lineage-specific PSMs which derive from ubiquitous coumarins such as umbelliferone (Fig. 1) (Munakata *et al.*, 2020). Interestingly, several furanocoumarin-biosynthetic enzymes were found to metabolize coumarins and/or to be closely related to coumarin-biosynthetic enzymes (Krieger *et al.*, 2018; Kruse *et al.*, 2008; Limones-Mendez *et al.*, 2020; Rajniak, Giehl, Chang, Murgia, von Wirén, *et al.*, 2018). For instance, within the CYP71AZ subfamily, CYP71AZ3 metabolizes esculetin (coumarin), CYP71AZ4 metabolizes scopoletin (coumarin) and psoralen (furanocoumarin), while CYP71AZ1 and CYP71AZ6 metabolize xanthotoxin (furanocoumarin) (Krieger *et al.*, 2018). In this study, we observed a similar pattern in the CYP71B subfamily, as it includes both UMBH (coumarin) and 5OHXS (furanocoumarin) activities. Moreover, our phylogenetic analyses strongly suggest that CYP71B 5OHXS activity evolved from the UMBH activity, via duplications and functional divergence of the *UMBH* genes. As closely related P450s often metabolize similar compounds (D. Nelson & Werck-Reichhart, 2011b), CYP71B functional divergence was likely facilitated by the structural similarity between umbelliferone and xanthotoxin. The CYP71B subfamily therefore perfectly exemplifies how furanocoumarin-biosynthetic enzymes can stem from coumarin-biosynthetic ones.

In previous studies, CYP71AZ1, CYP71AZ6, CYP82C2, CYP82C4, CYP82D64v1 and CYP82D64v2 were described as 5OHXS enzymes (Krieger *et al.*, 2018; Kruse *et al.*, 2008; Limones-Mendez *et al.*, 2020). With the characterization of CYP71B130 and CYP71B131a, the 5OHXS activity has now been identified in four P450 subfamilies. To better understand how multiple P450 subfamilies acquired the ability to metabolize xanthotoxin, we decided to perform SRS analyses. The SRSs are small regions of P450 enzymes known to regulate substrate specificity. Hence, amino acid substitutions affecting the SRSs often result in significant activity changes such as substrate range, affinity, regio- and stereoselectivity (Gotoh, 1992; Villard *et al.*, 2021). First, we focused on the CYP71B subfamily and identified seven SRS- amino acids that were well-

conserved in most Moraceous CYP71Bs but substituted in all putative 5OHXSs from subclade B3 (i.e., substitutions L116V, Q211H, A313S, V318I, P372A, A373G, T492S). In CYP76F112, similar evolutionary patterns were described for four SRS- amino acids that were experimentally shown to impact substrate specificity and affinity (Villard *et al.*, 2021). It can thus reasonably be assumed that (some of) the seven 5OHXS-specific SRS- substitutions identified in the CYP71Bs played a key role in the substrate specificity switch from umbelliferone to xanthotoxin. Then, we extended our analysis to the CYP71B, CYP71AZ, CYP82C and CYP82D subfamilies and identified an SRS-amino acids displaying a unique pattern: all 5OHXSs from the four subfamilies possessed a glycine in position 373, while non-5OHXS enzymes possessed a different amino acid at this position. Experimental results showed that increased steric hindrance at position 373 lowered CYP71B131a affinity for xanthotoxin and allowed umbelliferone metabolism. These results are consistent with the work of Krieger *et al.* (2018), in which the simultaneous substitution of seven SRS-amino acids, including position 373, impacted CYP71AZ4 substrate specificity. This suggests residue 373 may contribute to substrate specificity in multiple P450 subfamilies. In the CYP71B subfamily, UMBH enzymes possess an alanine in position 373, while 5OHXS possess a glycine (**Fig. 5a**). However, substituting CYP71B129 A373 by a glycine did not allow xanthotoxin metabolism, while substituting CYP71B131a G373 by an alanine was not enough to suppress the 5OHXS activity. Taken together, our results allow us to conclude that the acquisition of the 5OHXS activity in the CYP71Bs, CYP71AZs, CYP82Cs and CYP82Ds was made possible by the substitution of different sets of amino acids in each subfamily, tailored per enzymatic environment. Amongst these substitutions, the only constant through the four subfamilies is G373. This means that each enzyme subfamily took a different evolutive path leading to the 5OHXS activity, but along each path, G373 was found to be favorable to xanthotoxin metabolism. Although not indispensable, G373 has thus played a key role in the independent emergence of the 5OHXS activity in different P450 subfamilies.

From a different perspective, phylogenetic analyses indicated that the 5OHXS activity emerged recently in the Moraceae family, through a lineage-specific CYP71Bs diversification. This evolutionary pattern is similar to that of CYP76F112, the *F. carica*

marmesin synthase (Villard *et al.*, 2021a). When CYP76F112 was identified, the only Moraceous genomes available were from *Ficus* and *Morus*. By including recent *Artocarpus* and *Antiaris* genomes, we could strengthen and complete previous hypotheses. Indeed, the Moraceae family is divided into two clades (**Fig. 5b**): the first one consists of the Moreae (e.g., *Morus*) and Artocarpeae (e.g., *Artocarpus*) tribes (Gardner *et al.*, 2021), in which furanocoumarins were never reported. The second clade comprises five tribes, including the Dorstenieae (e.g., *Dorstenia*), Olmedieae (e.g., *Antiaris*) and Ficeae (e.g., *Ficus*), in which furanocoumarins were identified (Gottlieb *et al.*, 1972; Heinke *et al.*, 2011; Rovinski *et al.*, 1987; Shi *et al.*, 2014; Takahashi *et al.*, 2017). This clade will be defined as the furanocoumarin-producing clade. The lack of 5OHXS genes in both *Morus* and *Artocarpus* supports the hypothesis that *Ficus* furanocoumarin-biosynthetic genes emerged after the *Ficus-Morus* divergence (Villard *et al.*, 2021). The presence of orthologous 5OHXS genes in *Ficus* and *Antiaris* indicates the 5OHXS activity emerged before the Ficeae-Olmedieae divergence. This strongly suggests that furanocoumarins were absent in the most recent *Ficus-Morus* ancestor but produced by the Ficeae-Olmedieae one. Based on these patterns, it is reasonable to infer that furanocoumarins emerged only once in Moraceae, during the early evolution of the furanocoumarin-producing clade. However, as metabolic and genetic data is still scarce in this plant family, we cannot yet rule out the possibility of a multiple and independent emergence of furanocoumarins in the Dorstenieae tribe and in the Ficeae-Olmedieae subclade, respectively. Additional metabolomic and genetic resources will thus be necessary to completely elucidate furanocoumarin emergence in Moraceae.

### Acknowledgements

The authors warmly thank Jessica Amaral (Universidade Federal de São Carlos, Brasil), Sakihito Kitajima, (Kyoto Institute of Technology, Japan), Jérémie Grosjean and Marwa Roumani (Université de Lorraine, France) for their precious help on the project, including supply in pyranocoumarins standards, access to *Ficus carica* transcriptomic data, and support in phytochemical analysis and molecular biology. The authors also address special thanks to Élénore Lacoste, Anne-Flore Didelot, Coralie Deschamps, Yann Moreno, Hugo Schitterer, Juliette Braganti-Coral, Camille Chassin, Jeanne Couderc and Bleuenn Suard students of the Biotechnology specialization of ENSAIA, for their

investment in the phylogenetic studies, CYP71B129-131 cloning, and their inputs on P450 expression. Plants were grown on the Plant Experimental Platform in Lorraine (PEPLor, Université de Lorraine, France), metabolomic analyses were conducted on the Metabolomic and Structural Analytic Platform (PASM, Université de Lorraine, France). AB was financially supported by a PhD grant provided by the French Government and Grand Est Region. AB benefited from the mobility program DrEAM provided by “Lorraine Université d’Excellence”, funded by the ANR “Investissements d’avenir” [grant number 15-004].

### **Competing interests**

The authors declare there is no competing interest.

### **Author contributions**

AB, AO and CC realized the biochemical experiments. CV identified the coding sequences. CV and DRN performed phylogenetic analyses. AB and RK realized the computational structural biology. AB, CV, RL and AH wrote the article. JT, RL and AH supervised the project.

### **Data availability**

Nucleotide sequences coding for CYP71B129, CYP71B130 and CYP71B131a are openly available on EnsemblPlants (<https://plants.ensembl.org/index.html>) under accession numbers FCD\_00002714, FCD\_00002719 and FCD\_00002721 from the genome assembly UNIPI\_FiCari\_1.0.

### **References**

- Abegaz BM, Ngadjui BT, Folefoc GN, Fotso S, Ambassa P, Bezabih M, Dongo E, Rise F, Petersen D. 2004. Prenylated flavonoids, monoterpenoid furanocoumarins and other constituents from the twigs of *Dorstenia elliptica* (Moraceae). *Phytochemistry* 65: 221–226.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Research* 41: D36–D42.
- Daniel L, Buryska T, Prokop Z, Damborsky J, Brezovsky J. 2015. Mechanism-Based Discovery of Novel Substrates of Haloalkane Dehalogenases Using in Silico Screening. *Journal of Chemical Information and Modeling* 55: 54–62.
- DeLano WL. 2002. PyMOL: An Open-Source Molecular Graphics Tool.
- Dueholm B, Krieger C, Drew D, Olry A, Kamo T, Tabouret O, Weitzel C, Bourgaud F, Hehn A, Simonsen HT. 2015. Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evolutionary Biology* 15: 122.

- Gardner EM, Garner M, Cowan R, Dodsworth S, Epitawalage N, Arifiani D, Sahromi, Baker WJ, Forest F, Maurin O, et al. 2021. Repeated parallel losses of inflexed stamens in Moraceae: Phylogenomics and generic revision of the tribe Moreae and the reinstatement of the tribe Olmedieae (Moraceae). *TAXON* 70: 946–988.
- Gotoh O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry* 267: 83–90.
- Gottlieb OR, Da Silva ML, Maia JGS. 1972. Distribution of coumarins in Amazonian Brosimum species. *Phytochemistry* 11: 3479–3480.
- Gu M, Wang M, Guo J, Shi C, Deng J, Huang L, Huang L, Chang Z. 2019. Crystal structure of CYP76AH1 in 4-PI-bound state from *Salvia miltiorrhiza*. *Biochemical and Biophysical Research Communications* 511: 813–819.
- Hansen CC, Sørensen M, Veiga TAM, Zibrandtsen JFS, Heskes AM, Olsen CE, Boughton BA, Møller BL, Neilson EHJ. 2018. Reconfigured Cyanogenic Glucoside Biosynthesis in Eucalyptus cladocalyx Involves a Cytochrome P450 CYP706C55. *Plant Physiology* 178: 1081–1095.
- Hebsgaard S. 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Research* 24: 3439–3452.
- Hehmann M, Lukacin R, Ekiert H, Matern U. 2004. Furanocoumarin biosynthesis in *Ammi majus* L.: cloning of bergaptol O-methyltransferase. *European Journal of Biochemistry* 271: 932–940.
- Heinke R, Franke K, Porzel A, Wessjohann LA, Awadh Ali NA, Schmidt J. 2011. Furanocoumarins from *Dorstenia foetida*. *Phytochemistry* 72: 929–934.
- Hodgson H, De La Peña R, Stephenson MJ, Thimmappa R, Vincent JL, Sattely ES, Osbourn A. 2019. Identification of key enzymes responsible for protolimonoid biosynthesis in plants: Opening the door to azadirachtin production. *Proceedings of the National Academy of Sciences* 116: 17096–17104.
- Irmisch S, Clavijo McCormick A, Günther J, Schmidt A, Boeckler GA, Gershenzon J, Unsicker SB, Köllner TG. 2014. Herbivore-induced poplar cytochrome P450 enzymes of the CYP71 family convert aldoximes to nitriles which repel a generalist caterpillar. *The Plant Journal* 80: 1095–1107.
- Ji J, Han X, Zang L, Li Y, Lin L, Hu D, Sun S, Ren Y, Maker G, Lu Z, et al. 2024. Integrative multi-omics data elucidating the biosynthesis and regulatory mechanisms of furanocoumarins in *Angelica dahurica*.
- Karamat F, Olry A, Munakata R, Koeduka T, Sugiyama A, Paris C, Hehn A, Bourgaud F, Yazaki K. 2014. A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *The Plant Journal* 77: 627–638.
- Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 20: 1160–1166.
- Kitajima S, Aoki W, Shibata D, Nakajima D, Sakurai N, Yazaki K, Munakata R, Taira T, Kobayashi M, Aburaya S, et al. 2018. Comparative multi-omics analysis reveals diverse latex-based defense strategies against pests among latex-producing organs of the fig tree (*Ficus carica*). *Planta* 247: 1423–1438.
- Kokina A, Kibilds J, Liepins J. 2014. Adenine auxotrophy - be aware: some effects of adenine auxotrophy in *Saccharomyces cerevisiae* strain W303-1A. *FEMS Yeast Research* 14: 697–707.
- 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. *Frontiers in Plant Science* 9: 820.
- Kruse T, Ho K, Yoo H-D, Johnson T, Hippolyte M, Park J-H, Flavell R, Bobzin S. 2008. In *Planta Biocatalysis Screen of P450s Identifies 8-Methoxysoralen as a Substrate for the CYP82C Subfamily, Yielding Original Chemical Structures*. *Chemistry & Biology* 15: 149–156.
- Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research* 41: W22–W28.
- Larbat R, Kellner S, Specker S, Hehn A, Gontier E, Hans J, Bourgaud F, Matern U. 2007. Molecular Cloning and Functional Characterization of Psoralen Synthase, the First Committed Monooxygenase of Furanocoumarin Biosynthesis. *Journal of Biological Chemistry* 282: 542–554.
- Limones-Mendez M, Dugrand-Judek A, Villard C, Coqueret V, Bourgaud F,

- Olry A, Hehn A. 2020. Convergent evolution leading to the appearance of furanocoumarins in citrus plants. *Plant Science* 292: 110392.
- Ma B, Luo Y, Jia L, Qi X, Zeng Q, Xiang Z, He N. 2014. Genome-wide identification and expression analyses of cytochrome P450 genes in mulberry (*Morus nobilis*): Analyses of mulberry P450 genes. *Journal of Integrative Plant Biology* 56: 887–901.
- Mamoucha S, Fokialakis N, Christodoulakis NS. 2016. Leaf structure and histochemistry of *Ficus carica* (Moraceae), the fig tree. *Flora - Morphology, Distribution, Functional Ecology of Plants* 218: 24–34.
- Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nature Methods* 19: 679–682.
- Moore BM, Wang P, Fan P, Leong B, Schenck CA, Lloyd JP, Lehti-Shiu MD, Last RL, Pichersky E, Shiu S-H. 2019. Robust predictions of specialized metabolism genes through machine learning. *Proceedings of the National Academy of Sciences* 116: 2344–2353.
- Munakata R, Kitajima S, Nuttens A, Tatsumi K, Takemura T, Ichino T, Galati G, Vautrin S, Bergès H, Grosjean J, et al. 2020. Convergent evolution of the UbiA prenyltransferase family underlies the independent acquisition of furanocoumarins in plants. *New Phytologist* 225: 2166–2182.
- Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution: P450-centric evolution. *The Plant Journal* 66: 194–211.
- Ono E, Murata J. 2023. Exploring the Evolvability of Plant Specialized Metabolism: Uniqueness Out Of Uniformity and Uniqueness Behind Uniformity. *Plant And Cell Physiology* 64: 1449–1465.
- Pichersky E, Lewinsohn E. 2011. Convergent Evolution in Plant Specialized Metabolism. *Annual Review of Plant Biology* 62: 549–566.
- Pluskal T, Torrens-Spence MP, Fallon TR, De Abreu A, Shi CH, Weng J-K. 2019. The biosynthetic origin of psychoactive kavalactones in kava. *Nature Plants* 5: 867–878.
- Pompon D, Louerat B, Bronine A, Urban P. 1996. [6] Yeast expression of animal and plant P450s in optimized redox environments. In: *Methods in Enzymology*. Elsevier, 51–64.
- Rajniak J, Giehl RFH, Chang E, Murgia I, von Wirén N, Sattely ES. 2018. Biosynthesis of redox-active metabolites in response to iron deficiency in plants. *Nature Chemical Biology* 14: 442–450.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* 61: 539–542.
- Roselli S, Olry A, Vautrin S, Coriton O, Ritchie D, Navrot N, Krieger C, Vialart G, Bergès H, Bourgaud F, et al. 2016. Genomic approach to decipher specialized plant pathways: a case study with furanocoumarins in parsnip. *The plant Journal* Submitted.
- Rovinski JM, Tewalt GL, Sneden AT. 1987. Maquiroside A, a New Cytotoxic Cardiac Glycoside from Maquira calophylla. *Journal of Natural Products* 50: 211–216.
- Sarker SD, Nahar L. 2017. Progress in the Chemistry of Naturally Occurring Coumarins. In: Kinghorn AD, Falk H, Gibbons S, Kobayashi J, eds. *Progress in the Chemistry of Organic Natural Products. Progress in the Chemistry of Organic Natural Products* 106. Cham: Springer International Publishing, 241–304.
- Schuhegger R, Nafisi M, Mansourova M, Petersen BL, Olsen CE, Svatoš A, Halkier BA, Glawischning E. 2006. CYP71B15 (PAD3) Catalyzes the Final Step in Camalexin Biosynthesis. *Plant Physiology* 141: 1248–1254.
- Shi L-S, Kuo S-C, Sun H-D, Morris-Natschke SL, Lee K-H, Wu T-S. 2014. Cytotoxic cardiac glycosides and coumarins from *Antiaris toxicaria*. *Bioorganic & Medicinal Chemistry* 22: 1889–1898.
- Stevenson PC, Simmonds MSJ, Yule MA, Veitch NC, Kite GC, Irwin D, Legg M. 2003. Insect antifeedant furanocoumarins from *Tetradium daniellii*. *Phytochemistry* 63: 41–46.
- Takahashi T, Okiura A, Kohno M. 2017. Phenylpropanoid composition in fig (*Ficus carica* L.) leaves. *Journal of Natural Medicines* 71: 770–775.
- Trott O, Olson AJ. 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31: 455–461.
- Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D. 1997. Cloning, Yeast Expression, and Characterization of the Coupling of Two Distantly Related *Arabidopsis thaliana* NADPH-Cytochrome P450 Reductases with P450 CYP73A5. *Journal of Biological Chemistry* 272: 19176–19186.

Usai G, Mascagni F, Giordani T, Vangelisti A, Bosi E, Zuccolo A, Ceccarelli M, King R, Hassani-Pak K, Zambrano LS, et al. 2020. Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. *The Plant Journal* 102: 600–614.

Vieira IJC, Mathias L, Monteiro VDFF, Braz-Filho R, Rodrigues-Filho E. 1999. A New Coumarin from Brosimum Gaudichaudii Trecul. *Natural Product Letters* 13: 47–52.

Villard C, Munakata R, Kitajima S, Velzen R, Schranz ME, Larbat R, Hehn A. 2021. A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution. *New Phytologist* 231: 1923–1939.

Wang K, Zeng H, Dai Y, Wang Z, Tang H, Li J, Lu X, Jiang N, Xie G, Zhu Y, et al. 2024. Three types of enzymes complete the furanocoumarins core skeleton biosynthesis in *Angelica sinensis*. *Phytochemistry* 222: 114102.

Weng J-K, Lynch JH, Matos JO, Dudareva N. 2021. Adaptive mechanisms of plant specialized metabolism connecting chemistry to function. *Nature Chemical Biology* 17: 1037–1045.

Yano JK, Hsu M-H, Griffin KJ, Stout CD, Johnson EF. 2005. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nature Structural & Molecular Biology* 12: 822–823.

Yu J, Ke T, Tehrim S, Sun F, Liao B, Hua W. 2015. PTGBase: an integrated database to study tandem duplicated genes in plants. *Database* 2015.

Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Françoso E, Gallego-Paramo B, McGinnie C, et al. 2024. Phylogenomics and the rise of the angiosperms. *Nature*.

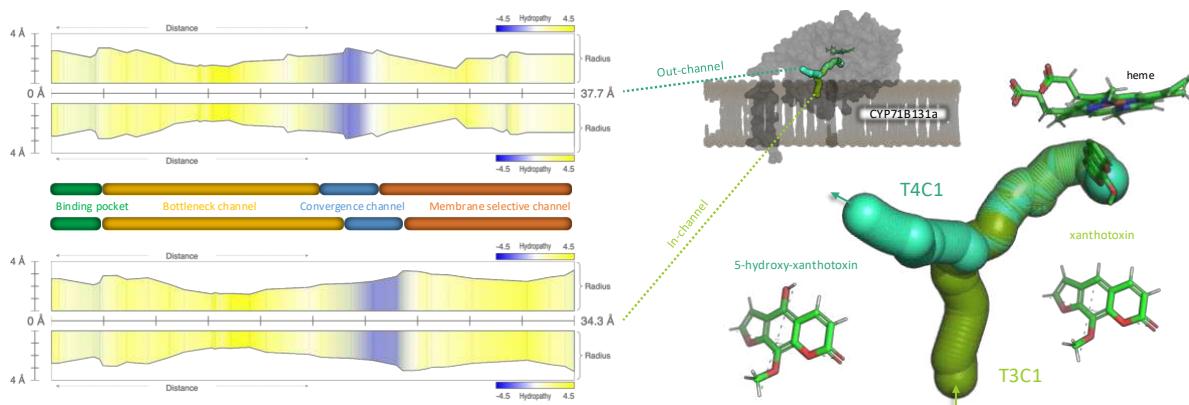


### 3.2.3- Étude complémentaire concernant CYP71B131a

#### A- Modélisation

Dans le prolongement des recherches effectuées pour cette publication, j'ai entrepris une modélisation des canaux d'entrée et de sortie probables de la xanthotoxine et de la 5OHX chez CYP71B131a (Figure 25). La prédition faite par ces modèles indique que i) la xanthotoxine accède probablement à la poche catalytique *via* un canal plongé dans la membrane du réticulum endoplasmique (RE) et que ii) le produit pourrait ressortir du site actif par un canal localisé à l'interface entre le cytosol et le RE, et enfin que iii) les deux canaux se rejoignent en amont de la poche catalytique. Cette structure génère 3 goulots d'étranglement successifs.

Dans cette structure, la glycine G373 se trouve à l'entrée de la poche catalytique. Ce positionnement suggère qu'en plus d'un rôle central dans la sélectivité du substrat, elle joue aussi potentiellement un rôle dans son maintien dès l'entrée de la poche catalytique lors de la réaction enzymatique. Cette vision statique gagnerait à être complétée par une modélisation de la dynamique de CYP71B131a pour décrire plus précisément le circuit de la xanthotoxine dans l'enzyme. En effet, la mesure en parallèle de la taille des canaux d'entrée et de sortie, et de la structure protonée de la xanthotoxine et de la 5OHX, suggère que CYP71B131a doive subir des modifications conformationnelles pour permettre les flux du substrat et du produit.



**Figure 25: Prédiction de la structure tridimensionnelle des canaux d'entrée (T3C1) et de sortie (T4C1) de CYP71B131a et du cheminement de la xanthotoxine dans le cadre de sa**

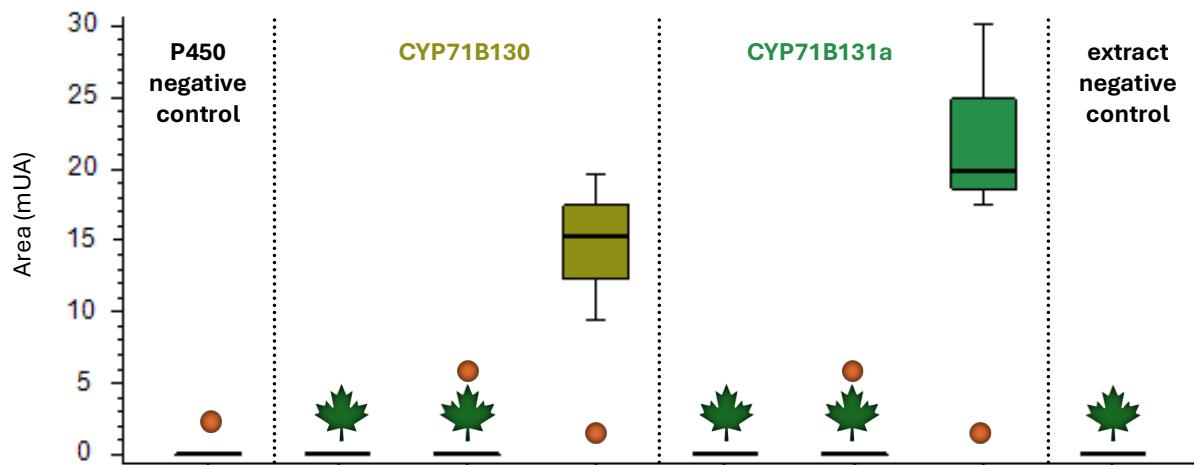
**fonction 5-hydroxyxanthotoxine synthase (5OHXS) (à droite).** Représentation schématique de l'évolution du rayon (en Å) et de l'hydropathie (-4,5 à +4,5) des canaux d'entrée et de sortie de CYP71B131a (à gauche) orientés de la partie la plus interne (de gauche) à la plus externe (à droite) du P450. Les canaux d'entrée et de sortie sont de taille similaire allant de 34,3 à 37,7Å. Ils présentent une hydropathie principalement positive qui témoigne du refuge hydrophobe qu'ils permettent pour l'entrée et la sortie de composés hémipolaires. La zone centrale légèrement moins hydrophobe représente la zone d'intersection des canaux d'entrée et de sortie. L'entrée de la poche catalytique présente un goulot d'étranglement de 2,3Å pour une hydropathie de -0,2 et une hydrophobicité de -0,39. Le goulot d'étranglement le plus étroit dans la zone intermédiaire dispose d'une taille de 1,5Å. Le canal d'entrée présente deux goulets d'étranglement successifs d'un rayon de 2,1Å puis 2Å de façon à restreindre l'entrée de métabolites. Une contrainte stérique qui n'est pas présente pour le canal de sortie qui possède un rayon plus large de 2,4Å au minimum pour une longueur plus courte. Le diamètre le plus imposant en largeur pour le substrat xanthotoxine mesure 7,1Å (distance maximale entre le groupement méthyl et le carbone C5) pour une longueur maximale de 9,1Å. Le diamètre le plus imposant en largeur pour le produit 5OHX mesure 7,7Å (distance maximale entre le groupement méthyl et le C5 hydroxylé). Cette modélisation a été obtenue en utilisant l'outil Channel DB (paramètres standards) sur la plateforme MoleOnline (Pravda *et al.*, 2018).

## B- Sélectivité de CYP71B129, CYP71B130, CYP71B131a

Le rôle physiologique des enzymes est une question très largement irrésolue dans le cadre de la multiplicité des substrats potentiels présents dans une cellule. Bien souvent, la caractérisation fonctionnelle des enzymes se fait dans des conditions *in vitro*, en présence d'un nombre restreint de molécules. Pour tenter d'apporter des éléments quant à la sélectivité de CYP71B129, CYP71B130 et CYP71B131a, j'ai incubé des microsomes<sup>188</sup> surexprimant chacune de ces enzymes avec un extrait méthanolique concentré produit à partir de feuilles de figuier. Par une approche de métabolomique non ciblée, j'ai cherché à identifier de nouveaux produits réactionnels par comparaison avec un extrait ayant subi le même traitement, mais sans enzyme. Les résultats obtenus n'ont pas permis d'identifier des produits nouveaux, ce qui pourrait s'expliquer par un effet inhibiteur de l'un ou l'autre des métabolites présents dans l'extrait. Cette hypothèse a été confirmée par l'absence de métabolisation de xanthotoxine lorsque celle-ci est rajoutée au mix réactionnel en présence de CYP71B130 et CYP71B131a (**Figure 26**).

---

<sup>188</sup> Microsome : Fraction membranaire dérivée du réticulum endoplasmique obtenue par centrifugation lors de la rupture des cellules.



**Figure 26: Effet inhibiteur du métabolome soluble foliaire sur CYP71B130 et CYP71B131a dans le cadre de la fonction 5-hydroxyxanthotoxine synthase (5OHXS).** Le diagramme représente l'aire de pic (en mUA) de l'accumulation du produit de réaction 5-hydroxyxanthotoxine. Les enzymes CYP71B129 (contrôle négatif), CYP71B130 et CYP71B131a ont été incubées en présence de xanthotoxine (orange) et/ou un extrait de feuilles de figuier (Cf=1X) (feuille verte) ainsi qu'avec le cofacteur NADPH. L'incubation a été réalisée à 28°C, pH=7 durant 30 minutes. Échantillonnage n=3. Graphique obtenu avec l'outil de comparaison métabolomique Compound Discoverer, ThermoScientific.

### 3.2.4- Caractérisation fonctionnelle de CYP71AN90, CYP71AN94 et CYP71A164 chez *F. carica*

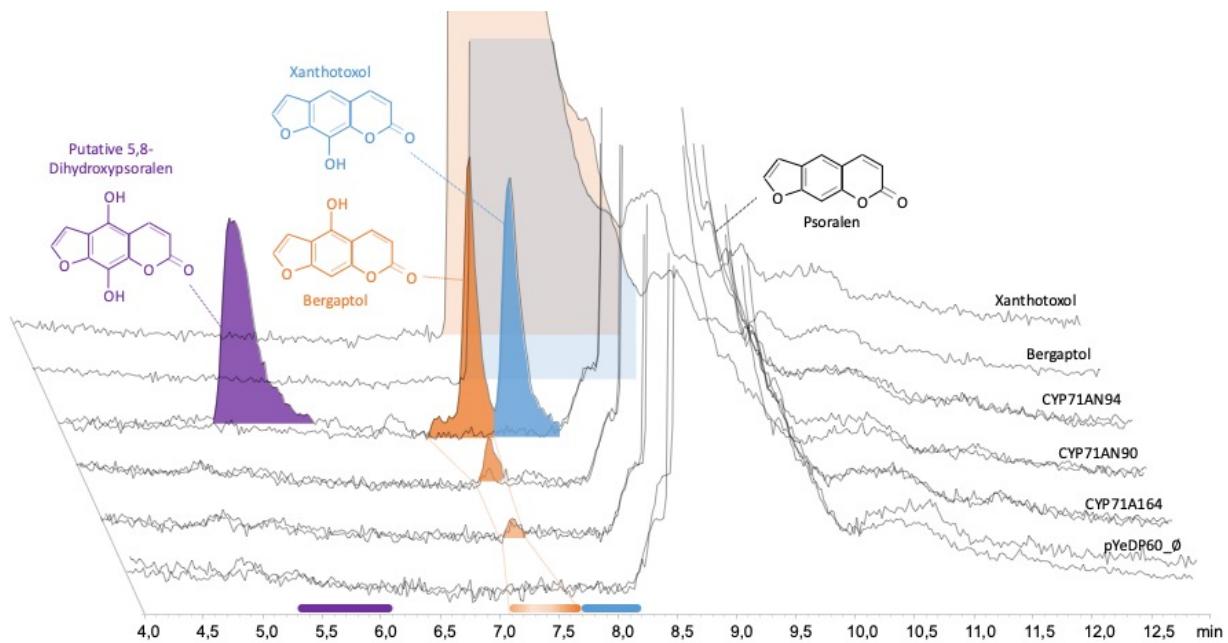
Au-delà de la sous-famille CYP71B, l'analyse de l'arbre phylogénétique réalisée sur les CYPomes des *Ficus carica* et de *Morus notabilis* a permis d'identifier d'autres sous-familles potentiellement impliquées dans la biosynthèse de FCs (Figure 2 de l'article). J'ai réalisé un certain nombre de travaux préliminaires sur certains membres des sous-familles CYP71ANs et CYP71As exprimés dans les feuilles de figuier.

Selon la même procédure que pour les CYP71Bs, j'ai cloné, exprimé (Figure 27) et caractérisé fonctionnellement CYP71AN90, CYP71AN94 et CYP71A164. Lorsque ces enzymes sont incubées en présence du cofacteur NADPH, CYP71AN94 transforme le psoralène en 3 produits (Figure 28). Par comparaison du temps de rétention avec des standards commerciaux, les produits semblent correspondre à du xanthotoxol ou du bergaptol. Le 3<sup>ème</sup> produit n'a pu être identifié par cette approche. À la vue de la capacité de CYP71AN94 à hydroxyler aussi bien la position C5 (e.g. bergaptol) et la position C8

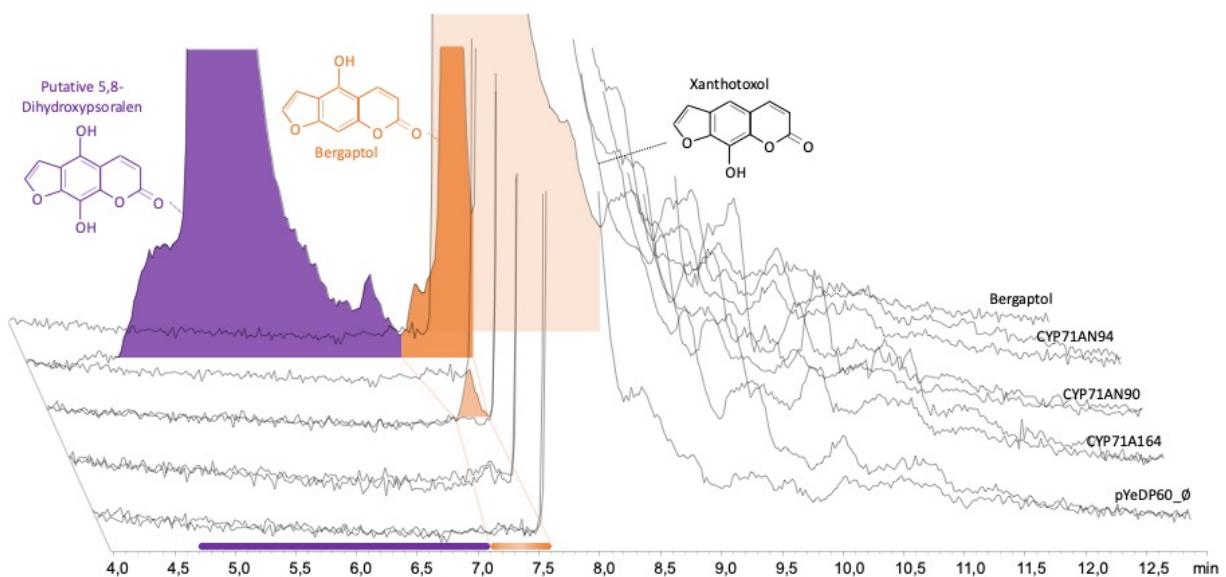
(e.g. xanthotoxol) du psoralène, il est possible que le 3<sup>ème</sup> produit puisse être du 5,8-dihydroxypsoralène. Cette hypothèse est renforcée par la capacité de cette enzyme de transformer le xanthotoxol en un second produit qui a le même temps de rétention que le 3<sup>ème</sup> produit de métabolisation du psoralène (**Figures 28-29**). La quantité de produit formé étant très faible, l'hypothèse d'un artefact peut difficilement être exclue. De fait, ces résultats et hypothèses devront cependant être confirmés par des analyses fines en spectrométrie de masse, voire en RMN.

Concernant CYP71AN90 et CYP71A164, les résultats préliminaires semblent indiquer qu'ils soient capables de métaboliser le psoralène en bergaptol.

Pour finir, et de façon surprenante, l'incubation de CYP71AN90 et CYP71AN94 en présence de xanthotoxol semble mener à la détection de pics ayant le même temps de rétention que le bergaptol (**Figure 28**). Ce résultat pourrait suggérer une potentielle activité d'isomérisation. Toutefois, la prudence est de mise, car plusieurs scénarios dont un artefact analytique pourrait expliquer cette détection de bergaptol : la présence, dans le standard de xanthotoxol, de traces de psoralène pourrait ainsi expliquer un tel profil dans la mesure où ces deux enzymes sont capables d'hydroxyler en C5 le psoralène. Alternativement, le pic du présumé 5,8-dihydroxypsoralène produit par CYP71AN94 est majoritaire et il n'est donc pas à exclure que la production de bergaptol en présence de xanthotoxol puisse provenir de la deshydroxylation du 5,8-dihydroxypsoralène. Une confirmation de ces données est donc nécessaire avant de conclure à une telle activité. Parrallélement, CYP71A164, CYP71AN90 et CYP71AN94 en présence de bergaptol et de NADPH n'ont pas permis la synthèse d'un produit en condition *in vitro*.



