Thèse

en vue de l'obtention du titre de

Docteur de l'Université de Lorraine en Sciences Agronomiques

présentée par

Cloé Villard

HOW TO PRODUCE YOUR FURANOCOUMARINS:

THE HIDDEN PATHWAY

~

From the characterisation of new P450s to the evolution of the furanocoumarin pathway and the development of tools allowing the study of furanocoumarins' metabolic cost.

A LA CROISEE DES VOIES,

OU COMMENT PRODUIRE DES FUROCOUMARINES

~

De la caractérisation de P450s à l'évolution de la voie des furocoumarines et au développement d'outils permettant l'étude du coût métabolique des furocoumarines.

"Pay attention.
Be astonished.
Tell about it."

Mary Oliver – *Sometimes*

~

Wait, isn't it the definition of science?

À Simone

ACKNOWLEDGMENTS

~

Merci   *Thank you !*   *Bedankt*

شكراً   *Mulțumesc*   *Trugarez*   *Tack*   *Grazie*   ありがとう

Bien évidemment, mes principaux remerciements vont à **tous les membres du Laboratoire Agronomie et Environnement**, qui m'ont accueillie pendant ces trois années de thèse, m'ont aidée dans mon projet de recherche et avec qui j'ai passé de très bons moments. En premier lieu, je tiens à remercier **Christophe Robin**, directeur du laboratoire, pour son implication dans ma formation doctorale et ses conseils.

Mais que serait une doctorante sans encadrants pour la guider sur les chemins tortueux de la recherche ? Pas grand-chose. Je tiens donc à exprimer ma plus profonde gratitude à **Alain Hehn** et **Romain Larbat**, mes directeur et co-directeur de thèse. Alain, Romain, vous m'avez souvent dis avoir été chanceuse pendant ma thèse, mais la plus grande chance que j'ai eue, c'était sans doute de vous avoir comme encadrants. Pendant ces trois années, vous avez toujours su vous rendre disponibles pour m'aider et me conseiller quand j'en avais besoin, tout en me poussant à voler de mes propres ailes. Toujours à l'écoute, pleins d'idées novatrices, un humour à l'épreuve de mes gribouillis (hé hé hé), mais aussi très complémentaires… ça a été un réel plaisir de travailler avec vous et de vous côtoyer au quotidien ! **Alain**, tu m'as poussée à prendre confiance en moi et à aller toujours plus loin dans mes travaux de thèse, tout en m'aidant à garder du recul. Tu as toujours su rester pédagogue et bienveillant, même lorsque tu me rendais des documents rouges de corrections. Avec toi, la critique est toujours positive et bienvenue. Tu as aussi toujours encouragé le petit côté « décalage artistique » que j'aime tant appliquer à mes travaux, et ça, ça compte beaucoup pour moi. Entre gribouilleurs, on se comprend bien ! **Romain**, pendant ces trois ans, un coin de ton canapé était toujours prêt à accueillir une doctorante qui avait besoin de conseils avisés ou d'un petit remontage de moral. Ta bonne humeur à toute épreuve rend les journées de grisaille nancéienne plus agréables, même si tu passes la moitié de ton temps à te moquer (gentiment) de moi ! Et non content de t'investir dans ma thèse au point de secouer des microsomes avec moi, tu m'as aussi accompagnée dans mon aventure de vulgarisation scientifique, et n'était jamais contre une petite discussion sur les dragons ! **Alain**, **Romain**, merci pour tout. Vraiment. Je n'aurais pas pu espérer avoir de meilleurs encadrants que vous deux.

Pour moi, parmi les personnes les plus fantastiques du LAE, il y a évidemment **Andréïna**, **Marwa** et **Yuka**. Commençons par la première que j'ai rencontrée : **Andréïna**, initialement « l'inconnue de ma soutenance » tu as vite été promue au statut de Lab'Sister et a récemment acquis le grade suprême de courgette molle du bulbe (déso pas déso). Trugarez pour tous les conseils mais aussi et surtout pour les fous rires, le soutien inconditionnel, la bonne ambiance (Bretonne), les petits délires, les blagues censurées et les potins. Beaucoup de potins ! Passons à **Marwa**, ma jumelle (maléfique), toujours là dans les meilleurs comme dans les pires moments au labo, et toujours partante pour un burger. Comme je ne veux pas (trop) te faire pleurer, je me contenterai de te dire que tu n'es qu'une patate rose, mais que tout le monde a besoin d'une patate dans sa vie. Alors شكراً à toi d'être ma patate. And now, **Yuka**. We spend three years in the same office, and I really enjoyed it! You are always nice, always willing to help, to share your Japanese culture with us (while eating not very Japanese pizzas), and we had great conversations! 本当にありがとう Yuka! Thank you so much **my friends**, you are the best! Cœur cœur love love !
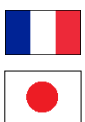
Parmi toutes les autres personnes du LAE qui m'ont aidée et avec qui j'ai passé de bons moments, je souhaite évidemment remercier **Gianni**, aka le docteur pannais, qui m'a refourgué une partie de son sujet de thèse, mais aussi les nombreux conseils qui allaient avec. 本当にありがとう **Ryosuke**, you may have been the most « childish » researcher of the

lab (he he he), but you always had an answer for every question, a solution for every problem, and enough humour to let me make fun of you! **Clément**, milles mercis pour les millions de questions pratico-pratiques auxquelles tu as répondu, et les innombrables services rendus. **Aude**, un immense merci à toi, mes bébés tomates sont aussi les tiens, et tu as même un super diplôme qui le prouve, y-a-t-il quelque chose à rajouter après ça ? Pour rester dans le même bureau, merci à **Jérémy** pour toutes les analyses et ton aide sur l'UHPLC, et merci à **Julie** pour ta personnalité aussi haute en couleur que tes cheveux. Merci à **Alex** pour ses conseils, mais aussi à **Marion** et **Alan**, mes nouveaux compatriotes de bureau. **Gael**, **Jodie**, merci à vous deux pour les conseils sur mon avenir professionnel et tous les moments de rigolade. Et comme la gratitude ne s'arrête pas aux frontières d'un labo, merci aux filles de PAT : **Ludivine** (et oui, tu es PAT maintenant !!), un immense merci pour toute ton aide, tes conseils précieux, et toutes les discussions qu'on a pu avoir – qu'elles tournent autour de Disney ou non ! Merci aussi à **Sissi**, **Cindy** et **Carole** pour tous les bons moments et les anecdotes croustillantes que vous avez toujours en réserve ! Merci à **vous tous**, ne changez surtout pas !

Dans la catégorie « de passage au LAE », je souhaite également remercier **Clémentine** et **Simona**. **Clem**, je ne saurais te remercier assez pour ton écoute, ta compréhension, ton grain de folie, ta critique artistique, les Escape Game, les histoires de défibrilateurs et de chats. Miiiiiiii ! Tu as beau être partie en direction de la Marwanie, les milliers de kilomètres ne nous séparerons pas. Mulțumesc ! **Simona**, it really was a pleasure to meet you ! You were like a little piece of Italian sun in our lab, and je vais toujours me ricorda de tuo fantastic mix of language. Grazie mille my friend. Enfin, je tiens à remercier **Léa** et **Anjara**, mes deux stagiaires.

My PhD also gave me the opportunity to travel, and I was very lucky to spend three months in the Wageningen university, among the very nice Biosystematics people. My first and most sincere thanks will go to **Eric Schranz**, chair of the Biosystematics Group: Eric, not only you accepted me in your lab, but you also took care of providing me with all the tools and knowledge I needed for my analyses, gave me a lot of wise advices, and were always willing to help me despite your crazy schedule. I also want to thank **Robin** and **Freek**, who largely contributed to open my eyes to the wonders of phylogenetics. **Robin**, thank you so much for all the help on my PhD project, and for the opportunity of working with you very soon on "the most cliché plant in the Netherlands". **Freek**, thanks you for letting me join your courses and for showing me the subtleties of tree building – even though I still do not understand everything! A very special thank you to **Femke**, my buddy in front of the window, who was so supportive and caring, who made me discover so many things and gave me my first "Dutch bike ride". A more general thank you to **the entire Biosystematics group**: you were all very nice and made me feel welcomed in your country, which is something I will never forget. I could not have asked for a better group to spend these three months, and I am very happy to come back in your lab soon! Bedankt voor alles! I also want to thank **Janani** from the Bioinformatics group, who helped me with the modelling and docking analyses. Finally, another special thanks to **Karen** (and her aphids) from the entomology lab. Karen, you have always been so kind with me, and this Swedish internship with you have been the trigger that first aroused my curiosity and interest for plant defences. It was and it will be a pleasure to see you again! Tack tack!

I also want to thank **Sakihito Kitajima**, from the University of Kyoto, who allowed us to access the *F. carica* RNAseq library I used during my PhD.

🇫🇷 Enfin, il me reste à remercier mes **amis** et ma **famille** qui, bien que n'ayant pas directement contribué à mon travail de thèse et parfois loin de moi, m'ont offert leur affection et leur soutien inconditionnel, m'ont embarquée dans des aventures incroyables et m'ont permis de toujours garder le moral au beau fixe – ne serait-ce que par leur présence, leurs sourires, les discussions enthousiastes, les rires et/ou les verres partagés.

Commençons par les **amis**. Les « vieux » amis qui restent toujours présents et fidèles malgré les années qui passent et l'éloignement, comme les amis plus récents qui, je n'en doute pas, m'accompagneront aussi pendant de longues années. **Clara-Lou**, même si on se voit IRL moins d'une fois par ans, tu réponds toujours présente quand j'en ai besoin. Merci pour ton petit grain de folie, ô toi ma seconde moitié de cerveau simiesque (tu t'en souviens de ça ?) ; BOOYA ! **Kirsten**, ma traductrice officielle, merci pour les millions de livres que tu m'as fait découvrir, les balades pied nues dans la montagne, et le soutien par-delà les frontières. **Yann**, toi le machin bizarre qui a animé tant de cours monotones et m'a suivie dans mes aventures peinturluresques, et toi **Lisa**, ma fillotte qui va un jour conquérir la lune grâce à ses hamsters de traineau fluorescents… Merci de votre amitié et de votre soutien indéfectible qui ne rouille pas malgré l'éloignement ; POUIC ! Dans la catégorie des amitiés à distance, je tiens aussi à remercier **Adeline** et **Camille** : les filles, on doit toujours se planifier ce Skype à trois, n'est-ce pas ? On a de quoi discuter pendant des heures ! Du côté des amis plus récent, je ne peux évidemment pas passer à côté d'**Eléa** et de **Nanou**. **Eléa**, ma « copine plantiste et gribouilleuse », merci de m'avoir fait confiance avant même de m'avoir rencontré, et de m'avoir poussée sur les chemins de la vulgarisation scientifique ! On se revoit quand tu veux autour d'un verre ! A travers toi, je tiens aussi à remercier toute la team incroyable et adorable de **Podcast Science**, qui m'a invitée pas moins de trois fois pour partager ma passion pour les plantes (ces petites choses innocentes) – avec des remerciements particuliers pour **Pierre** (mon modèle de vulga, rien que ça) et **Claire**. *Je sers la science et c'est ma joie !* **Nanou**, merci pour ta gentillesse, ton écoute (tu es mon meilleur public !), tes encouragements, et ton soutien sans faille ! C'est quand qu'on se refait un Escape Game ? Et tant que je suis dans la vulga, un grand merci à **Lisa** (la deuxième du paragraphe) et toute la team de **Pint of Science Nancy**, avec qui j'ai partagé des aventures 50% bière et 50% science pour un total 100% fun. Parmi ceux que j'ai croisé pendant ma thèse et qui ont fait un bout de chemin avec moi, je tiens à remercier **Sarah-Louise**, **Rodolphe**, les **doctorants INRAE** avec qui j'ai pu boire des coups, et les **ami-le-temps-d'un-séminaire-mais-en-fait-on-garde-quand-même-contact** rencontrés dans le cadre de l'EIR-A. Et bien sûr, **Justine**, merci à toi de m'avoir fait découvrir Amsterdam et d'avoir été mon repère français aux Pays-Bas !

Enfin, merci à toute ma **famille**, et en particulier à mes **parents**, mon **frangin** et mon **Pioupiou**. Sans vous, sans votre soutien indéfectible, je ne serais pas là aujourd'hui. Vous m'avez toujours encouragée quels que soient mes choix, m'avez supportée dans tous les sens du terme, m'avez houspillée quand j'en avais besoin, aidée et remotivée quand il le fallait… Les mots me manquent, alors je me contenterais d'un merci et de tout ce qu'il implique. Merci du fond du cœur !

~

Donc encore une fois, MERCI. Merci à tous ceux qui ont croisé mon chemin
et ont contribué à faire de moi celle que je suis devenue. Vous êtes les meilleurs !

Once again, THANK YOU. Thanks to all those who have crossed my path,
and who contributed to make me who I have become. You are the best!

TABLE OF CONTENTS

~

~

TABLE OF FIGURES

~

TABLE OF TABLES

~

TABLE OF SUPPLEMENTAL TABLES

~

TABLE OF SUPPLEMENTAL FIGURES

~

ABBREVIATIONS

~

ATP: Adenosine triphosphate

cDNA: Complementary deoxyribonucleic acid

CDS: Coding sequence

CO: Carbon monoxide

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

C4H: Cinnamate 4-hydroxylase

DAMP: Damage-associated molecular pattern

DIBOA: 2.4-dihydroxy-2*H*-1,4-benzoxazin-3(4*H*)-one

DMS: Demethylsuberosin (6-dimethylallylumbelliferone)

DNA: Deoxyribonucleic acid

FAD: Flavin adenine dinucleotide

FcPT: *Ficus carica* prenyltransferase

FMN: Flavin mononucleotide

GB: GoldenBraid

GDBH: Growth-differentiation balance hypothesis

LAE: Laboratory of agronomy and environment

LC-NMR: Liquid chromatography and nuclear magnetic resonance

NADH: Nicotinamide adenine dinucleotide (reduced form)

NADP: Nicotinamide adenine dinucleotide phosphate (oxidised form)

NADPH: Nicotinamide adenine dinucleotide phosphate (reduced form)

NFC: Nitrogen fixing clade

NMR: Nuclear magnetic resonance

ORF: Open reading frame

O_2: Dioxygen

PcPT: *Petroselinum crispum* prenyltransferase

PCR: Polymerase chain reaction

PsDiox: *Pastinaca sativa* dioxygenase

PsPT: *Pastinaca sativa* prenyltransferase

PT: prenyltransferase

P450: Cytochrome P450

RNA: Ribonucleic acid

RT-PCR: Reverse transcription polymerase chain reaction

SRS: Substrate recognition site

TU: transcriptional unit

UDT: Umbelliferone dimethylallyltransferase

UHPLC-MS: Ultra-high performance liquid chromatography – tandem mass spectrometer

UV: Ultraviolet

3D: Three-dimensional

Vernacular names

~

Colour code: Plant/Algae – Animal – Microorganism

Agrobacterium tumefaciens: Agent of the crown gall disease
Ammi majus: Bishop's weed (greater ammi)
Apium graveolens: Celery
Arabidopsis thaliana: Thale cress
Arnebia euchroma: Zi cao
Boehmeria nivea: Ramie
Brosimum Gaudichaudii: Mama-cadela
Cajanus cajan: Pigeon pea
Cannabis sativa: Cannabis
Catharanthus roseus: Pink periwinkle
Chlamydomonas reinhardtii (no known vernacular name)
Cicer arietinum: Chickpea
Citrus × paradisi : Grapefruit
Croton stellatopilosus (no known vernacular name)
Cucumis sativus: Cucumber
Dipteryx odorata: Cumaru (kumaru)
Escherichia coli (no known vernacular name)
Fatoua pilosa: Malbas-damo
Ficus capensis: Cape fig
Ficus carica: Common fig
Ficus coronata: Sandpaper fig
Ficus cyathistipula: African fig
Ficus erecta: Japanese fig
Ficus nervosa: Nerved fig (eechamaram)
Ficus palmata: Punjab fig (bedu)
Ficus salicifolia: Wonderboom
Ficus sycomorus: Sycamore fig (fig-mulberry)
Ficus religiosa: Sacred fig
Ficus ruficaulis (no known vernacular name)
Fragaria vesca: Wild strawberry
Glycine max: Soybean
Helianthus tuberosus: Jerusalem artichoke
Heracleum sosnowskyi: Hogweed
Humulus lupulus: Common hop
Ipomea batatas: Sweat potato
Juglans regia: Persian walnut
Malus domestica: Apple tree
Maquira calophylla (no known vernacular name)
Medicago truncatula: barrel medick (barrelclover)
Morus notabilis: Mulberry tree