**Figure 27: Produits de métabolisation des enzymes CYP71AN94, CYP71AN90 et CYP71A164 en présence du cofacteur NADPH et du substrat psoralène.** Chromatogramme UV PDA filtré à 320 nm des extraits d'incubation des microsomes de levures en présence de psoralène et en présence ou absence de NADPH. Les temps de rétention (RT) pour le produit sont pour le xanthotoxol entre 7,7 et 8,2 minutes ; pour le bergaptol entre 7 et 7,7 minutes et pour le 3<sup>ème</sup> produit (potentiellement le 5,8-dihydroxpsoralène) entre 5,3 et 6,1 minutes. Aucune métabolisation n'a été détectée pour les enzymes en absence de NADPH et pour le contrôle pYeDP60\_Ø avec ou sans NADPH. Ces produits de métabolisation sont appuyés par le criblage métabolique de deux séries de microsomes indépendantes.



**Figure 28: Produits de métabolisation des enzymes CYP71AN94, CYP71AN90 et CYP71A164 en présence du cofacteur NADPH et du substrat xanthotoxol.** Chromatogramme UV PDA filtré à 320 nm des extraits d'incubation des microsomes de levures en présence de xanthotoxol et en présence ou absence de NADPH. Le temps de rétention (RT) pour le produit sont pour le bergaptol entre 7 et 7,6 minutes et pour le 3<sup>ème</sup> produit (potentiellement le 5,8-dihydroxypsoralène) entre 4,7 et 7 minutes. Aucune métabolisation n'a été détectée pour les enzymes en absence de NADPH et pour le contrôle pYeDP60\_ø avec ou sans NADPH. Ces produits de métabolisation sont appuyés par le criblage métabolique de deux séries de microsomes indépendantes.

### 3.2.5- À la recherche d'autres P450s impliqués dans la synthèse des FCs

Des travaux antérieurs ont démontré que les FCs s'accumulaient fortement dans le latex de pétiole (Mamoucha *et al.*, 2016; Munakata *et al.*, 2020). L'analyse comparative des transcriptomes de différents latex a ainsi contribué à l'identification de gènes impliqués dans la voie de biosynthèse des FCs linéaires (e.g. PT1a/b, CYP76F112, CYP71B130, CYP71B131a) (Bouillé *et al.*, 2024; Munakata *et al.*, 2020; Villard *et al.*, 2021).

Parmi les 133 P450s identifiés comme appartenant au clan 71 chez *F. carica*, 26 sont transcrits dans le latex et, parmi eux, 15, le sont préférentiellement dans le latex de pétiole (**Tableau 3**). Parmi les P450s préférentiellement exprimés dans le latex de pétiole, 5 n'ont pour le moment pas encore été étudiés (à ma connaissance) comme étant des candidats potentiellement impliqués dans la voie de biosynthèse des FCs : CYP71A163, CYP71AN87, CYP701A177a, CYP701A177b, et CYP736A406. Se rajoutent à cette liste, 3 séquences (CYP736A411\_pg CYP73A410\_pg et CYP98A231\_pg) considérées comme des pseudogènes. Cette annotation de « pseudogène » nécessiterait cependant d'être confirmée, car elle pourrait résulter de problèmes d'assemblage au moment de la construction de la banque de données (**Tableau 3**). À titre d'exemple, lors du clonage de CYP71AN94 (annoté initialement en pseudogène) et en utilisant des amorces externes au gène potentiel, j'ai pu cloner 2 versions de ce P450. L'une correspondait à une séquence codante et l'autre à un pseudogène, ce qui pourrait assez facilement être expliqué par le fait que les plantes utilisées ne sont vraisemblablement pas des plantes homozygotes.

**Tableau 3: Expression des P450s du clan 71 dans le latex de figuier et classification en niveau d'intérêt en tant que candidats pour la biosynthèse de FCs.** Niveaux moyens, écart type standard et différentiels entre compartiments de l'expression dans le latex des fruit, pétiole et tronc de 26 P450s (d'après Kitajima *et al.*, 2018). Les P450s du clan 71 chez le figuier ont été blastés localement sur la banque transcriptomique partagée par R. Munakata (RISH, Kyoto University) (e-value < 0,1) via le logiciel Bioedit. Les transcrits ont été comparés un à un avec leurs séquences génomiques pour confirmer leurs annotations. Le statut de duplication a été défini par comparaison avec le CYPome de *M. notabilis*. Abréviation : \_pg = séquence définie comme pseudogène d'après sa séquence génomique (e.g. pas de codon start ou stop, séquences tronquées). Le niveau d'intérêt mineur (1) ou majeur (2) est défini par le fait que le P450 fasse partie d'une sous-famille récemment dupliquée chez *F. carica* comparativement au CYPome de *M. notabilis* ou (1) / et (2) est préférentiellement exprimé dans le latex de pétiole.

Attribution	Fruit		Pétiole		Trunk		Latex compartment preference	Duplication status in comparison with the <i>M. notabilis</i> CYPome	Gene interest level
	Average	SD	Average	SD	Average	SD			
Clan 51	<b>CYP51G1</b>	12,7	1,0	17,0	2,2	15,8	0,4	Ubiquitous	
	<b>CYP71BG22</b>	7,3	0,6	5,2	1	6,4	0,4	Fruit	Duplication 1
	<b>CYP71A163</b>	0,0	0,0	1,0	0,2	0,2	0,1	Pétiole	1
	<b>CYP71A164</b>	16,8	1,6	21,2	1,7	24,9	1,0	Ubiquitous	0
	<b>CYP71AP90</b>	0,1	0,1	1,3	0,4	0,6	0,1	Pétiole	Triplification 2
	<b>CYP71AN87</b>	0,0	0,0	1,4	0,3	0,3	0,3	Pétiole	2
	<b>CYP71AN91_pg</b>	4,3	0,5	4,7	0,4	5,9	0,6	Ubiquitous	1
	<b>CYP71AN94_pg</b>	0,0	0,0	2,2	0,5	0,3	0,3	Pétiole	2
	<b>CYP71GH2</b>	10,6	0,7	11,8	0,4	11,9	1,6	Ubiquitous	0
	<b>CYP71GH3</b>	246,6	10,7	208,1	8,8	215,4	19,6	Ubiquitous	0
	<b>CYP71B130</b>	0,0	0,0	2	0,5	0,4	0,3	Pétiole	Sextuplication 2
	<b>CYP71B131b</b>	0,0	0,0	1,6	0,4	0,2	0,1	Pétiole	2
	<b>CYP73A410_pg</b>	6,9	0,8	38,7	12,9	5,1	0,7	Pétiole	Conservation 1
	<b>CYP75B225_pg</b>	0,1	0,0	1,4	0,4	0,4	0,1	Pétiole	Conservation 1
Clan 71	<b>CYP76F112</b>	1,8	0,4	16,0	4,1	1,1	0,2	Pétiole	Duplication 2
	<b>CYP78A620</b>	2,3	0,3	1,3	0,2	1,9	0,5	Ubiquitous	0
	<b>CYP78A622</b>	0,2	0,2	0,3	0,1	0,4	0,1	Trunk	0
	<b>CYP81BN28</b>	0,1	0,0	0,6	0,1	1,2	0,8	Trunk	Conservation 0
	<b>CYP82C169</b>	0,0	0,0	0,4	0,1	2,3	1,9	Trunk	Conservation 0
	<b>CYP84A16</b>	0,0	0,0	0,8	0,3	0,6	0,2	Pétiole	Conservation 0
	<b>CYP98A231_pg</b>	0,1	0,0	1,4	0,5	0,7	0,3	Pétiole	Triplification 2
	<b>CYP701A177a</b>	0,0	0,0	0,8	0,3	0,6	0,2	Pétiole	2
	<b>CYP701A177b</b>	0,0	0,0	0,7	0,1	0,6	0,3	Pétiole	2
	<b>CYP706C178</b>	2,0	0,2	1,1	0,3	1,4	0,2	Ubiquitous	Conservation 0
	<b>CYP736A406</b>	0,1	0,1	1,1	0,4	0,6	0,2	Pétiole	2
	<b>CYP736A411_pg</b>	0,0	0,0	1,4	0,5	0,7	0,4	Pétiole	Duplication 2

### 3.2.6- Organisation génomique : Présence d'un cluster de biosynthèse pour les FCs ?

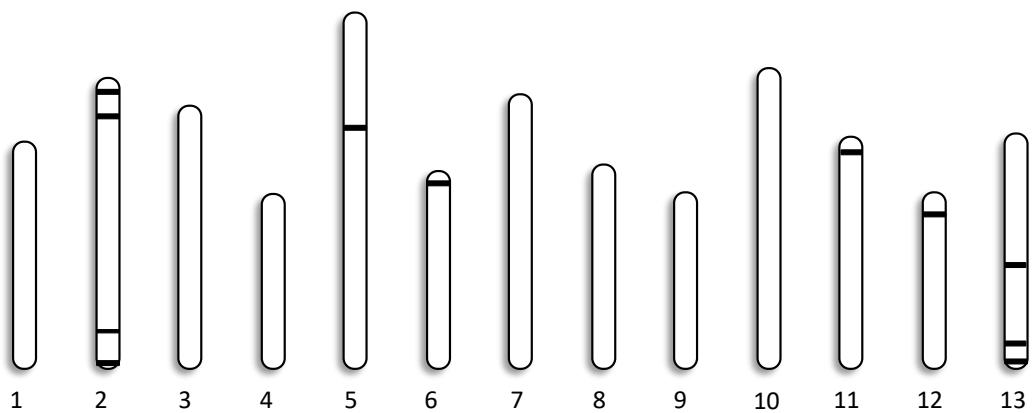
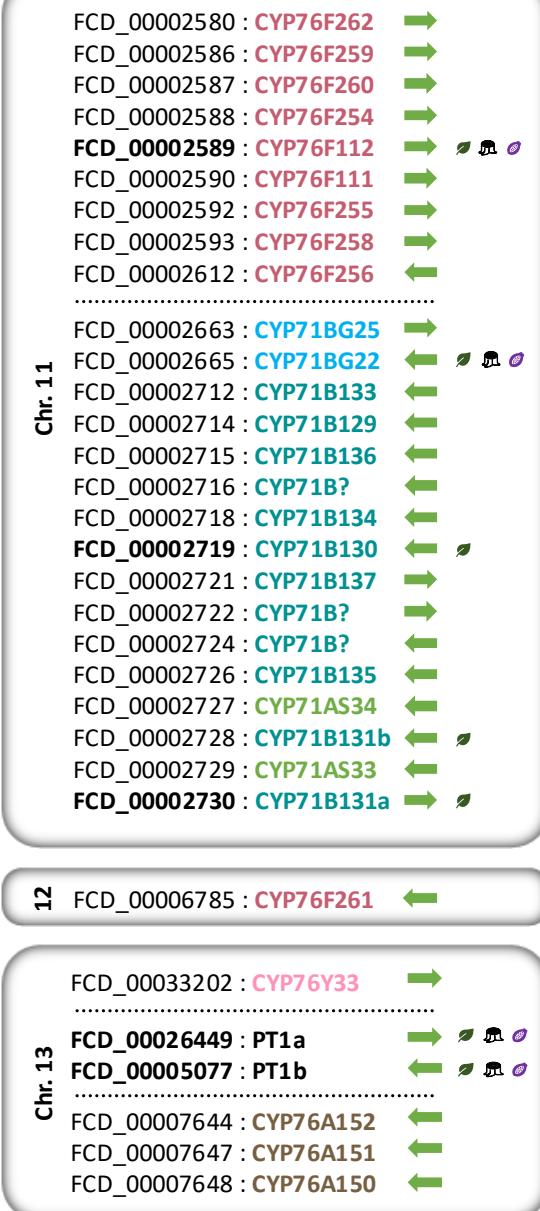
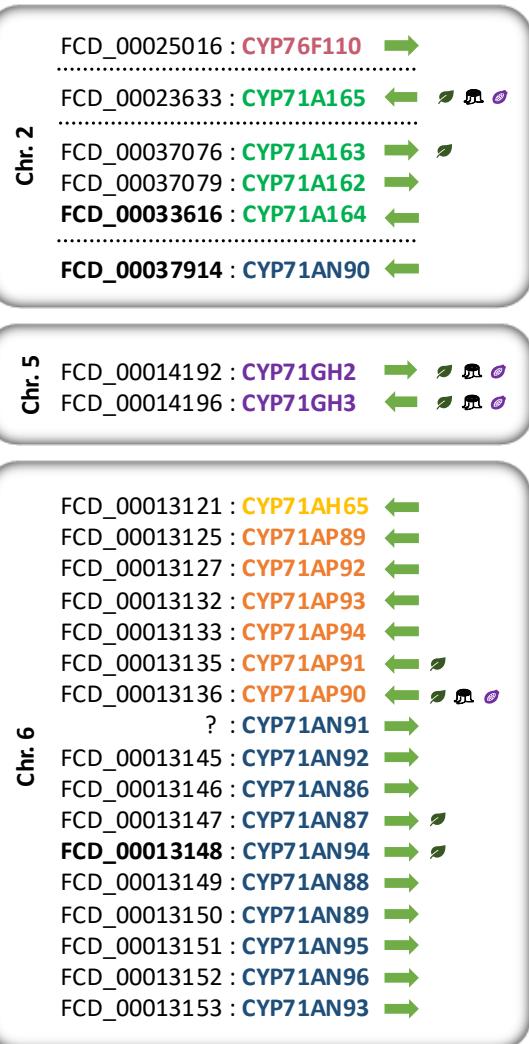
Un certain nombre des travaux ont démontré que plusieurs gènes (parfois tous) impliqués dans la synthèse d'un MS pouvaient être regroupés sur le génome pour former un cluster de biosynthèse comme synthétisé par Field *et al.* (2011) ou H. Nützmann *et al.* (2016). Lorsqu'on est en présence de ce type d'organisation, il peut être relativement aisés d'identifier l'intégralité de la voie de biosynthèse (**Figure 29**). Pour aller plus loin dans

la compréhension de la voie de biosynthèse des FCs et trouver de nouveaux gènes candidats, je me suis intéressé au positionnement des séquences codantes de certains P450s dans le génome du figuier. L'analyse de l'organisation génomique des CYP71Bs a permis d'identifier leur proximité chromosomique avec les CYP76Fs déjà impliqués dans la voie de biosynthèse des FCs linéaires chez le figuier (Villard *et al.*, 2021).

Les familles CYP71s et CYP76s sont les plus dupliquées chez le figuier comparativement au murier noble. Les CYP71s sont présents sur les chromosomes 2, 6 et 11, tandis que les CYP76s sont présents sur les chromosomes 2, 11, 12 et 13. L'analyse de l'organisation des CYP71s révèle un regroupement par sous-famille à l'exception de quelques paralogues de CYP71As et CYP71ANs qui sont également présents sur le chromosome 2. La même organisation est retrouvée pour les CYP76s. Cet agencement révèle l'impact de la duplication en tandem dans le déploiement des CYP71s notamment pour les sous-familles CYP71Bs, CYP71ANs et CYP71APs. Ces duplications peuvent suivre une orientation sens ou antisens avec une orientation généralement conservée comparativement aux séquences d'origines. L'analyse croisée de la phylogénie et de l'organisation chromosomique des sous-familles CYP71As et CYP71ANs nous révèle que les séquences détachées des clusters de paralogues impliquent des duplications récentes (e.g. CYP71AN90, CYP71A165). En effet, ces séquences présentent des redondances fonctionnelles et une grande similarité de séquence intragénique avec certains des membres du cluster de parologue. Ces sous-familles les plus dupliquées sont localisées dans des régions subtélomériques.

### Transcripts profil in latex

- Petiole
- Trunk
- Fruit



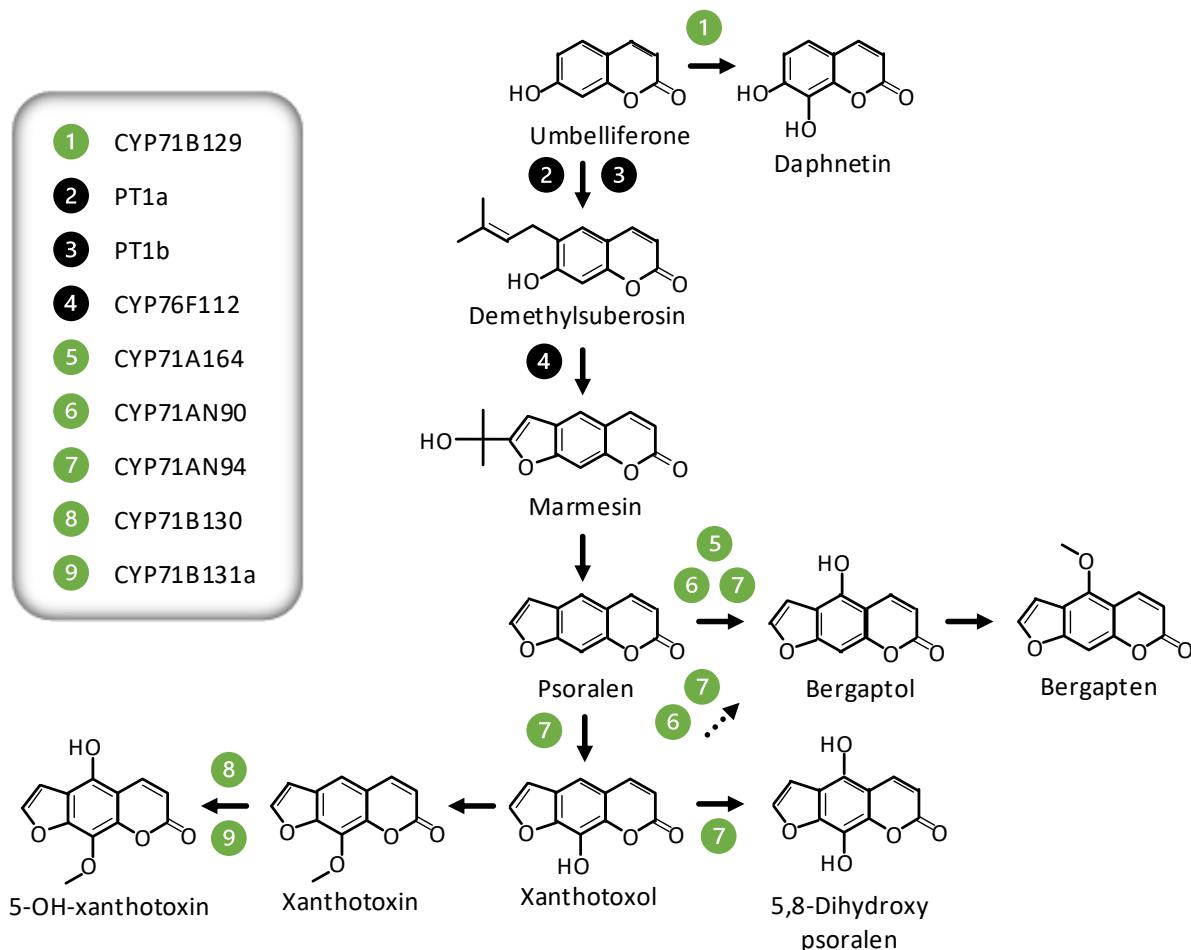
**Figure 29: Schéma de l'organisation chromosomique et du profil d'expression dans le latex des cytochromes P450s (P450s) des familles CYP71s et CYP76s dans le génome de *Ficus carica* (Moracées).** Les sous-familles CYP71 -A, -AH, -AN, -AP, -AS, -B, -BG, -GH et CYP76 -F, -Y, -A sont décrites en fonction d'un code couleur. Les noms officiels des gènes sont précédés de la mention : FCD\_ ; plateforme Ensembl Plant. Les noms en gras représentent les enzymes fonctionnellement caractérisées dans la voie de biosynthèse des FCs linéaires chez le figuier. Les profils d'expression des gènes dans le latex de pétiole, de tronc et de fruit chez le figuier sont dépeints par des symboles (d'après la banque de données de nos collaborateurs Japonais du RISH de l'Université de Kyoto) (Kitajima *et al.*, 2018). Les flèches vertes décrivent l'orientation (sens ou antisens) des gènes sur le génome.

### 3.3- Discussion et perspectives

#### 3.3.1- Caractérisation de la voie de biosynthèse des FCs chez *F. carica*

L'un de mes objectifs a été d'identifier de nouveaux P450s impliqués dans la voie de biosynthèse des FCs chez *F. carica*. Ma contribution a permis d'identifier la fonction de 6 enzymes impliquées dans la voie de biosynthèse des coumarines et des FCs (i.e. CYP71B129, CYP71B130, CYP71B131a, CYP71AN90, CYP71AN94 et CYP71A164) (**Figure 30**). Ces enzymes catalysent, en conditions *in vitro*, 5 à 6 réactions enzymatiques différentes dont potentiellement 3 nouvelles (i.e. xanthotoxol isomérase, xanthotoxol hydroxylase et daphnétine synthase). Je discute dans cette partie de l'apport de ces identifications fonctionnelles dans la compréhension de la voie de biosynthèse des FCs chez *F. carica*.

- 1 CYP71B129
- 2 PT1a
- 3 PT1b
- 4 CYP76F112
- 5 CYP71A164
- 6 CYP71AN90
- 7 CYP71AN94
- 8 CYP71B130
- 9 CYP71B131a



**Figure 30: Schéma simplifié des étapes de biosynthèses identifiées dans la voie des coumarines et des furocoumarines linéaires chez *Ficus carica* (Moracées).** Les enzymes en vert sont décrites dans le cadre de ce manuscrit. Les enzymes en noir sont décrites d'après (Munakata *et al.*, 2020; Villard *et al.*, 2021).

**Les CYP71B** - Parmi les CYP71Bs étudiés, CYP71B130 et CYP71B131a permettent la conversion de la xanthotoxine en 5-hydroxyxanthotoxine (5OHX) (Figure 30). Dans le cadre de la voie de biosynthèse des FCs chez le figuier, ces deux enzymes pourraient faciliter la production de dérivés de xanthotoxine (e.g. 5OHX, isopimpinelline, cnidiline, etc.). Je n'ai identifié dans la littérature la présence d'aucun de ces dérivés potentiels directs. Il est donc possible que l'activité 5OHXS en condition *in vitro* de ces deux enzymes ne soit pas le reflet d'une fonction physiologique chez le figuier, malgré leur haut degré d'affinité avec la xanthotoxine (CYP71B130 :  $Km_{app}$  de  $0.9 \pm 0.7 \mu\text{M}$  ; CYP71B131a :  $Km_{app}$  de  $9.4 \pm 1.2 \mu\text{M}$ ). Toutefois, leur niveau d'affinité pour la xanthotoxine est relativement similaire en conditions *in vitro* à celui d'autres 5OHXS pour des espèces

capables d'accumuler des dérivés de la xanthotoxine comme l'isopimpinelline et la cnidiline chez *Citrus paradisi* et *Citrus hystrix* (Rutacées) (Dugrand-Judek *et al.*, 2015; Limones-Mendez *et al.*, 2020), et l'isopimpinelline chez *Pastinaca sativa* (Ariacées) (Krieger *et al.*, 2018; Kviesis *et al.*, 2019). Une analyse quantitative de la présence de 5OHX et d'autres dérivés potentiels de la xanthotoxine, ainsi que de l'expression de CYP71B130 et CYP71B131a dans le latex de *F. carica* pourrait permettre de mieux comprendre le rôle de ces enzymes dans la physiologie du figuier.

Par ailleurs, CYP71B129 est capable de convertir l'umbelliférone en daphnétine (ou en 3,7-dihydroxycoumarine) (**Figure 30**). Une analyse bibliographique fine n'a pas permis d'identifier de cas où la présence de daphnétine chez *F. carica* est rapportée. CYP71B129 présente un niveau d'affinité moyen avec l'umbelliférone en conditions *in vitro* ( $Km_{app}$  de  $28.5 \pm 9.3 \mu\text{M}$ ). Il est donc possible que l'activité umbelliférone hydroxylase en condition *in vitro* ne soit pas le reflet d'une fonction physiologique chez le figuier. Néanmoins, la daphnétine a déjà été identifiée chez *Brosimum rubescens*, une espèce de Moracées relativement proche du genre *Ficus* (Braz filho *et al.*, 1972). Les données obtenues dans le cadre de mes travaux ne permettent pas de conclure quant à un quelconque rôle physiologique de CYP71B129 chez *F. carica*.

**Les CYP71AN et CYP71A** - En condition *in vitro*, CYP71AN94 semble capable d'hydroxyler le psoralène pour donner du bergaptol (carbone C5) et du xanthotoxol (carbone C8), mais également un produit qui pourrait correspondre à du 5,8-dihydroxypsoralène (**Figure 30**). Le fait que CYP71AN94 permette la synthèse d'un produit en présence de xanthotoxol mais pas de bergaptol suggère une activité catalytique séquentielle impliquant une hydroxylation d'abord en C8 puis en C5. Lorsqu'il est incubé avec du xanthotoxol, CYP71AN94 semble produire, en plus du potentiel 5,8-dihydroxy-psoralène, une quantité limitée de bergaptol. Ce produit révèle une possible activité d'isomérisation du xanthotoxol en bergaptol. Ce résultat nécessite cependant des investigations complémentaires et pourrait être un artéfact expérimental. En effet, la production de bergaptol pourrait également résulter d'une deshydroxylation du 5,8-dihydroxy-psoralène médiée par CYP71AN94, ou bien résulter de la présence infime de psoralène dans le standard purifié de xanthotoxol. Je note néanmoins, que cette activité ne semble pas provenir d'une activité enzymatique liée aux microsomes de

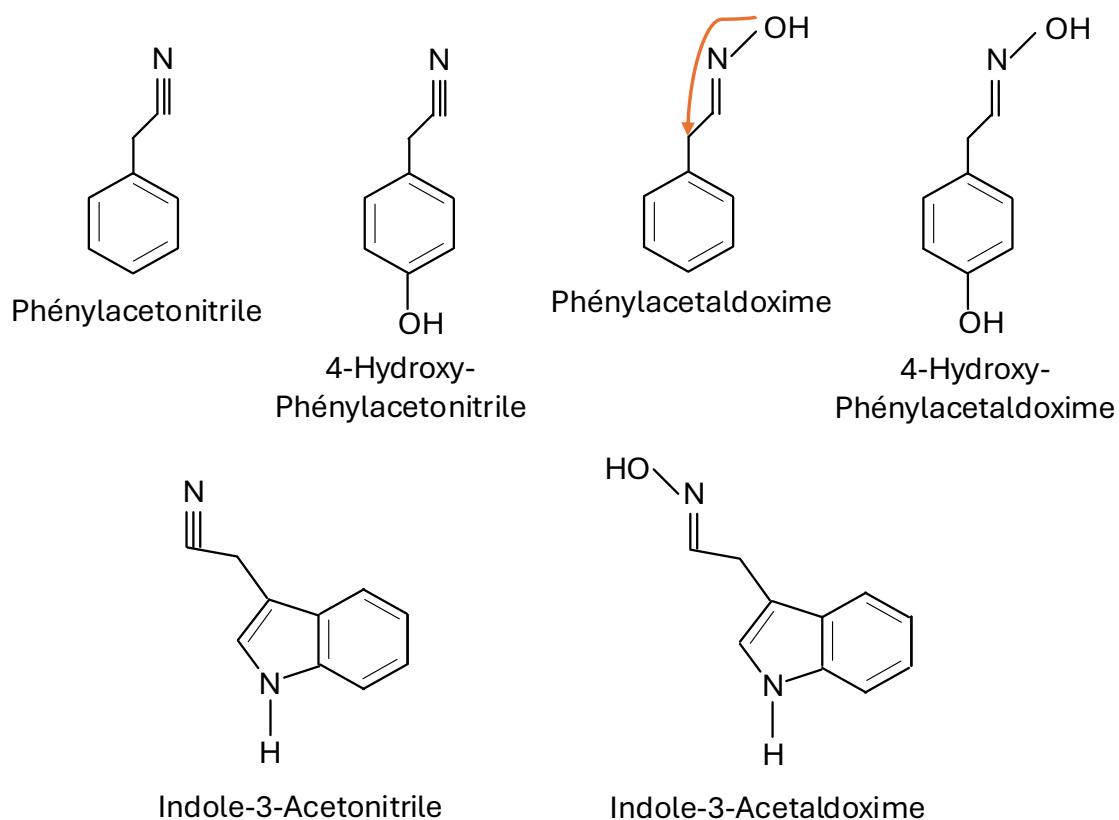
levures (témoin pYeDP60 vide). Je note également que cette bioconversion n'intervient qu'en présence du cofacteur NADPH, ce qui est typique des P450s eucaryotes (Omura & Sato, 1964). Dans les mêmes conditions d'incubation que pour CYP71AN94, les enzymes CYP71AN90 et CYP71A164 semblent capables en présence de NADPH de produire du bergaptol à partir de psoralène. CYP71AN90, comme CYP71AN94, serait aussi en mesure de produire du bergaptol en présence de xanthotoxol. Pour tous ces P450s et à ce stade, les analyses réalisées n'ont pas permis de trancher formellement sur la nature des produits. Des investigations plus poussées (MS/MS et/ou par RMN) seront nécessaires pour permettre une identification formelle.

Dans l'éventualité où ces fonctions enzymatiques venaient à être confirmées, plusieurs d'entre elles pourraient s'avérer inédites, tandis que d'autres ont déjà été décrites chez d'autres espèces et sous-familles de P450s. Par exemple, la fonction bergaptol synthase a très récemment été décrite pour CYP71AZ18 chez *Angelica dahurica* (Ji et al., 2024). Cette fonction pourrait néanmoins être inédite chez les CYP71ANs. Parmi cette sous-famille, seule CYP71AN24 a déjà été caractérisé comme capable de bioconvertir 6 aldoximes de tailles variables chez *Prunus mume* (Rosacées) (Yamaguchi et al., 2014). Ces structures sont proches de celle de l'auxine et d'hydroxycinnamates et partagent pour partie une certaine similarité structurale avec le psoralène et le xanthotoxol (**Figure 31**).

D'autre part, la fonction xanthotoxol synthase a déjà été décrite pour CYP71AZ4 chez *Pastinaca sativa* (Ariacées) (Krieger et al., 2018) et très récemment pour CYP71AZ19 et CYP83F95 chez *Angelica dahurica* (Ariacées) (Ji et al., 2024). Cette fonction pourrait néanmoins être inédite pour des CYP71ANs.

Enfin, la potentielle fonction xanthotoxol isomérase reste une fonction très surprenante et, en même temps, peu réaliste. Une attention particulière à l'identification de potentiels biais analytiques ayant conduit à la détection de cette activité enzymatique devra être portée pour confirmer ou non cette fonction. En effet, l'isomérisation est une fonction très rarement décrite chez des P450s (Hansen et al., 2021), bien que quelques exemples existent pour la voie des aldoximes et des terretonines (meroterpénoides) (Clausen et al., 2015; Matsuda et al., 2015). Toutefois, je relève que CYP71AN24 est capable de transférer un groupement hydroxyle sur un autre carbone (Yamaguchi et al.,

2014) (Figure 31). Cette spécificité catalytique pourrait être similaire avec la fonction isomérase potentielle de CYP71AN94.

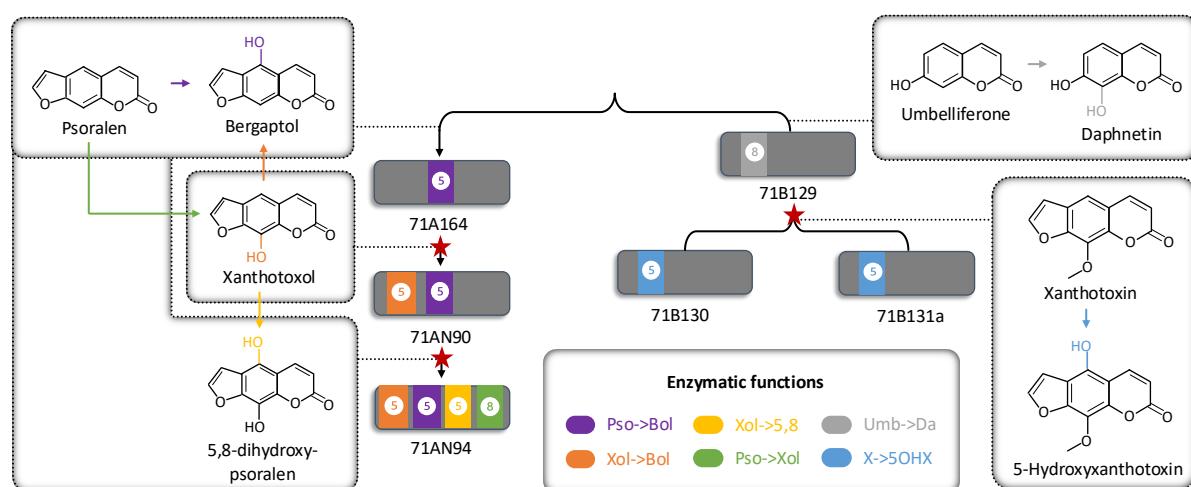


**Figure 31: Substrats métabolisables par CYP71AN24 en condition *in vitro*.** Cette enzyme provient de l'abricotier du japon (*Prunus mume*, Rosacées). Les structures décrites correspondent à des aldoximes et sont des intermédiaires de la voie des glycosides cyanogéniques. La flèche en orange décrit le déplacement du groupement hydroxyle médié par CYP71AN24 pour produire du mandelonitrile à partir de phénylacetaldoxime. D'après (Yamaguchi *et al.*, 2014).

### 3.3.2- Mieux comprendre l'évolution de la voie des FCs pour mieux la décrire

La description de la voie de biosynthèse des FCs a permis de démontrer que ses origines dépendaient d'une évolution convergente chez les angiospermes (Limones-Mendez *et al.*, 2020; Munakata *et al.*, 2020), et par conséquent d'une acquisition récente des gènes de biosynthèse chez les Moracées (Villard *et al.*, 2021). Dans le cadre de ma thèse, la

recherche de P450s candidats a principalement été axée sur l'étude de sous-familles récemment dupliquées donc potentiellement soumises à des processus de néofonctionnalisation ou de subfonctionnalisation (Panchy *et al.*, 2016). Cette stratégie a porté ses fruits et a permis l'identification de 4 à 5 nouvelles activités enzymatiques dans la voie de biosynthèse des FCs au sein de la famille CYP71 chez le figuier. L'étude des fonctions catalytiques UH et 5OHXS ont permis de confirmer un évènement de néofonctionnalisation apparu au cours de la diversification des CYP71Bs, et d'en proposer 2 autres potentiels parmi les CYP71As et les CYP71ANs (**Figure 32**). Dans la continuité de cette stratégie, je suggère de considérer les gènes CYP71A163, CYP71AN87, CYP736A406 et CYP736A411\_pg comme de potentiels candidats à la voie des FCs (**Tableau 1**). En effet, ces isoformes sont préférentiellement exprimées dans le latex de pétiole chez le figuier et appartiennent à des sous-familles abondamment dupliquées comparativement à *M. notabilis*. L'appartenance de CYP71A163, CYP71AN87 et CYP71AN91\_pg aux mêmes sous-familles que CYP71A164, CYP71AN90 et CYP71AN94 pourrait suggérer que ces P450s portent des fonctions redondantes à celles déjà partiellement identifiées. Enfin, je porte une attention particulière sur CYP736A406 et CYP736A411\_pg qui appartiennent à la sous-famille des CYP736As pour laquelle des isoformes ont très récemment été identifiées comme impliquées dans le début de la voie des FCs linéaires et angulaires chez *Angelica sinensis* (Apiacées) (K. Wang *et al.*, 2024).