Nicotiana tabacum: Tobacco
Oryza sativa: Rice (Asian rice)
Papilio brevicauda: Short-tailed swallowtail
Papilio polyxenes: Black swallowtail
Parasponia andersonii (no known vernacular name)
Paris polyphylla: Himalayan paris
Parthenium argentatum: Guayule
Pastinaca sativa: Parsnip
Petroselinum crispum: Parsley
Petunia hybrida: Common garden petunia
Physcomitrella patens: Spreading earthmoss
Phytophtora sojae (no known vernacular name)
Populus trichocarpa: Black cottonwood (western balsam-poplar)
Prunus avium : Wild cherry
Pseudomyrmex ferruginea: Acacia ant
Psoralea cinerea: Annual verbine
Pyrus x bretschneideri: Pearple
Ricinus communis: Castor oil plant (castor bean)
Rosa chinensis: Chinese rose
Ruta graveolens: Rue
Saccharomyces cerevisiae: Baker's yeast
Saccharum spontaneum: Wild sugarcane
Salvia miltiorrhiza: Red sage
Santalum album: Indian sandalwood
Schizosaccharomyces pombe: Fission yeast
Selaginella moellendorffii (no known vernacular name)
Solanum lycopersicum: Tomato
Sorghum bicolor: Sorghum
Swertia mussotii (no known vernacular name)
Trema orientale: Charcoal-tree (Oriental trema)
Trigonella foenum-graecum: Fenugreek
Vachellia cornigera: Swollen-thorn acacia (bullhorn acacia)
Vitis vinifera: Common grape vine
Ziziphus jujuba: Jujuba
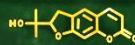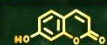
How to produce your furanocoumarins: the hidden pathway

~

From the characterisation of new P450s to the evolution of the furanocoumarin pathway and the development of tools allowing the study of furanocoumarins' metabolic cost
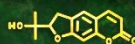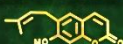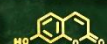
LAY SUMMARY / RÉSUMÉ VULGARISÉ

~

During the course of evolution, plants have developed many defences that allow them to resist the attacks of herbivores. For instance, the fig tree produces some toxic compounds called furanocoumarins. The goal of this project was to better understand the production of furanocoumarins in plants. In a first time, we inquired the molecular tools (enzymes) set up by the fig tree to produce these toxic compounds. This led us to identify three enzymes which activities were never described before. Among them, one enzyme (the marmesin synthase) strongly suggests that furanocoumarins appeared several times throughout evolution, in distant plant families. This enzyme also opens many new perspectives for the study and use of furanocoumarins. Therefore, in a second time, we used this enzyme to develop new tools that will allow to assess the amount of energy a plant has to invest to be able to produce defensive compounds such as furanocoumarins.

Au cours de l'évolution, les plantes ont développé de nombreuses défenses qui leur permettent de survivre aux attaques d'herbivores. Par exemple, le figuier produit des composés toxiques appelés furocoumarines. Ce projet avait pour but de mieux comprendre la production de furocoumarines chez les plantes. Dans un premier temps, nous nous sommes intéressés aux outils moléculaires (enzymes) mis en place par le figuier pour produire ces composés. Nous avons identifié trois enzymes aux activités encore jamais décrites, dont une (la marmésine synthase) qui suggère fortement que les furocoumarines sont apparues plusieurs fois au cours de l'évolution, chez des plantes éloignées. Cette enzyme ouvre également de nombreuses perspectives quant à l'étude et l'utilisation des furocoumarines. Ainsi, dans un second temps, nous avons utilisé cette enzyme pour développer des outils qui nous permettront d'évaluer la dépense énergétique que représente la production de furocoumarines pour une plante.

FOREWORD

~

Plant defence is a fascinating world.
Yet, like many fascinating things, it is also very complex.

~

To fully appreciate this study, many concepts related to plant defence are required.
But because of the vastness of this topic, some choices had to be made. Therefore, in the first
bibliographic chapter, I offer to dive in the world of plant defence with you, through the following
approach:

I will start with a general introduction to plant defence, from the establishment of
various defensive mechanisms to the cost of these defences. Then, I will focus on one specific plant,
Ficus carica, and one specific family of defensive compounds: the furanocoumarins.

In a second part, I will focus on the tools developed by the plants to produce furanocoumarins. In
particular, I will detail the role of one essential enzyme superfamily: the cytochromes P450.

~

Finally, while science is a serious thing, it does not have to be sad.
On the contrary, science is captivating and inspiring.

For this reason, I might have hidden a few popular references throughout this whole study,
with the secret hope they will make you smile. But of course, all references – even the hidden ones –
are listed at the end of this document.

~

Enjoy your reading!

Chapter I
State of Art

I.
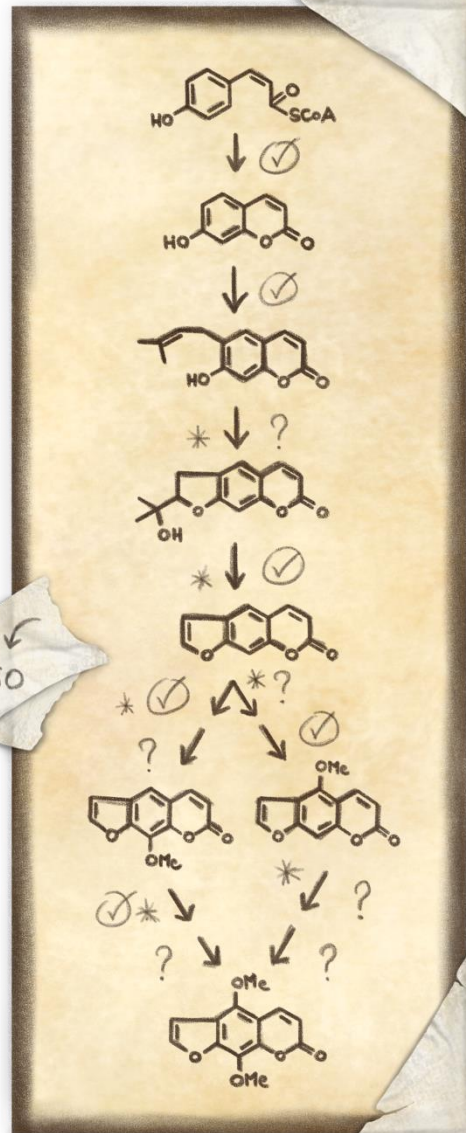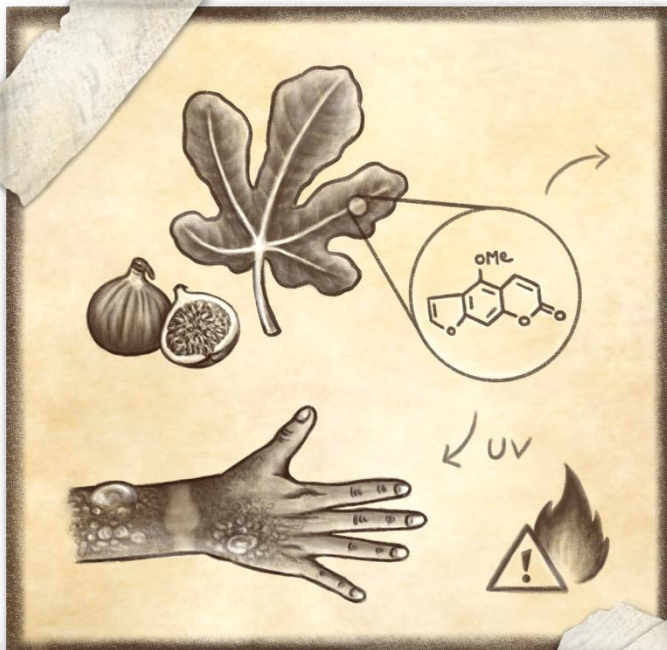
JOURNEY TO THE CENTRE OF PLANT DEFENCE

~

All you have always wanted to know about plant defence in general, and furanocoumarins in particular

A. THROUGH THE LOOKING-GLASS: OPEN YOUR EYES, DON'T BE PLANT BLIND!

When asked about plants, many people immediately picture delicate and harmless little green things that seem to be only good at being eaten, being decorative, or being stepped on. As they cannot move and do not make any sound, plants look boring and vulnerable: they are at the mercy of anyone who would want to eat or pick them, and they cannot do anything against it.

…

Or, can they?

Contrary to this popular misconception, plants are amazing organisms and heavily defended fortresses that can give a hard time to unprepared hungry enemies.

But let us start from the beginning: land plants appeared approximatively 450 million years ago (Steemans et al. 2009). They are brilliant chemists able to convert light energy and mineral resources into organic matter. Yet, this carbon autotrophy makes them the first link in the food chain, which attracts many enemies. In addition, plants are sessile organisms: their roots, which are very useful to draw resources from the soil, have the unfortunate consequence of preventing any escape attempt. As a result, plants cannot flee the many bioaggressors that try to eat them, nor can they move around when environmental conditions are not optimal. So, for hundreds of millions of years, plants have been relentlessly opposed to mortal threats they had to face without being able to move. But for hundreds of millions of years, they have survived. And not only they survived, but they also diversified so much that they can now be found in great abundance in most environments.

In the endless war that oppose them to the rest of the world, plants owe their survival to the many defensive strategies they evolved. For instance, they are equipped with physical structures that can repel their bioaggressors and prevent their infestation, but also with dangerous chemical arsenals that might incapacitate or kill their enemies. Moreover, in a war, every protagonist is perpetually evolving new strategies to dominate its opponents: thus, over thousand and million years, defensive innovations from plants have been repeatedly countered by innovations from the enemy – stagnating and falling behind would have mean death. This escalating arms race has impressively driven the diversification of plant defences, which is now a very complex yet finely regulated machinery composed of surveillance and alarm systems as well as strong armours and deadly weapons.

In this study, I will mainly focus on the fig tree and one specific family of plant defensive chemicals called furanocoumarins. Yet, to truly understand and appreciate the defensive role of furanocoumarins, but also to remember they are only one piece of the complex puzzle that allow plants to defend themselves, it is essential to start with a general overview of plant defences. And if this introduction catches your attention, you may enjoy reading the detailed yet wonderfully written book by Dale Walters "*Fortress Plant: How to survive when everything wants to eat you*" (2017) – which is one of the main sources of inspiration of the following paragraphs.

B. THE DEFENCE IN OUR PLANTS: FROM MARTIAL STRATEGIES TO THE COST OF WAR

In a war, even the most elaborated weapon might be useless if not used properly, at the right time, or against the right enemy. Therefore, to establish efficient defences, it is essential to know the enemy, to quickly detect its attacks, and to ring the alarm bell to warn the fortress that the fight has begun.

B.1. Defence starts before the establishment of defensive mechanisms

B.1.a. Various enemies imply various defences

Plants are attacked by various bioaggressors that range from big mammals to microscopic pathogens. On the one hand, there are the animals that eat entire plant parts (mammals or gastropods), macerate leaf tissues (arthropods such as insects and myriapods), ingest the content of epidermal cells (such as spider mites), consume phloem sap (such as aphids or leafhoppers), or parasitize plants (such as nematodes). On the other hand, there are numerous microscopic pathogens that cause various diseases. Pathogens include biotrophic and necrotrophic fungi (such as *Puccinia triticina* or *Botrytis cinerea* – rust and gray mold), pathogenous bacteria (such as *Pectobacterium carotovorum* – soft rot), viruses (such as the *Tobacco mosaic virus*) and viroids (such as the *Coconut cadang-cadang viroid*). But plants can also be the bioaggressors of other plants through parasitism or competition for the resources (Walters 2017).

In response to their various enemies, plants have evolved a multitude of defences that can be classified into two categories: constitutive and inducible defences. Constitutive defences are always present in the plants and provide a basal protection. Other defences are inducible, which means they are initiated in reaction to a stress: as they are only activated when needed, in response to a precise attack, they might be more specific and economic than constitutive ones (Tollrian and Harvell 1999; Walters 2017). However, inducibility implies that plants should be able to perceive the attacks and identify their enemies, which means they need "surveillance" and "recognition" systems.

B.1.b. How to realise you are under attack?

Plants can perceive attacks through the detection of physical and chemical stimuli. The objective here is not to list all these detection mechanisms but to give a quick overview of the diversity of signals plants can perceive and use to trigger their defences.

Physical detection is mainly based on the perception of mechanical stimuli such as herbivore "touch": for instance, the insects moving on plants can disrupt the trichomes they step on, leading to the induction of defence mechanisms (Peiffer et al. 2009). Soft undamaging mechanical stimulations, such as a gentle rubbing of leaf surfaces, might also contribute to activate defence mechanisms (Benikhlef et al. 2013). In addition, plants have the ability to perceive acoustic vibrations: indeed, the sound of a caterpillar eating a leaf can be perceived – "heard" – by plants and elicit defensive responses that other sounds such as wind or insect songs do not trigger (Appel and Cocroft 2014). However, if physical stimuli can make plants aware of a danger, they do not themselves provide information about the identity of the bioaggressor. But some chemical stimuli can do so.

For instance, thanks to ligand-receptor-like interactions, plants can distinguish molecular patterns associated to self-damages or specific bioaggressors. Damage-Associated Molecular Patterns (DAMPs) are molecules from damaged plants or molecules that are present in unusual locations and/or concentrations, reflecting that something is going wrong. As an example, fragments of broken cuticle, unusual release of methanol, or extracellular ATP can be recognised as DAMPs in different plants (Tanaka et al. 2014; Hann et al. 2014; Heil and Land 2014). Beside these non-specific signals, plants can also identify molecular patterns belonging to specific bioaggressors. These second kind of molecules are called HAMPs when they are associated to herbivores, PAMPs for pathogens, MAMPs for microbes and NAMPs for nematodes. For instance, some molecules described in the oral secretions of insects, the chitin present in the fungi cell wall, the flagellin from bacteria or the ascarosides from nematodes can be considered, respectively, as HAMPs, PAMPs, MAMPs, and NAMPs (Choi and Klessig 2016).

B.1.c. Signalling pathways to trigger defences

Once plants have detected an attack, they have to transform this information into an appropriate response, which involves complex signalling pathways and various messengers such as salicylic or jasmonic acid. In the cells surrounding the attack, the response of the plant generally consists in the induction of defence mechanisms. More distal parts can either be induced (systemic resistance) or primed. Priming is a physiological process that puts plants into alert and prepares them to respond more quickly or aggressively to future attacks. In other words, as it can be useless and wasteful to immediately activate defences far from the attack, priming only makes distal parts of the plants ready to face the next move of their bioaggressors (Frost et al. 2008; Walters 2017).

Depending on the nature of the bioaggressors, different signalling pathways are activated and induce specific responses. For instance, the response to herbivores is mostly mediated by the jasmonic acid pathway while the response to biotrophic pathogens involves the salicylic acid pathway. In general, these pathways lead to the accumulation of high level of phytohormones around the injury, which induces defence reactions, and low levels of phytohormones at distal locations that mainly triggers a priming (White 1979; Heil 2009; Barrett and Heil 2012; Janda and Ruelland 2015; Singh et al. 2016).

Yet, this description is far from being representative of the real complexity of plant signalling pathways and their hormonal crosstalk. Indeed, as plants might have to face simultaneous attacks at different locations, they have to prioritise their needs for defence and to trigger different type of responses. Therefore, plants' signalling pathways are finely regulated according to multiple factors such as the nature of the bioaggressor(s), the seriousness of the threat(s), the (relative) importance of the attacked

zone(s), the quantity of resources that can be invested in defence at a given instant, and even circadian clocks. Plant signalling pathways and their regulation have been extensively studied and reviewed, but I will not further detail them here (Frost et al. 2008; Spoel and Dong 2008; Pieterse et al. 2012; Hou et al. 2013; Hevia et al. 2015; Ingle et al. 2015; Walters 2017).

> In summary, surveillance and detection constitute the essential starting point of any induced defence: these complex and finely regulated systems aim to detect the attack(s) and gather enough information about the bioaggressor(s) to prepare an adapted defence strategy(ies).

B.2. Various defence mechanisms to respond to various threats

Whether they are constitutive or inducible, plant defences can be separated into direct and indirect mechanisms, which can be ubiquitously found in higher plants or limited to restricted taxa. Direct defences are additionally subdivided into physical and chemical mechanisms. Once again, the following paragraphs aim to show the diversity of plant defence, but not to provide an exhaustive list (**Figure 1**).
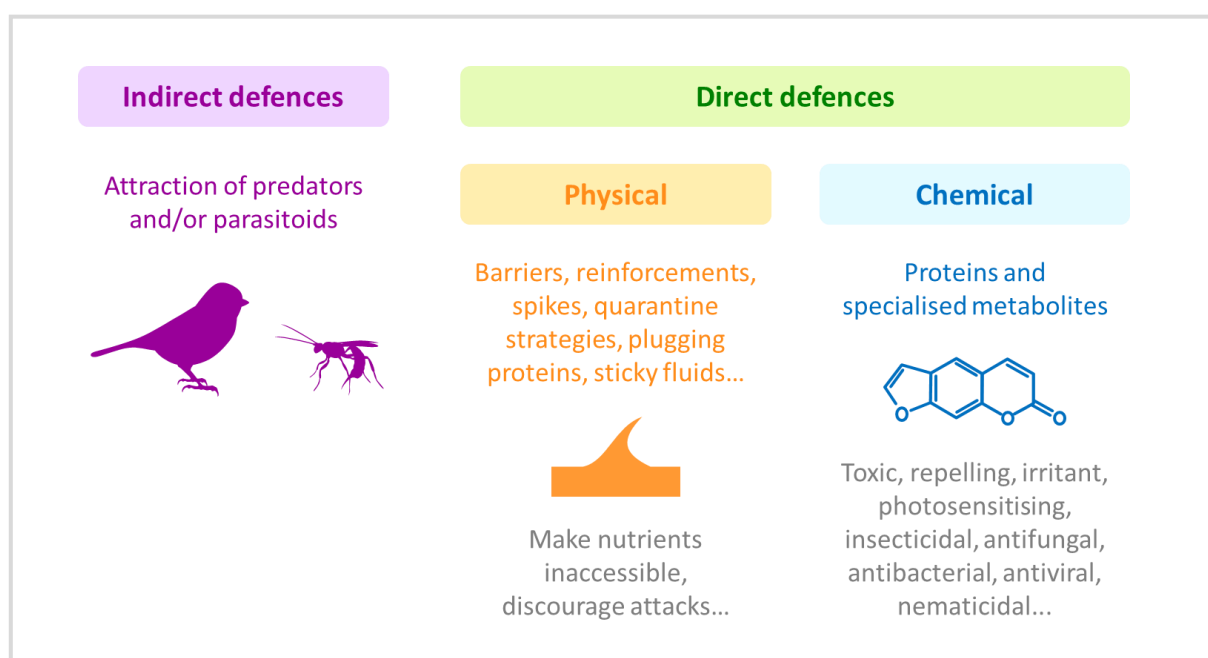
Figure 1 The different types of plant defence mechanisms.

B.2.a. Indirect defences: asking for help?

"The enemy of my enemy is my friend", that is the spirit of indirect defences.

To get rid of herbivorous arthropods, plants might enlist bodyguards that are natural carnivorous or parasitoid enemies of their own enemies (**Figure 1**). Indeed, when they perceive an attack or detect eggs on their leaves, plants can release cocktails of volatile compounds that might be perceived by

predators and/or parasitoids, such as predatory mites, parasitoid wasps, or even birds. These cocktails of chemicals might represent the promise of a good meal or a nice egg-laying site for those which perceive them: specific carnivores and/or parasitoids might therefore be quickly attracted (within hours) by the infested plants, where they start slaughtering their prey (Dicke et al. 1990; Hilker et al. 2002; Amo et al. 2013; Walters 2017). Indirect defences are usually inducible, but it should be noted that plants can also establish more permanent relationships through symbiosis. For instance, the swollen-thorn acacia (*Vachellia cornigera*) hosts and feeds some aggressive ants (*Pseudomyrmex ferruginea*) that protect the acacia against herbivores and neighbouring plants (Janzen 1966).

B.2.b. Direct defences: physical mechanisms to build a reinforced armour

The first defence line that plant enemies have to face are generally physical barriers (**Figure 1**) that can make nutrients inaccessible or indigestible, prevent or discourage attacks (Moore and Johnson 2017).

Indeed, the first contact point between plants and their enemies is usually the cuticle or the bark that cover plant surfaces, forming a physical barrier. Breaking the cuticle is not challenging for mammals, but making their way through it is not easy for microorganisms. In addition, as the cuticle is waxy, it can prevent insects to get a good grip on the leaves they land on (Whitney et al. 2009; Dale 2017). Bioaggressors that get through the cuticle then have to face a second barrier, the plant cell wall, which can be strengthened by structural reinforcements: for instance, lignin impregnation makes it more rigid and impermeable to pathogens, increased fibre contents and mineralisation contributes to increase plant toughness and to deter herbivores, while the formation of papillae helps to repair damages and to slow pathogen infection (Coley et al. 1985; Stone and Clarke 1992; Maher et al. 1994; Ruiz et al. 2002; Korth et al. 2006; Bhuiyan et al. 2009; Voigt 2014; Walters 2017).

A multitude of other physical defences exists in various plants. For example, spiky structures such as thorns, spines and trichomes are particularly adapted to repel animals ranging from big mammals to small insects. Trichomes, in particular, are divided into two categories: non-glandular trichomes are little hairs that increase the roughness of the plant and act as physical obstacles to insect movements. On the contrary, glandular trichomes are small structures that accumulate and secrete various substances such as defensive chemicals (Levin 1973; Wagner 1991). Other very different physical mechanisms include "quarantine" strategies such as the production of cork that might stop a pathogen invasion (Walters 2011, 2017), "plugging" proteins such as the forisomes (Fabaceae) that can physically stop the feeding of an aphid by clogging its stylets (Medina-Ortega and Walker 2015), or even the production of defensive fluids such as resin or latex. These fluids are stored in laticifers and resin ducts and can be exuded upon tissue damages. They then become sticky and can immobilise insects. In addition to this physical protection, latex and resin also provide a chemical defence for they contains many toxic chemicals that are directly delivered at the site of the injury (Farrell et al. 1990; Agrawal and Konno 2009).

B.2.c. Direct defences: chemical mechanisms to concoct poison pills

Behind their physical defences that mainly aim to prevent or repel the attacks, plants also hide a chemical arsenal made of various proteins and specialised metabolites that can discourage, incapacitate or kill their enemies (**Figure 1**).

B.2.c.1. Defensive proteins

Plants can produce a large diversity of defensive proteins exhibiting various activities such as irritant, insecticidal, nematicidal, antifungal, antibacterial, antiviral, or toxic properties. This diversity thus protects plants against a large range of bioaggressors (Schlumbaum et al. 1986; Broadway and Duffey 1986; Huynh et al. 1992; Terras et al. 1993; Peumans and Van Damme 1995; Edreva 2005). As an example, chitinases can degrade fungal cell walls (Schlumbaum et al. 1986; Melchers et al. 1994) while proteases and proteases inhibitors interfere with the digestion of the insects that ingest them (Felton and Gatehouse 1996; Konno et al. 2004). A more drastic example is the highly cytotoxic ricin produced by *Ricinus communis,* that can cause cell death by inhibiting protein synthesis (Lord et al. 1994).

B.2.c.2. Specialised metabolites

The major defensive chemicals produced by plants are compounds called specialised or secondary metabolites, in contrast to primary metabolites. Specialised metabolites contribute to survival and fitness by playing a major role in the adaptation to the environment and the interaction with the ecosystems. Therefore, even if they are present in low concentration in plants, they fulfil diverse functions such as the attraction of pollinators, the protection against abiotic stresses and the defence against bioaggressors (Fraenkel 1959; Wink 2008; Moore et al. 2014; Walters 2017). Considering that the present study focuses on a family of specialised metabolites – the furanocoumarins – it is worth describing them with more details.

Specialised metabolites are characterised by a huge molecular diversity: the plant kingdom might indeed contain a few hundred thousand different compounds, distributed in large or limited taxa. Therefore, every species produces its own combination of molecules (Wink 2008; Moore et al. 2014). Specialised metabolites fall under three main categories: terpenoids, alkaloids and phenolics. Terpenoids are derived from units of isopentenyl pyrophosphate, alkaloids contain nitrogen atom(s), and phenolics contain at least one phenol group. Terpenoids, alkaloids and phenolics each contains numerous families of compounds, with diverse structures and complexity, but also various properties including defensive ones (Croteau et al. 2000). And because of these multiple bioactivities, specialised metabolites can be used by humans in fields such as agronomy, cosmetic or medicine – even though many compounds and functions are still to be discovered.

Here are a few characteristics of defensive specialised metabolites:

Firstly, each compound may affect a small or a large range of enemies, and act with its own mode of action. For instance, some molecules have repelling effects, while others modify redox conditions, have photosensitising properties, or exhibit a more general toxicity (Fowlks et al. 1958; Pathak et al. 1962; Hebeish et al. 2008; Marrelli et al. 2012; Debib et al. 2014). A drastic example of deadly alkaloid is the curare produced by a few plants such as *Chondrodendron tomentosum*: curare is a neurotoxin that affects animals and leads to paralysis, respiratory paralysis and potential death (Carl et al. 2014).

Secondly, small structural variations might result in different properties, which may (partially) explain the great diversity of specialised metabolites. It looks indeed "easier" to modify or decorate a pre-existing molecule, leading to new toxic properties, rather than to build a completely new one (Appel

1993; Salminen et al. 2011; Moore et al. 2014). Structure-activity patterns will be exemplified later with furanocoumarins, for which linear and angular isomers display different defensive properties.

Thirdly, within one plant, the defensive specialised metabolites are unequally distributed and every compound might follow its own patterns: various metabolites might therefore be found within all plant tissues, be excreted, or be localised in specific organs or compartments such as laticifers, trichomes or the vacuole (Zobel and Brown 1988, 1989; Zobel et al. 1990; Innocenti et al. 1997b; Reinold and Hahlbrock 1997; Mamoucha et al. 2016; Walters 2017; Weryszko-Chmielewska and Chwil 2017).

Lastly, in a given plant, the composition and concentration of the specialised metabolite mixture vary a lot among tissues, but it also changes over time. These temporal variations apply to all type of defences and allow plants to modulate their protection according to many environmental and internal factors such as abiotic stresses (drought, cold), the pressure of the enemies (inducibility), seasonality, circadian circles, the age of the plant, the developmental stage of an organ, its importance, and the amount of resources that can be invested into defence (Zaynoun et al. 1984; Kim et al. 2003; Fujita et al. 2006; Marrelli et al. 2012; Moore et al. 2014; Hevia et al. 2015; Walters 2017).

> Simply put, plants are equipped with a large array of defence mechanisms: they can get rid of herbivores by recruiting predators, establish physical barriers and traps to prevent or stop attacks, and produce a huge diversity of chemicals with repelling or deadly effects.

B.3. The cost of defence: growing or fighting?

The production of any compound or structure requires resources that are not unlimited in the plant. Therefore, using resources to serve a given function such as defence is an investment that comes at the expense of other potential utilisations of these resources. In other words, defences have a cost.

B.3.a. Defence is a resource-consuming process

Plants draw their resources from their environment (photoassimilated carbon, minerals and water) and can invest them in three major processes that are critical for their overall reproductive success, also called plant fitness (**Figure 2**). The first process is growth, which consists in the production of biomass through the development of new organs such as leaves and roots. To grow new tissues, a plant needs to convert primary metabolites into cellular building blocks, which allows cell division and expansion. Growing is therefore resource-consuming, but new tissues allow the plants to draw more resources from their environment. The second process is reproduction: to ensure the transmission of their genetic heritage, plants require resources to produce offspring. The third process is defence, which is also costly: for instance, producing toxic compounds requires energy and molecular precursors. Moreover, additional resources are required to establish the cellular machinery and structures involved in the synthesis, transport and storage of these compounds, but also to establish the wound-detection systems and the signalling pathways that trigger their production. Yet, despite their cost, defences are essential to prevent the destruction of plant tissues and to protect plants and their offspring from bioaggressors (Herms and Mattson 1992; Züst and Agrawal 2017; Guo et al. 2018).
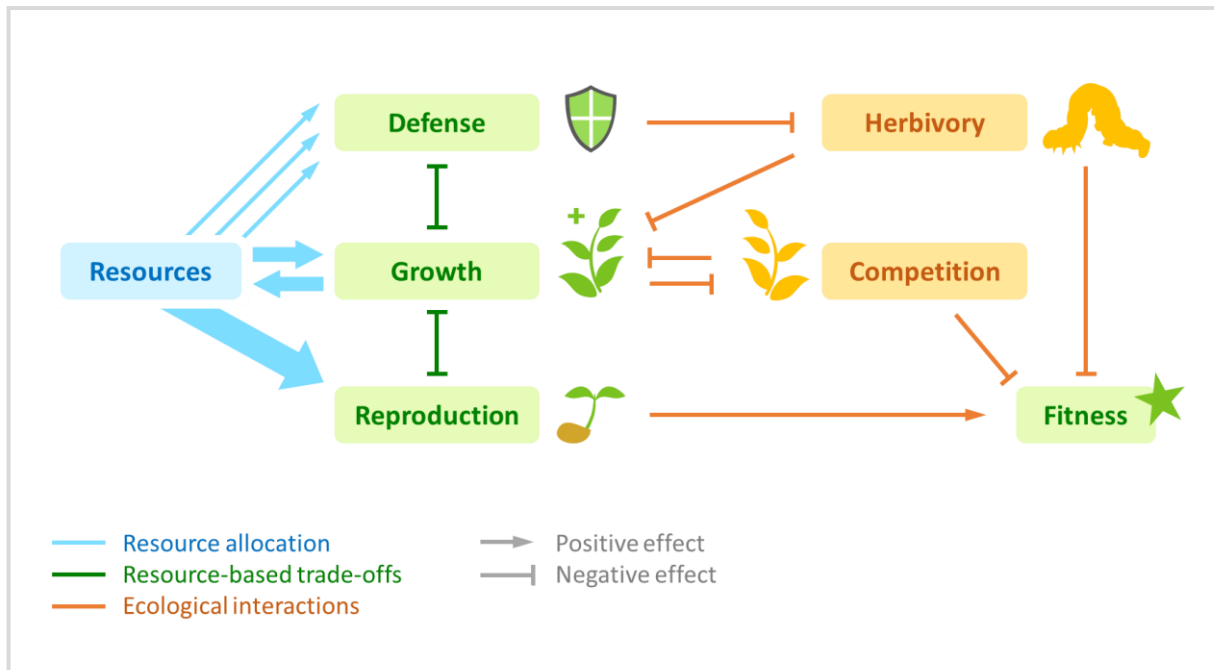
Figure 2 **Conceptual diagram of plant fitness determination via resource allocation, resource-based trade-offs and ecological interactions.** Adapted from Züst and Agrawal (2017).

So, to be competitive, plants must grow and produce offspring while maintaining enough defences to survive (Herms and Mattson 1992). However, as these three processes use the same pool of resources, they are in competition, which means that any resource allocation to defence might represent a redirection from primary metabolism that would negatively affect growth and reproduction (Züst and Agrawal 2017). Of course, the investment in defence might be partially offset by a reduced herbivory, but plants have also evolved strategies to reduce the cost of defence. For instance, defences can be restricted to particular tissues and limited in time (inducibility) (Heil and Baldwin 2002; Karasov et al. 2017; Züst and Agrawal 2017; Guo et al. 2018). Some metabolic intermediates can also be used in both primary and specialised pathways, which limits the required cellular machinery, but creates a competition between the primary and specialised metabolic pathways that use these intermediates (Herms and Mattson 1992). Therefore, plants need to adjust the allocation of their resources to meet their physiological and reproductive needs while adapting to fluctuating environmental constraints (resource availability and stresses). In other words, the optimisation of plant fitness requires resource-utilisation trade-offs between growth and the need for defence (Karasov et al. 2017) (**Figure 2**).