**Figure 32: Schéma de synthèse de l'évolution potentielle des fonctions de quelques P450s dans la voie des FCs linéaires chez *F. carica*.** Les boîtes grises correspondent à des cytochromes P450s. Les couleurs à l'intérieur représentent les fonctions enzymatiques

associées. Les chiffres 5 ou 8 représentent le carbone ramifié. Les étoiles rouges correspondent à des évènements de néofonctionnalisation. Abréviations : Umb : umbelliférone ; Da : Daphnétine ; Pso : Psoralène ; Xol : Xanthotoxol ; 5,8 : 5,8-dihydroxypsoralène ; Bol : Bergaptol. À noter que les fonctions attribuées aux CYP71As et CYP71ANs doivent encore être confirmées par l'identification formelle des produits de biosynthèse.

Tous les P450s identifiés dans la voie des FCs correspondent à des enzymes appartenant à des sous-familles richement dupliquées. La vision émergente d'une promiscuité enzymatique<sup>189</sup> importante chez les P450s et notamment dans le clan 71 (Werck-Reichhart, 2023), suggère d'élargir le champ de recherche de candidats au profit d'enzymes peu dupliquées. En effet, la duplication est un phénomène qui peut induire un relâchement de pression de sélection sur les enzymes présentant une fonction redondante (Flagel & Wendel, 2009; Panchy *et al.*, 2016). Ainsi, l'acquisition d'une nouvelle fonction pourrait également intervenir par un processus de subfonctionnalisation au sein de P450s présentant une fonction redondante avec les enzymes récemment dupliquées. Le contexte de forte duplication du CYPome du figuier constitue un élément en faveur de cette hypothèse de travail. Ainsi, je suggère d'élargir la recherche de candidats à la voie des FCs en considérant également les P450s de familles peu ou pas dupliquées et dont certaines isoformes sont exprimées préférentiellement dans le latex de pétiole. Ce cible amène à considérer CYP98A231\_pg, CYP701A177a, CYP701A177b, CYP73A410\_pg et CYP84A16 comme des candidats potentiels pour la biosynthèse de FCs (**Tableau 3**).

L'organisation des CYP71s et des CYP76s dans le génome de *F. carica* a permis de révéler que les enzymes impliquées dans la voie de biosynthèse des FCs linéaires chez le figuier n'étaient pas rassemblées sous la forme d'un cluster unique. Le constat est le même pour les deux diméthylsubérosine synthases impliquées dans le début de la voie des FCs linéaires (Munakata *et al.*, 2020). Ainsi, comme chez les Apiacées (X.-C. Huang *et al.*, 2024), les étapes de la voie de biosynthèse des FCs linéaires chez *F. carica* ne sont pas rassemblées au sein d'un cluster de biosynthèse unique. Je note néanmoins que la plupart des clusters de paralogues dupliqués en tandem se trouvent à des locus subtélomériques (e.g. chromosomes 6, 11) (**Figure 29**). Ces régions constituent des

---

<sup>189</sup> Promiscuité enzymatique : Capacité d'une enzyme à catalyser plusieurs réactions différentes.

zones dynamiques pouvant favoriser le rassemblement en cluster de biosynthèse par recombinaisons ectopiques (Field *et al.*, 2011; H. Nützmann *et al.*, 2016). Actuellement, aucune donnée n'est disponible concernant la biosynthèse de FCs angulaire chez *F. carica*. Chez certaines Apiacées comme *Peucedanum praeruptorum*, les 12 gènes impliqués dans cette voie (*PpC2'Hs*, *PpPTs* et *PpOC*) sont rassemblés sous forme d'un cluster de biosynthèse (X.-C. Huang *et al.*, 2024). L'identification de P450s impliqués dans cette voie chez *F. carica* pourrait permettre de mieux comprendre l'organisation génomique des P450s dans le génome du figuier.

## 4-Conclusion du chapitre 3

En conclusion, dans ce chapitre j'ai décrit 6 nouvelles enzymes impliquées dans la voie de biosynthèse des coumarines et des FCs et 5 à 6 fonctions catalytiques en conditions *in vitro*. Cette caractérisation fonctionnelle permet de mieux comprendre la voie de biosynthèse des FCs linéaires chez le figuier. J'ai ensuite décrit l'évolution structurale de certains CYP71Bs comparativement à celle d'autres 5-hydroxyxanthotoxine synthases et de proches paralogues. Ce travail a permis de décrire comment des P450s impliqués dans la voie de biosynthèse des coumarines pouvaient être cooptés vers celle des FCs. En complément, un travail préliminaire de caractérisation fonctionnel de CYP71As et CYP71ANs m'a permis de proposer de nouvelles fonctions qui pourraient compléter des étapes centrales dans la voie des FCs chez *F. carica*. Mon travail a ainsi confirmé que considérer des P450s exprimés préférentiellement dans le latex de pétiole et appartenant à des sous-familles récemment dupliquées chez le figuier pouvait s'avérer être une stratégie efficace dans l'identification de gènes impliqués dans la biosynthèse de FCs linéaires.

# Conclusions générales et perspectives

Dans le cadre de ce travail de thèse, je me suis intéressé au modèle de la voie de biosynthèse des FCs (i) d'une part pour mieux comprendre son évolution chez le figuier à travers l'étude fonctionnelle de P450s et (ii) d'autre part pour mieux appréhender ses impacts sur le phénotype de la tomate en me restreignant à la voie du psoralène.

Les résultats obtenus dans le cadre de ce travail indiquent que considérer l'évolution récente du génome peut s'avérer être un outil utile pour décrire l'évolution récente d'une voie de biosynthèse. En ce sens, par une analyse comparative des CYPome de *F. carica* et de *M. notabilis* j'ai pu caractériser fonctionnellement 6 enzymes dans la voie des FCs linéaires. Si je devais poursuivre l'étude de la voie des FCs chez les Moracées, je tenterais de développer une image plus précise sur le plan génétique et métabolomique. Pour ce faire, je développerais une approche de phytochimie comparative à des espèces capables d'accumuler une plus ou moins grande diversité de FCs (e.g. genres *Dorstenia*, *Fatoua* et *Broussonetia*) et à d'autres encore non décrites (e.g. genres *Morus*, *Artocarpus* et famille des Urticacées). J'établirais une quantification absolue des métabolites identifiés de façon à pouvoir faire un parallèle entre le taux de duplication de certaines sous-familles de P450s et la quantité de FCs accumulées. Enfin, de la même façon que pour l'approche de génomique comparative entre *F. carica* et *M. notabilis*, je m'intéresserais à comparer l'évolution récente du CYPome entre *F. carica* et une espèce du genre *Dorstenia*. Le genre *Dorstenia* dispose d'une très grande diversité de FCs et sans doute l'une des plus importante décrite chez les angiospermes (**Tableau 1**). La mise en œuvre d'une telle analyse demanderait de séquencer un génome de *Dorstenia sp* (aucun disponible actuellement), d'analyser la phylogénie de son CYPome et d'identifier les duplications récentes entre *F. carica* et *Dorstenia sp*.

Si l'identification de candidats P450s chez *F. carica* par des considérations évolutives et d'expression localisée nous est devenue très accessible, tester la fonction de ces enzymes reste très complexe. Elle est souvent limitée par la production d'enzymes

fonctionnelles et par le criblage métabolique. En ce sens, si je devais poursuivre dans ce projet, j'essaierais d'améliorer l'approche de production enzymatique. La production hétérologue de P450s en levure est bien souvent faible, ce qui rend difficile le criblage fonctionnel et oblige à restreindre l'analyse à seulement quelques gènes d'intérêt. Je proposerais d'utiliser une approche de production acellulaire d'enzyme notamment via l'utilisation de kit (e.g. ALiCE®). Cette approche permet généralement d'obtenir des quantités très abondantes de protéines correctement repliées. Cette approche nécessiterait quelques optimisations comme l'expression simultanée de la coenzyme cytochromes P450s réductase, l'ajout du groupement prosthétique protoporphyrine IX, de membranes de synthèse, etc. En contrepartie, et dans le cas où cette approche serait satisfaisante, cela pourrait permettre de cribler plus de séquences d'intérêt, ce qui semble nécessaire pour mieux comprendre la voie des FCs, tant le nombre d'enzymes candidates est important. Des séquences qui pourraient être synthétisées par voie chimique et clonées par un prestataire pour passer directement du gène candidat sur le génome au criblage fonctionnel. Ensemble, ces pistes d'amélioration pourraient permettre de s'intéresser plus rapidement à la fonction des P450s tout en limitant le temps opérationnel.

La création de la lignée PP est une première étape vers la meilleure compréhension du rôle physiologique des FCs chez les plantes. L'obtention d'une seule lignée transformante qui accumule seulement sous forme de trace le psoralène ouvre la porte à de nombreuses questions, comme par exemple : (i) L'accumulation hétérologue de FCs nécessite-t-elle une optimisation du circuit de production dans l'environnement cellulaire de la tomate ? (ii) L'accumulation de psoralène ou l'un de ses intermédiaires est-il toxique pour la tomate ? Si je devais poursuivre ou repartir de zéro sur ce projet, je testerai premièrement la toxicité du psoralène et des intermédiaires de sa voie à différentes concentrations (e.g. 0.1, 1, 10, 100µM et 1mM) sur la germination et la survie de tomates de la même variété que la lignée PP en conditions de culture *in vitro*. Si les tomates croissent à ces concentrations, alors j'essayerais dans un deuxième temps d'optimiser les flux de la voie du psoralène dans le cadre d'une approche de bioingénierie métabolique plus précise. En guise d'amélioration, je tenterais d'orienter la localisation subcellulaire de la diméthylsubérosine synthase vers le RE en lui ajoutant une séquence

d'adressage et en remplaçant sa séquence d'ancrage avec celui d'un P450 par exemple. Cette enzyme est chez *Catharanthus roseus* intégrée dans les plastes (Munakata *et al.*, 2016). J'essaierais également d'ajouter les enzymes *PsC4H* et *Ps4CL* en provenance du panais au transgène pour renforcer la biosynthèse d'acide *p*-coumarique et de *p*-coumaroyl-CoA. Enfin, je remplacerais le gène CYP76F112 codant la marmesine synthase chez le figuier avec l'homologue de CYP736A360 chez le panais par comparaison avec celui identifié très récemment chez *A. sinensis* (K. Wang *et al.*, 2024). Ensemble, ces modifications permettront d'avoir uniquement des gènes en provenance du panais, ce qui pourrait éventuellement court-circuiter le channeling endogène de l'environnement cellulaire de la tomate en faveur de l'accumulation de psoralène à partir de *p*-coumarate. Enfin, ces modifications pourraient permettre de concentrer les enzymes de biosynthèse à l'interface entre le cytosol et le réticulum endoplasmique et potentiellement faciliter l'accumulation de psoralène. À l'échelle de la tomate et dans le cas où la biosynthèse de FCs s'avèrerait non toxique et permettrait une forte accumulation, je pourrais alors tenter de mieux comprendre comment les FCs impactent la physiologie de la gestion de ses ressources à l'interface croissance/défense. L'un des objectifs initiaux du projet lié à la lignée PP.

La caractérisation transcriptomique, métabolomique et morphophysiologique de la lignée PP a permis de décrire plusieurs impacts potentiellement liés à l'accumulation de coumarines sur la croissance de la tomate. Cette caractérisation constitue une cartographie assez large des répercussions de l'expression du transgène PP sur la physiologie et le phénotype de la tomate dans des conditions non stressantes. Soumettre la lignée PP à des stress biotiques et abiotiques pourrait permettre de mieux comprendre le rôle de la PP et/ou des coumarines et notamment de la scopolétine sur la résilience de la tomate. En ce sens, il serait intéressant de poursuivre la description du rôle d'une carence en fer sur la croissance et le développement de la lignée PP. D'autre part, les coumarines peuvent induire une restructuration du microbiote racinaire (Cosme *et al.*, 2021). Si je devais poursuivre ce projet et en considérant que la lignée PP accumule principalement de la scopolétine dans ses parties supérieures, j'essaierais d'identifier le rôle potentiel des coumarines dans la régulation du microbiote des parties supérieures qui à ma connaissance n'est pas décrit.

Enfin, d'un point sociétal, mieux décrire la biosynthèse de coumarines et FCs chez le figuier peut permettre d'identifier de nouvelles enzymes impliquées dans la biosynthèse de MS à fort potentiels (e.g. santé, biomatériaux) (Dumur, 2022; Y.-P. Liu *et al.*, 2020; Sumorek-Wiadro *et al.*, 2020). La production de molécules d'intérêt en cellules-usines semble être sujette à un regain d'intérêt récent (Rodrigues *et al.*, 2022; Z. Wang *et al.*, 2023; S.-M. Yang *et al.*, 2015).

# Matériel et Méthodes

Cette section est dédiée aux matériels et méthodes qui n'ont pas été décrits dans les articles.

## Matériel

### 1- Espèces modèles

#### A- Figuier

Les plants de *Ficus carica* utilisés pour l'extraction d'ARN proviennent du centre INRAE de Colmar (France). Les tissus ont été récoltés à partir de plants bouturés d'environ un an et cultivés dans du terreau.

#### B- Tomate

Les graines de *Solanum lycopersicum* var. *cerasiformae* 'West Virginia 106' (WVa106) utilisées pour la transformation des plantes ont été fournies par le laboratoire Biologie du Fruit et Pathologie (UMR 1332, INRAE - Bordeaux, France).

#### C- Tabac

Les graines de *Nicotiana benthamiana* proviennent de l'Institut de Biologie Moléculaire des Plantes (UPR 2357, CNRS - Strasbourg, France).

#### D- *Escherichia coli*

La souche d'*E. coli* MC1022 (génotype : *araD139*,  $\Delta$ (*ara*, *leu*)7697,  $\Delta$ (*lacZ*)M15, *galU*, *galK*, *strA*) est utilisée pour le clonage et l'amplification de plasmides. Elle est dérivée de la souche M182 (génotype :  $\Delta$ (*lacI*POZY) X74, *galK*, *galU*, *strA*) et permet une transformation rapide et efficace ainsi que l'amplification de grands plasmides.

La souche d'*E. coli* ccdB Survival™ (génotype : *F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG tonA::Ptrc-ccdB*) a été achetée chez Invitrogen™. Elle est utilisée pour le clonage et l'amplification de plasmides. Elle possède une résistance au produit toxique du gène *ccdB*, elle permet l'amplification de vecteurs tels que pYeDP60\_GW®.

## E- Agrobactéries

La souche *Agrobacterium tumefaciens* EHA105 (Hood *et al.*, 1993) est utilisée pour l'expression hétérologue transitoire en tabac. Cette souche hypervirulente d'agrobactéries porte dans son génome le gène *RifR*, qui confère une résistance à la rifampicine.

## F- Levures

La souche *Saccharomyces cerevisiae* WAT21 est utilisée pour produire tous les P450s de *Ficus*. Elle est dérivée de la souche W303 et a été modifiée pour permettre l'expression hétérologue de P450s de plantes. Le gène codant pour la NADPH cytochrome P450 réductase (CPR) endogène a été remplacé par l'un de ces homologues chez *A. thaliana*, le gène *AthR2*. Ce gène est placé sous le contrôle transcriptionnel du promoteur GAL10-CYC1, qui est inductible par le galactose, mais réprimé par le glucose. Ainsi, lorsque cette souche est transformée avec un plasmide pYeDP60 contenant un P450 et cultivée sur du galactose, elle surexprime à la fois le P450 et la CPR, permettant l'étude de l'activité du P450 dans la fraction microsomale. Comme la souche W303, la souche WAT21 possède une mutation sur le gène ADE2 qui l'empêche de croître sur un milieu dépourvu d'adénine (Pompon *et al.*, 1996; Urban *et al.*, 1997).

# 2- Plasmides

## a. pCR®8/GW/TOPO™

Le kit de clonage TA pCR®8/GW/TOPO™ (Invitrogen™) (Figure 33) a été utilisé pour le clonage des P450s. Il contient le vecteur pCR®8/GW/TOPO™ linéarisé et prêt à l'emploi. Ce vecteur possède des extrémités 3'-T (résidu de désoxythymidine) empêchant la circularisation et permettant la ligation directe des produits PCR amplifiés et rallongés d'une base A (résidu de désoxyadénosine) complémentaire aux extrémités 3' grâce à une *Taq* DNA polymérase. Ce procédé est connu sous le nom de clonage TA. De façon complémentaire, une topoisomérase I est liée de manière covalente au vecteur pCR®8/GW/TOPO™ et complète la ligation en libérant l'énergie nécessaire. Ce vecteur permet ainsi un clonage rapide et efficace des produits PCR. Le plasmide contient une origine pUC permettant une réplication à haute copie dans *E. coli*, ainsi qu'un gène de résistance à la spectinomycine pour une sélection efficace. Enfin, comme il possède les sites *attL1* et *attL2*, ce plasmide permet également l'accès à la technologie Gateway®, permettant le transfert par recombinaison du produit PCR dans d'autres vecteurs de destination Gateway®. Ce système est utilisé pour le sous-clonage des P450s dans le vecteur pYeDP60\_GW® (Krieger, 2014).

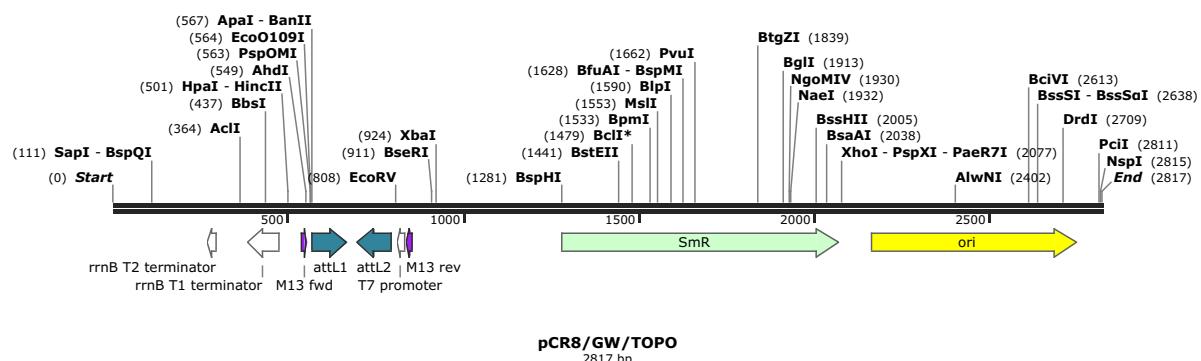
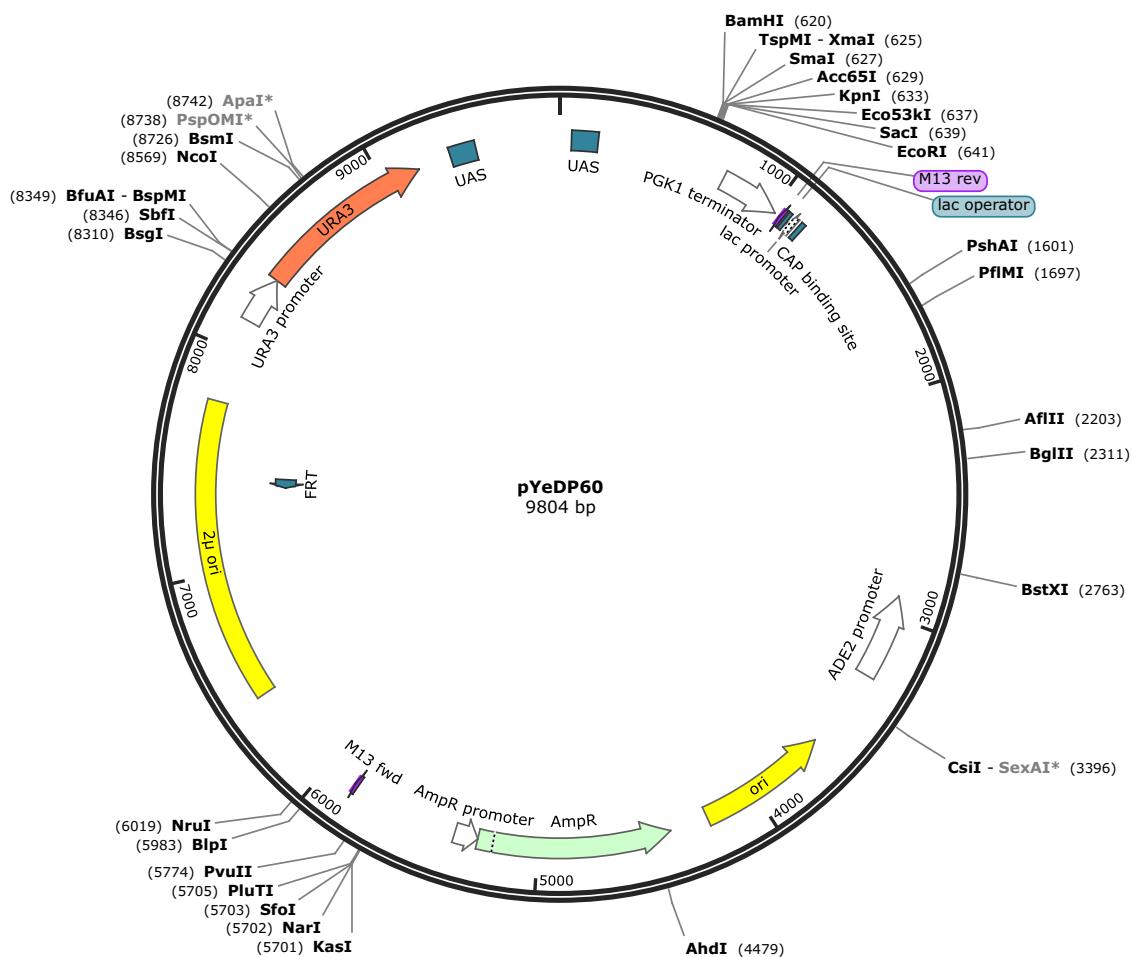


Figure 33: Carte du plasmide linéaire pCR®8/GW/TOPO™.

### b. pYeDP60\_GW

Le vecteur pYeDP60 (Figure 34) a été fourni par le Dr D. Pompon du Laboratoire d'Ingénierie des Protéines Membranaires (CNRS - Gif-sur-Yvette, France). Ce plasmide a été spécifiquement conçu pour l'expression de cytochromes P450 dans diverses souches de *Saccharomyces cerevisiae*, telle que WAT21 (Pompon *et al.*, 1996; Urban et

al., 1997). Il contient une origine de réplication 2 $\mu$  permettant la réplication dans la levure, ainsi que les marqueurs de sélection URA3 et ADE2 pour compléter l'auxotrophie de l'hôte levure. Il contient également une cassette d'expression constituée d'un site multiple de clonage *Bam*HI/*Sma*I/*Kpn*I/*Sac*I/*Eco*RI encadré par le promoteur GAL10-CYC1 en amont et par le terminateur PGK en aval. Le promoteur GAL10-CYC1 est induit par le galactose et réprimé par le glucose, ce qui permet de contrôler l'expression d'un gène d'intérêt cloné dans le site multiple de clonage. La présence d'une origine de réplication ColE1 et d'un gène de résistance à l'ampicilline permet également le sous-clonage, l'amplification et la sélection du vecteur pYeDP60 dans *E. coli*. Le vecteur pYeDP60 est utilisé pour exprimer les mutants de CYP71B131a.



**Figure 34: Carte plasmidique de pYeDP60.**

Le vecteur pYeDP60\_GW® (Figure 35) est dérivé du vecteur pYeDP60. Il a été construit par le Dr C. Krieger lors de sa thèse au LAE (Krieger, 2014) et résulte de l'ajout de séquences au vecteur pYeDP60 préexistant et permettant d'utiliser la technologie Gateway®. À cet effet, la cassette de recombinaison RfB, portant les sites *attR1* et *attR2*, a été insérée dans le site multiple de clonage en utilisant le site de restriction *BamHI*. Cette cassette contient également un gène de résistance au chloramphénicol et le gène *ccdB*. L'amplification du plasmide pYeDP60\_GW® vide doit donc être réalisée dans des bactéries résistantes à la toxine *ccdB*, telles que la souche *E. coli ccdB Survival*™. L'ajout de la technologie Gateway® permet un sous-clonage des gènes d'intérêt à partir de divers vecteurs de destination. Le vecteur pYeDP60\_GW® est utilisé pour exprimer divers cytochromes P450 clonés à partir de *F. carica*.

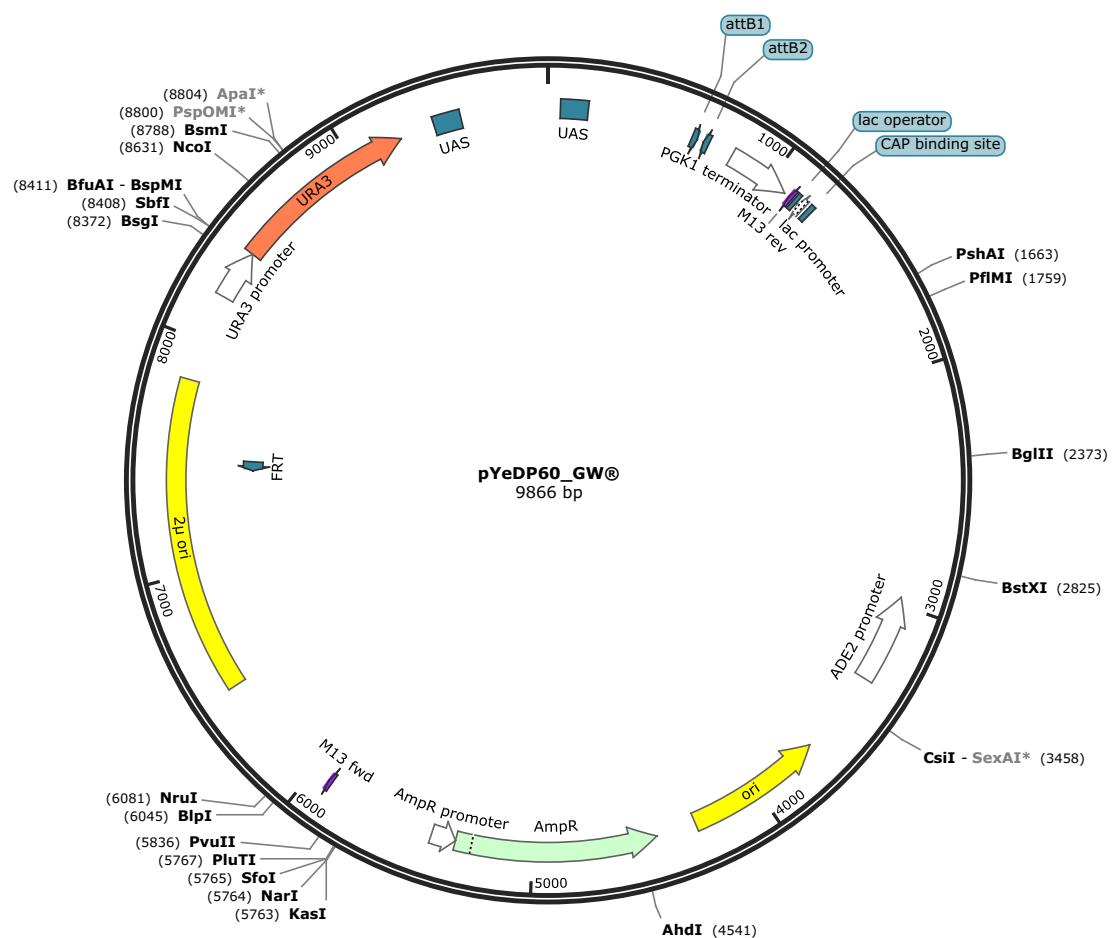


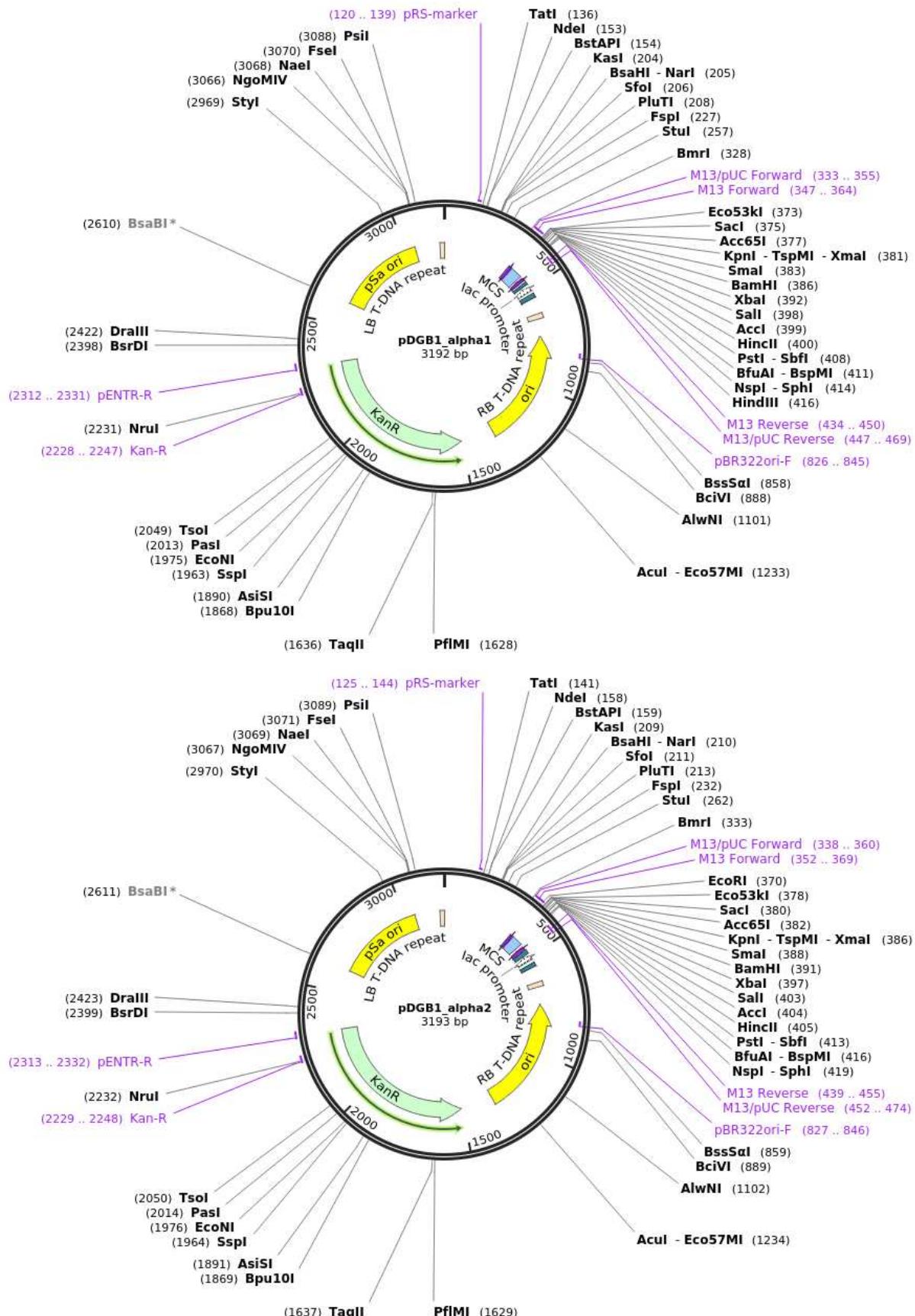
Figure 35: Carte plasmidique de pYeDP60\_GW®.

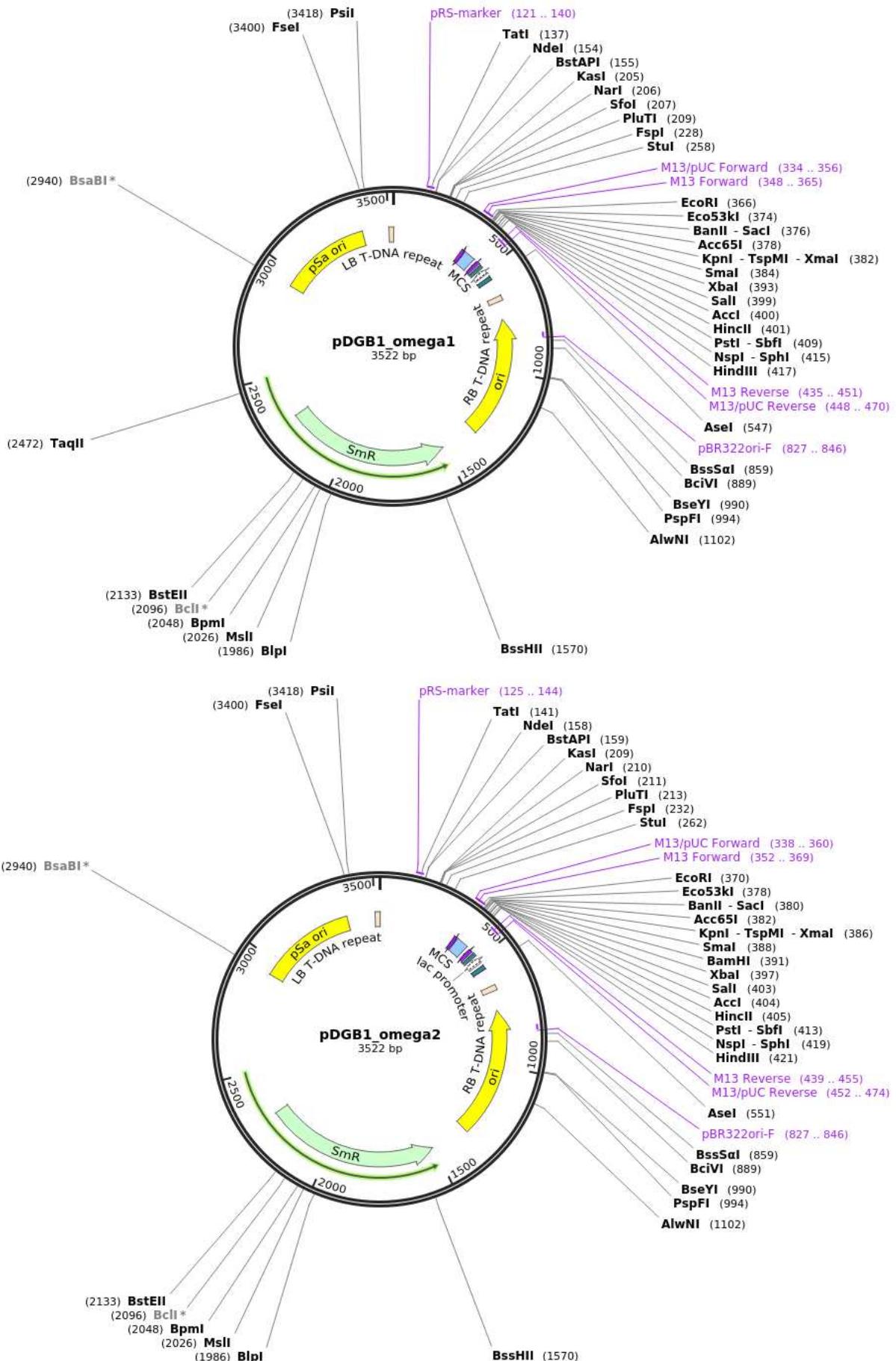
### c. Les vecteurs pDGB1 $\alpha$ et $\Omega$ du système GoldenBraid 2.0

Les vecteurs suivants sont des plasmides GoldenBraid commerciaux standardisés (Sarrion-Perdigones *et al.*, 2011, 2013). Ils ont été fournis par Addgene (laboratoire de Diego Orzaez - Cambridge, MA, USA) et sont utilisés pour la construction Goldenbraid (GB) et la transformation stable de la tomate.

Les plasmides pDGB1\_α1/pDGB1\_α2 et pDGB1\_Ω1/pDGB1\_Ω2 (**Figure 36**) sont des vecteurs binaires de niveaux  $\alpha$  et  $\Omega$  utilisés pour l'assemblage GoldenBraid (GB 2.0). Ils peuvent être directement utilisés pour la transfection de plantes via *Agrobacterium* ou simplement amplifiés en *E. coli*. Leurs marqueurs de sélection correspondent respectivement aux gènes de résistance à la kanamycine (niveau  $\alpha$ ) ou à la spectinomycine (niveau  $\Omega$ ). Tous contiennent le gène LacZ, encadré en amont et en aval par des sites de restriction d'enzymes de type IIS : *Bs*al et *Bsm*BI ; ils se distinguent par les extrémités laissées après la restriction par *Bsm*BI (niveau  $\alpha$ ) ou *Bs*al (niveau  $\Omega$ ). Les sites de reconnaissance *Bs*al et *Bsm*BI permettent l'assemblage de plusieurs éléments GB / unités transcriptionnelles en modules d'ordres supérieurs au sein du plasmide, en remplacement du gène LacZ. Ce site de reconnaissance est détruit lors du clonage. À titre d'exemple, les sites de reconnaissance *Bsm*BI permettent l'assemblage ordonné d'un module du plasmide donneur pDGB1\_α1 et d'un module du plasmide donneur pDGB1\_α2 dans un vecteur receveur de niveau pDGB1\_Ω. De la même façon, les sites de reconnaissance *Bs*al permettent l'assemblage ordonné d'un module du plasmide donneur pDGB1\_Ω1 et d'un module du plasmide donneur pDGB1\_Ω2 dans un vecteur receveur de niveau pDGB1\_α. Ces vecteurs contiennent des bordures gauche et droite (LB et RB) qui délimitent l'ADN-T pouvant être inséré dans le génome de la plante par une souche d'*Agrobacterium tumefaciens* telle que EHA105. La réPLICATION des plasmides GB dans *Agrobacterium* nécessite la présence d'un plasmide auxiliaire tel que le plasmide pSoup (**Figure 39**). La séquence complète du vecteur pDGB1\_α1 est disponible sur [https://gbcloning.upv.es/feature/pDGB1\\_alpha1/](https://gbcloning.upv.es/feature/pDGB1_alpha1/) (Plasmide Addgene #68224 ; <http://www.addgene.org/68224/> ; RRID:Addgene\_68224). La séquence complète du vecteur pDGB1\_α2 est disponible sur [https://gbcloning.upv.es/feature/pDGB1\\_alpha2/](https://gbcloning.upv.es/feature/pDGB1_alpha2/) (Plasmide Addgene #68225 ; <http://www.addgene.org/68225/> ; RRID:Addgene\_68225). La séquence complète du vecteur pDGB1\_Ω1 est disponible sur

[https://gbcloning.upv.es/feature/pDGB1\\_omega1/](https://gbcloning.upv.es/feature/pDGB1_omega1/) (Plasmide Addgene #68234 ; <http://www.addgene.org/68234/> ; RRID:Addgene\_68234). La séquence complète du vecteur pDGB1\_Ω2 est disponible sur [https://gbcloning.upv.es/feature/pDGB1\\_omega2/](https://gbcloning.upv.es/feature/pDGB1_omega2/) (Plasmide Addgene #68225 ; <http://www.addgene.org/68235/> ; RRID:Addgene\_68225).





**Figure 36: Cartes plasmidiques des vecteurs pDGB1\_α1, pDGB1\_α2, pDGB1\_Ω1 et pDGB1\_Ω2.**

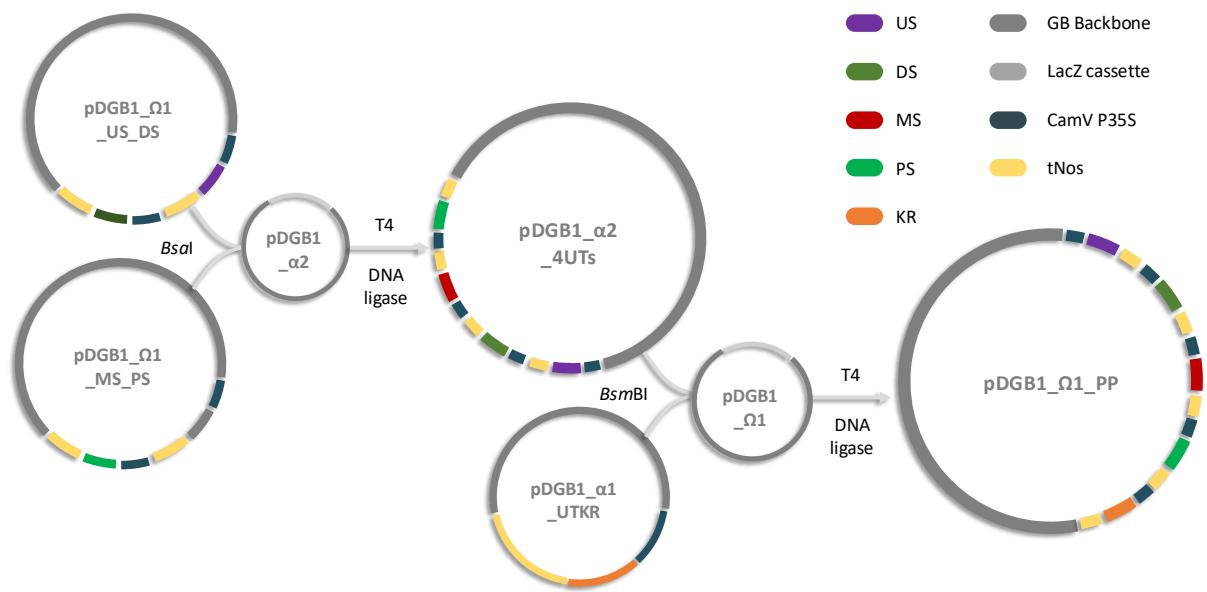
#### **d. Assemblage de la construction pDGB1\_Ω1\_PP**

Lors de travaux de thèse précédents, G. Galati et C. Villard ont généré les unités de transcriptions (UTs) de bases à la reconstitution de la voie du psoralène dans un système GB 2.0 (Galati, 2019; Villard, 2020). Plus spécifiquement les gènes de biosynthèse *PsUS* (ncbi: KY081678) (Roselli *et al.*, 2017), *PsDS* (ncbi : KM017083) (Munakata *et al.*, 2016), *FcMS* (ncbi : MW348922) (Villard *et al.*, 2021), *PsPS* (ncbi : C0SJS2) (Larbat *et al.*, 2009) et le gène de résistance à la kanamycine KR (Néomycine phosphotransférase II :NPTII) ont été assemblés sous le contrôle du promoteur P35S et du terminateur tNos. Ces briques ont par la suite été assemblées deux à deux pour former des constructions de niveaux Ω complémentaires 2UTs : pDGB1\_Ω1\_PsUS\_PsDS et pDGB1\_Ω2\_FcMS\_PsPS. En parallèle C. Villard a assemblé la construction pDGB1\_α1\_KR pour ajouter la résistance à la kanamycine dans le processus de sélection de la lignée transfectée. À partir de ces constructions, j'ai assemblé les 2 constructions 2UTs pour former une construction 4UTs dans un vecteur de niveau α (Figure 37). Enfin, j'ai assemblé les constructions pDGB1\_α1\_KR et pDGB1\_α2\_4UTs pour former la construction finale pDGB1\_Ω1\_PP (PP pour Psoralen Pathway) dans un vecteur de niveau Ω (Figures 37, 38 et 42). L'assemblage des constructions se fait par restriction/ligation. Ci-après je décris le mix réactionnel ayant permis l'assemblage à titre d'exemple de la construction finale pDGB1\_Ω1\_PP.

Le mélange réactionnel est composé de 1 µL de T4 DNA ligase (5 U/µL), 1 µL de tampon T4 DNA ligase (10x), 1 µL de *BsmBI* (10 U/µL), 100 ng du plasmide donneur pDGB1\_α1\_KR, 100 ng du plasmide donneur PDGB1\_α2\_4UTs, et 100 ng du plasmide receveur PDGB1\_Ω1 vide. Le tout est dilué dans un volume final de 10 µL avec de l'eau mQ. Le mélange réactionnel est incubé pendant 25 cycles composés d'une étape de digestion (37°C pendant 2 minutes) et d'une étape de ligation (16°C pendant 5 minutes).

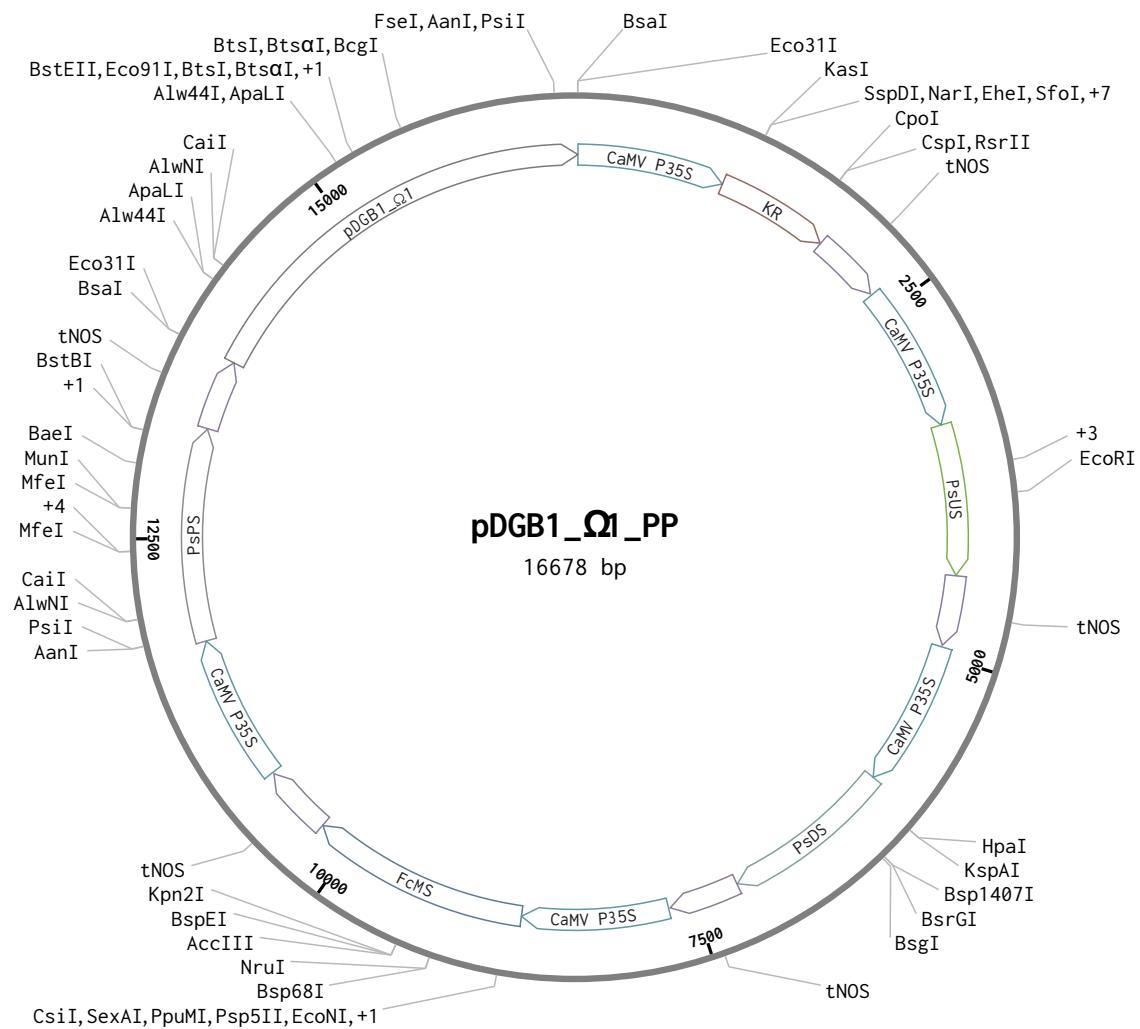
Lors des étapes de digestion, *BsmBI* clive les plasmides d'entrée et de destination, permettant l'assemblage ordonné de la construction KR suivi de celles contenant les 4UTs de biosynthèse, avec seulement 12 nucléotides entre chaque UT. Cette partie

composite de la construction pDGB1\_Q1\_PP est flanquée de sites de reconnaissance *Bsal* provenant du plasmide PDGB1\_Q1. De cette façon la construction pourrait être complétée à l'avenir d'autres UTs. Le plasmide recombinant résultant peut être introduit dans *E. coli*, pour être amplifié. La présence des 5 UTs a été confirmée par PCR, analyse de restriction et séquençage des UTs.



**Figure 37: Schéma simplifié de l'assemblage du plasmide pDGB1\_Q1\_PP.**

Abréviations : US : Umbelliférone Synthase ; DS : Diméthylsubérosine Synthase ; MS : Marmesine Synthase ; PS : Psoraléne Synthase ; KR : Résistance à la Kanamycine ; GB GoldenBraid ; UT : Unité de Transcription.



**Figure 38: Carte plasmidique du vecteur pDGB1\_Ω1\_PP.**

### e. pSOUP

Le vecteur pSoup (**Figure 39**) a été fourni par Mark Smedley (John Innes Center - Norwich, Royaume-Uni). Il est dérivé du vecteur pRK2 et est utilisé pour la co-transformation avec des plasmides de la famille pGreen, tels que les vecteurs pDGB. Il porte le gène de la répliqueuse pSa qui interagit en trans avec l'origine de réPLICATION pSa des vecteurs pDGB1- $\alpha$  et pDGB1- $\Omega$ , permettant leur réPLICATION dans les souches d'*Agrobacterium*. Le vecteur pSoup contient également un gène de résistance à la tétracycline comme marqueur de sélection.

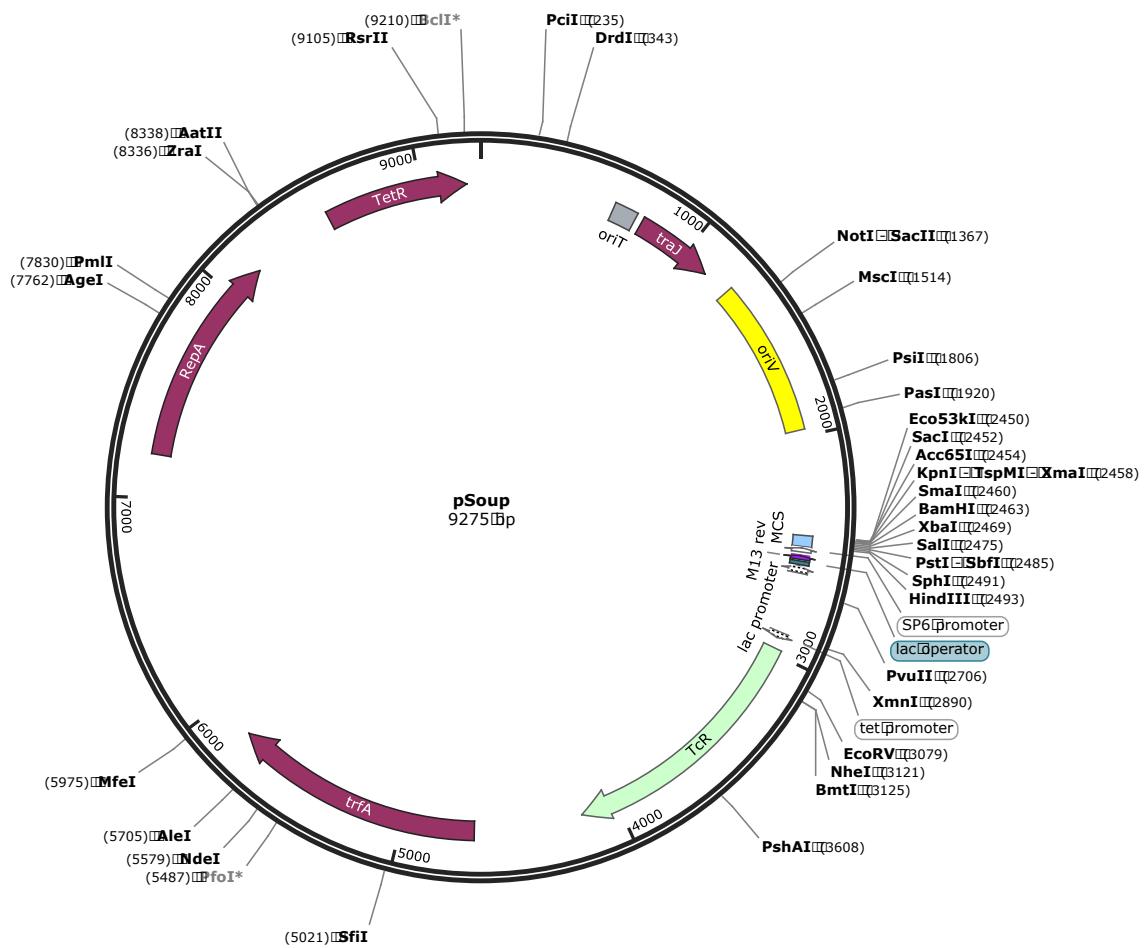
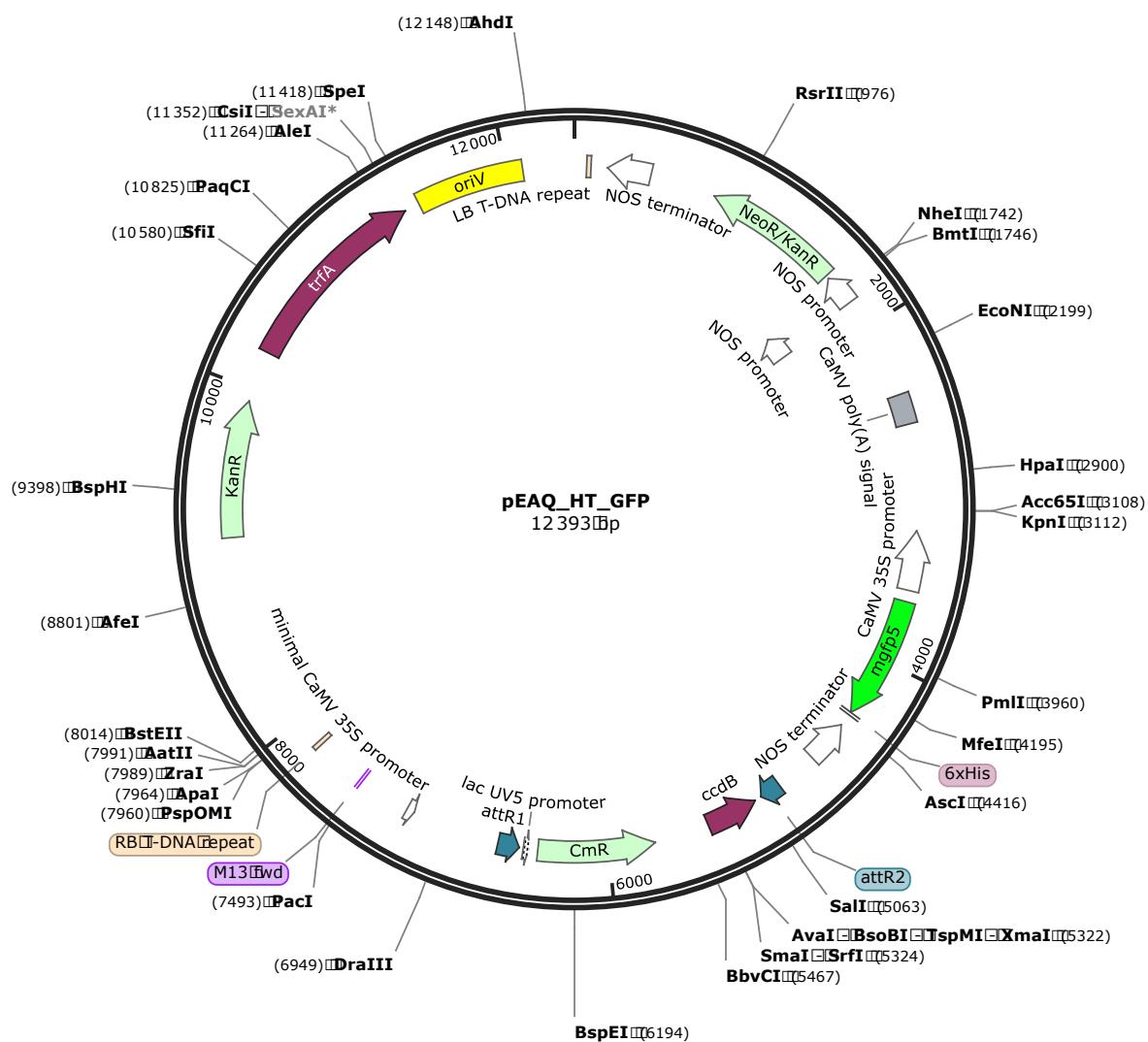


Figure 39: Carte plasmidique du vecteur pSOUP.

#### f. pEAQ-HT-GFP

Le vecteur pEAQ-HT-GFP (Figure 40) est un vecteur d'expression optimisé pour la production de protéines recombinantes par expression transitoire dans la plante et provient initialement du vecteur pEAQ-HT\_DEST1 additionné de la GFP5 (Peyret & Lomonossoff, 2013). Il porte un gène conférant la résistance à la kanamycine. La multiplication du plasmide est facilitée en *E. coli* via l'origine OriV et le gène TrfA codant la protéine d'initiation de la réplication. Le vecteur possède les sites de recombinaison *attR1* et *attR2* et une cassette de clonage pour le transfert optimisé de séquences d'intérêt via la technologie Gateway®. Il possède également le gène codant la protéine p19 sous le 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 "Tomato Bushy Stunt Virus" (TBSV) permet de contrer l'extinction d'expression d'ARNm surnuméraires, ce qui améliore considérablement l'expression transitoire de protéines chez les plantes (Lakatos *et al.*, 2004; Pumplin & Voinnet, 2013). Le vecteur pEAQ\_HT\_GFP est utilisé comme construction contrôle de l'expression transitoire en tabac. Il est également utilisé comme élément facilitateur de l'expression transitoire de la construction pDGB1\_Q1\_PP du fait de la coexpression de la protéine p19.



**Figure 40: Carte plasmidique du vecteur pEAQ\_HT\_GFP.**

## Méthodes

### 1- Expression hétérologue dans *Nicotiana benthamiana*

#### A- Culture de *N. benthamiana*

Des graines (~100) de *N. benthamiana* vernalisées (au moins 2 mois) sont semées sur du terreau. Les pots sont arrosés par dessous et recouverts d'un papier aluminium (pendant 4-5 jours) pour favoriser une hygrométrie élevée. Les graines sont mises à germer en chambre climatique à 24°C, en jours longs (16hJ/8hN), pour une hygrométrie de 70%. Au stade 4-6 feuilles (~30 jours), les plantules sont rempotées individuellement. Les plantes sont cultivées dans les mêmes conditions avec un arrosage régulier deux fois par semaine. Une attention particulière est portée sur le statut hydrique des plants pour éviter une fermeture des stomates notamment avant et pendant l'agroinfiltration.

#### B- Transformation et culture des agrobactéries

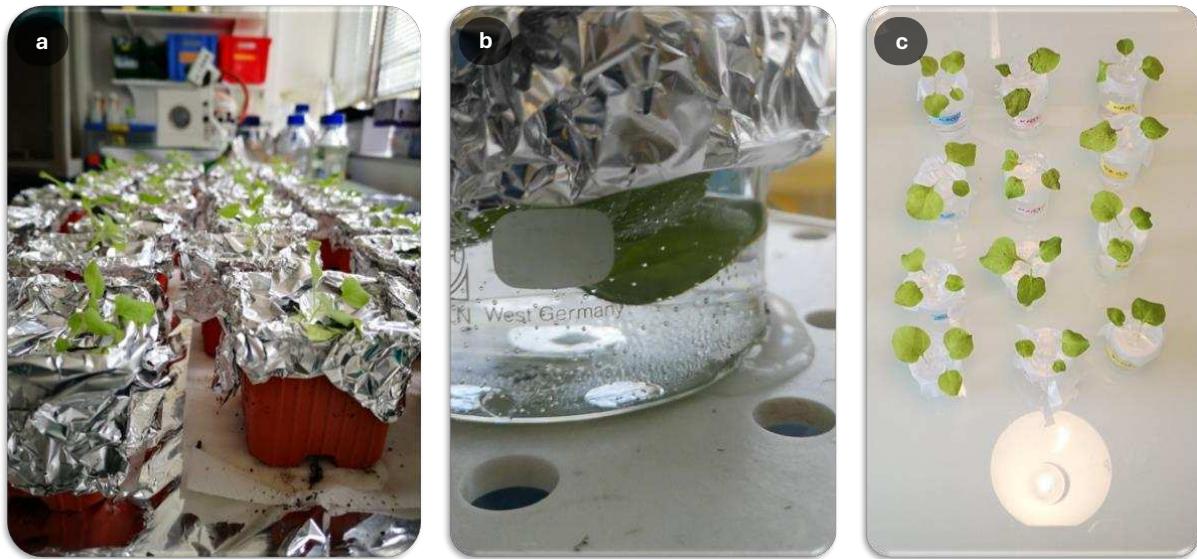
Six jours avant agroinfiltration, les agrobactéries sont transformées avec (A) pEAQ-HT\_GFP, ou co-transformées avec les plasmides (B) pDGB1\_Ω1\_PP et pSOUP. Les cultures sont incubées à 28°C durant 72h sur milieux sélectifs. Le milieu LB est complémenté avec (A) de la spectinomycine (100mg/L), de la tétracycline (5mg/L) et de la rifampicine (20 mg/L) et (B) de la kanamycine (50mg/L) et de la rifampicine (20 mg/L). Deux colonies indépendantes sont précultivées 16h dans 10 mL de LB en tubes de culture stérile de 50 mL (de type Falcon) avec antibiotiques à 28°C avec une agitation à 180 rpm. La présence des plasmides est ensuite vérifiée dans les agrobactéries transformantes. Pour ce faire, les plasmides amplifiés sont récoltés par une approche de lyse alcaline de 4 mL de préculture (kit Nucleospin Plasmid, Macherey-Nagel). Le tampon de resuspension est complété avec 2mg/mL de lysozyme pour faciliter la lyse des agrobactéries. Les plasmides sont quantifiés au Nanodrop. 1-3 µL de plasmide (300-700ng/µL) sont digérés pendant 30 minutes à 37°C en présence de 2 µL GreenBuffer, 1 µL enzyme de restriction, H<sub>2</sub>O ultrapure qsp 20µL. La digestion par *Bam*HI de pDGB1\_Ω1\_PP *Bam*HI génère des fragments de 7086, 5877, 2903, 812pb. La digestion

par *BamHI* de pSOUP ne génère qu'un fragment de 9275 (linéarisation). La digestion par *Sall*, *Ndel*, *PacI* de pEAQ-HT\_GFP génère des fragments de 5579, 3645, 2430pb. Deux jours avant agroinfiltration, 5 mL de précultures sont ensemencés dans 200 mL de LB complémentés par les antibiotiques *ad hoc* à 28°C avec une agitation à 180 rpm pendant 48h.

### C- Préparation de la suspension bactérienne et agroinfiltration

Les 200mL de cultures sont centrifugés pendant 4min, 7500g à température ambiante. Le surnageant est supprimé et le culot resuspendu très soigneusement par pipetage-refoulement dans 4 mL d'H<sub>2</sub>O mQ stérile puis complété à 45 mL. Les culots sont rassemblés par conditions et le lavage dans l'eau stérile est répété. Au final, le culot lavé est resuspendu avec 750 µL d'H<sub>2</sub>O mQ stérile. La suspension bactérienne est quantifiée par spectrophotométrie à DO<sub>600nm</sub>. 400mL de suspensions bactériennes sont préparés en mélangeant des agrobactéries contenant pEAQ-HT\_GFP (DO<sub>finale</sub> = 0,8) et pDGB1\_Q1\_PP (DO<sub>finale</sub> = 0,4) selon l'équation V = (Cf (=0,8) x 400 mL) / DO<sub>réelle</sub>. Le ratio est 2/1 en faveur de la protéine p19 contenue dans pEAQ-HT\_GFP (**Figure 40**). La suspension repose 1h minimum (et jusqu'à 3h) avant agroinfiltration.

En guise de préparation, les pots des plants de tabacs sont essorés, et un film d'aluminium est ajouté pour limiter les transferts de sol sur les feuilles (**Figure 41a**). Les plants sont sélectionnés de façon à avoir des lots avec des plantes de tailles comparables entre chaque condition d'infiltration. Les pots sont ensuite inversés sur le cristallisoir pour immerger toutes les feuilles. Les feuilles sont soumises à une pression sous vide de 50 mbar dans un système de cloche à vide (**Figure 41b**). Le vide est cassé progressivement, pour favoriser au maximum la pénétration de la suspension bactérienne dans les feuilles. Les plants sont ensuite incubés en caissons de confinement transparent et aérés 4 à 6 jours à température ambiante et à la lumière du jour. Les plants sont arrosés à l'eau osmosée stérilisée régulièrement.



**Figure 41: Photographies de la préparation des plants (a), des dispositifs d'agroinfiltration pour l'expression transitoire de jeunes plants de tabac (b), et de supplémentation (c).**

#### D- Vérification de l'expression de la GFP

Lors de la récolte, l'expression de la GFP est vérifiée grâce à une loupe binoculaire couplée à une source UV ( $\lambda_{\text{excitation}} = 395\text{nm} = \text{UV-bleu}$ ). Sa détection est visualisée par une couleur verte ( $\lambda_{\text{émission}} = 504\text{ nm} = \text{vert}$ ) en parallèle de la chlorophylle a ( $\lambda_{\text{émission}} = 669\text{nm} = \text{rouge}$ ) et de la lumière naturelle.

#### E- Approche de supplémentation et récolte

La solution de supplémentation à 10mM est préparée extemporanément à partir de poudre de substrat diluée dans du DMSO puis diluée dans l'eau. Le DMSO permet la dissolution à concentration élevée et limite la précipitation du substrat au contact de l'eau.

Le dispositif de supplémentation est constitué de pots en polypropylène de 50mL remplis de solution de supplémentation et recouvert d'un parafilm (**Figure 41c**). Trois à quatre trous permettent l'insertion des pétioles dans la solution. Les feuilles sont éclairées en continu pendant 24h et mises sous ventilation pour faciliter l'évapotranspiration et les flux entrants de substrats dans la plante. La base des pétioles

est taillée après 16h d'incubation pour enlever un éventuel bouchon de callose qui limiterait la captation du substrat. Le niveau de liquide est réajusté 2 fois durant l'expérience pour que les pétioles soient constamment immersés.

Lors de la récolte, les feuilles sont congelées instantanément dans de l'azote liquide. Les tissus sont broyés en fine poudre au mortier et pilon à l'azote liquide et sont conservés à -80°C pour les futures analyses métaboliques (100mg) et transcriptionnelles (80 mg).

## 2- Cultures *in vitro* de tomates

Pour ce qui concerne l'expérimentation de cultures *in vitro* de tomates, trois milieux de culture réalisés en interne directement à partir de micro- et macro -éléments avec des concentrations biodisponibles en fer ont été utilisés : milieu complet (40  $\mu\text{M}$   $\text{Fe}^{2+}$ ), carencé (10  $\mu\text{M}$   $\text{Fe}^{2+}$ ) et absent (0  $\mu\text{M}$   $\text{Fe}^{2+}$ ) en fer (**Tableau 4**). Chacune des catégories d'éléments comme la solution de fer est mélangée à partir de solutions stocks. L'ensemble est mélangé avec le tampon MES, le saccharose et l'agarose (**Tableau 4**). Le pH est ajusté à 7,3 avec du KOH pour être légèrement alcalin, puis les milieux sont autoclavés. Juste avant que la gélose ne fige, les vitamines sont rajoutées puis les milieux sont coulés en pots bébés (**Tableau 4**).

**Tableau 3: Liste des éléments de composition des milieux de culture *in vitro* de tomate dans le cadre d'une induction de la carence en fer.** La quantité de fer est modifiée en fonction des milieux pour une concentration finale : milieu complet (40  $\mu\text{M}$   $\text{Fe}^{2+}$ ), carencé (10  $\mu\text{M}$   $\text{Fe}^{2+}$ ) (rose) et absent (0  $\mu\text{M}$   $\text{Fe}^{2+}$ ) (vert) d'après (Siwinska *et al.*, 2018). Le pH est ajusté à 7,3 (légèrement alcalin) avec du KOH avant autoclave et les vitamines sont rajoutées après autoclave juste avant de couler les milieux en pots bébés. Abréviations : MES : acide 2-(N-morpholino) éthano-sulfonique ; EDTA : acide éthylène-diamine-tétra-acétique.

		complet	Fournisseur	REF
Macroéléments	NH4NO3	1,65	g/L	VWR
	KN03	1,9		VWR
	CaCL2, 2H2O	0,44		VWR
	MgS04, 7H2O	0,37		VWR

	KH <sub>2</sub> PO <sub>4</sub>	0,17		VWR	26936.260
Autres	Tampon MES	0,5		Sigma	M2933
	Agar	8		Kalys	HP696-1
	Saccharose	30		Duchefa	S0809
Microéléments	Kl	0,83	mg/L	Sigma	30315
	H3B03	6,2		Sigma	60541
	MnS04, H <sub>2</sub> O	16,9		ACCROS ORGANIC	205905000
	ZnSO <sub>4</sub> , 7H <sub>2</sub> O	8,6		VWR	29247.292
	Na <sub>2</sub> MoO <sub>4</sub> , 2H <sub>2</sub> O	0,25		VWR	1751
	CuSO <sub>4</sub> , 5H <sub>2</sub> O	0,025		CARLOS ERBA	476097
	CoCl <sub>2</sub> , 6H <sub>2</sub> O	0,025		ACCROS ORGANIC	206220500
	Inositol	100		Sigma	I5125
	Acide nicotinique	0,5		Sigma- Aldrich	N4126
Vitamines	Pyridoxine HCL	0,5		Sigma- Aldrich	P9755
	Thiamine HCL	0,1		Sigma- Aldrich	T-3902
	Glycine	2		Euromedex	26-128- 6405-C
	FeSO <sub>4</sub>	27,8 <i>/ 7,7 / 0</i>		Vel	1371
	Na <sub>2</sub> EDTA, 2H <sub>2</sub> O	37,3 <i>/ 9,2 / 0</i>		SIGMA	E5134

## Annexes

**Figure 42: Séquence du plasmide pDGB1\_Ω1\_PP.** Les séquences sont colorisées selon la légende ci-dessous. Séquences : Promoteur CaMV 35S (P35S) ; terminateur de la Nopaline synthase (tNos) ; Umbelliférone Synthase de *Pastinaca sativa* (PsUS) ; Diméthylsubérosine synthase de *Pastinaca sativa* (PsDS) ; Marmesine synthase de *Ficus carica* (FcMS) ; Psoralène synthase de *Pastinaca sativa* (PsPS).