B.3.b. The multiple theories of trade-off between growth and defence

The trade-offs between growth and defence are widely studied and are of agronomical interest: indeed, by breeding our crop to improve their yield, we have progressively reduced their defences, leading to an increased use of phytosanitary products. More generally, these trade-offs have an ecological cost that might strongly impact the natural evolution of plant defences (Vries et al. 2019).

The trade-offs between growth and defence are complex and there are many ways to study them by using various traits and measures. As an example, Züst and Agrawal (2017) define costs – such as allocation but also genetic and ecological costs – as the consequences of trade-offs. The allocation

costs consist in a direct limitation by resource competition: they correspond to the reductions of growth that directly result from the diversion and reallocation of limited resources to defence. Genetic costs refer to the genetic variation and links such as pleiotropy and linkage disequilibrium that underly growth and defence traits. Ecological costs consist in penalising side effects on ecological interactions, such as defence mechanisms that would negatively impact pollinators and/or mutualists. Other costs such as metabolic, phenotypic or opportunity cost can also be defined (Züst and Agrawal 2017).

As a result of the complexity to define and study the trade-offs between growth and defences, several theories have been proposed to explain these trade-offs (Stamp 2003). Among them, we can mention the optimal defence hypothesis (McKey 1974), the carbon-nutrient balance hypothesis (Bryant et al. 1983; Tuomi 1992), the growth rate hypothesis (Coley et al. 1985) or the growth-differentiation balance hypothesis (GDBH), which is often considered as the most complete hypothesis (Loomis 1932, 1953; Herms and Mattson 1992; Stamp 2003, 2004).

The GDBH predicts how plants balance the allocation of resources between growth and differentiation-related processes (such as defence) depending on environmental conditions. "Resources" refer to nutrients and water, but not to the light that conditions photosynthesis. In particular, the GDBH states that growth is prioritised above defence, but distinguishes three zones depending on the availability of resources (**Figure 3**). The first case corresponds to drastic environmental conditions: when plants experience a very low level of resources, they are directly limited in both their growth and photosynthetic capacity. In this case, growth, photosynthesis and defence are directly correlated to
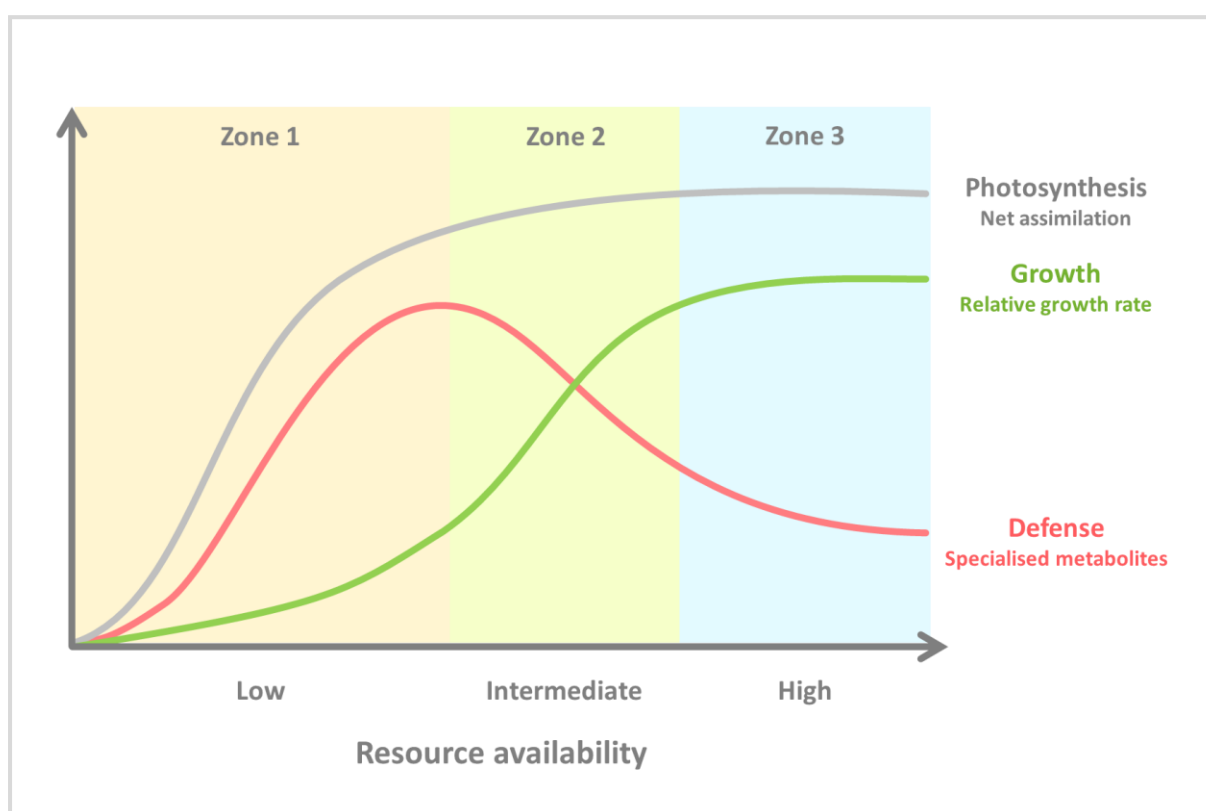
Figure 3 Schematic representation of the Growth-Differentiation Balance Hypothesis (GDBH), drawn according to Herms and Mattson (1992), Stamp (2003) and Le Bot *et al.* (2009)

resource availability. In the second zone, the resource availability increases to an intermediate level that still limits growth but has little to no impact on photosynthesis: in this situation, as the photoassimilated carbon cannot be entirely used for the growth (limited by resources), it is reallocated to defence with only a low cost to plant fitness. In the third zone, environmental conditions are optimal and plants experience a high resource availability that do not limit photosynthesis nor growth. Thus, the majority of carbon assimilated through photosynthesis is allocated to growth, rather than to defence (Loomis 1932, 1953; Herms and Mattson 1992; Stamp 2003).

In brief, plant defences are complex: to cope with various bioaggressors, plants have evolved a multitude of defences that range from indirect to physical and chemical direct mechanisms. These defences can either be constitutive and/or induced by the detection of an attack, but their establishment always requires resources, which might negatively impact growth and reproduction. Therefore, plants have also evolved strategies to limit the cost of defence, together with trade-offs that balance the allocation of resource between growth and defence.

To add another level of complexity, it should not be forgotten that every plant species possesses a combination of defences of its own. This study will mainly focus on one plant that is well defended by both physical and chemical mechanisms: ***Ficus carica*, the fig tree**.

C. FEED ME KILL ME: AGRONOMICAL INTEREST AND DEFENCE OF *FICUS CARICA*

C.1. The fig tree, a plant of agronomical interest

Ficus carica, the common fig tree, is a deciduous tree that can reach a height ranging from two to five meters and produces a popular fruit – the fig (**Figure 4**). *F. carica* is thought to be one of the earliest cultivated fruit trees: its figs are rich in vitamins, minerals and fibres, and they can be consumed fresh, dried of processed. Therefore, the fig tree has become an important crop species. *F. carica* is believed to originate from the Middle East, but it can now be found in many warm regions worldwide, and it is mainly cultivated in the Mediterranean region (Solomon et al. 2006; Oliveira et al. 2009; Bonamonte et al. 2010). In 2016, the world fig production was close to one million tons, and more than half of this production originated from Turkey, Egypt and Algeria (FAO 2017).

However, if *F. carica* is a species of agronomical and economical importance, it is only one of the 850 species that (approximatively) compose the *Ficus* genus (Janzen 1979; Pierantoni et al. 2018). Phylogenetically speaking, the *Ficus* genus belongs to the Moraceae family, from the Rosales order (**Figure 4**); it is therefore included in the Rosid I group of the Nitrogen Fixing Clade (Stevens 2001; The Angiosperm Phylogeny Group 2016). Diverse *Ficus* species are adapted to various climatic and geographic regions, in which they occur under many lifeforms such as trees (*F. carica*) but also vines, stranglers, shrubs or epiphytes (Janzen 1979; Pierantoni et al. 2018).

The *Ficus* genus is characterised and easily recognisable by its fruit: the fig (**Figure 4**). But in fact, a fig is not just a fruit, it is an inflorescence called syconium that forms a cavity enclosing tiny flowers, and that turns into an infructescence (Anstett 2001). The fig is also associated to a beneficial pollination mutualism: *Ficus* species are dependent of some tiny pollinating wasps named agaonids (Chalcidoidea, Agaonidae) that represent their only pollen vector. On their side, agaonids lay their eggs in the inflorescence of their *Ficus* host and their larvae only feed on *Ficus* flowers. *Ficus* and agaonids species are generally associated with a single or a limited number of hosts (Cruaud et al. 2012).

Photo credits:
A.a) by WAV12 - personnal work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=89686770
A.b) by AxelRohdeElias, CC BY 2.0, https://commons.wikimedia.org/w/index.php?curid=25334809
A.c) CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=263529
A.d) from the public domain, https://commons.wikimedia.org/w/index.php?curid=86246
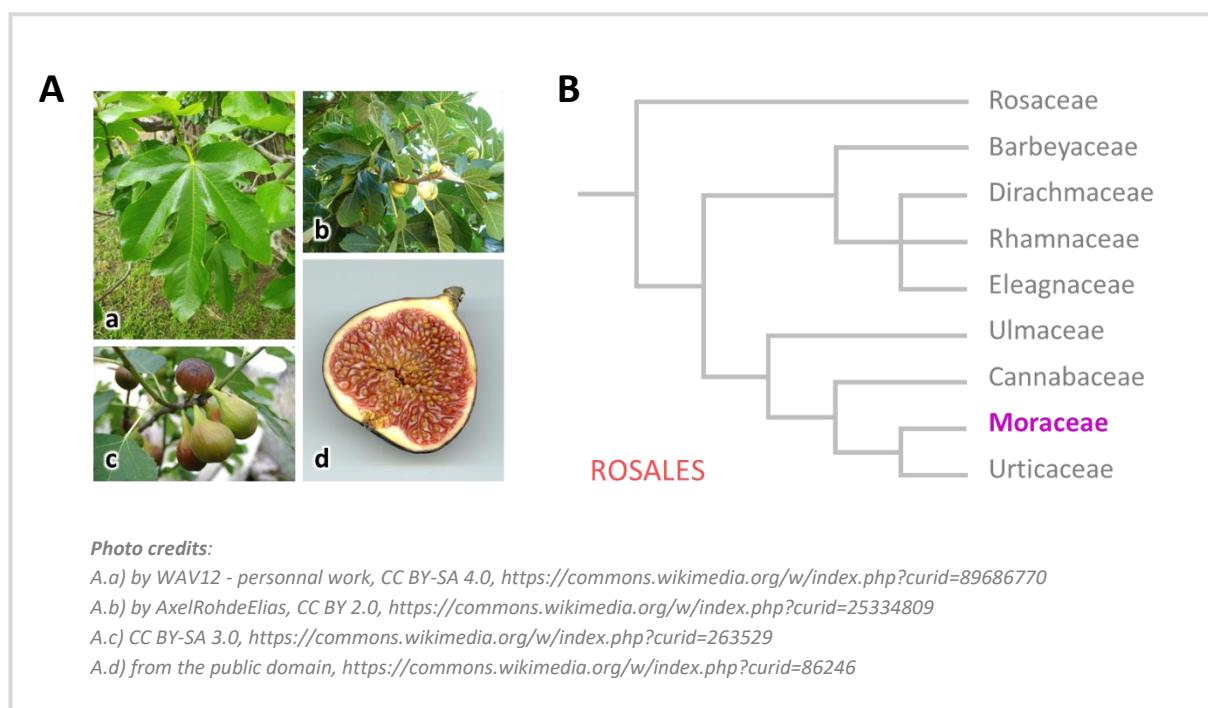
Figure 4 **Presentation of *Ficus carica*, a species from the Moraceae family. A** Photos of *F. carica* illustrating, a) the typical shape of a leaf, b) the general appearance of a branch bearing figs, c) a bunch of maturing figs, and d) the half of a ripe fig. **B** Phylogenetic tree showing the families contained in the Rosales order, based on Stevens (2001) and The Angiosperm Phylogeny Group (2016).

Like every plant, the *Ficus* can be attacked by many bioaggressors that include a large diversity of generalist and specialist insects (Basset and Novotny 1999; Novotny et al. 2010; Volf et al. 2018), and pathogens such as fungi and viruses (McKenzie 1986; Elbeaino et al. 2011; Hosomi et al. 2012; Bayoudh et al. 2017). In Europe, the fig wax scale, the fig canker and the fig rust are among the main pests and diseases that affect *F. carica*, leading to yield reduction. Yet, aside from these, *F. carica* and the *Ficus* genus in general are usually excellent at countering the attacks of their enemies (Chamont 2014).

C.2. *Ficus* defence mechanisms

In response to the selective pressure exerted by pests and pathogens, *Ficus* species have evolved pyramiding strategies that include a large variety of defence mechanisms: on one hand, *Ficus* direct defences are well known and have been reviewed in Villard *et al.* (2019). On the other hand, *Ficus* indirect defences are poorly documented. Here, I will summarise the major direct mechanisms to provide a quick overview of *Ficus* defences diversity (Villard et al. 2019).

C.2.a. Tough mineralised leaves and specialised tissues

Ficus physical defence mechanisms include mineralisation, non-glandular trichomes and latex that make the leaves tougher, rougher and sticky upon injury (**Figure 5**).

Firstly, minerals such as calcium oxalate crystals and silica can be deposited in *Ficus* tissues (Wu and Kuo-Huang 1997; Chantarasuwan et al. 2014; Pierantoni et al. 2018). These mineral deposits have been associated with increased mechanical protection and abrasiveness: they therefore contribute to establishing a physical barrier, deterring herbivores, and they can even abrade insects' and mammals' mouthparts (Yoshihara et al. 1980; Ward et al. 1997; Ruiz et al. 2002; Hudgins et al. 2003; Korth et al. 2006; Massey et al. 2007a, b; Hunt et al. 2008; Massey and Hartley 2009; Müller et al. 2014).
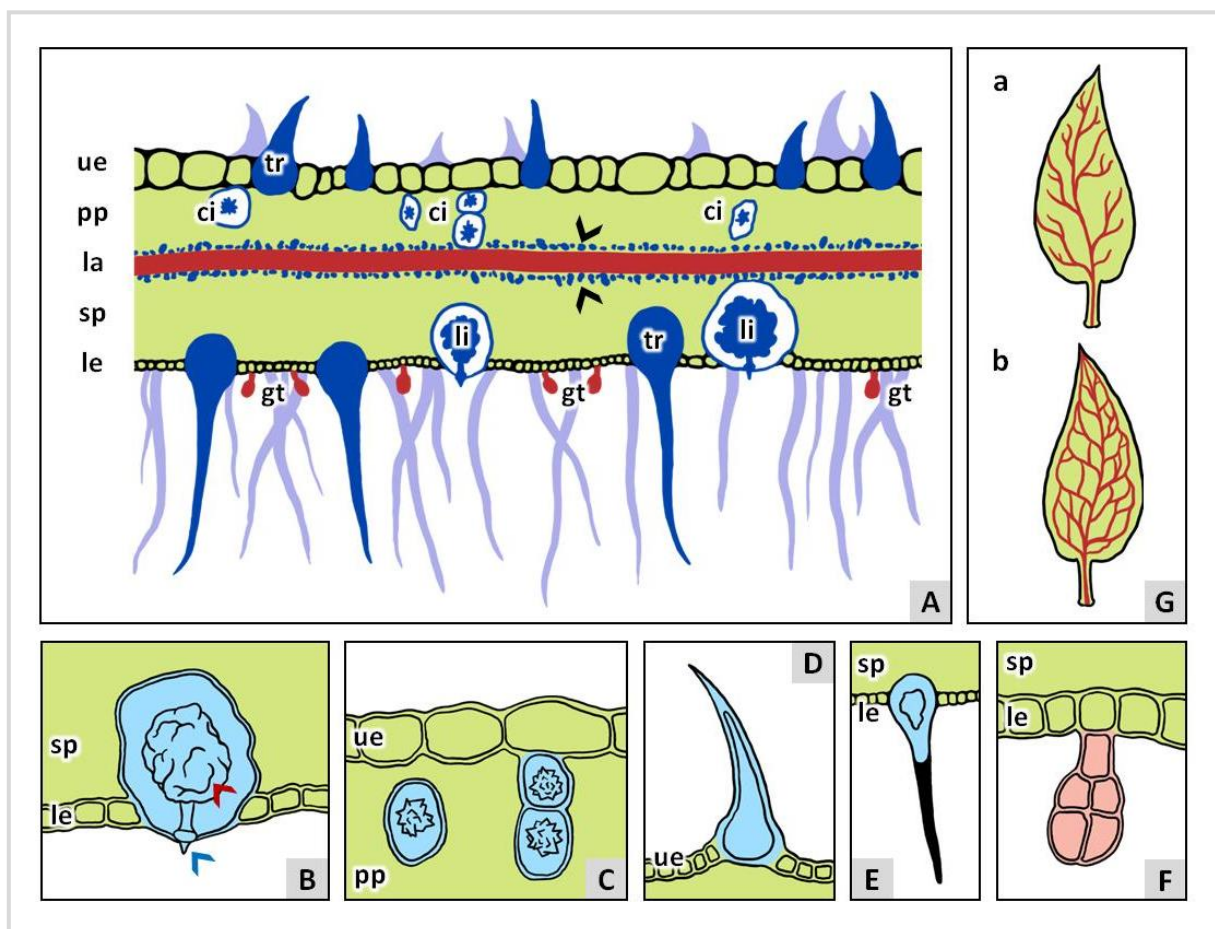
Figure 5 **Defensive structures described in *Ficus carica* leaves, from Villard *et al.* (2019).** The mineralised structures (physical defences) are in blue while the secretory structures (chemical defences) are in red. **A** Cross section of a leaf that show calcium oxalate crystals (black arrowheads) deposited parallel to a laticifer (la), lithocysts (li) located in the lower epidermis (le) and spongy parenchyma (sp), and crystal idioblasts (ci) in the palisade parenchyma (pp). (Silicified) non-glandular trichomes (tr) are present on the upper (ue) and lower (le) epidermises while glandular trichomes (gt) are only found under the nerves. These different structures are detailed in **B-G. B** Lithocyst that hosts a calcified cystolith (red arrowhead) and a silicified spike (blue arrowhead). **C** Crystal idioblasts containing spiny crystals of indeterminate nature. **D** Non-glandular conic silicified trichome. **E** Non-glandular hairy silicified trichome. **F** Capitate trichome (glandular). **G** Laticifers can a) present a tree-like shape (non-articulated), like in *Ficus*, or b) be merged (articulated).