**Légende :** P35S / tNos / KR / PsUS / PsDS / FcMS / PsPS / Cicatrices d'assemblage

>pDGB1\_Q1\_PP  
TaacggatccggctcaGgagactagagccaaagctgatctccctgccccggagatcaccatggacacttcttatctacgat  
cttagaagaaagttcgacggagaaggtagcagataccatgttaccaccgataatgagaagattagccttcaattcagaaagaat  
gctgaccacagatggtagagaggcctacgcggcaggctgtcaagacgatctaccggataataatctccaggagatcaaata  
ccttccaaagaaggtaaagatgcgtcaaaagattcaggactaactgcataagaacacagagaaagatataattctcaagatca  
gaagtactattccagtaggacgattcaaggctgctcataaaccaggcaagtaatagagattggagtctctaagaaagtagttcc  
tactgaatcaaaggccatggagtcaaaattcagatcgaggatctaacaaggactcgcgtgaagactggcaacagttcatacag  
agtctttacgactcaatgacaagaagaaaatctcgtcaacatggtagcagcacactctgtctactccaagaataatcaaagat  
acagtctcagaagaccaaaggctattgagactttcaacaaaggtaatatcgaaaacccctcggattccattgcccagctatc  
tgtcacttcatcaaaggacagtagaaaaggaaaggtggcacctacaatgcctatgcgataaaggaaaggctatcgtaaga  
tgcccctccgacagtggccaaagatggcaccccccacccacgaggagcatgtggaaaaagaagacgttccaaccacgtctc  
aaagcaagtggattgtatgtatctccactgcgttaggatgcgcacaatccactatcctcgaagaccctccttatata  
aggaagtcttcatggagaggactccgtatccatcacaataaccacaacaaacaacaaacattacaatttact  
attctagtcgaAatgggtgaacaagatggattgcacgcaggctctccggccctggtagagggctattcgctatgactggcac  
aacagacaatcggtctgtatggcccggttccggctgtcagcgcagggccgggttctttgtcaagaccgaccctgtcc  
gtgcccctgaatgaactgcaggacggcaggcagcggctatgtggctggccacgcacggcgttctgacgtgtctcgtacgtt  
tcactgaagcggaaaggactggctcttggcgaagtgcggggcaggatctctgtcatctcaccttgcctccgagaaag  
tatccatcatggctgtatgcataatggccgtcatacgcgtatccgggtacccattcgaccaccacggaaacatcgcatc  
gagcggacgtactcgatggaaagccgttgcgtcatacggtatctggacgcagagcatcagggctcgccagccgaaac  
tggtcgccaggctcaaggcgcataccggcggcggctgggtgtggccgaccgctatcggacatgcgttgctaccctgatatt  
aaaatggccgtttctggattcatcgactgtggccggctgggtgtggccgaccgctatcggacatgcgttgctaccctgatatt  
gctgaagagcttggccgaaatggctgaccgttctcggttacggatcgccgtcccgatcgccgtcccgatcgccgttctatc  
gccttctgacgagttcttgcGcttggaaatggatctcgatcccgatcgtaaaacatttgcataaaagttcttaagattgaatcc  
gttgccggcttgcgacgattatcatataattctgttgaattacgttaagcatgtataatcatgtatgcgttattttagaga  
tgggtttatgatttagagtcggcaattatacatttaacgcgtatggaaacaaatagcgcgaaacttaggataaaatttcg  
cgccgggtcatctatgttacttagatcggttgcatttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt  
atgtcatgataataatggttcttagacgtcagggttgcatttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt  
caaataatgtatccgctcatgagacaataaccctgataatgttcaataatggaccgactcgccgtgtcaGGAGACTAGAGC  
CAAGCTGATCTCCCTTGCCCCGGAGATCACCATGGACGACTTCTCTATCTACGATCTAGGAAGAA  
AGTCGACGGAGAAGGTGACGATACCATGTTACCACCGATAATGAGAAGATTAGCCTCTCAATTCA  
GAAAGAATGCTGACCCACAGATGGTTAGAGAGGCCTACGCCAGGTCTGATCAAGACGATCTACC  
CGAGTAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTAAAGATGCAGTCAAAGACTTCCAGTATGGACGAT  
CTAAGGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTTAAGAAAGTAGTTCCACTGTA  
ATCAAAGGCCATGGAGTCAAAATTCAAGATCGAGGATCTAACAGAACTCGCCGTAAAGACTGGCGAA  
CAGTTCATACAGAGTCTTACGACTCAATGACAAGAAGAAAATCTCGTCAACATGGTGGAGCAGCAG  
ACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTCAA  
CAAAGGGTAATATCGGGAAACCTCTCGGATTCATTGCCAGCTATGTCACCTCATCAAAGGAC  
AGTAGAAAAGGAAGGTGGCACCTACAAATGCCATATTGCGATAAAGGAAAGGCTATCGTCAAGATG

CCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCGAGGAGCATCGTGGAAAAAGAAGAC  
GTTCCAACCACGTCTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACA  
ATCCCACATCCTCGCAAGACCCCTCCTATATAAGGAAGTTCAATTGAGAGGACTCCGGT  
ATTTTACAACAATACCACAACAAAACAACAAACATTACAATTACTATTCTAGTCGAAATGG  
CTCCATCTCTACTGATGATTACAGAAACTCGCTTACAAAGGGACATGGTAAAGGACTTCAG  
ATCTTAAATTGGATGCTTGCAGAGCAATACATTCAACCTGTTGAAGAGAGGGCTGATATGACTAAGGT  
TTGAAGAAAGAATCTATCCCAGTTATTGATATGTCAAATCTGATGATCCTAAAGTTGCTGATCAAATTGC  
TGCTGCTGCTGAAAGTGGGATTTCCAAATTATAATCATGGTCTTCAATCGAAGTTCTGAGAATG  
TTAAAGAGGCTACTAGAAGATTTGCTTGCCTGTTGAAGAGAAGATTAAACACTCAAGAACAAATC  
TCCAACAAACTCAGTTAGACTACTACATCTTCTCTAAAGGTGATAAGGTTGGAGTGGAGGATT  
ATCTTCTATCTGGTTTCAGATAAGAAATCTTCAGAAATTGCCCCTACTTGTAAGAATGATGTTAAAGA  
ATACGTTGAGAAGTCAGAATTGTTGAAGTGGCTTGAAAGCTCTTGAAGGGATTGAATGTTGATA  
TGGATTCTAAGGAGTCAATCCTATGGGTTCAACAAAGAATTATGAAATTATTACCCCTGTTGCCAAAT  
CCTGAACTTGCTATTGGAGTTGCTAGGCATTCTGATGTTCAACTCTTACATTCTTGCAGATAACGT  
TGGAGGTTGCATATTAGAAAGATGGACTTGATACATGGATTGTTCCACCTGTTGAGGGAGCTATTG  
TTATTAAATATCGGAGATGCTCTCAAATCTGTCTAATGGAAATATAAGTCAGCTAACATTGTTGCTG  
CTAATGGTCTAACGATAGAATCTCAGTCCAATTACTAACCCATCTCTGATGATATTATGGACCA  
CTCCTGAGCTTGAAGAATGGTAAAAGCCTATATATAAGCATGTTGACTCTGATTACGTTAACGAT  
TTCTACAGGAAATCACATGAGGAAAGCATACTGGATTGCTCAAATGTGAGCTTGGAAATGGATCTT  
CGATCCCCGATCGTCTCAAACATTGGCAATAAAGTTCTTAAAGATTGAAATCCTGTTGCCGGCTTGC  
GATTATCATATAATTCTGTTGAATTACGTTAACGATGTAATAATTAAACATGTAATGCATGACGTTATT  
GATGGGTTTATGATTAGAGTCCCGCAATTACATTAAACCGTAGAAAACAAATAGCGCGC  
AAACTAGGATAAATTATCGCGCGCGTGTCACTATGTTACTAGATCGGGATTGCCAACGTAATTCTG  
AAGACGAAAGGGCCTCGTGTACGCCATTGTTAGGTTAATGTCATGATAATAATGGTTCTAGACGT  
CAGGTGGCACTTTCGGGAAATGTGGCGGAAACCCCTATTGTTATTCTAAATACATTCAAATAT  
GTATCCGCTCATGAGACAATAACCCGTATAATGCTCAATAATGGACCGACTCGCGTCAAG  
ACTAGAGCCAAGCTGATCTCTTGCCTTGGAGATCACCATGGACACTTCTCTATCTACGATCTA  
GGAAGAAAGTCGACGGAGAAGGTGACGATACCATGTTACCCACCGATAATGAGAAGATTAGCCTCTT  
CAATTTCAGAAAGAATGCTGACCCACAGATGGTAGAGAGGGCTACGGGGAGGTCTGATCAAGACG  
ATCTACCCGAGTAATACTCCAGGAGATCAAATACCTCCAAGAAGGTTAAAGATGCAGTCAAAGA  
TTCAGGACTAACTGCATCAAGAACACAGAGAAAGATATTCTCAAGATCAGAAGTACTATTCCAGTAT  
GGACGATTCAAGGCTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTCAAGAAAGTAGTT  
CTACTGAATCAAAGGCCATGGAGTCAAAATTCAAGATCGAGGATCTAACAGAACTCGCCGTGAAGACT  
GGCGAACAGTTCATACAGAGTCTTACGACTCAATGACAAGAAGAAATCTCGTCAACATGGTGA  
GCACGACACTCTCGTCACTCCAAGAATACAAAGATACTGAGGACTACAGAAGACCAAAGGGCTATT  
CTTCAACAAAGGTAATATCGGAAACCTCCTCGGATTCCATTGCCAGCTATGTCACATTCA  
AAAGGACAGTAGAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGT  
TCAAGATGCCCTGCCGACAGTGGCCAAAGATGGACCCCCACCGAGGAGCATCGTGGAAA  
AAGAAGACGTTCAACCACGCTTCAAAGCAAGTGGATTGATGTGATATCTCAACTGACGTAAGGGAT  
GACGCACAATCCCACATCCTCGCAAGACCCCTCCTATATAAGGAAGTTCAATTGAGAGGG  
ACTCCGGTATTAAACAAATACCACAACAAAACAACAAACACATTACAATTACTATTCTAGTC  
GAAATGGCTCAAACAAATTGATTGATTCAAAGACTTCTCAGGATTGTCATCTCAAAGAGATAAGGG  
TAGGACTTGCACACAAAGAAGGCATGCTAAAGTTGTTAACGGAGATCAAGAATTGCTTGG  
TGTTCATGTGATAAGAATCTGATTCTACTAAGAACCTCTCAGGTTCTGCGAGAAGCCTATCAGAACTC  
ATACAAATAAGCTTGCACAAACTATTCAAGCTACATCTGATAGGAAAGCTATTATTCAACCAAAGGATGAT  
TATGAGGCTCCTGGCAAATACTCTAGAAGGAAATGGGATGCTTTGTACATTGGAAAGACCAACT  
CAGCTATCTGCACTATTATTGGTATCTCTCAGTTCTCTTGCCTTGCCATTGACATCAGTTAAGGATTTCTG  
CTCCTTATTGTTGGACTTGCAGCTTGTGATCCCTTCTTGTGCTAACATCTACATTCAAGGTATTAA  
TCAACTTGTGATGTTGATATCGATAAGGATAATAAGCCATATTGCCCTTGTGCTTGGAGAATTCTT  
GGAGAGGGTAGAGCTATTGTTCTGCTTGGCTTATGTGCCCTGCTGTTGGTATTGTGACATTCTAC  
ACCACTTTCGTTGGAGTTGGTTATTCTTATCGGTACAGCTACTCAGTTGAATTGCCACTTTGAG



ccaaacttagataaattatcgccgcgggtcatctatgttactagatcggaattccaaggtaattctgaagacgaaaggccct  
cgtgatcgcctattttataggtaatgtcatgataataatggttcttagacgtcaggcactttcggggaaatgtgcgcggacc  
cctattgttattttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgctcaataatggaccgactcg  
cgctgtcaggagactagagccaagctgatcccttgcggagatcaccatggacgacttctatctctacgatctaggaaga  
aagttcgcggagaaggtagcgtacccatgtcaccaccgataatgagaagattagccttcattcagaaagaatgctgaccca  
cagatgttagagaggcctacgcccaggtctgatcaagacatctaccggagaataatctccaggagatcaaatacctccaa  
gaaggttaaagatcgtcaaaagattcaggactactgcatcaagaacacagagaagatatttctcaagatcagaact  
ttccagtagtggacgattcaaggcttcataaaaccaggactaatgagaagattggactctcaagaaagttagtccactgaatc  
aaaggccatggagtcaaaaattcagatcgaggatctaacagaactcgccgtgaagactggcgaacagttcatacagacttac  
gactcaatgacaagaagaaaatctcgtaacatggggacgacactctcgctactccaagaatatacagatcgtca  
gaagaccaaagggtatttagactttcaacaaaggtaatatggggaaacctcccgattccattgcccagctatctgcacttc  
tcaaaaggacagtagaaaaggaaggtagccatcataatgcccattgcgatataaggaaaggctatcgtcaagatgcctcgcc  
gacagtgtcccaaagatggaccccccacccacgaggagcatgtggaaaaagaagacgttccaaccacgtctcaagcaagt  
ggattgtatctccactgacgtaaaggatgacgcacaatcccactatcctcgcaagaccctcttatataaggatgtca  
ttcatttggagaggactccgttataacaataccacaacaaaacaacaaacacattacaatttactattctgtcgaa  
AATGAAGATGCTTGAGCAAAACCCACTTATTGTATTCTTCCTTTCTTGTACAATTCTTATA  
AGTGGTTGGTAAGAAAACCTCCATCTAAAATTGCCACCTCACCACCTAGACTTCCTATTGGAAA  
TTTGCATCAAATTGGTCCAGATCCTCAAATTCTCTAGAGATTGGCTAGGAAGTACGGACCAGTTAG  
CATCTTAAATTGGTTCAGTCCCTGTTGGTCTTCAGCTGATGGAGCTAGAGAGATTAAAGACA  
CATGATCTGCTTGCTGATAGGCCACTCTCAGTTGCTAACAGAATTCTATAATGGTAGGGATAT  
GGTTTCGCTAGATACTGAATACTGGAGACAGGTTAAGTCTACTGTGTTACACAACCTTGTCTGTT  
AAAGAGTTCAATCATTCCATAACGTTAGGGAGAGGAAGTTGCTTTGCTGATAACATCGAGAACTC  
TAAGTCAAAAGTTATAATCTTCTGAGATGCTTATTGAAATTGACAGGAAATGGTGTGAGCTGCTT  
GGGATCAGGTACAACGTTGATTCTACAAGTCATTGCTTGTCAAATTGGATATGGGGATACTCTA  
GGTCAATTGAAGATTCTTCCATCTTGGTGGGGTGGATTACTGGATTGAAGGGTAAAGTTGAG  
AAAGCTGCTAATGGTGTGATGCTTCTGAAGGTGTTGAAGAACCATACAAACCTCTACTTCTC  
AGCTAATAAGGATTGTTCAATCCTTGGAGATCCAAGAAGCTGATGCTGGTCTCTATGGATAAGG  
AATGTATTAATCACTTATTGGGATATGTTGGAGCTGGTACTGAGACAATTGCTACAGCTTGAGTGG  
ACTATTGGAGCTTGATTAAGTCTCCAGATGCTATGCTAAAGCTTAAAAGAGGTTAGAGAAATTGGAA  
AGGGTAAATCTAGGATTGAGGAAGGGAGATCTGTTAAGATGGATTATTGAAAGGCTGTTATGAAAGAGTC  
AATGAGATTGACTTCACAGCTCCACTTGGTCTTAGAGAAGCTAGGCAAGATGTTAAGTTATGGGA  
TACGATATTAATCTGGTACTCAAGTCTTATTAAATGCTGGCTATTGCTAGGGATCCATCTCATGGGAT  
AATCCAGAGGAGTTAGACCTGAGAGGTTTGAACCTCCTATCGATTACAAGGGTTTAATTACGAATA  
CATCCCATTGGAGCTGGTAGAAGGGGATGCCCTGGTATCCAATCGCTATCTCAGTTAATGAGCTGT  
TGTGCTAACGTTATAAGTTAATTGCAACTCCTGATGGAAAAAGATTGGAGGAGATGGATATGA  
CTGCTTCTACAGGTATCATTCCATAAGAAATCACCACCTTGGTGTGCTAGGCCTCATGTTGATGA  
GCTTggatggatctcgatcccgatcgtaaatacattggcaataaagttcttaagattgaatctgtgcgggtctgcgacgatta  
tcatataattctgtgaaattacgttaagcatgtataataacatgtaatgcgtacgatcgttattatgagatgggtttatgattagatgtcc  
gcaattatacatttaatacgcgtatagaaaacaaaatacgccgcaacttaggataaaattatcgccgcgggtcatctatgttacta  
gatcggtatggccaaagcttaattctgtgaaagacaaaggccctcgatcgcctattttataggttaatgtcatgataataatggttct  
tagacgtcagggtggactttcgggaaatgtgcggaaacccctatttttctaaatacattcaaaataatgtatccgctcatga  
gacaataacccctgataaaatgttcataataatggaccgactcgctgtcatgagaccggatcctgacaggatattggccggtaaa  
ctaagtgcgttatgtttgatctcatgtgagcaaaaaggccagcaaaaaggccaggaaaccgtaaaaaggccgcgtgctgg  
cgttttccataggctccgccccctgacgaggcatcacaatgcacgctcaagtgcggcggcggcggatccgtccgccttctc  
aagataccaggcggtttccctgaaagctccctgtgcgtctctgtccgaccctggcgttaccggatccgtccgccttctc  
cctcggggagcgtggcgtttctcatagctcagcgttaggtatctcagttcggttaggtcgtccagttccagctggctgtgcac  
gaaccccccgtcagccgaccgctgcgccttacggtaactatgcgtcttgatggccggatcctgacaggatattggccggtaaa  
ggcagcggccactggtaacaggattagcagagcggatgttaggcgggtctacagatgttgcgttgcgtccactacggcta  
cactagaagaacagtattggatctgcgtctgtgaaagccagttacccggaaagaagagttgttagcttgcgtccatccggcaacaa  
accaccgtggtagccgtggtttttgcgtccatgcgtcggcggatccgtccatggccggtaaaaggatctcaagaagatcccttgcgtcc  
acgggggtctgcgtccatgtggaaacactcgcgttaaggatgttgcgtccatggccggatccgtccatggccggtaaa

**Tableau 5: Liste de FCs linéaires et angulaires identifiées chez les Moracées.** Cette liste décrit le nom des FCs d'après leurs auteurs, les ramifications sur le squelette, l'espèce d'origine, la méthode d'acquisition des données d'identification. La légende des couleurs est au bas du tableau et représente les tissus analysés, ainsi que la confiance à porter sur l'analyse en fonction de la méthode d'acquisition. Abréviations : RMN (résonance magnétique nucléaire), CG-HR-SM (chromatographie gazeuse à haute résolution et couplée à la spectrométrie de masse), CCM (chromatographie sur couche mince).



		Angular furanocoumarins	Species
	45	Angelicin	<i>Ficus carica</i>
C5	46	Isobergapten	<i>Dorstenia bryoniifolia</i>
	47	Pimpinelin	<i>Dorstenia bryoniifolia</i>
	48	6-methoxy-5-(3-methyl-2,3-dihydroxybutyl)-angelicin	<i>Dorstenia gigas</i>
	49	6-methoxy-5-[3-( <i>b</i> -glucopyranosyloxy)-2-hydroxy-3-methyl-butyl]-angelicin	<i>Dorstenia gigas</i>
	50	6-methoxy-5-(3-hydroxy-2-oxo-3-methyl-butyl)-angelicin	<i>Dorstenia gigas</i>
C5 + C6	51	6-methoxy-5-(4-hydroxy-3-methylbut-2-enyl)-angelicin	<i>Dorstenia gigas</i>
	52	5-O- <i>β</i> -D-glucopyranosyl-6-hydroxyangelicin	<i>Ficus ruficaulis</i>
	53	6-O- <i>β</i> -D-glucopyranosyl-5-hydroxyangelicin	<i>Ficus ruficaulis</i>
	54	5,6-O- <i>β</i> -D-diglucopyranosylangelicin	<i>Ficus ruficaulis</i>

Tissue
Whole plant
Figs
Leaves and twigs
Leaves
Twigs, trunk wood, stem bark
Rhizomes, roots, bark roots

# Bibliographie

- Abegaz, B. M., Ngadjui, B. T., Dongo, E., Ngameni, B., Nindi, M. N., & Bezabih, M. (2002). Chalcones and other constituents of *Dorstenia prorepens* and *Dorstenia zenkeri*. *Phytochemistry*, 59(8), 877-883. [https://doi.org/10.1016/S0031-9422\(01\)00483-6](https://doi.org/10.1016/S0031-9422(01)00483-6)
- Abegaz, B. M., Ngadjui, B. T., Folefoc, G. N., Foto, S., Ambassa, P., Bezabih, M., Dongo, E., Rize, F., & Petersen, D. (2004). Prenylated flavonoids, monoterpenoid furanocoumarins and other constituents from the twigs of *Dorstenia elliptica* (Moraceae). *Phytochemistry*, 65(2), 221-226. <https://doi.org/10.1016/j.phytochem.2003.10.028>
- Abenavoli, M. R., De Santis, C., Sidari, M., Sorgonà, A., Badiani, M., & Cacco, G. (2001). Influence of coumarin on the net nitrate uptake in durum wheat. *New Phytologist*, 150(3), 619-627. <https://doi.org/10.1046/j.1469-8137.2001.00119.x>
- Abley, K., & Locke, J. C. W. (2021). Noisy transcription under the spotlight. *Nature Plants*, 7(8), 996-997. <https://doi.org/10.1038/s41477-021-00987-x>
- Adamowicz, S., Le Bot, J., Huanosto Magaña, R., & Fabre, J. (2012). Totomatix : A novel automatic set-up to control diurnal, diel and long-term plant nitrate nutrition. *Annals of Botany*, 109(1), 309-319. <https://doi.org/10.1093/aob/mcr254>
- Adler, F. R., & Karban, R. (1994). Defended Fortresses or Moving Targets? Another Model of Inducible Defenses Inspired by Military Metaphors. *The American Naturalist*, 144(5), 813-832. <https://doi.org/10.1086/285708>
- Aksøy, E., Yılmaz, H., & Kayıhan, C. (2022). The Revolution of Omics Technology in Plant Science. In C. S. Prakash, S. Fiaz, & S. Fahad (Éds.), *Principles and Practices of OMICS and Genome Editing for Crop Improvement* (p. 23-56). Springer International Publishing. [https://doi.org/10.1007/978-3-030-96925-7\\_2](https://doi.org/10.1007/978-3-030-96925-7_2)
- Alseekh, S., & Fernie, A. R. (2018). Metabolomics 20 years on : What have we learned and what hurdles remain? *The Plant Journal*, 94(6), 933-942. <https://doi.org/10.1111/tpj.13950>
- Araniti, F., Scognamiglio, M., Chambery, A., Russo, R., Esposito, A., D'Abrosca, B., Fiorentino, A., Lupini, A., Sunseri, F., & Abenavoli, M. R. (2017). Highlighting the effects of coumarin on adult plants of *Arabidopsis thaliana* (L.) Heynh. By an integrated -omic approach. *Journal of Plant Physiology*, 213, 30-41. <https://doi.org/10.1016/j.jplph.2017.02.013>
- Bailey, D. G., Malcolm, J., Arnold, O., & Spence, J. D. (1998). *Grapefruit juice-drug interactions*. 10.
- Bailey, D., Spence, J., Munoz, C., & Arnold, J. M. O. (1991). Interaction of citrus juices with felodipine and nifedipine. *The Lancet*, 337(8736), 268-269. [https://doi.org/10.1016/0140-6736\(91\)90872-M](https://doi.org/10.1016/0140-6736(91)90872-M)
- Bak, S., Kahn, R. A., Olsen, C. E., & Halkier, B. A. (1997). Cloning and expression in *Escherichia coli* of the obtusifoliol 14α-demethylase of *Sorghum bicolor* (L.) Moench, a cytochrome P450 orthologous to the sterol 14α-demethylases (CYP51) from fungi and mammals. *The Plant Journal*, 11(2), 191-201. <https://doi.org/10.1046/j.1365-313X.1997.11020191.x>
- Baldwin, M. A., & McLafferty, F. W. (1973). Liquid chromatography-mass spectrometry interface-I : The direct introduction of liquid solutions into a chemical ionization mass spectrometer. *Organic Mass Spectrometry*, 7(9), 1111-1112. <https://doi.org/10.1002/oms.1210070913>
- Barbehenn, R. V., & Peter Constabel, C. (2011). Tannins in plant-herbivore interactions. *Phytochemistry*, 72(13), 1551-1565. <https://doi.org/10.1016/j.phytochem.2011.01.040>
- Baskin, J. M., Ludlow, C. J., Harris, T. M., & Wolf, F. T. (1967). Psoralen, an inhibitor in the seeds of *Psoralea subacaulis* (Leguminosae). *Phytochemistry*, 6(9), 1209-1213. [https://doi.org/10.1016/S0031-9422\(00\)86083-5](https://doi.org/10.1016/S0031-9422(00)86083-5)

- Ben-Hur, E., & Song, P.-S. (1984). The Photochemistry and Photobiology of Furocoumarins (Psoralens). In *Advances in Radiation Biology* (Vol. 11, p. 131-171). Elsevier. <https://doi.org/10.1016/B978-0-12-035411-5.50009-5>
- Bennet, & Bonner. (1953). *Isolation of Plant Growth Inhibitors from Thamnosma montana*. 6.
- Bennetzen, J. L., & Wang, H. (2014). The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Annual Review of Plant Biology*, 65(1), 505-530. <https://doi.org/10.1146/annurev-arplant-050213-035811>
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 41(D1), D36-D42. <https://doi.org/10.1093/nar/gks1195>
- Benton, M. J., Wilf, P., & Sauquet, H. (2022). The Angiosperm Terrestrial Revolution and the origins of modern biodiversity. *New Phytologist*, 233(5), 2017-2035. <https://doi.org/10.1111/nph.17822>
- Berenbaum, M., & Feeny, P. (1981). Toxicity of Angular Furanocoumarins to Swallowtail Butterflies: Escalation in a Coevolutionary Arms Race? *Science*, 212(4497), 927-929. <https://doi.org/10.1126/science.212.4497.927>
- Berenbaum, M. R., Zangerl, A. R., & Nitao, J. K. (1986). CONSTRAINTS ON CHEMICAL COEVOLUTION: WILD PARSNIPS AND THE PARSNIP WEBWORM. *Evolution*, 40(6), 1215-1228. <https://doi.org/10.1111/j.1558-5646.1986.tb05746.x>
- Blázquez, M. A., Nelson, D. C., & Weijers, D. (2020). Evolution of Plant Hormone Response Pathways. *Annual Review of Plant Biology*, 71(1), 327-353. <https://doi.org/10.1146/annurev-arplant-050718-100309>
- Bloch, F. (1946). Nuclear Induction. *Physical Review*, 70(7-8), 460-474. <https://doi.org/10.1103/PhysRev.70.460>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bouillé, A., Rashmi Kumari, R., Olry, A., Charles, C., Nelson, D. R., Larbat, R., Thornton, J., Villard, C., & Hehn, A. (2024). Using lineage-specific patterns to understand convergence of enzymatic functions led to the identification of Moraceae-specific P450s involved in furanocoumarin biosynthesis. <https://doi.org/10.1101/2024.08.19.608558>
- Bourgaud, F., Olry, A., & Hehn, A. (2014). Recent Advances in Molecular Genetics of Furanocoumarin Synthesis in Higher Plants. In C. Jacob, G. Kirsch, A. Slusarenko, P. G. Winyard, & T. Burkholz (Éds.), *Recent Advances in Redox Active Plant and Microbial Products* (p. 363-375). Springer Netherlands. [https://doi.org/10.1007/978-94-017-8953-0\\_14](https://doi.org/10.1007/978-94-017-8953-0_14)
- Bowman, J. L. (2022). The origin of a land flora. *Nature Plants*, 8(12), 1352-1369. <https://doi.org/10.1038/s41477-022-01283-y>
- Braz filho, R., Farias, A., & Gottlieb, O. R. (1972). COUMARINS FROM BROSILMIUM RUBESCENS. [https://doi.org/10.1016/S0031-9422\(00\)86395-5](https://doi.org/10.1016/S0031-9422(00)86395-5)
- Brian, P. w. (1951). Antibiotics produced by fungi. *The Botanical Review*, 17(6).
- Brignac-Huber, L. M., Park, J. W., Reed, J. R., & Backes, W. L. (2016). Cytochrome P450 Organization and Function Are Modulated by Endoplasmic Reticulum Phospholipid Heterogeneity. *Drug Metabolism and Disposition*, 44(12), 1859-1866. <https://doi.org/10.1124/dmd.115.068981>
- Bruno, L., Talarico, E., Cabeiras-Freijanes, L., Madeo, M. L., Muto, A., Minervino, M., Lucini, L., Miras-Moreno, B., Sofo, A., & Araniti, F. (2021). Coumarin Interferes with Polar Auxin Transport Altering Microtubule Cortical Array Organization in *Arabidopsis thaliana* (L.) Heynh. Root Apical Meristem. *International Journal of Molecular Sciences*, 22(14), 7305. <https://doi.org/10.3390/ijms22147305>
- Bryant, J. P., Chapin, F. S., & Klein, D. R. (1983). Carbon/Nutrient Balance of Boreal Plants in Relation to Vertebrate Herbivory. *Oikos*, 40(3), 357. <https://doi.org/10.2307/3544308>
- Burkart, G. M., & Brandizzi, F. (2021). A Tour of TOR Complex Signaling in Plants. *Trends*

- in *Biochemical Sciences*, 46(5), 417-428.  
<https://doi.org/10.1016/j.tibs.2020.11.004>
- Caceres, A., Rastrelli, L., Simone, F. D., Martino, G. D., Saturnino, C., Saturnino, P., & Aquino, R. (2001). *Furanocoumarins from the aerial parts of*.
- Campos, M. D., Félix, M. D. R., Patanita, M., Materatski, P., Albuquerque, A., Ribeiro, J. A., & Varanda, C. (2022). Defense Strategies: The Role of Transcription Factors in Tomato–Pathogen Interaction. *Biology*, 11(2), 235. <https://doi.org/10.3390/biology11020235>
- Candolle, A. P. D. (1816). *Essai sur les propriétées mädicales des plantes, comparées avec leurs formes extâerieures et leur classification naturelle /*. Chez Crochard, Libraire, <https://doi.org/10.5962/bhl.title.112422>
- Cannell, N., Emms, D. M., Hetherington, A. J., MacKay, J., Kelly, S., Dolan, L., & Sweetlove, L. J. (2020). Multiple Metabolic Innovations and Losses Are Associated with Major Transitions in Land Plant Evolution. *Current Biology*, 30(10), 1783-1800.e11. <https://doi.org/10.1016/j.cub.2020.02.086>
- Chae, L., Kim, T., Nilo-Poyanco, R., & Rhee, S. Y. (2014). Genomic Signatures of Specialized Metabolism in Plants. *Science*, 344(6183), 510-513. <https://doi.org/10.1126/science.1252076>
- Chaney, R. L. (1984). Diagnostic practices to identify iron deficiency in higher plants. *Journal of Plant Nutrition*, 7(1-5), 47-67. <https://doi.org/10.1080/01904168409363174>
- Chang, M.-S., Yang, Y.-C., Kuo, Y.-C., Kuo, Y.-H., Chang, C., Chen, C.-M., & Lee, T.-H. (2005). Furocoumarin Glycosides from the Leaves of *Ficus ruficaulis* Merr. Var. *Antaoensis*. *Journal of Natural Products*, 68(1), 11-13. <https://doi.org/10.1021/np0401056>
- Chassagne, F., Cabanac, G., Hubert, G., David, B., & Marti, G. (2019). The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products®. *Phytochemistry Reviews*, 18(3), 601-622. <https://doi.org/10.1007/s11101-019-09606-2>
- Chiang, C., Cheng, M., Peng, C., Huang, H., & Chen, I. (2010). A Novel Dimeric Coumarin Analog and Antimycobacterial Constituents from *Fatoua pilosa*. *Chemistry & Biodiversity*, 7(7), 1728-1736. <https://doi.org/10.1002/cbdv.200900326>
- Choi, I., Ahn, C. S., Lee, D.-H., Baek, S.-A., Jung, J. W., Kim, J. K., Lee, H.-S., & Pai, H.-S. (2022). Silencing of the Target of Rapamycin Complex Genes Stimulates Tomato Fruit Ripening. *Molecules and Cells*, 45(9), 660-672. <https://doi.org/10.14348/molcells.2022.2025>
- Chong, J., Poutaraud, A., & Hugueney, P. (2009). Metabolism and roles of stilbenes in plants. *Plant Science*, 177(3), 143-155. <https://doi.org/10.1016/j.plantsci.2009.05.012>
- Christenhusz, M. J. M., & Byng, J. W. (2016). The number of known plant species in the world and its annual increase. *Phytotaxa*, 261(3), 201. <https://doi.org/10.11646/phytotaxa.261.3.1>
- Christoffoleti, P. J., Figueiredo, M. R. A. D., Peres, L. E. P., Nissen, S., & Gaines, T. (2015). Auxinic herbicides, mechanisms of action, and weed resistance: A look into recent plant science advances. *Scientia Agricola*, 72(4), 356-362. <https://doi.org/10.1590/0103-9016-2014-0360>
- Chu, P., & Agapito-Tenfen, S. Z. (2022). Unintended Genomic Outcomes in Current and Next Generation GM Techniques: A Systematic Review. *Plants*, 11(21), 2997. <https://doi.org/10.3390/plants11212997>
- Clark, J. W., & Donoghue, P. C. J. (2018). Whole-Genome Duplication and Plant Macroevolution. *Trends in Plant Science*, 23(10), 933-945. <https://doi.org/10.1016/j.tplants.2018.07.006>
- Clark, J. W., Hetherington, A. J., Morris, J. L., Pressel, S., Duckett, J. G., Puttick, M. N., Schneider, H., Kenrick, P., Wellman, C. H., & Donoghue, P. C. J. (2023). Evolution of phenotypic disparity in the plant kingdom. *Nature Plants*.

- <https://doi.org/10.1038/s41477-023-01513-x>
- Clausen, M., Kannangara, R. M., Olsen, C. E., Blomstedt, C. K., Gleadow, R. M., Jørgensen, K., Bak, S., Motawie, M. S., & Møller, B. L. (2015). The bifurcation of the cyanogenic glucoside and glucosinolate biosynthetic pathways. *The Plant Journal*, 84(3), 558-573. <https://doi.org/10.1111/tpj.13023>
- Coley, P. D., Bryant, J. P., & Chapin, F. S. (1985). Resource Availability and Plant Antiherbivore Defenses. 230.
- Collett, M. G. (2019). Photosensitisation diseases of animals: Classification and a weight of evidence approach to primary causes. *Toxicon: X*, 3, 100012. <https://doi.org/10.1016/j.toxcx.2019.100012>
- Cosme, M., Fernández, I., Declerck, S., Van Der Heijden, M. G. A., & Pieterse, C. M. J. (2021). A coumarin exudation pathway mitigates arbuscular mycorrhizal incompatibility in *Arabidopsis thaliana*. *Plant Molecular Biology*, 106(4-5), 319-334. <https://doi.org/10.1007/s11103-021-01143-x>
- Dahmani, I., Qin, K., Zhang, Y., & Fernie, A. R. (2023). The formation and function of plant metabolons. *The Plant Journal*, 114(5), 1080-1092. <https://doi.org/10.1111/tpj.16179>
- Dall'Acqua, F., Marciani, S., Ciavatta, L., & Rodighiero, G. (1971). Formation of Inter-Strand Cross-Linkings in the Photoreactions between Furocoumarins and DNA. *Zeitschrift Für Naturforschung B*, 26(6), 561-569. <https://doi.org/10.1515/znb-1971-0613>
- Damjanić, & Akačić. (1974). FUROCOUMARINS IN *FICUS CARICA*.
- Daniel, L., Buryska, T., Prokop, Z., Damborsky, J., & Brezovsky, J. (2015). Mechanism-Based Discovery of Novel Substrates of Haloalkane Dehalogenases Using *in Silico* Screening. *Journal of Chemical Information and Modeling*, 55(1), 54-62. <https://doi.org/10.1021/ci500486y>
- Davis, G. V., & Glover, B. J. (2024). Characterisation of the R2R3 Myb subgroup 9 family of transcription factors in tomato. *PLOS ONE*, 19(3), e0295445. <https://doi.org/10.1371/journal.pone.0295445>
- De Vleesschauwer, D., Filipe, O., Hoffman, G., Seifi, H. S., Haeck, A., Canlas, P., Van Bockhoven, J., De Waele, E., Demeestere, K., Ronald, P., & Hofte, M. (2018). Target of rapamycin signaling orchestrates growth-defense trade-offs in plants. *New Phytologist*, 217(1), 305-319. <https://doi.org/10.1111/nph.14785>
- DeLano, W. L. (2002). PyMOL: An Open-Source Molecular Graphics Tool. <https://citeseerk.ist.psu.edu/document?repid=rep1&type=pdf&doi=ab82608e9a44c17b60d7f908565fba628295dc72#page=44>
- Delli-Ponti, R., Shivhare, D., & Mutwil, M. (2021). Using Gene Expression to Study Specialized Metabolism—A Practical Guide. *Frontiers in Plant Science*, 11, 625035. <https://doi.org/10.3389/fpls.2020.625035>
- Denisov, I. G., & Sligar, S. G. (2012). A novel type of allosteric regulation: Functional cooperativity in monomeric proteins. *Archives of Biochemistry and Biophysics*, 519(2), 91-102. <https://doi.org/10.1016/j.abb.2011.12.017>
- Diawara, M. M., Trumble, J. T., White, K. K., Carson, W. G., & Martinez, L. A. (1993). Toxicity of linear furanocoumarins to *Spodoptera exigua*: Evidence for antagonistic interactions. *Journal of Chemical Ecology*, 19(11), 2473-2484. <https://doi.org/10.1007/BF00980684>
- Didaran, F., Kordrostami, M., Ghasemi-Soloklui, A. A., Pashkovskiy, P., Kreslavski, V., Kuznetsov, V., & Allakhverdiev, S. I. (2024). The mechanisms of photoinhibition and repair in plants under high light conditions and interplay with abiotic stressors. *Journal of Photochemistry and Photobiology B: Biology*, 259, 113004. <https://doi.org/10.1016/j.jphotobiol.2024.113004>
- Dimitrov, D., Xu, X., Su, X., Shrestha, N., Liu, Y., Kennedy, J. D., Lyu, L., Nogués-Bravo, D., Rosindell, J., Yang, Y., Fjeldså, J., Liu, J., Schmid, B., Fang, J., Rahbek, C., & Wang, Z. (2023). Diversification of flowering plants in space and time. *Nature Communications*, 14(1), 7609. <https://doi.org/10.1038/s41467-023-43396-8>

- Dixon, R. A., & Dickinson, A. J. (2024). A century of studying plant secondary metabolism—From “what?” to “where, how, and why?” *Plant Physiology*, 195(1), 48-66. <https://doi.org/10.1093/plphys/kiad596>
- Dong, N., & Lin, H. (2021). Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *Journal of Integrative Plant Biology*, 63(1), 180-209. <https://doi.org/10.1111/jipb.13054>
- Drapal, M., Enfissi, E. M. A., & Fraser, P. D. (2021). Metabolic effects of agro-infiltration on *N. benthamiana* accessions. *Transgenic Research*, 30(3), 303-315. <https://doi.org/10.1007/s11248-021-00256-9>
- Dueholm, B., Krieger, C., Drew, D., Olry, A., Kamo, T., Tabouret, O., Weitzel, C., Bourgaud, F., Hehn, A., & Simonsen, H. T. (2015). Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. *BMC Evolutionary Biology*, 15(1), 122. <https://doi.org/10.1186/s12862-015-0396-z>
- Dugé de Bernonville, T., Papon, N., Clastre, M., O'Connor, S. E., & Courdavault, V. (2020). Identifying Missing Biosynthesis Enzymes of Plant Natural Products. *Trends in Pharmacological Sciences*, 41(3), 142-146. <https://doi.org/10.1016/j.tips.2019.12.006>
- Dugrand-Judek, A., Olry, A., Hehn, A., Costantino, G., Ollitrault, P., Froelicher, Y., & Bourgaud, F. (2015). The Distribution of Coumarins and Furanocoumarins in Citrus Species Closely Matches Citrus Phylogeny and Reflects the Organization of Biosynthetic Pathways. *PLOS ONE*, 10(11), e0142757. <https://doi.org/10.1371/journal.pone.0142757>
- Dumur, F. (2022). Recent advances on coumarin-based photoinitiators of polymerization. *European Polymer Journal*, 163, 110962. <https://doi.org/10.1016/j.eurpolymj.2021.110962>
- Durán-Medina, Y., Ruiz-Cortés, B. E., Guerrero-Largo, H., & Marsch-Martínez, N. (2021). Specialized metabolism and development: An unexpected friendship. *Current Opinion in Plant Biology*, 64, 102142. <https://doi.org/10.1016/j.pbi.2021.102142>
- Egerton, F. N. (2012). History of Ecological Sciences, Part 43: Plant Physiology, 1800s. *Bulletin of the Ecological Society of America*, 93(3), 197-219. <https://doi.org/10.1890/0012-9623-93.3.197>
- Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and Plants: A Study in Coevolution. *Evolution*, 18(4), 586. <https://doi.org/10.2307/2406212>
- Ellis, C. D., & Henderson, W. J. (1934). *Artificial radioactivity*.
- Elser, D., Pflieger, D., Villette, C., Moegle, B., Miesch, L., & Gaquerel, E. (2023). Evolutionary metabolomics of specialized metabolism diversification in the genus *Nicotiana* highlights N-acylnornicotine innovations. *SCIENCE ADVANCES*.
- Fasan, R., Mehareenna, Y. T., Snow, C. D., Poulos, T. L., & Arnold, F. H. (2008). Evolutionary History of a Specialized P450 Propane Monooxygenase. *Journal of Molecular Biology*, 383(5), 1069-1080. <https://doi.org/10.1016/j.jmb.2008.06.060>
- Feeny, P. (1976). Plant Appärenz and Chemical Defense. In J. W. Wallace & R. L. Mansell (Eds.), *Biochemical Interaction Between Plants and Insects* (p. 1-40). Springer US. [https://doi.org/10.1007/978-1-4684-2646-5\\_1](https://doi.org/10.1007/978-1-4684-2646-5_1)
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, 246(4926), 64-71. <https://doi.org/10.1126/science.2675315>
- Fiehn, O., Robertson, D., Griffin, J., Van Der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., Van Ommen, B., Lindon, J. C., & Sansone, S.-A. (2007). The metabolomics standards initiative (MSI). *Metabolomics*, 3(3), 175-178. <https://doi.org/10.1007/s11306-007-0070-6>
- Field, B., Fiston-Lavier, A.-S., Kemen, A., Geisler, K., Quesneville, H., & Osbourn, A. E. (2011). Formation of plant metabolic

- gene clusters within dynamic chromosomal regions. *Proceedings of the National Academy of Sciences*, 108(38), 16116-16121. <https://doi.org/10.1073/pnas.1109273108>
- Firn, R. D., & Jones, C. G. (2003). Natural products ? A simple model to explain chemical diversity. *Natural Product Reports*, 20(4), 382. <https://doi.org/10.1039/b208815k>
- Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*.
- Forman, R. T. T. (2014). *Urban ecology : Science of cities*. Cambridge university press.
- Fowlks, W. L. (1959). Part IV: Basic Considerations of the Psoralens: The Chemistry of the Psoralens From the Division of Dermatology, University of Oregon Medical School, Portland, Oregon. *Journal of Investigative Dermatology*, 32(2), 249-254. <https://doi.org/10.1038/jid.1959.42>
- Fowlks, W. L., Griffith, D. G., & Oginsky, E. L. (1958). Photosensitization of Bacteria by Furocoumarins and Related Compounds. *Nature*, 181(4608), 571-572. <https://doi.org/10.1038/181571a0>
- Francis, S., & Delgoda, R. (2014). A patent review on the development of human cytochrome P450 inhibitors. *Expert Opinion on Therapeutic Patents*, 24(6), 699-717. <https://doi.org/10.1517/13543776.2014.899583>
- Frank, S., Caffieri, S., Raffaelli, A., Vedaldi, D., & Dall'Acqua, F. (1998). Characterization of psoralen-oleic acid cycloadducts and their possible involvement in membrane photodamage. *Journal of Photochemistry and Photobiology B: Biology*, 44(1), 39-44. [https://doi.org/10.1016/S1011-1344\(98\)00103-1](https://doi.org/10.1016/S1011-1344(98)00103-1)
- Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grün, S., Winklmair, A., Eisenreich, W., Bacher, A., Meeley, R. B., Briggs, S. P., Simcox, K., & Gierl, A. (1997). Analysis of a Chemical Plant Defense Mechanism in Grasses. *Science*, 277(5326), 696-699. <https://doi.org/10.1126/science.277.5326.696>
- Fu, L., Wang, P., & Xiong, Y. (2020). Target of Rapamycin Signaling in Plant Stress Responses. *Plant Physiology*, 182(4), 1613-1623. <https://doi.org/10.1104/pp.19.01214>
- Fujiyama, K., Hino, T., Kanadani, M., Watanabe, B., Jae Lee, H., Mizutani, M., & Nagano, S. (2019). Structural insights into a key step of brassinosteroid biosynthesis and its inhibition. *Nature Plants*, 5(6), 589-594. <https://doi.org/10.1038/s41477-019-0436-6>
- Galán-Pérez, J. A., Gámiz, B., & Celis, R. (2022). Soil modification with organic amendments and organo-clays : Effects on sorption, degradation, and bioactivity of the allelochemical scopoletin. *Journal of Environmental Management*, 302, 114102. <https://doi.org/10.1016/j.jenvman.2021.114102>
- Galati, G. (2019). *Etude de la synthèse des furocoumarines chez le panais par des approches d'ingénierie métabolique et de multi-omique*.
- Gardner, E. M., Bruun-Lund, S., Niissalo, M., Chantarasuwan, B., Clement, W. L., Geri, C., Harrison, R. D., Hipp, A. L., Holvoet, M., Khew, G., Kjellberg, F., Liao, S., Pederneiras, L. C., Peng, Y.-Q., Pereira, J. T., Phillipps, Q., Ahmad Puad, A. S., Rasplus, J.-Y., Sang, J., ... Rønsted, N. (2023). Echoes of ancient introgression punctuate stable genomic lineages in the evolution of figs. *Proceedings of the National Academy of Sciences*, 120(28), e2222035120. <https://doi.org/10.1073/pnas.2222035120>
- Gardner, E. M., Garner, M., Cowan, R., Dodsworth, S., Epitawalage, N., Arifiani, D., Sahromi, Baker, W. J., Forest, F., Maurin, O., Zerega, N. J. C., Monro, A. K., & Hipp, A. L. (2021). Repeated parallel losses of inflexed stamens in Moraceae : Phylogenomics and generic revision of the tribe Moreae and the reinstatement of the tribe Olmedieae (Moraceae). *TAXON*, 70(5), 946-988. <https://doi.org/10.1002/tax.12526>
- Gardner, E. M., Sarraf, P., Williams, E. W., & Zerega, N. J. C. (2017). Phylogeny and biogeography of Maclura (Moraceae) and the origin of an anachronistic fruit. *Molecular Phylogenetics and Evolution*, 117, 49-59. <https://doi.org/10.1016/j.ympev.2017.06.021>

- Gayon, J. (2016). From Mendel to epigenetics : History of genetics. *Comptes Rendus Biologies*, 339(7-8), 225-230. <https://doi.org/10.1016/j.crvi.2016.05.009>
- Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO : A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8), 2628-2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Gentile, F., Agrawal, V., Hsing, M., Ton, A.-T., Ban, F., Norinder, U., Gleave, M. E., & Cherkasov, A. (2020). Deep Docking : A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS Central Science*, 6(6), 939-949. <https://doi.org/10.1021/acscentsci.0c00229>
- Gibernau, M., Buser, H. R., Frey, J. E., & Hossaert-McKey, M. (1997). Volatile compounds from extracts of figs of *Ficus carica*. *Phytochemistry*, 46(2), 241-244. [https://doi.org/10.1016/S0031-9422\(97\)00292-6](https://doi.org/10.1016/S0031-9422(97)00292-6)
- Giera, M., Aisporna, A., Uritboonthai, W., & Siuzdak, G. (2024). The hidden impact of in-source fragmentation in metabolic and chemical mass spectrometry data interpretation. *Nature Metabolism*. <https://doi.org/10.1038/s42255-024-01076-x>
- Gotoh, O. (1992). Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry*, 267(1), 83-90. [https://doi.org/10.1016/S0021-9258\(18\)48462-1](https://doi.org/10.1016/S0021-9258(18)48462-1)
- Gottlieb, O. R., Da Silva, M. L., & Maia, J. G. S. (1972). Distribution of coumarins in Amazonian Brosimum species. *Phytochemistry*, 11(12), 3479-3480. [https://doi.org/10.1016/S0031-9422\(00\)89841-6](https://doi.org/10.1016/S0031-9422(00)89841-6)
- Graña, E., Costas-Gil, A., Longueira, S., Celeiro, M., Teijeira, M., Reigosa, M. J., & Sánchez-Moreiras, A. M. (2017). Auxin-like effects of the natural coumarin scopoletin on *Arabidopsis* cell structure and morphology. *Journal of Plant Physiology*, 218, 45-55. <https://doi.org/10.1016/j.jplph.2017.07.007>
- Gravot, A., Larbat, R., Hehn, A., Lièvre, K., Gontier, E., Goergen, J.-L., & 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. *Archives of Biochemistry and Biophysics*, 422(1), 71-80. <https://doi.org/10.1016/j.abb.2003.12.013>
- Green, M. T. (2009). CH bond activation in heme proteins : The role of thiolate ligation in cytochrome P450. *Current Opinion in Chemical Biology*, 13(1), 84-88. <https://doi.org/10.1016/j.cbpa.2009.02.028>
- Greenberg, R. B., Alberti, M., Hearst, J. E., Chua, M. A., & Saffran, W. A. (2001). Recombinational and Mutagenic Repair of Psoralen Interstrand Cross-links in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 276(34), 31551-31560. <https://doi.org/10.1074/jbc.M103588200>
- Grossmann, K. (2010). Auxin herbicides : Current status of mechanism and mode of action. *Pest Management Science*, 66(2), 113-120. <https://doi.org/10.1002/ps.1860>
- Gu, M., Wang, M., Guo, J., Shi, C., Deng, J., Huang, L., Huang, L., & Chang, Z. (2019). Crystal structure of CYP76AH1 in 4-PI-bound state from *Salvia miltiorrhiza*. *Biochemical and Biophysical Research Communications*, 511(4), 813-819. <https://doi.org/10.1016/j.bbrc.2019.02.103>
- Guiochon, G. (2002). Preparative liquid chromatography. *Journal of Chromatography A*, 965(1-2), 129-161. [https://doi.org/10.1016/S0021-9673\(01\)01471-6](https://doi.org/10.1016/S0021-9673(01)01471-6)
- Guo, Q., Major, I. T., & Howe, G. A. (2018). Resolution of growth-defense conflict : Mechanistic insights from jasmonate signaling. *Current Opinion in Plant Biology*, 44, 72-81. <https://doi.org/10.1016/j.pbi.2018.02.009>
- Hale, A. L., Meepagala, K. M., Oliva, A., Aliotta, G., & Duke, S. O. (2004). Phytotoxins from the Leaves of *Ruta graveolens*. *Journal of Agricultural and Food*

- Chemistry*, 52(11), 3345-3349. <https://doi.org/10.1021/jf0497298>
- Halpert, J. R., Guengerich, F. P., Bend, J. R., & Correia, M. A. (1994). Selective Inhibitors of Cytochromes P450. *Toxicology and Applied Pharmacology*, 125(2), 163-175. <https://doi.org/10.1006/taap.1994.1061>
- Han, L., Zhang, L., He, Y., Liao, L., Li, J., Xu, S., Zhao, Y., Bian, X., & Xia, Y. (2023). Three carbon-/oxygen-prenyltransferases responsible for furanocoumarin synthesis in *Angelica dahurica*. *Industrial Crops and Products*, 200, 116814. <https://doi.org/10.1016/j.indcrop.2023.116814>
- Hansen, C. C., Nelson, D. R., Møller, B. L., & Werck-Reichhart, D. (2021). Plant cytochrome P450 plasticity and evolution. *Molecular Plant*, 14(8), 1244-1265. <https://doi.org/10.1016/j.molp.2021.06.028>
- Hansen, C. C., Sørensen, M., Veiga, T. A. M., Zibrandtsen, J. F. S., Heskes, A. M., Olsen, C. E., Boughton, B. A., Møller, B. L., & Neilson, E. H. J. (2018). Reconfigured Cyanogenic Glucoside Biosynthesis in *Eucalyptus cladocalyx* Involves a Cytochrome P450 CYP706C55. *Plant Physiology*, 178(3), 1081-1095. <https://doi.org/10.1104/pp.18.00998>
- Hartmann, T. (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*, 68(22-24), 2831-2846. <https://doi.org/10.1016/j.phytochem.2007.09.017>
- Hassan, Mohd. Z., Osman, H., Ali, M. A., & Ahsan, M. J. (2016). Therapeutic potential of coumarins as antiviral agents. *European Journal of Medicinal Chemistry*, 123, 236-255. <https://doi.org/10.1016/j.ejmech.2016.07.056>
- He, K., Iyer, K. R., Hayes, R. N., Sinz, M. W., Woolf, T. F., & Hollenberg, P. F. (1998). Inactivation of Cytochrome P450 3A4 by Bergamottin, a Component of Grapefruit Juice. *Chemical Research in Toxicology*, 11(4), 252-259. <https://doi.org/10.1021/tx970192k>
- Hebsgaard, S. (1996). Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Research*, 24(17), 3439-3452. <https://doi.org/10.1093/nar/24.17.3439>
- Hegnauer, R. (1986). Phytochemistry and plant taxonomy—An essay on the chemotaxonomy of higher plants. *Phytochemistry*, 25(7), 1519-1535. [https://doi.org/10.1016/S0031-9422\(00\)81204-2](https://doi.org/10.1016/S0031-9422(00)81204-2)
- Hehmann, M., Lukacinc, R., Ekiert, H., & Matern, U. (2004). Furanocoumarin biosynthesis in *Ammi majus* L. : Cloning of bergaptol O-methyltransferase. *European Journal of Biochemistry*, 271(5), 932-940. <https://doi.org/10.1111/j.1432-1033.2004.03995.x>
- Hehmann, M., Lukacinc, R., Ekiert, H., & Matern, U. (2004). Furanocoumarin biosynthesis in *Ammi majus* L. : Cloning of bergaptol O-methyltransferase. *European Journal of Biochemistry*, 271, 932-940.
- Heinke, R., Franke, K., Porzel, A., Wessjohann, L. A., Awadh Ali, N. A., & Schmidt, J. (2011). Furanocoumarins from *Dorstenia foetida*. *Phytochemistry*, 72(9), 929-934. <https://doi.org/10.1016/j.phytochem.2011.03.008>
- Heitman, J., Movva, N. R., & Hall, M. N. (1991). Targets for Cell Cycle Arrest by the Immunosuppressant Rapamycin in Yeast. *Science*, 253(5022), 905-909. <https://doi.org/10.1126/science.1715094>
- Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., & Mullineaux, P. M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Molecular Biology*, 42(6), 819-832. <https://doi.org/10.1023/A:1006496308160>
- Herms, D. A., & Mattson, W. J. (1992). The Dilemma of Plants : To Grow or Defend. *The Quarterly Review of Biology*, 67(3), 283-335. <https://doi.org/10.1086/417659>
- Hodgson, H., De La Peña, R., Stephenson, M. J., Thimmappa, R., Vincent, J. L., Sattely, E. S., & Osbourn, A. (2019). Identification of key enzymes responsible for protolimonoid biosynthesis in plants : Opening the door to azadirachtin production. *Proceedings of the National Academy of Sciences*, 116(34), 17096-17104. <https://doi.org/10.1073/pnas.1906083116>

- Hoengenaert, L., Wouters, M., Kim, H., De Meester, B., Morreel, K., Vandersyppe, S., Pollier, J., Desmet, S., Goeminne, G., Ralph, J., Boerjan, W., & Vanholme, R. (2022). Overexpression of the scopoletin biosynthetic pathway enhances lignocellulosic biomass processing. *Science Advances*, 8(28), eab05738. <https://doi.org/10.1126/sciadv.ab05738>
- Hood, E. E., Gelvin, S. B., Melchers, L. S., & Hoekema, A. (1993). NewAgrobacterium helper plasmids for gene transfer to plants. *Transgenic Research*, 2(4), 208-218. <https://doi.org/10.1007/BF01977351>
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M. Y., Nakanishi, H., Ikeda, K., ... Nishioka, T. (2010). MassBank: A public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7), 703-714. <https://doi.org/10.1002/jms.1777>
- Hou, A.-J., Fukai, T., Shimazaki, M., Sakagami, H., Sun, H.-D., & Nomura, T. (2001). Benzophenones and Xanthones with Isoprenoid Groups from *Cudrania cochinchinensis*. *Journal of Natural Products*, 64(1), 65-70. <https://doi.org/10.1021/np000406p>
- Hrycay, E. G., & Bandiera, S. M. (2015). Monooxygenase, Peroxidase and Peroxygenase Properties and Reaction Mechanisms of Cytochrome P450 Enzymes. In E. G. Hrycay & S. M. Bandiera (Eds.), *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450* (Vol. 851, p. 1-61). Springer International Publishing. [https://doi.org/10.1007/978-3-319-16009-2\\_1](https://doi.org/10.1007/978-3-319-16009-2_1)
- Huang, A. C., Jiang, T., Liu, Y.-X., Bai, Y.-C., Reed, J., Qu, B., Goossens, A., Nützmann, H.-W., Bai, Y., & Osbourn, A. (2019). A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science*, 364(6440), eaau6389. <https://doi.org/10.1126/science.aau6389>
- Huang, X.-C., Tang, H., Wei, X., He, Y., Hu, S., Wu, J.-Y., Xu, D., Qiao, F., Xue, J.-Y., & Zhao, Y. (2024). The gradual establishment of complex coumarin biosynthetic pathway in Apiaceae. *Nature Communications*, 15(1), 6864. <https://doi.org/10.1038/s41467-024-51285-x>
- Huang, X.-Q., & Dudareva, N. (2023). Plant specialized metabolism. *Current Biology*, 33(11), R473-R478. <https://doi.org/10.1016/j.cub.2023.01.057>
- Hudson, J., Fong, R., Altamirano, M., & Towers, G. (1987). Comparative Anti-Bacteriophage Activity of Naturally-Occurring Photosensitizers. *Planta Medica*, 53(06), 536-539. <https://doi.org/10.1055/s-2006-962804>
- Irmisch, S., Clavijo McCormick, A., Günther, J., Schmidt, A., Boeckler, G. A., Gershenzon, J., Unsicker, S. B., & Köllner, T. G. (2014). Herbivore-induced poplar cytochrome P450 enzymes of the CYP71 family convert aldoximes to nitriles which repel a generalist caterpillar. *The Plant Journal*, 80(6), 1095-1107. <https://doi.org/10.1111/tpj.12711>
- Jacobowitz, J. R., & Weng, J.-K. (2020). Exploring Uncharted Territories of Plant Specialized Metabolism in the Postgenomic Era. *Annual Review of Plant Biology*, 71(1), 631-658. <https://doi.org/10.1146/annurev-aplant-081519-035634>
- James, A. T., & Martin, A. J. P. (1951). *Gas-liquid Partition Chromatography: The Separation and Micro-estimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid*. 50.
- Jamil, I. N., Remali, J., Azizan, K. A., Nor Muhammad, N. A., Arita, M., Goh, H.-H., & Aizat, W. M. (2020). Systematic Multi-Omics Integration (MOI) Approach in Plant Systems Biology. *Frontiers in Plant Science*, 11, 944. <https://doi.org/10.3389/fpls.2020.00944>
- Jeřábek, P., Florián, J., & Martínek, V. (2016). Lipid molecules can induce an opening of membrane-facing tunnels in cytochrome P450 1A2. *Physical Chemistry Chemical Physics*, 18(44), 30344-30356. <https://doi.org/10.1039/C6CP03692A>

- Ji, J., Han, X., Zang, L., Li, Y., Lin, L., Hu, D., Sun, S., Ren, Y., Maker, G., Lu, Z., & Wang, L. (2024). Integrative multi-omics data elucidating the biosynthesis and regulatory mechanisms of furanocoumarins in Angelica dahurica. <https://doi.org/10.1101/2024.07.23.604792>
- Jian, X., Zhao, Y., Wang, Z., Li, S., Li, L., Luo, J., & Kong, L. (2020). Two CYP71AJ enzymes function as psoralen synthase and angelicin synthase in the biosynthesis of furanocoumarins in *Peucedanum praeruptorum* Dunn. *Plant Molecular Biology*, 104(3), 327-337. <https://doi.org/10.1007/s11103-020-01045-4>
- Jones, C. G., & Firn, R. D. (1991). *On the evolution of plant secondary chemical diversity*.
- Jumpur, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kaessmann, H., Vinckenbosch, N., & Long, M. (2009). RNA-based gene duplication: Mechanistic and evolutionary insights. *Nature Reviews Genetics*, 10(1), 19-31. <https://doi.org/10.1038/nrg2487>
- Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., Hehn, A., Bourgaud, F., & Yazaki, K. (2014). A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. *The Plant Journal*, 77(4), 627-638. <https://doi.org/10.1111/tpj.12409>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160-1166. <https://doi.org/10.1093/bib/bbx108>
- Kessler, A. (2015). The information landscape of plant constitutive and induced secondary metabolite production. *Current Opinion in Insect Science*, 8, 47-53. <https://doi.org/10.1016/j.cois.2015.02.002>
- Kessler, A., & Kalske, A. (2018). Plant Secondary Metabolite Diversity and Species Interactions. *Annual Review of Ecology, Evolution, and Systematics*, 49(1), 115-138. <https://doi.org/10.1146/annurev-ecolsys-110617-062406>
- Kiba, T., Kudo, T., Kojima, M., & Sakakibara, H. (2011). Hormonal control of nitrogen acquisition: Roles of auxin, abscisic acid, and cytokinin. *Journal of Experimental Botany*, 62(4), 1399-1409. <https://doi.org/10.1093/jxb/erq410>
- Kim, H. W., Wang, M., Leber, C. A., Nothias, L.-F., Reher, R., Kang, K. B., Van Der Hooft, J. J. J., Dorrestein, P. C., Gerwick, W. H., & Cottrell, G. W. (2021). NPClassifier : A Deep Neural Network-Based Structural Classification Tool for Natural Products. *Journal of Natural Products*, 84(11), 2795-2807. <https://doi.org/10.1021/acs.jnatprod.1c00399>
- Kim, J., & DellaPenna, D. (2006). Defining the primary route for lutein synthesis in plants: The role of *Arabidopsis* carotenoid  $\beta$ -ring hydroxylase CYP97A3. *Proceedings of the National Academy of Sciences*, 103(9), 3474-3479. <https://doi.org/10.1073/pnas.0511207103>
- Kitajima, J., Kimizuka, K., & Tanaka, Y. (1998). New Sterols and Triterpenoids of *Ficus pumila* Fruit. *Chemical and Pharmaceutical Bulletin*, 46(9), 1408-1411. <https://doi.org/10.1248/cpb.46.1408>
- Kitajima, S., Aoki, W., Shibata, D., Nakajima, D., Sakurai, N., Yazaki, K., Munakata, R., Taira, T., Kobayashi, M., Aburaya, S., Savadogo, E. H., Hibino, S., & Yano, H. (2018). Comparative multi-omics analysis reveals diverse latex-based defense strategies against pests among latex-producing organs of the fig tree (*Ficus carica*). *Planta*, 247(6), 1423-1438. <https://doi.org/10.1007/s00425-018-2880-3>
- Kitamura, N., Kohtani, S., & Nakagaki, R. (2005). Molecular aspects of furocoumarin reactions: Photophysics, photochemistry, photobiology, and structural analysis. *Journal of Photochemistry and Photobiology C*:

- Photochemistry Reviews*, 6(2-3), 168-185.  
<https://doi.org/10.1016/j.jphotochemre> v.2005.08.002
- Knoll, A., Fauser, F., & Puchta, H. (2014). DNA recombination in somatic plant cells: Mechanisms and evolutionary consequences. *Chromosome Research*, 22(2), 191-201. <https://doi.org/10.1007/s10577-014-9415-y>
- Kokina, A., Kibilds, J., & Liepins, J. (2014). Adenine auxotrophy - be aware: Some effects of adenine auxotrophy in *Saccharomyces cerevisiae* strain W303-1A. *FEMS Yeast Research*, 14(5), 697-707. <https://doi.org/10.1111/1567-1364.12154>
- Kossel, A. (1891). *DIE METHODEN DER MIKROSKOPISCHEN UNTERSUCHUNG*.
- Krieger, C. (2014). Identification moléculaire et caractérisation fonctionnelle d'une nouvelle sous-famille de cytochromes P450, CYP71AZ, impliquée dans la synthèse de furanocoumarines et de coumarines chez *Pastinaca sativa*.
- Krieger, C., Roselli, S., Kellner-Thielmann, S., Galati, G., Schneider, B., Grosjean, J., Olry, A., Ritchie, D., Matern, U., Bourgaud, F., & Hehn, A. (2018). The CYP71AZ P450 Subfamily: A Driving Factor for the Diversification of Coumarin Biosynthesis in Apiaceous Plants. *Frontiers in Plant Science*, 9, 820. <https://doi.org/10.3389/fpls.2018.00820>
- Kruse, T., Ho, K., Yoo, H.-D., Johnson, T., Hippely, M., Park, J.-H., Flavell, R., & Bobzin, S. (2008). In *Planta Biocatalysis Screen of P450s Identifies 8-Methoxysoralen as a Substrate for the CYP82C Subfamily, Yielding Original Chemical Structures. Chemistry & Biology*, 15(2), 149-156. <https://doi.org/10.1016/j.chembiol.2008.01.008>
- Kumar, S., Korra, T., Thakur, R., Arutselvan, R., Kashyap, A. S., Nehela, Y., Chaplygin, V., Minkina, T., & Keswani, C. (2023). Role of plant secondary metabolites in defence and transcriptional regulation in response to biotic stress.
- Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, 41(W1), W22-W28. <https://doi.org/10.1093/nar/gkt389>
- Kuster, R. M., Bernardo, R. R., Da Silva, A. J. R., Parente, J. P., & Mors, W. B. (1994). Furocoumarins from the rhizomes of *Dorstenia brasiliensis*. *Phytochemistry*, 36(1), 221-223. [https://doi.org/10.1016/S0031-9422\(00\)97041-9](https://doi.org/10.1016/S0031-9422(00)97041-9)
- Kutchan, T. M. (2001). Ecological Arsenal and Developmental Dispatcher. The Paradigm of Secondary Metabolism. *Plant Physiology*, 125(1), 58-60. <https://doi.org/10.1104/pp.125.1.58>
- Kviesis, J., Klímenkovs, I., Arbidans, L., Podjava, A., Klaviniš, M., & Liepiņš, E. (2019). Evaluation of furanocoumarins from seeds of the wild parsnip (*Pastinaca sativa* L. s.l.). *Journal of Chromatography B*, 1105, 54-66. <https://doi.org/10.1016/j.jchromb.2018.12.012>
- Lacchini, E., & Goossens, A. (2020). Combinatorial Control of Plant Specialized Metabolism: Mechanisms, Functions, and Consequences. *Annual Review of Cell and Developmental Biology*, 36(1), 291-313. <https://doi.org/10.1146/annurev-cellbio-011620-031429>
- Lakatos, L., Szittyá, G., Silhavy, D., & Burgýán, J. (2004). Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *The EMBO Journal*, 23(4), 876-884. <https://doi.org/10.1038/sj.emboj.7600096>
- Lallemand, B., Erhardt, M., Heitz, T., & Legrand, M. (2013). Sporopollenin Biosynthetic Enzymes Interact and Constitute a Metabolon Localized to the Endoplasmic Reticulum of Tapetum Cells. *Plant Physiology*, 162(2), 616-625. <https://doi.org/10.1104/pp.112.213124>
- Lansky, E. P., Paavilainen, H. M., Pawlus, A. D., & Newman, R. A. (2008). *Ficus* spp. (fig): Ethnobotany and potential as anticancer and anti-inflammatory agents. *Journal of Ethnopharmacology*, 119(2), 195-213. <https://doi.org/10.1016/j.jep.2008.06.025>
- Larbat, R., Adamowicz, S., Robin, C., Han, P., Desneux, N., & Le Bot, J. (2016). Interrelated responses of tomato plants and the leaf miner *Tuta absoluta* to

- nitrogen supply. *Plant Biology*, 18(3), 495-504.  
<https://doi.org/10.1111/plb.12425>
- Larbat, R., Hehn, A., Hans, J., Schneider, S., Juggé, H., Schneider, B., Matern, U., & Bourgaud, F. (2009). Isolation and Functional Characterization of CYP71AJ4 Encoding for the First P450 Monooxygenase of Angular Furanocoumarin Biosynthesis. *Journal of Biological Chemistry*, 284(8), 4776-4785.  
<https://doi.org/10.1074/jbc.M807351200>
- Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., Bourgaud, F., & Matern, U. (2007). Molecular Cloning and Functional Characterization of Psoralen Synthase, the First Committed Monooxygenase of Furanocoumarin Biosynthesis. *Journal of Biological Chemistry*, 282(1), 542-554.  
<https://doi.org/10.1074/jbc.M604762200>
- Larbat, R., Paris, C., Le Bot, J., & Adamowicz, S. (2014). Phenolic characterization and variability in leaves, stems and roots of Micro-Tom and patio tomatoes, in response to nitrogen limitation. *Plant Science*, 224, 62-73.  
<https://doi.org/10.1016/j.plantsci.2014.04.010>
- Laursen, T., Borch, J., Knudsen, C., Bavishi, K., Torta, F., Martens, H. J., Silvestro, D., Hatzakis, N. S., Wenk, M. R., Dafforn, T. R., Olsen, C. E., Motawia, M. S., Hamberger, B., Møller, B. L., & Bassard, J.-E. (2016). Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science*, 354(6314), 890-893.  
<https://doi.org/10.1126/science.aag2347>
- Lee, C., Lu, C., Kuo, Y., Chen, J., & Sun, G. (2004). New Prenylated Flavones from the Roots of *Ficus Beecheyana*. *Journal of the Chinese Chemical Society*, 51(2), 437-441.  
<https://doi.org/10.1002/jccs.200400067>
- Lee, D.-S., Nioche, P., Hamberg, M., & Raman, C. S. (2008). Structural insights into the evolutionary paths of oxylipin biosynthetic enzymes. *Nature*, 455(7211), 363-368.  
<https://doi.org/10.1038/nature07307>
- Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in Genetics*, 20(3), 116-122.  
<https://doi.org/10.1016/j.tig.2004.01.007>
- Li, D., & Gaquerel, E. (2021). Next-Generation Mass Spectrometry Metabolomics Revives the Functional Analysis of Plant Metabolic Diversity. *Annual Review of Plant Biology*, 72(1), 867-891.  
<https://doi.org/10.1146/annurev-aplant-071720-114836>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.  
<https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Liu, G., & Fu, D. (2023). Transcriptional regulation of tomato fruit quality. *Postharvest Biology and Technology*, 202, 112393.  
<https://doi.org/10.1016/j.postharvbio.2023.112393>
- Li, J., Nagpal, P., Vitart, V., McMorris, T. C., & Chory, J. (1996). A Role for Brassinosteroids in Light-Dependent Development of *Arabidopsis*. *Science*, 272(5260), 398-401.  
<https://doi.org/10.1126/science.272.5260.398>
- Li, Q., Dai, Y., Huang, X.-C., Sun, L., Wang, K., Guo, X., Xu, D., Wan, D., An, L., Wang, Z., Tang, H., Qi, Q., Zeng, H., Qin, M., Xue, J.-Y., & Zhao, Y. (2024). The chromosome-scale assembly of the *Notopterygium incisum* genome provides insight into the structural diversity of coumarins. *Acta Pharmaceutica Sinica B*, S2211383524001333.  
<https://doi.org/10.1016/j.apsb.2024.04.005>
- Li, R.-T., Zhao, A.-H., Sheng, Y.-H., Na, Z., & Sun, H.-D. (2005). Chemical constituents from *Schisandra plena*. *Journal of Asian Natural Products Research*, 7(6), 847-852.  
<https://doi.org/10.1080/1028602042000204045>
- Li, X., Schuler, M. A., & Berenbaum, M. R. (2007). Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics. *Annual Review of Entomology*, 52(1), 231-253.