Secondly, *Ficus* species are dotted with trichomes. Non-glandular trichomes usually cover both sides of *Ficus* leaves. They can be silicified and contribute to physical defence by increasing the roughness of the leaves and the resistance to herbivores (Klimko and Truchan 2006; Mamoucha et al. 2016; Sosnovsky 2016; Pierantoni et al. 2018). On the contrary, in *F. carica*, glandular trichomes – which accumulate chemicals – are localised under the nerves (Mamoucha et al. 2016).

Thirdly, the *Ficus* genus is equipped with a network of laticifers that contains latex (Lewinsohn 1991; Konno 2011). Upon injury of the plant, latex is quickly mobilised and delivered at the site of the injury. Once exposed to air, it becomes sticky: latex can thus form a barrier, close wounds, glue insects' mouthparts, and trap small insects (Dussourd and Eisner 1987; Farrell et al. 1990; Dussourd and Denno 1991, 1994; Konno et al. 2004; Agrawal and Konno 2009; Konno 2011). Latex also contributes to chemical defence mechanisms since it contains many defensive proteins and metabolites.

C.2.b. Ficus defensive proteins and metabolites

Ficus species produce a vast array of chemicals that exhibit various biological activities such as antimicrobial, nematicidal, antipyretic or irritant properties (Saeed and Sabir 2002; Jeong et al. 2009; Patil Vikas et al. 2010; Liu et al. 2011). *Ficus* defensive chemicals comprise proteins that are restricted to the latex, and specialised metabolites distributed in various tissues including latex and glandular trichomes (**Figure 5** – Konno 2011; Mamoucha et al. 2016; Kitajima et al. 2018).

More than 50 proteins including proteases, protease inhibitors, chitinases, oxidases, and other miscellaneous enzymes have been identified in *Ficus* latexes (Sgarbieri et al. 1964; Taira et al. 2005; Konno 2011; Kitajima et al. 2018). Some of them were described as very stable enzymes whose activities were not impacted by the digestive juice of herbivores: such proteins might therefore be involved in the defence against herbivorous pests and pathogens (Zhu-Salzman et al. 2008; Konno 2011). The major defensive proteins found in *Ficus* are ficin and trypsin inhibitors: ficin is described as a key enzyme in the defence against insects (Sgarbieri et al. 1964; Konno et al. 2004; Konno 2011; Kitajima et al. 2018) while trypsin inhibitors are well known for their antifungal and insecticidal activities (Broadway and Duffey 1986; Hilder et al. 1987; Ryan 1990; Huynh et al. 1992; Terras et al. 1993; Felton and Gatehouse 1996; Kitajima et al. 2018). *Ficus* latex also contain chitinases with antifungal and potential insecticidal effects (Schlumbaum et al. 1986; Mauch et al. 1988; Roberts and Selitrennikoff 1988; Melchers et al. 1994; Kitajima et al. 2018), and oxidases which might affect latex stickiness (Kon and Whitaker 1965; Kim et al. 2003; Taira et al. 2005; Lawrence and Novak 2006; Wahler et al. 2009; Mohamed et al. 2011).

In addition, *Ficus* species contain hundreds of specialised metabolites belonging to the alkaloid, terpenoid and phenolic families (**Figure 6**). Ficus alkaloids include indolizidine alkaloids, chlorophenanthroindolizidine alkaloids, septicine-type alkaloids and furoquinoline alkaloids that exhibit diverse defensive properties (Baumgartner et al. 1990; Smyth et al. 2012; Yap et al. 2015, 2016; Al-Khdhairawi et al. 2017). The alkaloids that have been the most extensively studied in *Ficus* are the phenanthroindolizidine alkaloids which have cytotoxic, antifungal, antibacterial, antiamoebic and antiviral properties (Baumgartner et al. 1990; Wu et al. 2002; Damu et al. 2005, 2009; Yap et al. 2015, 2016). *Ficus* terpenoids include monoterpenoids, triterpenoids, sesquiterpenoids, steroids, cardiac steroids, norisoprenoids and esters with various bioactivities such as antibacterial, insecticidal,

antiproliferative and irritant properties (Rubnov et al. 2001; Saeed and Sabir 2002; Oliveira et al. 2010a, b; Ojo et al. 2014). It should be noted that rubber is the terpenoid responsible for latex stickiness (Konno 2011). Lastly, *Ficus* phenolics have been extensively studied: they include various compounds such as phenolic acids, flavonoids, tannins and phlobatannins, coumarins, pyranocoumarins and furanocoumarins – which each possess various properties such as antioxidant, antimycobacterial or nematicidal activities (Ojala et al. 2000; Oliveira et al. 2009; Chen et al. 2010; Takahashi et al. 2014; Debib et al. 2014).
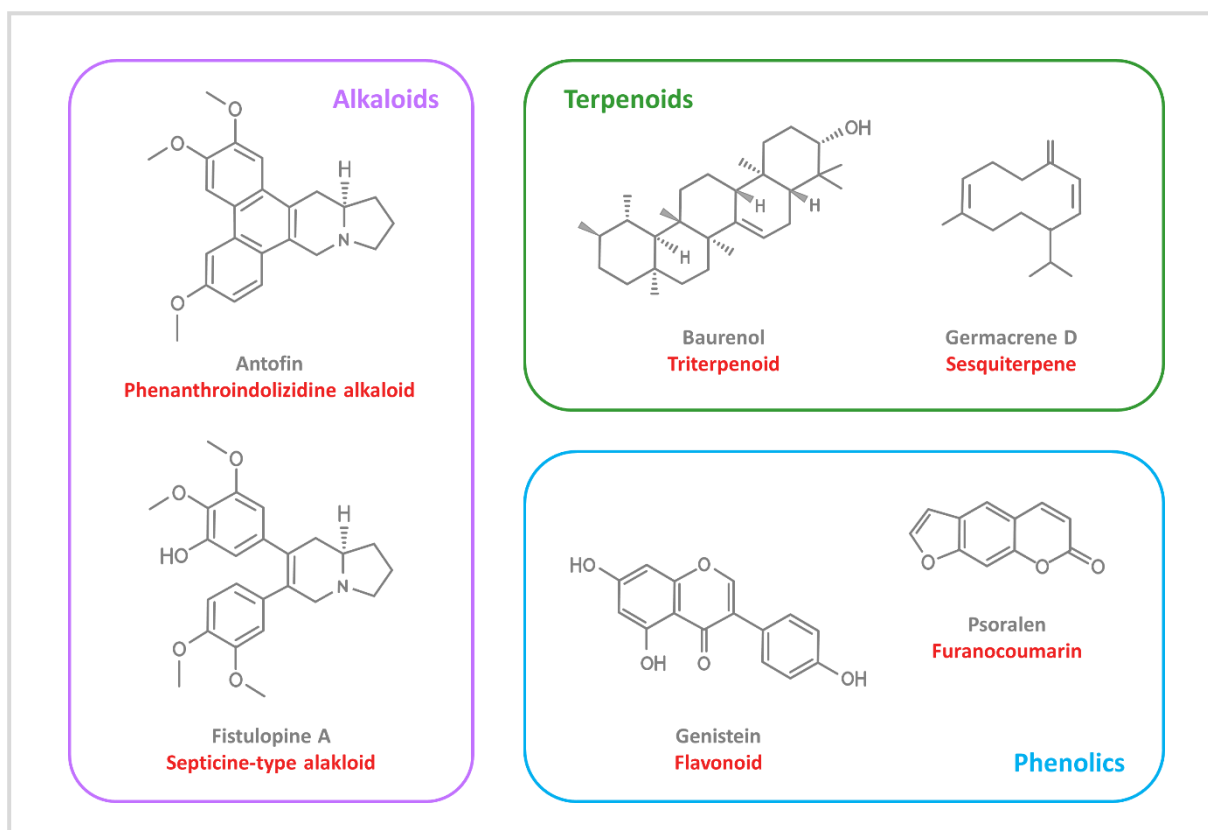
Figure 6 Structure of some *Ficus* specialised metabolites, adapted from Villard *et al.* (2019).

Finally, all these physical and chemical defences vary with many factors such as the genotype, seasonality, plant age, organs developmental stage, and some can be induced by environmental stresses (Zaynoun et al. 1984; Kim et al. 2003; Xiang and Chen 2004; Konno 2011; Marrelli et al. 2012, 2014; Pierantoni et al. 2018). They are also differentially distributed between plant parts, which might mirror the differential repartition of bioaggressors (Kitajima et al. 2018; Villard et al. 2019).

As a summary, *Ficus* defensive strategies are multiple and variable: they combine physical and chemical mechanisms that take part in the fascinating arms race between plants and pathogens.

However, as interesting these mechanisms are, I will now focus on a single family of phenolics produced by *Ficus* species and some other plants: **the furanocoumarins**.

D. Smells like toxic furanocoumarins

D.1. Furanocouma-what?

D.1.a. Furanocoumarins constitute one of the four classes of coumarins

Coumarins form a large group of phenolic compounds that is widespread in plants but also present in bacteria and fungi (Murray et al. 1982; Venugopala et al. 2013). Coumarin itself was initially found in *Dipteryx odorata*, the cumaru that produces tonka beans (Vogel 1820). Since then, coumarins have been described in more than 150 plant species from 30 families (Venugopala et al. 2013).

The core structure of all coumarins consists in a heterocycle resulting from the fusion of a benzene and an α-pyrone (benzo-α-pyrone – **Figure 7**). Then, this core coumarin can be modified, giving four classes of compounds: simple coumarins, furanocoumarins, pyranocoumarins and phenylcoumarins (Murray et al. 1982; Estévez-Braun and González 1997; Bourgaud et al. 2006; Venugopala et al. 2013). Simple coumarins are benzo-α-pyrone derivatives that underwent minor modifications. Furanocoumarins are formed through the addition of a furan ring to the coumarin core: depending on the attachment of this ring at the 6,7 or 7,8 position on the coumarin core, furanocoumarins are subclassified into linear and angular isomers. Similarly, pyranocoumarins are formed by the addition of a pyran ring on the coumarin core, and the position of the ring also leads to linear and angular isomers. Lastly, phenylcoumarins are formed via the addition of a phenyl cycle on the coumarin core. If the first three classes of coumarins derive from the phenylpropanoid pathway, phenylcoumarins derive from the isoflavone metabolism (Bourgaud et al. 2006). The basic structure of these four classes of coumarins can subsequently be modified via diverse reactions, giving a large variety of compounds.
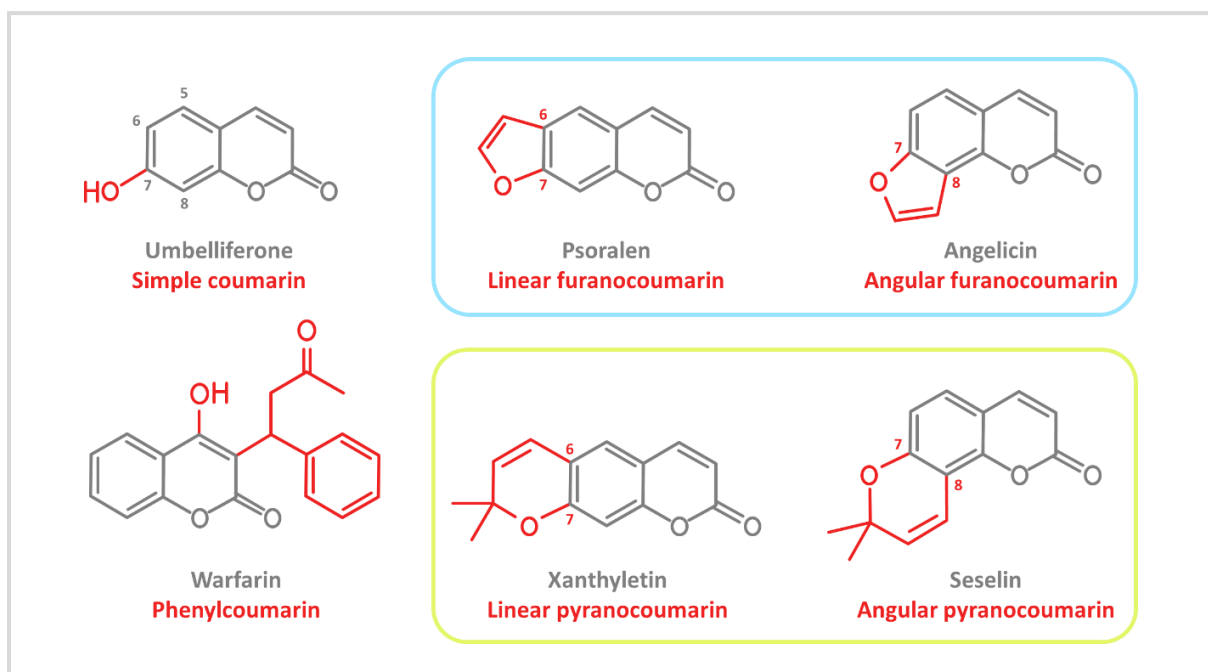
Figure 7 Typical structures of the four classes of coumarins: simple coumarins, furanocoumarins phenylcoumarins and pyranocoumarins. The coumarin core is in grey, the modifications are in red.

D.1.b. Distribution of furanocoumarins in higher plants

Unlike simple coumarins that are ubiquitous in higher plants, furanocoumarins are restricted to limited taxa scattered among angiosperms. The major furanocoumarin-producing species are found in four distant families: Apiaceae (such as parsnip), Fabaceae (such as *Coronilla*), Moraceae (such as the fig tree) and Rutaceae (such as bergamot) (Pathak et al. 1962; Murray et al. 1982; Bourgaud et al. 1989, 2006; Seigler 1998; Venugopala et al. 2013). Minor levels of furanocoumarins have also been reported in some species from the orders of Fagales, Malpighiales, Caryophyllales, Asterales, Dipsacales, Solanales, and even in Poales, a monocot order (Munakata et al. 2020). The furanocoumarins commonly found in higher plants include psoralen, xanthotoxin and bergapten for linear isomers. The common angular ones include angelicin, pimpinellin, sphondin and isobergapten (Bruni et al. 2019).

Linear and angular furanocoumarins are unequally distributed between plant taxa: indeed, some species are known to produce only linear furanocoumarins but, as far as we know, angular isomers seem to only exist in plants that also produce linear ones. For instance, high concentrations of both linear and angular furanocoumarins have been reported in the Apiaceae and Fabaceae families. On the contrary, in the Moraceae and Rutaceae families, only minor levels of angular furanocoumarins have been found in species that contain high levels of linear isomers (Bourgaud et al. 1989, 2006; Bruni et al. 2019; Munakata et al. 2020). The distribution of furanocoumarins is therefore heterogenous between plant families, but it is also inequal within them. As an example, in the Apiaceae family, all species do not produce furanocoumarins. And among the furanocoumarin-producing Apiaceae, some species such as parsnip produce both linear and angular isomers while others such as parsley or celery only produce linear ones (Berenbaum 1983; Peroutka et al. 2007; Karamat et al. 2014).

Last, furanocoumarin concentration varies among species: Apiaceae, Fabaceae, Moraceae and Rutaceae are all considered to be significant furanocoumarin producers, but, so far, the highest concentrations have been found in apiaceous species such as parsnip (145 μg.g^{-1} FW) or parsley (112 μg.g^{-1} FW). In comparison, the concentration of furanocoumarins reported in grapefruit juices (Rutaceae) is about 6 μg.g^{-1} FW (Bourgaud et al. 1989; Peroutka et al. 2007; Dugrand et al. 2013).

D.1.c. Repartition and variation of furanocoumarins within plants

Within a given plant, the chemical composition and concentration of furanocoumarins can vary a lot between plant parts: different compounds can indeed be found in varying levels in the leaves, flowers, fruits, seeds, lignified part and the roots of one single plant (Zaynoun et al. 1984; Zangerl and Rutledge 1996; Milesi et al. 2001; Oliveira et al. 2009; Marrelli et al. 2014). Furanocoumarins may also vary between the different tissues of a given organ, be localised in particular structures such as oil glands or laticifers, or be exuded on the surface of the plant (Zobel and Brown 1988, 1989; Zobel et al. 1990; Zobel and Brown 1991a; Reinold and Hahlbrock 1997; Mamoucha et al. 2016; Weryszko-Chmielewska and Chwil 2017). For instance, in *Citrus* fruits, furanocoumarins are generally more diversified and more concentrated in the peel rather than in the pulp (Dugrand-Judek et al. 2015). Yet, the repartition of furanocoumarins within a plant is highly dependent on the plant genotype – which includes species but also varieties (Oliveira et al. 2009; Alam et al. 2015; Takahashi et al. 2017). In *Citrus* fruits, it has thus been reported that the exact diversity and location of furanocoumarins, as well as their quantity and relative quantity, were highly variable between *Citrus* varieties (Dugrand-Judek et al. 2015).

In addition, the concentration and composition of furanocoumarins within a given tissue might vary across time, depending on factors such as seasonality, the age or the developmental stage of the organ (Innocenti et al. 1982; Zaynoun et al. 1984; Nitao and Zangerl 1987; Zobel and Brown 1991b; Trumble et al. 1992; Marrelli et al. 2012, 2014). These variations concern the constitutive furanocoumarin contents, but it has also been clearly demonstrated that the production of furanocoumarins can sometimes be induced by biotic and abiotic stresses (Berenbaum 2002; Bruni et al. 2019). For instance, increases of furanocoumarin contents have been reported in various plants after mechanical or insects damages (Zangerl and Berenbaum 1990; Zangerl and Rutledge 1996; Berenbaum 2002; Galati et al. 2019), UVs, pesticide or pollution exposure (Zobel and Brown 1993; Nigg et al. 1997; Reitz et al. 1997; Galati et al. 2019). But once again, some plant parts are more inducible than others, which is somehow linked to the toxic and defensive properties of furanocoumarins (Zangerl and Rutledge 1996).

> In brief, furanocoumarins are a class of coumarins restricted to a few distant plant taxa. Various linear and angular compounds are heterogeneously distributed, localised and concentrated in different species and tissue – where they also vary upon time according to many factors.
>
> From now on, I will mainly focus on linear furanocoumarins, which are the most abundant in plants, have been more studied, and are the main isomers found in the *Ficus* genus.

D.2. Defensive properties of furanocoumarins

Linear furanocoumarins exhibit toxic properties that make them an efficient defence against a large range of organisms that include pathogens, herbivores and plants. Indeed, furanocoumarins are well known for their antimicrobial (Ojala et al. 2000), antifungal (Fracarolli et al. 2016) and antiviral activities (Palú et al. 1984), as well as for their repulsive and toxic effects against various animals such as arthropods (Berenbaum and Feeny 1981; Diawara et al. 1993; Li et al. 2003), mammals (Bonamonte et al. 2010; Son et al. 2017) and nematodes (Liu et al. 2011; Caboni et al. 2015; Guo et al. 2016). Furanocoumarins also act as phytoalexins and are involved in allelopathic processes such as delaying seed germination and growth (Baskin et al. 1967; Junttila 1976; Beier and Oertli 1983).

Furanocoumarins defensive properties can be divided into toxic and phototoxic effects that seem to mainly result – respectively – from their ability to inhibit some enzymes such as cytochrome P450s (P450s), and to become reactive molecules upon UV excitation.

D.2.a. Toxic properties and P450 inhibition

Furanocoumarins can interact with P450s from various organisms such as insects, mammals and plants. In particular, linear furanocoumarins inhibit P450s by irreversibly binding to them, acting as "suicide inhibitors". The inhibition of P450s lead to a decrease in enzymatic activities, which can result in toxic effects depending on the initial roles of the inhibited P450s (Neal and Wu 1994; He et al. 1998; Koenigs and Trager 1998; Ohta et al. 2002; Gravot et al. 2004; Messer et al. 2012; Lin et al. 2012).

For instance, in humans, furanocoumarins can inhibit CYP3A4, a P450 found in the intestine and liver that is essential for metabolising several drugs. This inhibition might lead to an increase of the bioavailability of some medicines, which can cause serious adverse effects such as respiratory depression. This result is often called the "grapefruit juice effect" since it was first discovered with the ingestion of furanocoumarins through the consumption of grapefruit juice (Bailey et al. 1991, 2013; Wilkinson 2005).

D.2.b. Phototoxic properties and phytophotodermatitis

When exposed to ultraviolet (UV) radiations, furanocoumarins can also display phototoxic activities. Indeed, linear furanocoumarins can be activated by UVA ranging from 320 to 400 nm: when they receive and absorb a photon, they form a triplet excited state that can bind and react with molecules such as oxygen or the pyrimidine bases in DNA and RNA molecules. The reaction with oxygen leads to the formation of reactive oxygen species that can subsequently damage various macromolecules such as proteins, lipids, or nucleic acids. The reaction with pyrimidine bases results in the formation of monoadducts and/or cross-linked of complementary strands of DNA (diaducts), which might lead to genetic damages (mutations, recombinations, double-stand breaks) and block the replication. All these mechanisms can result in cell death. The ability of furanocoumarins to alter DNA upon UV exposure is assumed to be the main cause of their toxicity (Musajo et al. 1965; Diawara and Trumble 1997; Dardalhon et al. 1998; Berenbaum 2002; Melough and Chun 2018; Bruni et al. 2019).

The furanocoumarins are therefore potential photosensitisers, and they are well known to cause sever contact dermatitis to humans. Indeed, when applied on the skin and exposed to light, furanocoumarins might cause cell damages that manifest through the appearance of injuries such as burns, blisters, blights or erythema (**Figure 8**). This is called phytophotodermatitis (Pathak et al. 1962; Bonamonte et al. 2010; Son et al. 2017).

Figure 8 Cases of phytophotodermatitis caused by the furanocoumarins contained in *Ficus carica*, adapted from Mandalia *et al.* (2008). Hand and leg of children showing partial thickness burns with blistering.

The furanocoumarin photosensitising effects can cause various problems to humans: for instance, in the cosmetic industry, some of the *Citrus* essential oils that are widely used for cosmetics and perfumes contain furanocoumarins (Dugrand-Judek et al. 2015). In addition, people that work with furanocoumarin-producing plants, such as fig pickers, can experience phytophotodermatitis (Bonamonte et al. 2010). Domestic accidents involving furanocoumarin-producing plants are also more frequent that one might think: as an example, some children playing with fig tree branches in their garden ended up with severe burns and bullaes (**Figure 8**) (Mandalia et al. 2008). Worse, some furanocoumarin-producing plants are used in traditional medicine that promotes their direct

application on the skin. Here again, the fig tree is a good example for it is widely used in popular medicine, in attempts to cure skin diseases: as a result, cases of sever phytophotodermatitis have been reported after people voluntary (and repeatedly) soaked their feet in fig leaf decoctions (Son et al. 2017). Phytophotodermatitis are therefore common after contact followed by UV exposure, but, in rare cases, they have also been reported after ingestion of furanocoumarin-producing plants followed by UV exposure (Calka et al. 2005).

D.2.c. From defensive properties to interesting bioactivities

Thanks to their defensive properties, furanocoumarins exhibit a large range of bioactivities such as photosensitiser, P450-inhibitors, antifungal, antibacterial and antiviral activities, but also antioxidant, anti-inflammatory, anticancer, anticonvulsant, antiproliferative, anti-hyperglycemic, cytoprotective, apoptotic agent or even anti-coagulant activities (Bourgaud et al. 1990; Venugopala et al. 2013; Bruni et al. 2019). Some of these properties have been known for hundreds or thousands of years (Scott et al. 1976), which explains why furanocoumarin-producing plants have been widely used by humans in many domains such as medicine and pharmaceutics (Bourgaud et al. 1990).

For instance, linear furanocoumarins have been historically used against skin disorders, skin depigmentation and other diseases such as psoriasis or vitiligo. Unfortunately, the toxic properties of furanocoumarins also apply to humans and these medical uses have been linked to side effects such as increased skin cancers (Diawara and Trumble 1997; Sarker and Nahar 2004; Melough and Chun 2018). Yet, despite these side effects, furanocoumarin bioactivities are still very interesting for both academic and industrial research, including the medical field. As an example, furanocoumarins are still tested and evaluated for their cytotoxicity against cancer (Chauthe et al. 2015), or for a safe use against diseases such as vitiligo (Zabolinejad et al. 2020).

> As a summary, furanocoumarins exhibit various bioactivities that include a toxicity linked to P450 inhibition and a phototoxicity that, among other things, leads to mutations and cell death. While these properties may prove useful for humans, in a first instance, they make furanocoumarins fearsome defensive compounds effective against animals, pathogens and plants – which had some consequences on the plants that started to produce them.

D.3. Consequences of these defensive properties for the plants themselves

D.3.a. Emergence of linear and angular furanocoumarins

Insects are not the only plant bioaggressors, but they exert a pressure that is an important driver of plant evolution and which might have played a key role in the emergence and diversification of furanocoumarins (Berenbaum and Feeny 1981). Indeed, the distribution of furanocoumarins in phylogenetically distant plant families suggest that these compounds might have independently appeared several times in different plant taxa, by convergent evolution, in response to environmental pressure (Berenbaum 1983; Berenbaum and Zangerl 2008). In addition, angular furanocoumarins seem

to only exist in plants that also produce linear isomers: angular compounds might therefore have emerged latter, in plants that already produced linear isomers, in response to the apparition of insects able to detoxify linear furanocoumarins (Berenbaum 1978, 1983; Berenbaum and Feeny 1981; Stanjek and Boland 1998). For this reason, the emergence and diversification of linear and angular furanocoumarins in higher plants is often described as an example of co-evolution between some herbivorous insects and their host plants (Ehrlich and Raven 1964; Berenbaum 1983).

Indeed, the majority of herbivores are sensitive to furanocoumarins (unadapted species), but some specialist insects (mainly butterflies) can feed on furanocoumarin-producing plants. Specialists such as *Papilio polyxenes* can ingest and tolerate linear isomers but are sensitive to angular furanocoumarins (adapted oliphagous), while a few species such as *Papilio brevicauda* can feed on plants that contain both linear and angular isomers (adapted polyphagous) (Ehrlich and Raven 1964; Berenbaum 1983; Berenbaum and Zangerl 1993, 1998). Insect resistance against furanocoumarins involve two mechanisms. The first one is a biochemical resistance based on cytochrome P450s: in insects, P450s are involved in the biosynthesis and catabolism of various compounds, which includes the detoxification of exogenous toxic compounds. Therefore, some insect P450s can metabolise furanocoumarins. In addition, insects that feed on furanocoumarin-producing plants might be less sensitive to P450 inhibition than other insects (Zumwalt and Neal 1993; Neal and Wu 1994; Schuler 1996; Berenbaum 2002). The second resistance is behavioural. It consists in avoiding the ingestion of a lethal dose of furanocoumarin and/or preventing their toxicity, through strategies such as leaf-rolling that allows to avoid UV exposure. However, this resistance is not total because the detoxification of furanocoumarins is costly for the insects and have been linked to reduced growths (Berenbaum 1983, 2002; Berenbaum and Zangerl 1992; Schuler 1996; Calla et al. 2020).

On the basis of the data and hypothesis found in the literature, we may propose the following coevolutive story, summarised in **Figure 9**.

1) Let us start from a time when simple coumarins were widespread in plants. Through gene duplication and diversification events, some plants evolved a new pathway that modifies simple coumarins (umbelliferone) into linear furanocoumarins. **2)** The retention of this pathway might be due to the (photo)toxic properties of linear furanocoumarins, that provided a better defence against insects and other bioaggressors than their coumarin precursors (Ehrlich and Raven 1964; Berenbaum 1983). At some point, some lepidopteran species have evolved mechanisms that were able to counter the toxicity of linear furanocoumarins (Berenbaum and Feeny 1981; Berenbaum 1983, 2002; Berenbaum and Zangerl 1992, 1993; Li et al. 2003). **3)** These caterpillars were therefore able to feed on furanocoumarin-producing plants and, by exploiting resources that could not be exploited by the majority of herbivores, they should have entered a new adaptative zone (Ehrlich and Raven 1964; Berenbaum 1983). However, by feeding on the plants that produced linear furanocoumarins, these newly specialised insects might have exerted an increased selective pressure that, in a few furanocoumarin producers, drove the emergence of angular isomers (Stanjek and Boland 1998). **4)** Angular furanocoumarins seem to display weaker phototoxicity than their linear isomers (Wamer et al. 1995; Bruni et al. 2019). Yet, as they can interfere with the P450-mediated detoxification process of insects, they contribute to reduce the detoxification of linear isomers. Therefore, the combination of both linear and angular furanocoumarins happened to be more toxic than pure linear isomers, which created a new protection against specialists (Berenbaum and Zangerl 1993; Stanjek and Boland 1998).