- <https://doi.org/10.1146/annurev.ento.51.110104.151104>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923-930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lichman, B. R. (2020). Gene and genome duplications in the evolution of chemodiversity: Perspectives from studies of Lamiaceae. *Current Opinion in Plant Biology*.
- Limones-Mendez, M., Dugrand-Judek, A., Villard, C., Coqueret, V., Froelicher, Y., Bourgaud, F., Olry, A., & Hehn, A. (2020). Convergent evolution leading to the appearance of furanocoumarins in citrus plants. *Plant Science*, 292, 110392. <https://doi.org/10.1016/j.plantsci.2019.110392>
- Lindon, J. C., Holmes, E., Bolland, M. E., Stanley, E. G., & Nicholson, J. K. (2004). Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers*, 9(1), 1-31. <https://doi.org/10.1080/13547500410001668379>
- Lisch, D. (2013). How important are transposons for plant evolution? *Nature Reviews Genetics*, 14(1), 49-61. <https://doi.org/10.1038/nrg3374>
- Liu, K., Hughes, J. M. X., Hay, S., & Scrutton, N. S. (2017). Liver microsomal lipid enhances the activity and redox coupling of colocalized cytochrome P450 reductase-cytochrome P450 3A4 in nanodiscs. *The FEBS Journal*, 284(14), 2302-2319. <https://doi.org/10.1111/febs.14129>
- Liu, Y.-P., Yan, G., Xie, Y.-T., Lin, T.-C., Zhang, W., Li, J., Wu, Y.-J., Zhou, J.-Y., & Fu, Y.-H. (2020). Bioactive prenylated coumarins as potential anti-inflammatory and anti-HIV agents from Clausena lenis. *Bioorganic Chemistry*, 97, 103699. <https://doi.org/10.1016/j.bioorg.2020.103699>
- Llabres, G., Baiwir, M., Vilegas, W., Pozetti, G. L., & Yariwake Vilegas, J. H. (1992). A 1H and 13C NMR study of a novel naturally occurring furocoumarin from Dorstenia cayapia. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 48(10), 1347-1353. [https://doi.org/10.1016/0584-8539\(92\)80140-R](https://doi.org/10.1016/0584-8539(92)80140-R)
- Lopes, D., deOliveira, R. R., Kaplan, M. A., Lage, C. S., & Leitão, A. C. (2001). Photosensitization and Mutation Induced in *Escherichia coli* and *Saccharomyces cerevisiae* Strains by Dorstenin, a Psoralen Analog Isolated from *Dorstenia bahiensis*. *Planta Medica*, 67(9), 820-824. <https://doi.org/10.1055/s-2001-18858>
- Louca, S., Mazel, F., Doeblei, M., & Parfrey, L. W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLOS Biology*, 17(2), e3000106. <https://doi.org/10.1371/journal.pbio.3000106>
- Lynch, J. H., Huang, X.-Q., & Dudareva, N. (2021). Silent constraints: The hidden challenges faced in plant metabolic engineering. *Current Opinion in Biotechnology*, 69, 112-117. <https://doi.org/10.1016/j.copbio.2020.12014>
- Machado, A. F. P., Rønsted, N., Bruun-Lund, S., Pereira, R. A. S., & Paganucci De Queiroz, L. (2018). Atlantic forests to the all Americas: Biogeographical history and divergence times of Neotropical *Ficus* (Moraceae). *Molecular Phylogenetics and Evolution*, 122, 46-58. <https://doi.org/10.1016/j.ympev.2018.01.015>
- Maeda, H. A., & Fernie, A. R. (2021). Evolutionary History of Plant Metabolism. *Annual Review of Plant Biology*, 72(1), 185-216. <https://doi.org/10.1146/annurev-arplant-080620-031054>
- Makarov, A. (2000). Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry*, 72(6), 1156-1162. <https://doi.org/10.1021/ac991131p>
- Mamoucha, S., Fokialakis, N., & Christodoulakis, N. S. (2016). Leaf structure and histochemistry of *Ficus carica* (Moraceae), the fig tree. *Flora - Morphology, Distribution, Functional Ecology of Plants*, 218, 24-34. <https://doi.org/10.1016/j.flora.2015.11.003>
- Margalha, L., Confraria, A., & Baena-González, E. (2019). SnRK1 and TOR: Modulating

- growth-defense trade-offs in plant stress responses. *Journal of Experimental Botany*, 70(8), 2261-2274. <https://doi.org/10.1093/jxb/erz066>
- Marks, R. A., Hotaling, S., Frandsen, P. B., & VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. *Nature Plants*. <https://doi.org/10.1038/s41477-021-01031-8>
- Matsuda, Y., Iwabuchi, T., Wakimoto, T., Awakawa, T., & Abe, I. (2015). Uncovering the Unusual D-Ring Construction in Terretonin Biosynthesis by Collaboration of a Multifunctional Cytochrome P450 and a Unique Isomerase. *Journal of the American Chemical Society*, 137(9), 3393-3401. <https://doi.org/10.1021/jacs.5b00570>
- Matsumoto, S., Mizutani, M., Sakata, K., & Shimizu, B.-I. (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. <https://doi.org/10.1016/j.phytochem.2011.11.009>
- Mc Lafferty, F. W. (1959). *Mass Spectrometric Analysis*.
- McKey, D. (1974). Adaptive Patterns in Alkaloid Physiology. *The American Naturalist*, 108(961), 305-320. <https://doi.org/10.1086/282909>
- McLean, K. J., Girvan, H. M., Mason, A. E., Dunford, A. J., & Munro\*, A. W. (2011). Chapter 8. Structure, Mechanism and Function of Cytochrome P450 Enzymes. In S. P. De Visser & D. Kumar (Éds.), *Iron-Containing Enzymes* (p. 255-280). Royal Society of Chemistry. <https://doi.org/10.1039/9781849732987-00255>
- Mérillon, J.-M., & Ramawat, K. G. (Éds.). (2020). *Co-Evolution of Secondary Metabolites*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-96397-6>
- Mestres, J. (2004). Structure conservation in cytochromes P450. *Proteins: Structure, Function, and Bioinformatics*, 58(3), 596-609. <https://doi.org/10.1002/prot.20354>
- Méteignier, L.-V., Nützmann, H.-W., Papon, N., Osbourn, A., & Courdavault, V. (2022). Emerging mechanistic insights into the regulation of specialized metabolism in plants. *Nature Plants*. <https://doi.org/10.1038/s41477-022-01288-7>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6), 679-682. <https://doi.org/10.1038/s41592-022-01488-1>
- Mizutani, M. (2012). Impacts of Diversification of Cytochrome P450 on Plant Metabolism. *Biological and Pharmaceutical Bulletin*, 35(6), 824-832. <https://doi.org/10.1248/bpb.35.824>
- Mizutani, M., & Ohta, D. (2010). Diversification of P450 Genes During Land Plant Evolution. *Annual Review of Plant Biology*, 61(1), 291-315. <https://doi.org/10.1146/annurev-arplant-042809-112305>
- Mo, Y., Nagel, C., & Taylor, L. P. (1992). Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proceedings of the National Academy of Sciences*, 89(15), 7213-7217. <https://doi.org/10.1073/pnas.89.15.7213>
- Modolo, L. V., Reichert, A. I., & Dixon, R. A. (2009). Introduction to the Different Classes of Biosynthetic Enzymes. In A. E. Osbourn & V. Lanzotti (Éds.), *Plant-derived Natural Products* (p. 143-163). Springer US. [https://doi.org/10.1007/978-0-387-85498-4\\_6](https://doi.org/10.1007/978-0-387-85498-4_6)
- Moghe, G., & Last, R. L. (2015). Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiology*, pp.00994.2015. <https://doi.org/10.1104/pp.15.00994>
- Molina-Venegas, R., Rodríguez, M. Á., Pardo-de-Santayana, M., & Mabberley, D. J. (2021). A global database of plant services for humankind. *PLOS ONE*, 16(6), e0253069. <https://doi.org/10.1371/journal.pone.0253069>
- Monson, R. K., Trowbridge, A. M., Lindroth, R. L., & Lerdau, M. T. (2022). Coordinated resource allocation to plant growth-defense tradeoffs. *New Phytologist*,

- 233(3), 1051-1066.  
<https://doi.org/10.1111/nph.17773>
- Moore, B. M., Wang, P., Fan, P., Leong, B., Schenck, C. A., Lloyd, J. P., Lehti-Shiu, M. D., Last, R. L., Pichersky, E., & Shiu, S.-H. (2019). Robust predictions of specialized metabolism genes through machine learning. *Proceedings of the National Academy of Sciences*, 116(6), 2344-2353.  
<https://doi.org/10.1073/pnas.18170741>  
16
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, 9(8), e1001127.  
<https://doi.org/10.1371/journal.pbio.1001127>
- Morikawa, T., Mizutani, M., Aoki, N., Watanabe, B., Saga, H., Saito, S., Oikawa, A., Suzuki, H., Sakurai, N., Shibata, D., Wadano, A., Sakata, K., & Ohta, D. (2006). Cytochrome P450 CYP710A Encodes the Sterol C-22 Desaturase in *Arabidopsis* and Tomato. *The Plant Cell*, 18(4), 1008-1022.  
<https://doi.org/10.1105/tpc.105.036012>
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H., & Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, 115(10).  
<https://doi.org/10.1073/pnas.17195881>  
15
- Muller, K. (1940). Experimentelle untersuchungen über die Phytophthoraresistenz der kartoffel. *Reichsanstalt Landw Forstw Berlin Arb Biol*, 23, 189-231.
- Munakata, R., Kitajima, S., Nuttens, A., Tatsumi, K., Takemura, T., Ichino, T., Galati, G., Vautrin, S., Bergès, H., Grosjean, J., Bourgaud, F., Sugiyama, A., Hehn, A., & Yazaki, K. (2020). Convergent evolution of the UbiA prenyltransferase family underlies the independent acquisition of furanocoumarins in plants. *New Phytologist*, 225(5), 2166-2182.  
<https://doi.org/10.1111/nph.16277>
- Munakata, R., Olry, A., Karamat, F., Courdavault, V., Sugiyama, A., Date, Y., Krieger, C., Silie, P., Foureau, E., Papon, N., Grosjean, J., Yazaki, K., Bourgaud, F., & Hehn, A. (2016). Molecular evolution of parsnip (*Pastinaca sativa*) membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. *New Phytologist*, 211(1), 332-344.  
<https://doi.org/10.1111/nph.13899>
- Murat, F., Peer, Y. V. D., & Salse, J. (2012). Decoding Plant and Animal Genome Plasticity from Differential Paleo-Evolutionary Patterns and Processes. *Genome Biology and Evolution*, 4(9), 917-928.  
<https://doi.org/10.1093/gbe/evs066>
- Murray, R. D. H. (2002). The Naturally Occurring Coumarins. In W. Herz, H. Falk, G. W. Kirby, & R. E. Moore (Eds.), *Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products* (Vol. 83, p. 1-619). Springer Vienna.  
[https://doi.org/10.1007/978-3-7091-6172-2\\_1](https://doi.org/10.1007/978-3-7091-6172-2_1)
- Musajo, L., Bordin, F., Caporale, G., Marciani, S., & Rigatti, G. (1967). PHOTOREACTIONS AT 3655Å BETWEEN PYRIMIDINE BASES AND SKIN-PHOTOSENSITIZING FUROCOUMARINS. *Photochemistry and Photobiology*, 6(10), 711-719.  
<https://doi.org/10.1111/j.1751-1097.1967.tb08736.x>
- Mutwil, M. (2020). Computational approaches to unravel the pathways and evolution of specialized metabolism. *Current Opinion in Plant Biology*, 55, 38-46.  
<https://doi.org/10.1016/j.pbi.2020.01.007>
- Nakanishi, K. (s. d.). *A Brief History of Natural Products Chemistry*.
- Nana, F., Sandjo, L. P., Keumedjio, F., Ambassa, P., Malik, R., Kuete, V., Rincheval, V., Choudhary, M. I., & Ngadjui, B. T. (2012). Ceramides and cytotoxic constituents from *Ficus glumosa* Del. (Moraceae). *Journal of the Brazilian Chemical Society*, 23(3), 482-487.  
<https://doi.org/10.1590/S0103-50532012000300015>
- Navrátilová, V., Palonciová, M., Berka, K., & Otyepka, M. (2016). Effect of Lipid Charge on Membrane Immersion of Cytochrome P450 3A4. *The Journal of Physical Chemistry B*, 120(43), 11205-11213.  
<https://doi.org/10.1021/acs.jpcb.6b10108>
- Negin, B., & Jander, G. (2023). Convergent and divergent evolution of plant chemical defenses. *Current Opinion in Plant Biology*, 73, 102368.

- <https://doi.org/10.1016/j.pbi.2023.102368>
- Neilson, E. H., Goodger, J. Q. D., Woodrow, I. E., & Møller, B. L. (2013). Plant chemical defense : At what cost? *Trends in Plant Science*, 18(5), 250-258. <https://doi.org/10.1016/j.tplants.2013.01.001>
- Nelson, D. R. (2018). Cytochrome P450 diversity in the tree of life. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1866(1), 141-154. <https://doi.org/10.1016/j.bbapap.2017.05.003>
- Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., & Nebert, D. W. (1996). P450 superfamily: Update on new sequences, gene mapping, accession numbers and nomenclature: *Pharmacogenetics*, 6(1), 1-42. <https://doi.org/10.1097/00008571-199602000-00002>
- Nelson, D., & Werck-Reichhart, D. (2011). A P450-centric view of plant evolution : P450-centric evolution. *The Plant Journal*, 66(1), 194-211. <https://doi.org/10.1111/j.1365-313X.2011.04529.x>
- Ngadjui, B. T., Dongo, E., Happi, E. N., Bezabih, M.-T., & Abegaz, B. M. (1998). Prenylated flavones and phenylpropanoid derivatives from roots of Dorstenia psilurus. *Phytochemistry*, 48(4), 733-737. [https://doi.org/10.1016/S0031-9422\(98\)00017-X](https://doi.org/10.1016/S0031-9422(98)00017-X)
- Ngameni, B., Touaibia, M., Patnam, R., Belkaid, A., Sonna, P., Ngadjui, B. T., Annabi, B., & Roy, R. (2006). Inhibition of MMP-2 secretion from brain tumor cells suggests chemopreventive properties of a furanocoumarin glycoside and of chalcones isolated from the twigs of Dorstenia turbinata. *Phytochemistry*, 67(23), 2573-2579. <https://doi.org/10.1016/j.phytochem.2006.09.017>
- Nisar, N., Li, L., Lu, S., Khin, N. C., & Pogson, B. J. (2015). Carotenoid Metabolism in Plants. *Molecular Plant*, 8(1), 68-82. <https://doi.org/10.1016/j.molp.2014.12.007>
- Niu, G., Guo, Q., Wang, J., Zhao, S., He, Y., & Liu, L. (2020). Structural basis for plant lutein biosynthesis from  $\alpha$ -carotene. *Proceedings of the National Academy of Sciences*, 117(25), 14150-14157. <https://doi.org/10.1073/pnas.2001806117>
- Norkunas, K., Harding, R., Dale, J., & Dugdale, B. (2018). Improving agroinfiltration-based transient gene expression in *Nicotiana benthamiana*. *Plant Methods*, 14(1), 71. <https://doi.org/10.1186/s13007-018-0343-2>
- Nützmann, H., Huang, A., & Osbourn, A. (2016). Plant metabolic clusters – from genetics to genomics. *New Phytologist*, 211(3), 771-789. <https://doi.org/10.1111/nph.13981>
- Nützmann, H.-W., & Osbourn, A. (2014). Gene clustering in plant specialized metabolism. *Current Opinion in Biotechnology*, 26, 91-99. <https://doi.org/10.1016/j.copbio.2013.10.009>
- Oliver, S. G., Winson, M. K., Douglas B. Kell, & Frank Baganz. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, 16(9), 373-378. [https://doi.org/10.1016/S0167-7799\(98\)01214-1](https://doi.org/10.1016/S0167-7799(98)01214-1)
- Omura, T., & Gotoh, O. (2017). Evolutionary origin of mitochondrial cytochrome P450. *The Journal of Biochemistry*, 161(5), 399-407. <https://doi.org/10.1093/jb/mvx011>
- Omura, T., & Sato, R. (1962). A New Cytochrome in Liver Microsomes. *Journal of Biological Chemistry*, 237(4), PC1375-PC1376. [https://doi.org/10.1016/S0021-9258\(18\)60338-2](https://doi.org/10.1016/S0021-9258(18)60338-2)
- Omura, T., & Sato, R. (1964). The Carbon Monoxide-binding Pigment of Liver Microsomes. *Journal of Biological Chemistry*, 239(7), 2370-2378. [https://doi.org/10.1016/S0021-9258\(20\)82244-3](https://doi.org/10.1016/S0021-9258(20)82244-3)
- Ono, E., & Murata, J. (2023). Exploring the Evolvability of Plant Specialized Metabolism: Uniqueness Out Of Uniformity and Uniqueness Behind Uniformity. *Plant And Cell Physiology*, 64(12), 1449-1465. <https://doi.org/10.1093/pcp/pcad057>
- Ortiz de Montellano, P. R. (Ed.). (2015). *Cytochrome P450*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-12108-6>

- Ortiz De Montellano, P. R., & De Voss, J. J. (2002). Oxidizing species in the mechanism of cytochrome P450. *Natural Product Reports*, 19(4), 477-493. <https://doi.org/10.1039/b101297p>
- Panchy, N., Lehti-Shiu, M., & Shiu, S.-H. (2016). Evolution of Gene Duplication in Plants. *Plant Physiology*, 171(4), 2294-2316. <https://doi.org/10.1104/pp.16.00523>
- Paxton, J. D. (1981). Phytoalexins—A Working Redefinition. *Journal of Phytopathology*, 101(2), 106-109. <https://doi.org/10.1111/j.1439-0434.1981.tb03327.x>
- Pederneiras, L. C., Gaglioti, A. L., Romaniuc-Neto, S., & Mansano, V. D. F. (2018). The role of biogeographical barriers and bridges in determining divergent lineages in *Ficus* (Moraceae). *Botanical Journal of the Linnean Society*, 187(4), 594-613. <https://doi.org/10.1093/botlinnean/boy036>
- Peters, N. K., Frost, J. W., & Long, S. R. (1986). A Plant Flavone, Luteolin, Induces Expression of *Rhizobium meliloti* Nodulation Genes. *Science*, 233(4767), 977-980. <https://doi.org/10.1126/science.3738520>
- Peyret, H., & Lomonossoff, G. P. (2013). The pEAQ vector series : The easy and quick way to produce recombinant proteins in plants. *Plant Molecular Biology*, 83(1-2), 51-58. <https://doi.org/10.1007/s11103-013-0036-1>
- Pichersky, E., & Lewinsohn, E. (2011). Convergent Evolution in Plant Specialized Metabolism. *Annual Review of Plant Biology*, 62(1), 549-566. <https://doi.org/10.1146/annurev-arplant-042110-103814>
- Pistelli, L., Chiellini, E. E., & Morelli, I. (2000). Flavonoids from *Ficus pumila*. *Biochemical Systematics and Ecology*, 28(3), 287-289. [https://doi.org/10.1016/S0305-1978\(99\)00064-2](https://doi.org/10.1016/S0305-1978(99)00064-2)
- Pluskal, T., Torrens-Spence, M. P., Fallon, T. R., De Abreu, A., Shi, C. H., & Weng, J.-K. (2019). The biosynthetic origin of psychoactive kavalactones in kava. *Nature Plants*, 5(8), 867-878. <https://doi.org/10.1038/s41477-019-0474-0>
- Pompon, D., Louerat, B., Bronine, A., & Urban, P. (1996). [6] Yeast expression of animal and plant P450s in optimized redox environments. In *Methods in Enzymology* (Vol. 272, p. 51-64). Elsevier. <https://linkinghub.elsevier.com/retrieve/pii/S0076687996720086>
- Pravda, L., Berka, K., Svbodová Vařeková, R., Sehnal, D., Banáš, P., Laskowski, R. A., Koča, J., & Otyepka, M. (2014). Anatomy of enzyme channels. *BMC Bioinformatics*, 15(1), 379. <https://doi.org/10.1186/s12859-014-0379-x>
- Pravda, L., Sehnal, D., Toušek, D., Navrátilová, V., Bazgier, V., Berka, K., Svbodová Vařeková, R., Koča, J., & Otyepka, M. (2018). MOLEonline : A web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Research*, 46(W1), W368-W373. <https://doi.org/10.1093/nar/gky309>
- Pumplin, N., & Voinnet, O. (2013). RNA silencing suppression by plant pathogens : Defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology*, 11(11), 745-760. <https://doi.org/10.1038/nrmicro3120>
- Qiu, J., Ni, L., Xia, X., Chen, S., Zhang, Y., Lang, M., Li, M., Liu, B., Pan, Y., Li, J., & Zhang, X. (2022). Genome-Wide Analysis of the Protein Phosphatase 2C Genes in Tomato. *Genes*, 13(4), 604. <https://doi.org/10.3390/genes13040604>
- Rabeh, K., Oubohssaine, M., & Hnini, M. (2024). TOR in plants : Multidimensional regulators of plant growth and signaling pathways. *Journal of Plant Physiology*, 294, 154186. <https://doi.org/10.1016/j.jplph.2024.154186>
- Rajniak, J., Giehl, R. F. H., Chang, E., Murgia, I., Von Wirén, N., & Sattely, E. S. (2018). Biosynthesis of redox-active metabolites in response to iron deficiency in plants. *Nature Chemical Biology*, 14(5), 442-450. <https://doi.org/10.1038/s41589-018-0019-2>
- Rajniak, J., Giehl, R. F. H., Chang, E., Murgia, I., von Wirén, N., & Sattely, E. S. (2018). Biosynthesis of redox-active metabolites in response to iron deficiency in plants. *Nature Chemical Biology*, 14(5), 442-450.

- <https://doi.org/10.1038/s41589-018-0019-2>
- Ray, A., & Langer, M. (2002). Homologous recombination: Ends as the means. *Trends in Plant Science*, 7(10), 435-440. [https://doi.org/10.1016/S1360-1385\(02\)02327-0](https://doi.org/10.1016/S1360-1385(02)02327-0)
- Razavi, S. M. (2010). Plant Coumarins as Allelopathic Agents. *International Journal of Biological Chemistry*, 5(1), 86-90. <https://doi.org/10.3923/ijbc.2011.86.90>
- Razavi, S. M., & Zarrini, G. (2010). Bioactivity of aviprin and aviprin-3"-O-glucoside, two linear furanocoumarins from Apiaceae. *Russian Journal of Bioorganic Chemistry*, 36(3), 359-362. <https://doi.org/10.1134/S1068162010030118>
- Ražná, K., Nôžková, J., Vargaová, A., Harenčár, L., & Bželková, M. (2021). Biological functions of lignans in plants. *Agriculture (Pol'nohospodárstvo)*, 67(4), 155-165. <https://doi.org/10.2478/agri-2021-0014>
- Robe, K., Conejero, G., Gao, F., Lefebvre-Legendre, L., Sylvestre-Gonon, E., Rofidal, V., Hem, S., Rouhier, N., Barberon, M., Hecker, A., Gaymard, F., Izquierdo, E., & Dubos, C. (2021). Coumarin accumulation and trafficking in *Arabidopsis thaliana*: A complex and dynamic process. *New Phytologist*, 229(4), 2062-2079. <https://doi.org/10.1111/nph.17090>
- Robe, K., Izquierdo, E., Vignols, F., Rouached, H., & Dubos, C. (2021). The Coumarins: Secondary Metabolites Playing a Primary Role in Plant Nutrition and Health. *Trends in Plant Science*, 26(3), 248-259. <https://doi.org/10.1016/j.tplants.2020.10.008>
- Robe, K., Stassen, M., Chamieh, J., Gonzalez, P., Hem, S., Santoni, V., Dubos, C., & Izquierdo, E. (2021). Uptake of Fe-fraxetin complexes, an IRT1 independent strategy for iron acquisition in *Arabidopsis thaliana* [Preprint]. *Plant Biology*. <https://doi.org/10.1101/2021.08.03.454955>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rodrigues, J. L., Gomes, D., & Rodrigues, L. R. (2022). Challenges in the Heterologous Production of Furanocoumarins in *Escherichia coli*. *Molecules*, 27(21), 7230. <https://doi.org/10.3390/molecules27217230>
- Rojas-Lima, S., Santillan, R. L., Domínguez, M.-A., & Gutiérrez, A. (1999). Furocoumarins of three species of the genus *Dorstenia*. *Phytochemistry*, 50(5), 863-868. [https://doi.org/10.1016/S0031-9422\(98\)00520-2](https://doi.org/10.1016/S0031-9422(98)00520-2)
- Romeo, J. T. (1999). *Phytochemicals in Human Health Protection, Nutrition, and Plant Defense*. Springer US: Imprint: Springer.
- Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572-1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3), 539-542. <https://doi.org/10.1093/sysbio/sys029>
- Roselli, S., Olry, A., Vautrin, S., Coriton, O., Ritchie, D., Navrot, N., Krieger, C., Vialart, G., Bergès, H., Bourgaud, F., & Hehn, A. (2016). Genomic approach to decipher specialized plant pathways: A case study with furanocoumarins in parsnip. *The plant Journal*, Submitted.
- Roumani, M. (2021). *Les phénolamides de la tomate : Développement d'une approche d'ingénierie métabolique pour l'étude de leurs fonctions in planta, et évaluation de leurs activités biologiques*.
- Roumani, M., Besseau, S., Hehn, A., & Larbat, R. (2024). Functional characterization of a small gene family coding for putrescine hydroxycinnamoyltransferases, involved in phenolamide accumulation, in tomato. *Phytochemistry*, 229, 114271. <https://doi.org/10.1016/j.phytochem.2024.114271>

- Rovinski, J. M., Tewalt, G. L., & Sneden, A. T. (1987). Maquiroside A, a New Cytotoxic Cardiac Glycoside from Maquira calophylla. *Journal of Natural Products*, 50(2), 211-216. <https://doi.org/10.1021/np50050a015>
- Royer, M., Larbat, R., Le Bot, J., Adamowicz, S., & Robin, C. (2013). Is the C:N ratio a reliable indicator of C allocation to primary and defence-related metabolisms in tomato? *Phytochemistry*, 88, 25-33. <https://doi.org/10.1016/j.phytochem.2012.12.003>
- Rutz, A., Sorokina, M., Galgonek, J., Mietchen, D., Willighagen, E., Gaudry, A., Graham, J. G., Stephan, R., Page, R., Vondrášek, J., Steinbeck, C., Pauli, G. F., Wolfender, J.-L., Bisson, J., & Allard, P.-M. (2022). The LOTUS initiative for open knowledge management in natural products research. *eLife*, 11, e70780. <https://doi.org/10.7554/eLife.70780>
- Sachs, J. (1873). *GRUNDZÜGE DER PFLANZEN-PHYSIOLOGIE*.
- Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., & Fernie, A. R. (2013). The flavonoid biosynthetic pathway in *Arabidopsis* : Structural and genetic diversity. *Plant Physiology and Biochemistry*, 72, 21-34. <https://doi.org/10.1016/j.plaphy.2013.02.001>
- Saito, R., Hayashi, K., Nomoto, H., Nakayama, M., Takaoka, Y., Saito, H., Yamagami, S., Muto, T., & Ueda, M. (2021). Extended JAZ degron sequence for plant hormone binding in jasmonate co-receptor of tomato SICO11-SUAZ. *Scientific Reports*, 11(1), 13612. <https://doi.org/10.1038/s41598-021-93067-1>
- Sakurai, T., Yamada, Y., Sawada, Y., Matsuda, F., Akiyama, K., Shinozaki, K., Hirai, M. Y., & Saito, K. (2013). PRIMe Update : Innovative Content for Plant Metabolomics and Integration of Gene Expression and Metabolite Accumulation. *Plant and Cell Physiology*, 54(2), e5-e5. <https://doi.org/10.1093/pcp/pcs184>
- Samanta, A., Das, G., & Das, S. K. (2011). *ROLES OF FLAVONOIDS IN PLANTS*. 6(1).
- Sarker, S. D., & Nahar, L. (2017). Progress in the Chemistry of Naturally Occurring Coumarins. In A. D. Kinghorn, H. Falk, S. Gibbons, & J. Kobayashi (Eds.), *Progress in the Chemistry of Organic Natural Products* 106 (Vol. 106, p. 241-304). Springer International Publishing. [https://doi.org/10.1007/978-3-319-59542-9\\_3](https://doi.org/10.1007/978-3-319-59542-9_3)
- Sarrión-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juárez, P., Fernández-del-Carmen, A., Granell, A., & Orzaez, D. (2011). GoldenBraid : An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules. *PLoS ONE*, 6(7), e21622. <https://doi.org/10.1371/journal.pone.0021622>
- Sarrión-Perdigones, A., Vazquez-Vilar, M., Palaci, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., & Orzaez, D. (2013). GoldenBraid 2.0 : A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *PLANT PHYSIOLOGY*, 162(3), 1618-1631. <https://doi.org/10.1104/pp.113.217661>
- Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., Akiyama, K., Sakurai, T., Matsuda, F., Aoki, T., Hirai, M. Y., & Saito, K. (2012). RIKEN tandem mass spectral database (ReSpect) for phytochemicals : A plant-specific MS/MS-based data resource and database. *Phytochemistry*, 82, 38-45. <https://doi.org/10.1016/j.phytochem.2012.07.007>
- Schelkunov, M. I., Shtratnikova, V. Yu., Klepikova, A. V., Makarenko, M. S., Omelchenko, D. O., Novikova, L. A., Obukhova, E. N., Bogdanov, V. P., Penin, A. A., & Logacheva, M. D. (2024). The genome of the toxic invasive species *Heracleum sosnowskyi* carries an increased number of genes despite absence of recent whole-genome duplications. *The Plant Journal*, 117(2), 449-463. <https://doi.org/10.1111/tpj.16500>
- Schlötterer, C. (2015). Genes from scratch – the evolutionary fate of de novo genes. *Trends in Genetics*, 31(4), 215-219. <https://doi.org/10.1016/j.tig.2015.02.007>
- Schmid, N. B., Giehl, R. F. H., Doll, S., Mock, H.-P., Strehmel, N., Scheel, D., Kong, X., Hider, R. C., & Von Wieren, N. (2014). Feruloyl-CoA 6'-Hydroxylase1-Dependent Coumarins Mediate Iron Acquisition from Alkaline Substrates in

- Arabidopsis. *PLANT PHYSIOLOGY*, 164(1), 160-172. <https://doi.org/10.1104/pp.113.228544>
- Schmid, N. B., Giehl, R. F. H., Doll, S., Mock, H.-P., Strehmel, N., Scheel, D., Kong, X., Hider, R. C., & von Wieren, N. (2014). Feruloyl-CoA 6'-Hydroxylase1-Dependent Coumarins Mediate Iron Acquisition from Alkaline Substrates in Arabidopsis. *PLANT PHYSIOLOGY*, 164(1), 160-172. <https://doi.org/10.1104/pp.113.228544>
- Schuhegger, R., Nafisi, M., Mansourova, M., Petersen, B. L., Olsen, C. E., Svatoš, A., Halkier, B. A., & Glawischnig, E. (2006). CYP71B15 (PAD3) Catalyzes the Final Step in Camalexin Biosynthesis. *Plant Physiology*, 141(4), 1248-1254. <https://doi.org/10.1104/pp.106.082024>
- Schuler, M. A. (2011). P450s in plant-insect interactions. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1814(1), 36-45. <https://doi.org/10.1016/j.bbapap.2010.09.012>
- Scott, B. R., Pathak, M. A., & Mohn, G. R. (1976). Molecular and genetic basis of furocoumarin reactions. *Mutation Research/Reviews in Genetic Toxicology*, 39(1), 29-74. [https://doi.org/10.1016/0165-1110\(76\)90012-9](https://doi.org/10.1016/0165-1110(76)90012-9)
- Seligmann, B., Liu, S., & Franke, J. (2024). Chemical tools for unpicking plant specialised metabolic pathways. *Current Opinion in Plant Biology*, 80, 102554. <https://doi.org/10.1016/j.pbi.2024.102554>
- Shahaf, N., Rogachev, I., Heinig, U., Meir, S., Malitsky, S., Battat, M., Wyner, H., Zheng, S., Wehrens, R., & Aharoni, A. (2016). The WEIZMASS spectral library for high-confidence metabolite identification. *Nature Communications*, 7(1), 12423. <https://doi.org/10.1038/ncomms12423>
- Shaw, R., & Cheung, C. Y. M. (2020). Multi-tissue to whole plant metabolic modelling. *Cellular and Molecular Life Sciences*, 77(3), 489-495. <https://doi.org/10.1007/s00018-019-03384-y>
- Shi, L.-S., Kuo, S.-C., Sun, H.-D., Morris-Natschke, S. L., Lee, K.-H., & Wu, T.-S. (2014). Cytotoxic cardiac glycosides and coumarins from *Antiaris toxicaria*. *Bioorganic & Medicinal Chemistry*, 22(6), 1889-1898. <https://doi.org/10.1016/j.bmc.2014.01.052>
- Shuai, H., Chen, T., Wlk, T., Rozhon, W., Pimenta Lange, M. J., Sieberer, T., Lange, T., & Poppenberger, B. (2022). SICESTA Is a Brassinosteroid-Regulated bHLH Transcription Factor of Tomato That Promotes Chilling Tolerance and Fruit Growth When Over-Expressed. *Frontiers in Plant Science*, 13, 930805. <https://doi.org/10.3389/fpls.2022.930805>
- Siwinska, J., Siatkowska, K., Olry, A., Grosjean, J., Hehn, A., Bourgaud, F., Meharg, A. A., Carey, M., Lojkowska, E., & Ihnatowicz, A. (2018). Scopoletin 8-hydroxylase : A novel enzyme involved in coumarin biosynthesis and iron-deficiency responses in Arabidopsis. *Journal of Experimental Botany*, 69(7), 1735-1748. <https://doi.org/10.1093/jxb/ery005>
- Sønderby, I. E., Geu-Flores, F., & Halkier, B. A. (2010). Biosynthesis of glucosinolates – gene discovery and beyond. *Trends in Plant Science*, 15(5), 283-290. <https://doi.org/10.1016/j.tplants.2010.02.005>
- Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics*, 13(1), 2. <https://doi.org/10.1186/s13321-020-00478-9>
- Soukup, R. W., & Soukup, K. (2015). The Series “Progress in the Chemistry of Organic Natural Products” : 75 Years of Service in the Development of Natural Product Chemistry. In A. D. Kinghorn, H. Falk, & J. Kobayashi (Eds.), *Progress in the Chemistry of Organic Natural Products 100* (Vol. 100, p. 453-588). Springer International Publishing. [https://doi.org/10.1007/978-3-319-05275-5\\_5](https://doi.org/10.1007/978-3-319-05275-5_5)
- Šrejber, M., Navrátilová, V., Palonciová, M., Bazgier, V., Berka, K., Anzenbacher, P., & Otyepka, M. (2018). Membrane-attached mammalian cytochromes P450 : An overview of the membrane’s effects on structure, drug binding, and interactions with redox partners. *Journal of Inorganic Biochemistry*, 183, 117-136. <https://doi.org/10.1016/j.jinorgbio.2018.03.002>

- Stahl, E. (1888). *PFLANZEN UND SCHNECKEN : Eine biologische studie a ber die schutzmittel der pflanzen gegen... schneckenfrass (classic reprint)*. FORGOTTEN BOOKS.
- Stahl, E. (Ed.). (1969). *Thin-Layer Chromatography*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-88488-7>
- Stamp, N. (2003). Out Of The Quagmire Of Plant Defense Hypotheses. *The Quarterly Review of Biology*, 78(1), 23-55. <https://doi.org/10.1086/367580>
- Stamp, N. (2004). Can the growth-differentiation balance hypothesis be tested rigorously? *Oikos*, 107(2), 439-448. <https://doi.org/10.1111/j.0030-1299.2004.12039.x>
- Stevenson, P. C., Simmonds, M. S. J., Yule, M. A., Veitch, N. C., Kite, G. C., Irwin, D., & Legg, M. (2003). Insect antifeedant furanocoumarins from *Tetradium daniellii*. *Phytochemistry*, 63(1), 41-46. [https://doi.org/10.1016/S0031-9422\(02\)00748-3](https://doi.org/10.1016/S0031-9422(02)00748-3)
- Stringlis, I. A., de Jonge, R., & Pieterse, C. M. J. (2019). The Age of Coumarins in Plant-Microbe Interactions. *Plant and Cell Physiology*, 60(7), 1405-1419. <https://doi.org/10.1093/pcp/pcz076>
- Sumorek-Wiadro, J., Zajac, A., Maciejczyk, A., & Jakubowicz-Gil, J. (2020). Furanocoumarins in anticancer therapy – For and against. *Fitoterapia*, 142, 104492. <https://doi.org/10.1016/j.fitote.2020.104492>
- Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E., & Mittler, R. (2014). Abiotic and biotic stress combinations. *New Phytologist*, 12. <https://doi.org/10.1111/nph.12797>
- Sweetlove, L. J., Nielsen, J., & Fernie, A. R. (2017). Engineering central metabolism – a grand challenge for plant biologists. *The Plant Journal*, 90(4), 749-763. <https://doi.org/10.1111/tpj.13464>
- Szekeres, M., Németh, K., Koncz-Kálmann, Z., Mathur, J., Kauschmann, A., Altmann, T., Rédei, G. P., Nagy, F., Schell, J., & Koncz, C. (1996). Brassinosteroids Rescue the Deficiency of CYP90, a Cytochrome P450, Controlling Cell Elongation and De-etiolation in *Arabidopsis*. *Cell*, 85(2), 171-182. [https://doi.org/10.1016/S0092-8674\(00\)81094-6](https://doi.org/10.1016/S0092-8674(00)81094-6)
- Takahashi, T., Okiura, A., & Kohno, M. (2017). Phenylpropanoid composition in fig (*Ficus carica* L.) leaves. *Journal of Natural Medicines*, 71(4), 770-775. <https://doi.org/10.1007/s11418-017-1093-6>
- Tautenhahn, R., Cho, K., Uritboonthai, W., Zhu, Z., Patti, G. J., & Siuzdak, G. (2012). An accelerated workflow for untargeted metabolomics using the METLIN database. *Nature Biotechnology*, 30(9), 826-828. <https://doi.org/10.1038/nbt.2348>
- Tautz, D., & Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. *Nature Reviews Genetics*, 12(10), 692-702. <https://doi.org/10.1038/nrg3053>
- Tawfik, O. K. and D. S. (2010). Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annual Review of Biochemistry*, 79(1), 471-505. <https://doi.org/10.1146/annurev-biochem-030409-143718>
- Terreaux, C., Maillard, M., Stoeckli-Evans, H., Gupta, M. P., Downum, K. R., Quirke, J. M. E., & Hostettmann, K. (1995). Structure revision of a furanocoumarin from *Dorstenia contrajerva*. *Phytochemistry*, 39(3), 645-647. [https://doi.org/10.1016/0031-9422\(95\)00092-L](https://doi.org/10.1016/0031-9422(95)00092-L)
- Thompson, J. N. (2005). Coevolution: The Geographic Mosaic of Coevolutionary Arms Races. *Current Biology*, 15(24), R992-R994. <https://doi.org/10.1016/j.cub.2005.11.046>
- Thorpe, T. A. (2007). History of plant tissue culture. *Molecular Biotechnology*, 37(2), 169-180. <https://doi.org/10.1007/s12033-007-0031-3>
- Tian, L., Musetti, V., Kim, J., Magallanes-Lundback, M., & DellaPenna, D. (2004). The *Arabidopsis LUT1* locus encodes a member of the cytochrome P450 family that is required for carotenoid ε-ring hydroxylation activity. *Proceedings of the National Academy of Sciences*, 101(1), 402-407. <https://doi.org/10.1073/pnas.2237237100>
- Tissier, A., Ziegler, J., & Vogt, T. (2014). Specialized Plant Metabolites: Diversity and Biosynthesis. In G. Krauss & D. H. Nies (Eds.), *Ecological Biochemistry* (1<sup>re</sup>

- éd., p. 14-37). Wiley.  
<https://doi.org/10.1002/9783527686063.ch2>
- Trott, O., & Olson, A. J. (2010). AutoDock Vina : Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455-461.  
<https://doi.org/10.1002/jcc.21334>
- True, J. R., & Carroll, S. B. (2002). Gene Co-Option in Physiological and Morphological Evolution. *Annual Review of Cell and Developmental Biology*, 18(1), 53-80.  
<https://doi.org/10.1146/annurev.cellbio.18.020402.140619>
- Tsai, H.-H., Rodríguez-Celma, J., Lan, P., Wu, Y.-C., Vélez-Bermúdez, I. C., & Schmidt, W. (2018). Scopoletin 8-Hydroxylase-Mediated Fraxetin Production Is Crucial for Iron Mobilization. *Plant Physiology*, 177(1), 194-207.  
<https://doi.org/10.1104/pp.18.00178>
- Tsugawa, H., Nakabayashi, R., Mori, T., Yamada, Y., Takahashi, M., Rai, A., Sugiyama, R., Yamamoto, H., Nakaya, T., Yamazaki, M., Kooke, R., Bac-Molenaar, J. A., Oztolan-Erol, N., Keurentjes, J. J. B., Arita, M., & Saito, K. (2019). A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nature Methods*, 16(4), 295-298.  
<https://doi.org/10.1038/s41592-019-0358-2>
- Tswett, M. (1906). 51. M. Tswett : Physikalisch-chemische Studien über das Chlorophyll. Die Adsorptionen. *Berichte der Deutschen Botanischen Gesellschaft*, 24(6), 316-323.  
<https://doi.org/10.1111/j.1438-8677.1906.tb06524.x>
- Uchiyama, T., Hara, S., Makino, M., & Fujimoto, Y. (2002). Seco-Adianane-type triterpenoids from Dorstenia brasiliensis (Moraceae). *Phytochemistry*, 60(8), 761-764.  
[https://doi.org/10.1016/S0031-9422\(02\)00180-2](https://doi.org/10.1016/S0031-9422(02)00180-2)
- Urban, P., Lautier, T., Pompon, D., & Truan, G. (2018). Ligand Access Channels in Cytochrome P450 Enzymes : A Review. *International Journal of Molecular Sciences*, 19(6), 1617.  
<https://doi.org/10.3390/ijms19061617>
- Urban, P., Mignotte, C., Kazmaier, M., Delorme, F., & Pompon, D. (1997). Cloning, Yeast Expression, and Characterization of the Coupling of Two Distantly Related *Arabidopsis thaliana*NADPH-Cytochrome P450 Reductases with P450 CYP73A5. *Journal of Biological Chemistry*, 272(31), 19176-19186.  
<https://doi.org/10.1074/jbc.272.31.19176>
- Usai, G., Mascagni, F., Giordani, T., Vangelisti, A., Bosi, E., Zuccolo, A., Ceccarelli, M., King, R., Hassani-Pak, K., Zambrano, L. S., Cavallini, A., & Natali, L. (2020). Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. *The Plant Journal*, 102(3), 600-614.  
<https://doi.org/10.1111/tpj.14635>
- Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.  
<https://doi.org/10.1007/s11192-009-0146-3>
- Van Der Meer, I. M., Stam, M. E., Van Tunen, A. J., Mol, J. N., & Stuitje, A. R. (1992). Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. *The Plant Cell*, 4(3), 253-262.  
<https://doi.org/10.1105/tpc.4.3.253>
- van der Hooft, J. J. J., Mohimani, H., Bauermeister, A., Dorrestein, P. C., Duncan, K. R., & Medema, M. H. (2020). Linking genomics and metabolomics to chart specialized metabolic diversity. *Chemical Society Reviews*, 49(11), 3297-3314.  
<https://doi.org/10.1039/D0CS00162G>
- Van Hoek, M. J. A., & Hogeweg, P. (2009). Metabolic Adaptation after Whole Genome Duplication. *Molecular Biology and Evolution*, 26(11), 2441-2453.  
<https://doi.org/10.1093/molbev/msp160>
- Vanholme, R., Sundin, L., Seetso, K. C., Kim, H., Liu, X., Li, J., De Meester, B., Hoengenaert, L., Goeminne, G., Morreel, K., Haustraete, J., Tsai, H.-H., Schmidt, W., Vanholme, B., Ralph, J., & Boerjan, W. (2019). COSY catalyses trans-cis isomerization and lactonization in the biosynthesis of coumarins. *Nature Plants*, 5(10), 1066-1075.  
<https://doi.org/10.1038/s41477-019-0510-0>

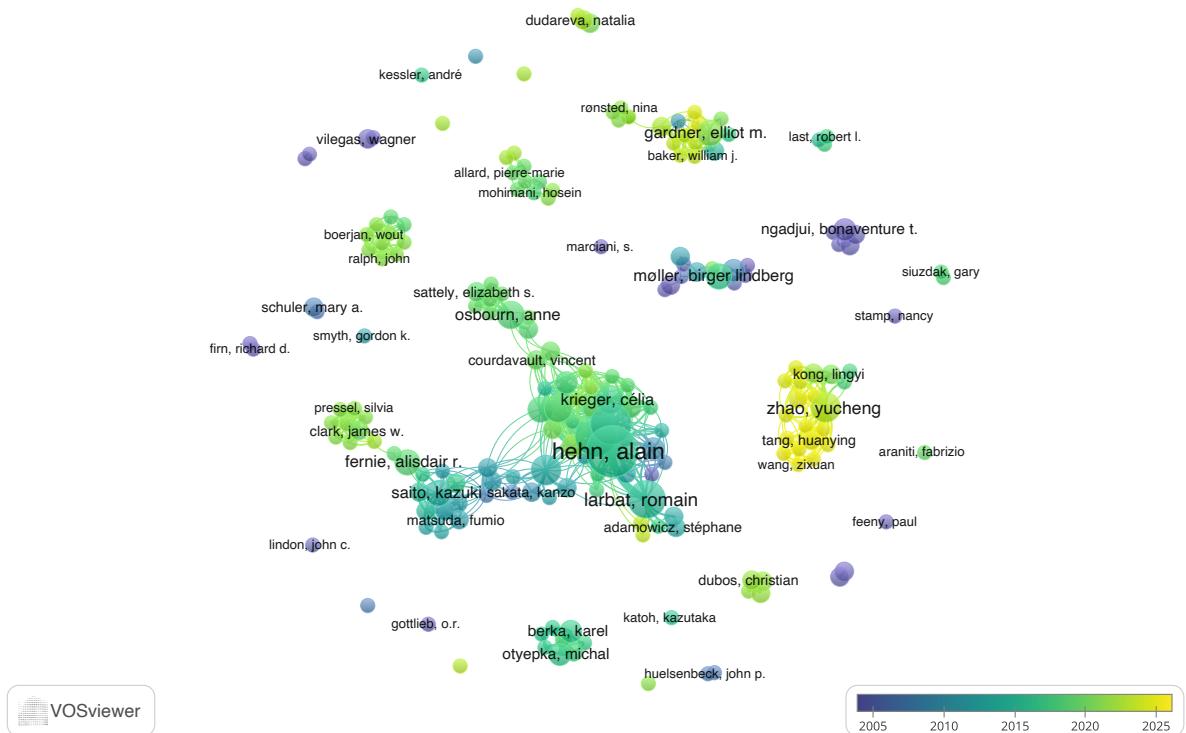
- van Santen, J. A., Poynton, E. F., Iskakova, D., McMann, E., Alsup, T. A., Clark, T. N., Fergusson, C. H., Fewer, D. P., Hughes, A. H., McCadden, C. A., Parra, J., Soldatou, S., Rudolf, J. D., Janssen, E. M.-L., Duncan, K. R., & Linington, R. G. (2022). The Natural Products Atlas 2.0: A database of microbially-derived natural products. *Nucleic Acids Research*, 50(D1), D1317-D1323. <https://doi.org/10.1093/nar/gkab941>
- Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., Paris, C., Shimizu, B., Sugimoto, Y., Mizutani, M., & Bourgaud, F. (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: C2'H involved in umbelliferone synthesis. *The Plant Journal*, 70(3), 460-470. <https://doi.org/10.1111/j.1365-313X.2011.04879.x>
- Vieira, I. J. C., Mathias, L., Monteiro, V. D. F. F., Braz-Filho, R., & Rodrigues-Filho, E. (1999). A New Coumarin from *Brosimum Gaudichaudii* Trecul. *Natural Product Letters*, 13(1), 47-52. <https://doi.org/10.1080/10575639908048490>
- Vilegas, J. H. Y., Lanças, F. M., Vilegas, W., & Pozetti, G. L. (1997). Further triterpenes, steroids and furocoumarins from brazilian medicinal plants of dorstenia genus (moraceae). *Journal of the Brazilian Chemical Society*, 8(5), 529-535. <https://doi.org/10.1590/S0103-50531997000500016>
- Vilegas, W., Pozetti, G. L., & Harumi Yariwake Vilegas, J. (1993). Coumarins from *Brosimum gaudichaudii*. *Journal of Natural Products*, 56(3), 416-417. <https://doi.org/10.1021/np50093a015>
- Villard, C. (2020). *HOW TO PRODUCE YOUR FURANOCOUMARINS: THE HIDDEN PATHWAY*.
- Villard, C., Munakata, R., Kitajima, S., Velzen, R., Schranz, M. E., Larbat, R., & Hehn, A. (2021). A new P450 involved in the furanocoumarin pathway underlies a recent case of convergent evolution. *New Phytologist*, 231(5), 1923-1939. <https://doi.org/10.1111/nph.17458>
- Vismans, G., Van Bentum, S., Spooren, J., Song, Y., Goossens, P., Valls, J., Snoek, B. L., Thiombiano, B., Schilder, M., Dong, L., Bouwmeester, H. J., Pétriacq, P., Pieterse, C. M. J., Bakker, P. A. H. M., & Berendsen, R. L. (2022). Coumarin biosynthesis genes are required after foliar pathogen infection for the creation of a microbial soil-borne legacy that primes plants for SA-dependent defenses. *Scientific Reports*, 12(1), 22473. <https://doi.org/10.1038/s41598-022-26551-x>
- Vries, S., Fürst-Jansen, J. M. R., Irisarri, I., Dhabalia Ashok, A., Ischebeck, T., Feussner, K., Abreu, I. N., Petersen, M., Feussner, I., & Vries, J. (2021). The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. *The Plant Journal*, 107(4), 975-1002. <https://doi.org/10.1111/tpj.15387>
- Waadt, R., Seller, C. A., Hsu, P.-K., Takahashi, Y., Munemasa, S., & Schroeder, J. I. (2022). Plant hormone regulation of abiotic stress responses. *Nature Reviews Molecular Cell Biology*, 23(10), 680-694. <https://doi.org/10.1038/s41580-022-00479-6>
- Wang, H., Wang, Q., Liu, Y., Liao, X., Chu, H., Chang, H., Cao, Y., Li, Z., Zhang, T., Cheng, J., & Jiang, H. (2021). PCPD: Plant cytochrome P450 database and web-based tools for structural construction and ligand docking. *Synthetic and Systems Biotechnology*, 6(2), 102-109. <https://doi.org/10.1016/j.synbio.2021.04.004>
- Wang, K., Zeng, H., Dai, Y., Wang, Z., Tang, H., Li, J., Lu, X., Jiang, N., Xie, G., Zhu, Y., Zhao, Y., & Qin, M. (2024). Three types of enzymes complete the furanocoumarins core skeleton biosynthesis in *Angelica sinensis*. *Phytochemistry*, 222, 114102. <https://doi.org/10.1016/j.phytochem.2024.114102>
- Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W.-T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., ... Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural

- Products Social Molecular Networking. *Nature Biotechnology*, 34(8), 828-837. <https://doi.org/10.1038/nbt.3597>
- Wang, X.-B., Li, G.-H., Li, L., Zheng, L.-J., Huang, R., & Zhang, K.-Q. (2008). Nematicidal coumarins from *Heracleum candicans* Wall. *Natural Product Research*, 22(8), 666-671. <https://doi.org/10.1080/14786410701766463>
- Wang, Z., Zhou, Y., Wang, Y., & Yan, X. (2023). Reconstitution and Optimization of the Marmesin Biosynthetic Pathway in Yeast. *ACS Synthetic Biology*, 12(10), 2922-2933. <https://doi.org/10.1021/acssynbio.3c00267>
- Warmund, M. R., Ellersieck, M. R., & Smeda, R. J. (2022). Sensitivity and Recovery of Tomato Cultivars Following Simulated Drift of Dicamba or 2,4-D. *Agriculture*, 12(9), 1489. <https://doi.org/10.3390/agriculture12091489>
- Weißenborn, S., & Walther, D. (2017). Metabolic Pathway Assignment of Plant Genes based on Phylogenetic Profiling—A Feasibility Study. *Frontiers in Plant Science*, 8, 1831. <https://doi.org/10.3389/fpls.2017.01831>
- Wen, Z., Berenbaum, M. R., & Schuler, M. A. (2006). Inhibition of CYP6B1-Mediated Detoxification of Xanthotoxin by Plant Allelochemicals in the Black Swallowtail (*Papilio polyxenes*). *Journal of Chemical Ecology*, 32(3), 507-522. <https://doi.org/10.1007/s10886-005-9014-7>
- Weng, J.-K., Lynch, J. H., Matos, J. O., & Dudareva, N. (2021). Adaptive mechanisms of plant specialized metabolism connecting chemistry to function. *Nature Chemical Biology*, 17(10), 1037-1045. <https://doi.org/10.1038/s41589-021-00822-6>
- Werck-Reichhart, D. (2023). Promiscuity, a Driver of Plant Cytochrome P450 Evolution? *Biomolecules*, 13(2), 394. <https://doi.org/10.3390/biom13020394>
- Werck-Reichhart, D., & Feyereisen, R. (2000). *Cytochromes P450: A success story*. 9. <https://doi.org/10.1186/gb-2000-1-6-reviews3003>
- Weryszko-Chmielewska, E., & Chwil, M. (2017). Localisation of furanocoumarins in the tissues and on the surface of shoots of *Heracleum sosnowskyi*. *Botany*, 95(11), 1057-1070. <https://doi.org/10.1139/cjb-2017-0043>
- Williams, E. W., Gardner, E. M., Harris, R., Chaveerach, A., Pereira, J. T., & Zerega, N. J. C. (2017). Out of Borneo: Biogeography, phylogeny and divergence date estimates of *Artocarpus* (Moraceae). *Annals of Botany*, mcw249. <https://doi.org/10.1093/aob/mcw249>
- Wink, M. (2008). Plant Secondary Metabolism: Diversity, Function and its Evolution. *Natural Product Communications*, 3(8), 1934578X0800300. <https://doi.org/10.1177/1934578X0800300>
- Wray, G. A. (2003). The Evolution of Transcriptional Regulation in Eukaryotes. *Molecular Biology and Evolution*, 20(9), 1377-1419. <https://doi.org/10.1093/molbev/msg140>
- Xiong, F., Tian, J., Wei, Z., Deng, K., Li, Y., & Zhang, Y. (2023). Suppression of the target of rapamycin kinase accelerates tomato fruit ripening through reprogramming the transcription profile and promoting ethylene biosynthesis. *Journal of Experimental Botany*, 74(8), 2603-2619. <https://doi.org/10.1093/jxb/erad056>
- Yamaguchi, T., Yamamoto, K., & Asano, Y. (2014). Identification and characterization of CYP79D16 and CYP71AN24 catalyzing the first and second steps in l-phenylalanine-derived cyanogenic glycoside biosynthesis in the Japanese apricot, *Prunus mume* Sieb. Et Zucc. *Plant Molecular Biology*, 86(1-2), 215-223. <https://doi.org/10.1007/s11103-014-0225-6>
- Yan, J., & Wang, X. (2022). Unsupervised and semi-supervised learning: The next frontier in machine learning for plant systems biology. *The Plant Journal*. <https://doi.org/10.1111/tpj.15905>
- Yan, Z., Wang, D., Cui, H., Zhang, D., Sun, Y., Jin, H., Li, X., Yang, X., Guo, H., He, X., Pan, L., Ren, X., Guo, K., & Qin, B. (2016). Phytotoxicity mechanisms of two coumarin allelochemicals from *Stellera chamaejasme* in lettuce seedlings. *Acta Physiologiae Plantarum*, 38(10), 248.

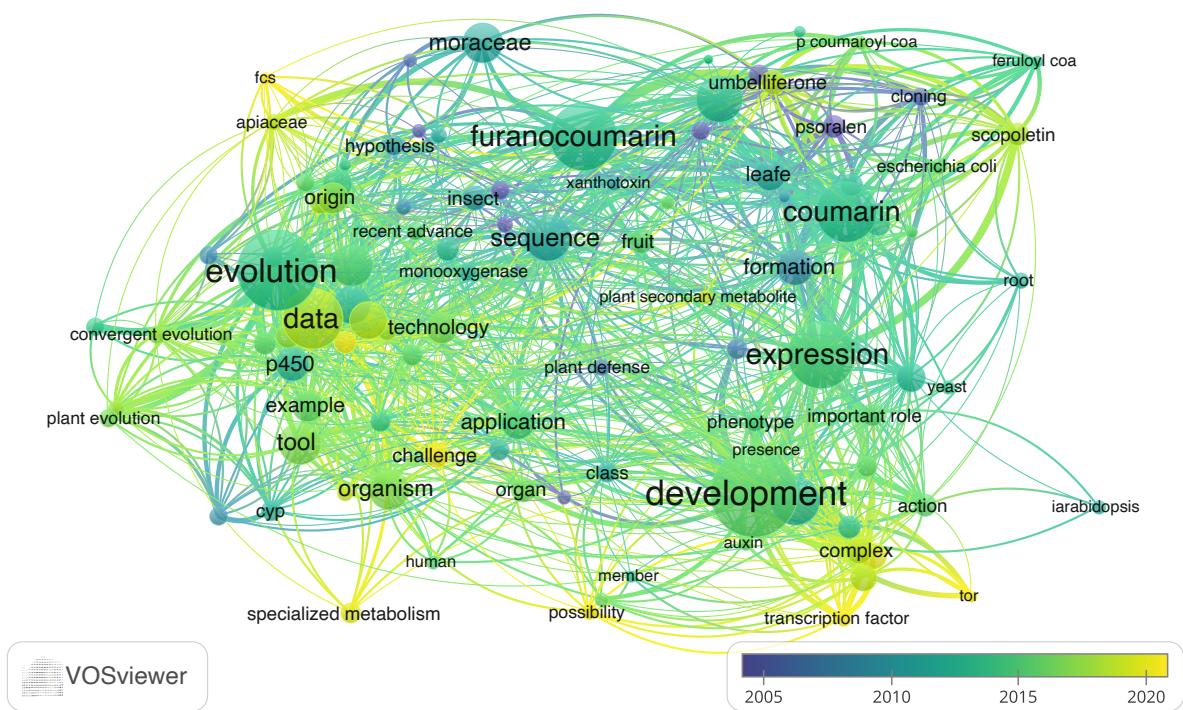
- <https://doi.org/10.1007/s11738-016-2270-z>
- Yang, C.-X., Liu, S.-Y., Zerega, N. J. C., Stull, G. W., Gardner, E. M., Tian, Q., Gu, W., Lu, Q., Folk, R. A., Kates, H. R., Guralnick, R. P., Soltis, D. E., Soltis, P. S., Wang, Y.-H., & Yi, T.-S. (2023). Phylogeny and Biogeography of *Morus* (Moraceae). *Agronomy*, 13(8), 2021. <https://doi.org/10.3390/agronomy13082021>
- Yang, J., Liao, M., Shou, M., Jamei, M., Yeo, K., Tucker, G., & Rostami-Hodjegan, A. (2008). Cytochrome P450 Turnover: Regulation of Synthesis and Degradation, Methods for Determining Rates, and Implications for the Prediction of Drug Interactions. *Current Drug Metabolism*, 9(5), 384-393. <https://doi.org/10.2174/138920008784746382>
- Yang, S.-M., Shim, G. Y., Kim, B.-G., & Ahn, J.-H. (2015). Biological synthesis of coumarins in *Escherichia coli*. *Microbial Cell Factories*, 14(1), 65. <https://doi.org/10.1186/s12934-015-0248-y>
- Yano, J. K., Hsu, M.-H., Griffin, K. J., Stout, C. D., & Johnson, E. F. (2005). Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nature Structural & Molecular Biology*, 12(9), 822-823. <https://doi.org/10.1038/nsmb971>
- Yarosh, É. A., & Nikonov, G. K. (1973). The coumarins of the genus *Ficus*. *Chemistry of Natural Compounds*, 9(2), 255-256. <https://doi.org/10.1007/BF00563357>
- Yosca, T. H., Ledray, A. P., Ngo, J., & Green, M. T. (2017). A new look at the role of thiolate ligation in cytochrome P450. *JBIC Journal of Biological Inorganic Chemistry*, 22(2-3), 209-220. <https://doi.org/10.1007/s00775-016-1430-3>
- Yu, J., Ke, T., Tehrim, S., Sun, F., Liao, B., & Hua, W. (2015). PTGBase: An integrated database to study tandem duplicated genes in plants. *Database*, 2015. <https://doi.org/10.1093/database/bav017>
- Zähringer, U., Ebel, J., & Grisebach, H. (1978). Induction of phytoalexin synthesis in Soybean. *Archives of Biochemistry and Biophysics*, 188(2), 450-455.
- [https://doi.org/10.1016/S0003-9861\(78\)80029-0](https://doi.org/10.1016/S0003-9861(78)80029-0)
- Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M., & Schell, J. (1983). Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *The EMBO Journal*, 2(12), 2143-2150. <https://doi.org/10.1002/j.1460-2075.1983.tb01715.x>
- Zangar, R. (2004). Mechanisms that regulate production of reactive oxygen species by cytochrome P450. *Toxicology and Applied Pharmacology*, 199(3), 316-331. <https://doi.org/10.1016/j.taap.2004.01.018>
- Zenk, M. H. (1991). 6. Chasing the enzymes of secondary metabolism: Plant cell cultures as a pot of gold. *Phytochemistry*, 30(12), 3861-3863. [https://doi.org/10.1016/0031-9422\(91\)83424-J](https://doi.org/10.1016/0031-9422(91)83424-J)
- Zerega, N. J. C., Clement, W. L., Datwyler, S. L., & Weiblen, G. D. (2005). Biogeography and divergence times in the mulberry family (Moraceae). *Molecular Phylogenetics and Evolution*, 37(2), 402-416. <https://doi.org/10.1016/j.ympev.2005.07.004>
- Zhang, H., Zhu, J., Gong, Z., & Zhu, J.-K. (2022). Abiotic stress responses in plants. *Nature Reviews Genetics*, 23(2), 104-119. <https://doi.org/10.1038/s41576-021-00413-0>
- Zhang, Q., Onstein, R. E., Little, S. A., & Sauquet, H. (2019). Estimating divergence times and ancestral breeding systems in *Ficus* and Moraceae. *Annals of Botany*, 123(1), 191-204. <https://doi.org/10.1093/aob/mcy159>
- Zhao, X., Zhao, Y., Gou, M., & Liu, C.-J. (2023). Tissue-preferential recruitment of electron transfer chains for cytochrome P450-catalyzed phenolic biosynthesis. *Science Advances*, 9(2), eade4389. <https://doi.org/10.1126/sciadv.ade4389>
- Zhao, Y., He, Y., Han, L., Zhang, L., Xia, Y., Yin, F., Wang, X., Zhao, D., Xu, S., Qiao, F., Xiao, Y., & Kong, L. (2024). Two types of coumarins-specific enzymes complete the last missing steps in pyran- and furanocoumarins biosynthesis. *Acta Pharmaceutica Sinica B*, 14(2), 869-880. <https://doi.org/10.1016/j.apsb.2023.10.016>

- Zhao, Y., Wang, N., Zeng, Z., Xu, S., Huang, C., Wang, W., Liu, T., Luo, J., & Kong, L. (2016). Cloning, Functional Characterization, and Catalytic Mechanism of a Bergaptol O-Methyltransferase from *Peucedanum praeruptorum* Dunn. *Frontiers in Plant Science*, 7. <https://doi.org/10.3389/fpls.2016.00722>
- Zhou, H., Jian, Y., Shao, Q., Guo, F., Zhang, M., Wan, F., Yang, L., Liu, Y., Yang, L., Li, Y., Yang, P., Li, Z., Li, S., & Ding, W. (2023). Development of Sustainable Insecticide Candidates for Protecting Pollinators: Insight into the Bioactivities, Selective Mechanism of Action and QSAR of Natural Coumarin Derivatives against Aphids. *Journal of Agricultural and Food Chemistry*, 71(47), 18359-18374. <https://doi.org/10.1021/acs.jafc.3c03493>
- Zuntini, A. R., Carruthers, T., Maurin, O., Bailey, P. C., Leempoel, K., Brewer, G. E., Epitawalage, N., Françoso, E., Gallego-Paramo, B., McGinnie, C., Negrão, R., Roy, S. R., Simpson, L., Toledo Romero, E., Barber, V. M. A., Botigué, L., Clarkson, J. J., Cowan, R. S., Dodsworth, S., ... Baker, W. J. (2024). Phylogenomics and the rise of the angiosperms. *Nature*. <https://doi.org/10.1038/s41586-024-07324-0>

# Analyse bibliométrique



**Figure 43: Analyse bibliométrique en réseau du co-auteurat dans le cadre de ce manuscrit pour des articles publiés entre 2005 et 2024.** Publications citées : 394. Co-auteurat : 2051 personnes. Les données ont été référencées et analysées via les logiciels Zotero et VOSviewer. La taille des ronds défini le nombre de citations dans des articles différents et cités dans le cadre de ce document pour un seuil minimal de co-occurrence de l'aut.eur.rice de 2 articles, soit 197 aut.eur.rice.s. La taille des liens définit le nombre de co-publications. Les noms affichés représentent les aut.eur.rice.s les plus cités dans ce document. La méthode de robustesse d'association a été utilisée. Les publications inférieures à 2005 prennent la couleur de 2005. D'après Van eck et Walkman, 2010. L'analyse décrit principalement les travaux d'Alain Hehn et Romain Larbat menés au Laboratoire Agronomie et Envirronement (LAE) (cluster central). Elle décrit également de nombreux petits clusters de co-autorats de personnes ayant travaillé sur l'évolution et les furanocoumarines (2005 et avant). Elle décrit enfin des clusters de tailles intermédiaires généralement par équipe qui travaille sur le métabolisme spécialisé et les P450s notamment. À noter que parmi les deux clusters les plus récents on retrouve les travaux portés sur l'analyse de l'évolution des Moracées (Gardner) et sur la voie des furanocoumarines chez les Apiaiacées portée par plusieurs équipes collaboratrices émergentes chinoises.



**Figure 44: Analyse bibliométrique en réseau de mots clés dans le cadre de ce manuscrit pour des articles publiés entre 2005 et 2024.** Publications citées : 394. Nombres de mots clés : 5680. Les données ont été référencées et analysées via les logiciels Zotero et VOSviewer. La taille des ronds définit le nombre de citations dans des articles différents et cités dans le cadre de ce document pour un seuil minimal de co-occurrence de 5 mots, soit 190 mots. La taille des liens définit le nombre de co-occurrences de mots. Les noms affichés représentent les mots clés les plus cités dans les titres et résumés d'articles pour les documents cités dans ce manuscrit. La méthode de robustesse d'association a été utilisée. Les publications inférieures à 2005 prennent la couleur de 2005. D'après Van Eck et Walkman, 2010. L'analyse décrit principalement des travaux portés sur l'évolution, le développement, les furocoumarines et les coumarines.

## Résumé

Les plantes régulent leurs ressources pour s'adapter à leurs environnements en constante évolution. L'un des moyens de résilience des plantes consiste à produire des molécules actives en réponse à ces fluctuations. Il existe chez les plantes entre 200 000 et 1 000 000 de ces molécules qui sont aussi appelées métabolites spécialisés (Chassagne *et al.*, 2019). Ces derniers sont apparus chez les plantes au cours de l'évolution et ont eu pour conséquence de modifier leur physiologie et leurs interactions avec l'environnement (Durán-Medina *et al.*, 2021; Weng *et al.*, 2021).

Parmi la multitude de molécules produites chez les plantes, on trouve les furocoumarines. Ces métabolites spécialisés dérivent des phénylpropanoïdes (aussi appelés polyphénols) et plus spécifiquement des coumarines. Elles ont été identifiées chez une vingtaine de familles de plantes et sont connus pour être fortement produites chez les Apiacées, Rutacées, Moracées et Fabacées (Bourgaud *et al.*, 2014; Sarker & Nahar, 2017). Des travaux antérieurs ont permis de caractériser une vingtaine d'enzymes impliquées dans la production de ces molécules. Les travaux réalisés dans le cadre de cette thèse ont mené à l'identification d'enzymes complémentaires impliquées dans la production de furocoumarines comme la 5-hydroxyxanthotoxine à partir de xanthotoxine, mais également de coumarines comme la daphnétine à partir d'umbelliférone. Les enzymes étudiées appartiennent à la famille des cytochromes P450s. Leurs structures a été passée au peigne fin, ce qui a permis de découvrir comment l'évolution a permis de passer d'enzymes qui produisent des coumarines à des enzymes qui produisent des furocoumarines.

La production de métabolites spécialisés dans les plantes mobilise un certain nombre de ressources qui ne peuvent pas être utilisées par ailleurs et notamment pour la croissance. Pour étudier ce détournement de ressources j'ai introduit une partie de la voie de biosynthèse des furocoumarines dans le génome de la tomate (qui n'est pas capable de produire ces molécules naturellement) et j'en ai analysé l'impact. Les résultats obtenus indiquent que le psoralène n'est pas accumulé comme cela était escompté, mais que la scopolétine, un produit probablement dérivé d'un des

intermédiaires de synthèse, était détectée en quantité non négligeable. Dans cette thèse est décrit l'impact de ces molécules sur la physiologie de la tomate.

L'ensemble des résultats obtenus dans ces travaux pose la question de l'émergence de nouvelles voies de biosynthèse chez les plantes. Cette question est notamment abordée ici par une analyse des processus évolutifs qui ont permis aux Moracées d'acquérir la capacité de produire des furocoumarines il y a plusieurs millions d'années durant le Crétacé.

# Summary

Plants regulate their resources to adapt to their constantly changing environments. One of their resilience mechanisms is the production of active molecules in response to these fluctuations. Between 200,000 and 1,000,000 such molecules exist in plants, also known as specialized metabolites (Chassagne *et al.*, 2019). These compounds emerged during plant evolution, altering their physiology and interactions with the environment (Durán-Medina *et al.*, 2021; Weng *et al.*, 2021).

Among the diverse molecules produced by plants are furanocoumarins. These specialized metabolites derive from phenylpropanoids (also called polyphenols), specifically coumarins. They have been identified in around twenty plant families and are strongly produced in Apiaceae, Rutaceae, Moraceae, and Fabaceae (Bourgaud *et al.*, 2014; Sarker & Nahar, 2017). Previous research has characterized about twenty enzymes involved in the production of these molecules. In this thesis, additional enzymes involved in furanocoumarin biosynthesis, such as 5-hydroxyxanthotoxin from xanthotoxin, and in coumarin biosynthesis, such as daphnetin from umbelliferone, have been identified. The studied enzymes belong to the cytochrome P450 family. Their structures were thoroughly analyzed, revealing how evolution enabled the transition from enzymes producing coumarins to those producing furanocoumarins.

The production of specialized metabolites in plants mobilizes resources that cannot be allocated elsewhere, particularly to growth. To study this resource allocation, I introduced part of the furanocoumarin biosynthetic pathway into the tomato genome (which does not naturally produce these molecules) and analyzed its impact. The results indicate that psoralen was not accumulated as expected, but scopoletin, likely derived from one of the synthetic intermediates, was detected in significant amounts. This thesis describes the impact of these molecules on tomato physiology.

Overall, the findings raise the question of how new biosynthetic pathways emerge in plants. This is addressed here through an analysis of the evolutionary processes that enabled Moraceae to acquire the ability to produce furanocoumarins millions of years ago during Cretaceous.