But once again, angular furanocoumarins were not immune to counteradaptation because a few lepidopteran species – probably ones that were already able to metabolise linear furanocoumarins – became able to feed on plants that produce both linear and angular isomers (Berenbaum and Feeny 1981; Berenbaum and Zangerl 1993, 1998; Berenbaum 2002; Li et al. 2003). **5)** This assumed coevolutive story, which is also an example of escalating arms race, results in the actual situation: on the one hand, most plants do not produce furanocoumarins, some species produce linear isomers, and a subset also produces angular ones. On the other hand, most herbivores cannot feed on furanocoumarin-producing plants, some specialists can resist to linear isomers, and a subset also resists to angular ones.
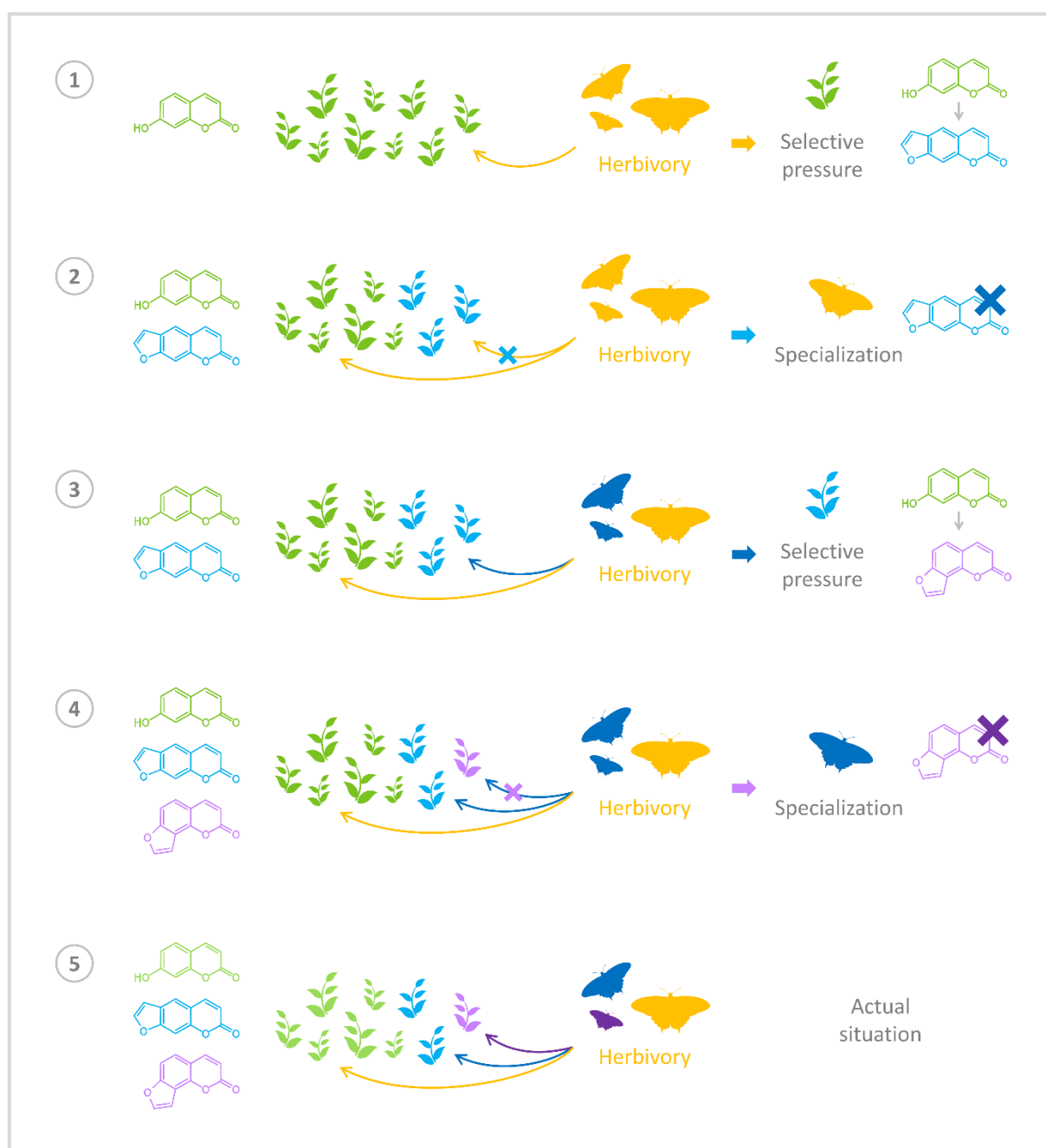
Figure 9 Hypothesised co-evolution between furanocoumarin-producing plants and insects.

D.3.b. Defensive properties and spatio-temporal variation within plants

As mentioned previously, in a given plant, furanocoumarins are unequally distributed between plants parts (organs, compartments) and vary over time (constitutive variations and inducibility). In fact, these variations are partly linked to the defensive nature of furanocoumarins.

On the one hand, furanocoumarins fall within the framework of the trade-offs between growth and defence, which means the production and storage of furanocoumarins might negatively impact plant growth. Plants might therefore have evolved strategies to reduce the cost of furanocoumarin production. For instance, plants may produce these toxic compounds only when and where there are needed. In particular, the spatial and temporal variations in the production of furanocoumarins might reflect evolutionary mechanisms that have often been linked to the optimal defence theory – a theory which predicts that plant tissues should be defended according to their contribution to plant fitness and the probability of attack (McKey 1974, 1979; Nitao and Zangerl 1987; Zangerl and Berenbaum 1990; Zangerl and Rutledge 1996; Innocenti et al. 1997a; Zangerl and Nitao 1998).

On the other hand, because of the high toxicity of furanocoumarins, plants have evolved strategies to synthesise, store and bring these defensive chemicals into contact with their bioaggressors, without poisoning themselves. Some common strategies to prevent autotoxicity consist in storing defensive compounds at specific locations, in order to isolate them from sensitive tissues. It might therefore explain why – depending on species – furanocoumarins are compartmented in structures such as laticifer cells, trichomes or oil ducts (Reinold and Hahlbrock 1997; Mamoucha et al. 2016; Weryszko-Chmielewska and Chwil 2017). Furanocoumarins can also be stored as inactivated glycosylated forms, in specific cellular compartments, separately from their activating enzymes. As an example, in *Psoralea* and *Coronilla*, glycosylated furanocoumarins are stored in the vacuole: upon attack, if the vacuole membrane (tonoplast) is disrupted, glycosylated furanocoumarins are brought into contact with endogenous β-glucosidases that hydrolyse them into activated free-form furanocoumarins (Innocenti et al. 1997b). Lastly, furanocoumarins might also be transported on the surface of the plant, where they represent a lower risk for the plant itself while serving as a barrier against bioaggressors. For instance, furanocoumarins have been found on the surface of the leaves, stems and/or trichomes of several Apiaceae such as parsley (*Petroselinum crispum*), the rue (*Ruta graveolens*) or the hogweed (*Heracleum sosnowskyi*). In these plants, surface furanocoumarins can even represent the major part of the total furanocoumarin content – which might explain many phytophotodermatitis (Zobel and Brown 1988; Reinold and Hahlbrock 1997; Weryszko-Chmielewska and Chwil 2017).

Finally, if furanocoumarins can partially inhibit plant P450s, the P450s of furanocoumarin-producing plants are usually less sensitive to furanocoumarin inhibition than average. This increased resistance might be an adaptative solution that allows furanocoumarin-producing plants to conserve a metabolic activity in presence of furanocoumarins (Gravot et al. 2004).

> In brief, the defensive properties of furanocoumarins might underly their emergence, retention and diversification. However, as these properties might also apply to the plants that produce them, furanocoumarins-producing plants have had to evolve strategies to avoid autotoxicity.

D.4. The furanocoumarins in the Moraceae family

In the Moraceae family, the presence of furanocoumarins have sometimes been reported in species such as *Maquira calophylla* (Rovinski and Sneden 1984), *Brosimum Gaudichaudii* (Vieira et al. 1999), *Fatoua pilosa* (Chiang et al. 2010), and three *Dorstenia* species (Abegaz et al. 2004; Heinke et al. 2012). Yet, it is in the *Ficus* genus that large amounts of furanocoumarins have mostly been described (**Figure 10**). As far as I know, other Moraceae including the *Morus* genus do not contain furanocoumarins.
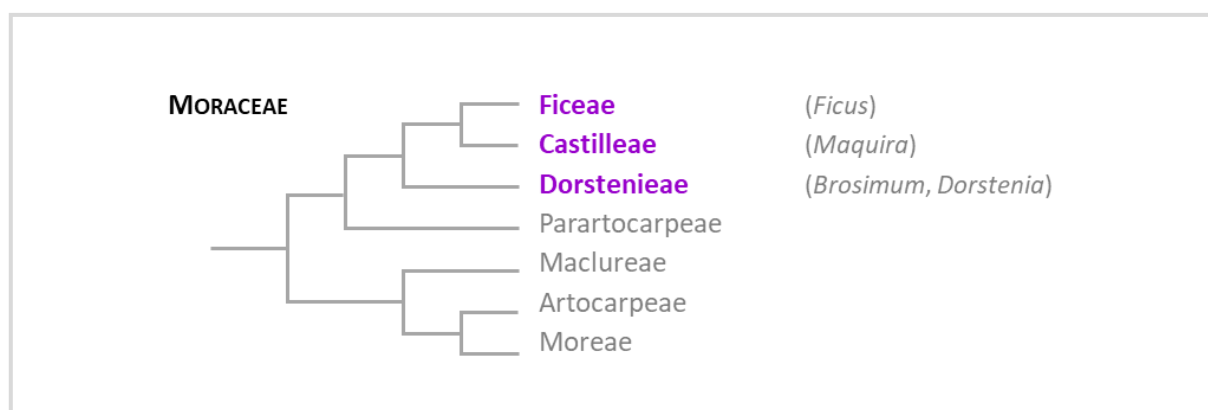
Figure 10 Distribution of furanocoumarins within the Moraceae family. The phylogeny of the Moraceae family has been drawn according to Zerega and Gardner (2019). The Moraceae tribes which are known to contains furanocoumarin producing species are in purple, the associated genera are detailed. The *Ficus*, *Maquira*, *Brosimum* and *Dorstenia* genera have been positioned in the tribes according to Clement and Weiblen (2009) and Zerega and Gardner (2019). As the position of the *Fatoua* genus was conflictuous, it has not been included (Clement and Weiblen 2009).

D.4.a. Furanocoumarins described in various *Ficus* species

Diverse furanocoumarins have been detected in various *Ficus* species such as *F. carica* (Späth et al. 1937; Athnasios et al. 1962) but also *F. capensis* (Sirisha et al. 2010), *F. coronata* (Smyth et al. 2012), *F. cyathistipula* (El-Sakhawy et al. 2016), *F. nervosa* (Chen et al. 2010), *F. palmata* (Alam et al. 2015), *F. ruficaulis* (Chang et al. 2005), *F. salicifolia* and *F. sycomorus* (Abu-Mustafa et al. 1963).

The major furanocoumarins described in *F. carica* (**Figure 11**) are psoralen and bergapten (Takahashi et al. 2014, 2017; Wang et al. 2017), two linear isomers which were highlighted for the first times in 1937 and 1962 (Späth 1937; Athnasios et al. 1962). Psoralen and bergapten are also among the main phenolics found *F. carica* leaves (Oliveira et al. 2012). Another compound found in large amount in *F. carica* leaves is psoralic acid glucoside (PAG), a glycosylated derivative of psoralen. PAG content can be equivalent to psoralen content, but its physiological function is still unknown (Takahashi et al. 2014, 2017; Wang et al. 2017). Among the other linear furanocoumarins found in *F. carica*, we can mention marmesin (Innocenti et al. 1982), xanthotoxol (Mamoucha et al. 2016), xanthotoxin and rutaretin (Marrelli et al. 2012), 4',5'-dihydropsoralen (Caporale et al. 1970; Innocenti et al. 1982) and 5-(1",1"-dimethylallyl)-8-methyl psoralen (Jaina et al. 2013). Angelicin and pimpinellin, two angular furanocoumarins, have even been detected in *F. carica* fruits (**Figure 11**) (Marrelli et al. 2012). Many

of these compounds were described in other *Ficus* species, together with additional molecules such as chalepin or rutamarin (Smyth et al. 2012; Luz et al. 2015). Last, *F. ruficaulis* seems to contain several glycosylated derivatives of both linear and angular furanocoumarins (Chang et al. 2005).
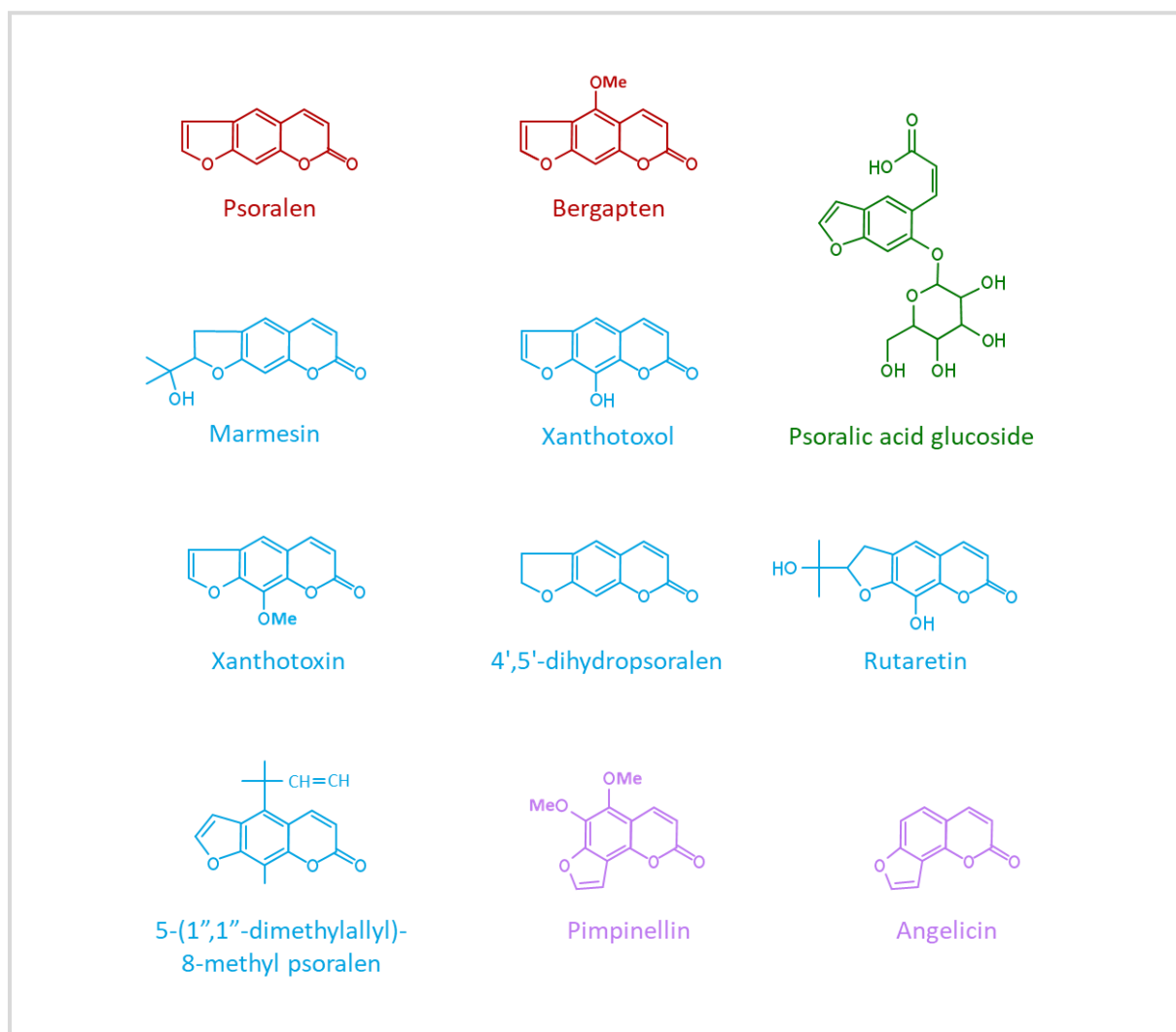
Figure 11 Structure of the furanocoumarins described in *Ficus carica*. The major (linear) furanocoumarins are in red, the linear isomers are in blue, the angular isomers are in purple, the glycosylated derivative is in green.

D.4.b. Repartition of furanocoumarins in *Ficus carica*

As for furanocoumarins in general, the composition and concentrations of furanocoumarins in the fig tree is highly variable since it depends on genotype, plant parts, and changes over time (Zaynoun et al. 1984; Oliveira et al. 2009; Takahashi et al. 2017). For instance, in a recent study done on 37 *F. carica* cultivars, furanocoumarins were detected in every cultivar but one, Grise de Tarascon. In addition, for the 36 furanocoumarin-producing cultivars, the concentrations measured in the leaves were variable: depending on the genotypes, psoralen and bergapten concentrations ranged – respectively – from 3.8 to 23.0 mg/g DW and 0.4 to 5.0 mg/g DW (Takahashi et al. 2017).

In *F. carica*, furanocoumarins have been detected in the leaves, stems, roots, bark, trunks, lignified parts and the fruits (Zaynoun et al. 1984; Oliveira et al. 2009; Jaina et al. 2013; Marrelli et al. 2014). Despite some variations, furanocoumarin concentrations are higher in the leaves rather than in the stems and trunk, which are themselves richer than the fruits. The concentration of furanocoumarins can indeed be 100 to 1000 times higher in the leaves rather than in the fruits (Oliveira et al. 2009). The furanocoumarin content is even so low in fruits that it has remained under analytical detection limits for a long time and it has long been said that figs were furanocoumarin-free (Zaynoun et al. 1984).

In addition, if many studies were performed on entire leaves, some recent work showed that, in *F. carica*, furanocoumarins might be mostly located around the nerves, in the laticifer cells and the glandular trichomes (Mamoucha et al. 2016). This distribution is consistent with the previously described mechanisms that allow to avoid autotoxicity while maintaining furanocoumarins ready for any attack. Finally, the differential repartition of furanocoumarins between plant parts has also been confirmed within latexes: furanocoumarin concentration levels are higher in the latex of the leaves (petioles) than in the latexes of the trunks and the fruits, and these differences might reflect adaptative strategies to counter the different bioaggressors found on every plant parts (Kitajima et al. 2018).

As a summary, furanocoumarins are defensive compounds found in distant plant taxa, such as the *Ficus* genus. They are effective against a large range of bioaggressors, but might also be toxic for the plants that produces them. Therefore, plants may have evolved strategies to reduce or avoid autotoxicity, which includes a spatio-temporal regulation of their production.

But, how do plants produce furanocoumarins?

E. THE FURANOCOUMARIN PATHWAY MUST GO ON

The biosynthesis pathway of furanocoumarins has been well studied and biochemically characterised for decades. Between the 1960's and 1990's, feeding experiments with radiolabelled precursors allowed the identification of the successive intermediates of the pathway. Yet, the study of the enzymes involved in this pathway is more recent (Floss and Mothes 1966; Brown and Steck 1973; Caporale et al. 1981; Hamerski and Matern 1988a; Stanjek et al. 1997). Indeed, the use of radiolabelled compounds, combined to cell culture approaches, suggested the involvement of cytochromes P450s in the biosynthesis of furanocoumarin in *Ammi majus* (Hamerski and Matern 1988a). However, it is only in the 2000's that these results were confirmed and that other enzyme families were identified.

To date, four enzyme families involved in the furanocoumarin pathway have been described: the cytochromes P450, but also dioxygenases, prenyltransferases, and methyltransferases (**Figure 12**). The first gene coding for one of these enzymes was cloned and characterised in 2004 (Hehmann et al. 2004). It has been followed by other biosynthetic genes, mainly identified in apiaceous species. Nevertheless, the molecular elucidation of the pathway is still incomplete and many genes are still to be discovered.

E.1. The different steps of the furanocoumarin biosynthesis pathway

The furanocoumarin pathway derives from the phenylpropanoid pathway. In particular, it starts with the ubiquitous *p*-coumaroyl-coA, a major branching point that can lead to various compounds such as flavonoids, stilbenes, monolignols or phenolamides. The sequential steps of the main linear furanocoumarin pathway, summarised in **Figure 12** are the following:

1) First, *p*-coumaroyl-coA is converted into 2-4-dihydroxycinnamate through an *ortho*-hydroxylation catalysed by a dioxygenase. It then spontaneously cyclises into umbelliferone (Bourgaud et al. 2006). The molecular elucidation of this step started in 2008 with some studies done on a 2-oxoglutarate-dependent dioxygenase (2-OGD) involved in the formation of scopoletin in *A. thaliana* (Kai et al. 2008). Yet, it is only in 2012 that the first dioxygenase involved in the formation of umbelliferone was described. This enzyme, the *p*-coumaroyl CoA 2'-hydroxylase (C2'H), is a 2-OGD which was identified and characterised in the rue (*Ruta graveolens*) (Vialart et al. 2012). Similar results have also been obtained in the sweat potato (*Ipomea batatas* – which does not produce furanocoumarins) with the identification of another C2'H (Matsumoto et al. 2012). More recently, in parsnip (*Pastinaca sativa*), an α-ketoglutarate-dependent dioxygenase named PsDiox has been described: it catalyses the same reaction than C2'H but shares only 52% identity with the gene from *R. graveolens* (Roselli et al. 2017).

2) Umbelliferone, which is still a simple coumarin, can subsequently be prenylated to open the way to linear and angular furanocoumarins. This prenylation, catalysed by a prenyltransferase (PT) called umbelliferone dimethylallyltransferases (UDT), is regiospecific: a prenylation on the C6 forms demethylsuberosin (DMS, 6-dimethylallylumbelliferone), which leads to linear isomers **(2)**, while a prenylation on the C8 forms osthenol and paves the way to angular compounds **(2')** (Brown and Steck 1973). The first UDT, identified in parsley (*Petroselinum crispum*), has been named PcPT – for *P. crispum* PT. It mainly catalyses the C6 prenylation of umbelliferone into DMS but also shows a minor activity at C8 (Karamat et al. 2014). Two years later, two other UDTs were found in parsnip (*Pastinaca sativa*): PsPT1 and PsPT2. If PsPT1 mainly catalyses the conversion of umbelliferone into DMS (C6), PsPT2 is more efficient at forming osthenol (C8). It should be noted that PsPT2 might have emerged through the duplication and neo-functionalisation of PsPT1 (Munakata et al. 2016), and that PsPT1 is clustered with PsDiox (Roselli et al. 2017). These apiaceous UDTs are all expressed in the plastids, which suggests it is where the furanocoumarin biosynthesis starts. More recently, another UDT specialised in the conversion of umbelliferon into DMS has been found in *F. carica* (FcPT1) (Munakata et al. 2020).

3) The pathway to linear furanocoumarins continues with the conversion of DMS into (+) marmesin through the cyclisation of the prenyl group on the hydroxyl group. This cyclisation is not spontaneous but catalysed by an enzyme named the marmesin synthase. In *Ammi majus*, evidences strongly suggest that the marmesin synthase is a cytochrome P450 (Hamerski and Matern 1988a) but, until now, this enzyme has never been characterised and its gene remains unknown.

4) The next step is the conversion of marmesin into psoralen, which is the first linear furanocoumarin that exhibits a strong toxicity. This reaction consists in a P450-mediated cleavage of marmesin that produces psoralen and a molecule of acetone (Hamerski and Matern 1988a). Several genes coding for psoralen synthases have been isolated in apiaceous species. The first one, *CYP71AJ1*, was isolated and characterised from *Ammi majus* cell cultures in 2007 – which made it the first known P450 from the

furanocoumarin biosynthesis pathway (Larbat et al. 2007). The search for CYP71AJ1 orthologs in other Apiaceae latter led to the identification of two other psoralen synthases: CYP71AJ2 from parsley (*P. crispum*) and CYP71AJ3 from parsnip (*P. sativa*) (Larbat et al. 2009).
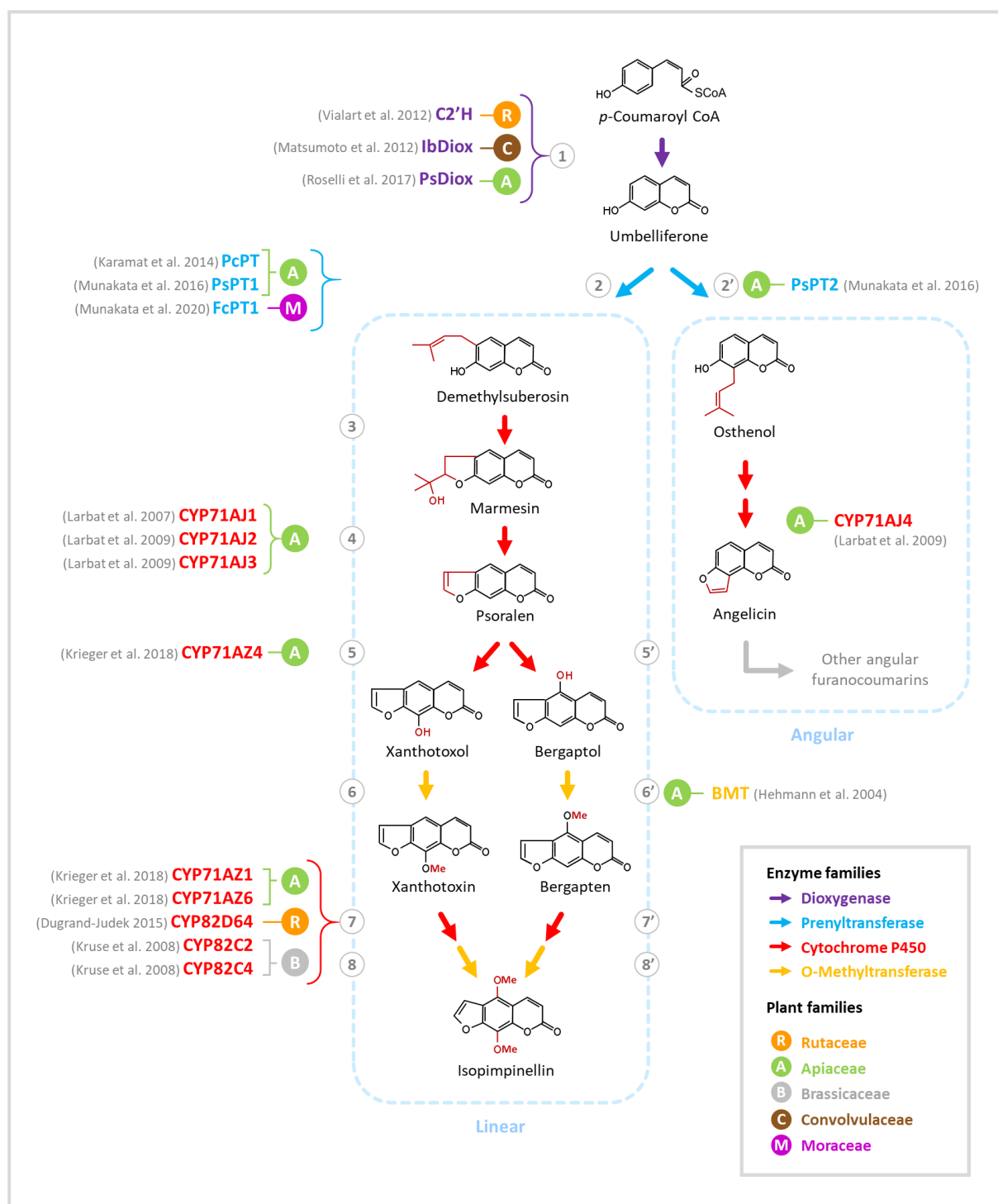
Figure 12 Simplified representation of the biosynthesis pathway of linear and angular furanocoumarins. Enzymatic reactions are pictured by an arrow which color reflects the nature of the enzyme. When the step has been molecularly elucidated, the name of the associated gene(s) is detailed, together with the plant family in which it has been isolated. The pathway of angular furanocoumarins mirrors the linear one, but as it is poorly known, it has not been detailed.

5) Then, psoralen can be hydroxylated at positions C5 **(5')** or C8 **(5)**, which produces, respectively, bergaptol (5-hydroxypsoralen) or xanthotoxol (8-hydroxypsoralen). If both reactions might be catalysed by P450s, the only enzyme characterised so far is CYP71AZ4, from *P. sativa*, that converts psoralen into xanthotoxol (Krieger et al. 2018).

6) The hydroxyl groups from bergaptol and xanthotoxol can subsequently be methylated to form bergapten **(6')** and xanthotoxin **(6)**. In *Ammi majus*, the synthesis of bergapten is catalysed by the bergaptol O-methyltransferase, which was the first gene of the furanocoumarin pathway to be cloned (Hehmann et al. 2004). The formation of xanthotoxin has still to be molecularly elucidated.

7) Then, bergapten and xanthotoxin might be hydroxylated at positions C8 and C5, which respectively forms 8-hydroxybergapten **(7')** and 5-hydroxyxanthotoxin **(7)**. Both hydroxylations might be catalysed by P450s, but it has only been demonstrated for xanthotoxin. Five P450s able to convert xanthotoxin into 5-hydroxyxanthotoxin **(7)** have indeed been identified in three species: CYP71AZ1 and CYP71AZ6 in *P. sativa* (Krieger et al. 2018), CYP82D64 in *Citrus* (Dugrand-Judek 2015; Limones-Mendez et al. 2020), CYP82C2 and CYP82C4 in *A. thaliana* (Kruse et al. 2008).

8) Finally, 8-hydroxybergapten **(8')** and 5-hydroxyxanthotoxin **(8)** both lead to the formation of isopimpinellin through the methylation of their respective hydroxy groups. But, as far as we know, there is no molecular evidence for any of these reactions.

Less is known about angular furanocoumarins: the pathway of angular isomers starts with the prenylation of umbelliferone into osthenol (PsPT2), as described above. The following steps might perfectly mirror the linear pathway, but with other enzymes: osthenol is converted into (+)-columbianetin and then into angelicin, the angular equivalent of psoralen. The conversion of (+)-columbianetin into angelicin is catalysed by a P450: in *P. sativa*, this P450 is CYP71AJ4 and it has been identified by similarity with CYP71AJ1 and CYP71AJ3, its linear equivalents (Larbat et al. 2009). It should also be noted that CYP71AJ3 and CYP71AJ4 (both from *P. sativa*) are arranged into a genomic cluster (Roselli et al. 2017). Angelicin can latter experience hydroxylation and methylation, similarly to psoralen, but the molecular elucidation of this part of the pathway is still to be done.
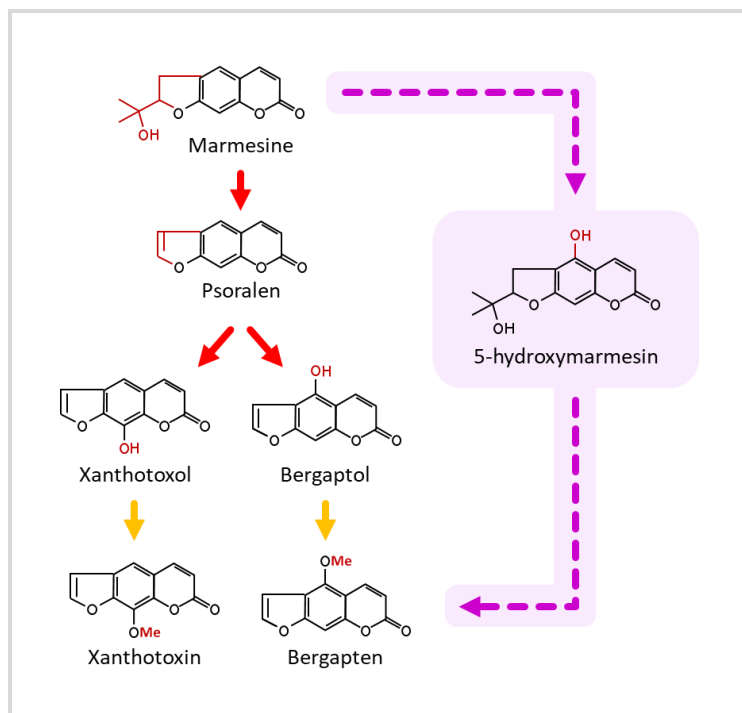
Figure 13 **Alternative conversion of marmesin into bergapten (purple), proposed in *Ficus carica*.**

Lastly, it should be noted that the steps described here are only the main steps of the pathway: they include all the reactions that are molecularly characterised, but other alternative reactions might exist in various plants. For instance, in the fig tree, some feeding experiments with radiolabelled precursors tended to confirm these main steps, but they also revealed potential alternative steps. Indeed, in *F. carica*, marmesin might be somehow converted into 5-hydroxymarmesin and then into bergaptol, without psoralen taking place in this synthesis (**Figure 13**) (Caporale et al. 1970, 1981; Dall'Acqua et al. 1979). However, this potential alternative pathway has never been confirmed.

E.2. Evolutionary perspectives and concluding remarks

Furanocoumarins are distributed in distant plant families. Yet, the furanocoumarin biosynthesis pathway seems to be very similar in all producing plants studied so far, which raises the question of the emergence of this pathway. Two alternative hypotheses have been proposed to explain the presence of a similar pathway in distant plant taxa: on the one hand, the pathway may have emerged only once, but it would have then been massively lost in many plant taxa. On the other hand, a similar pathway may have independently emerged several times in different plant taxa (Munakata et al. 2020). The molecular elucidation of the furanocoumarin pathway provided some clues in favour of a multiple origin – even though new data would be welcome to support and strengthen this hypothesis.

The most complete data in favour of this hypothesis results from the work done on the UDTs that open the way to linear or angular furanocoumarins: a gene-family phylogenetic analysis have been conducted on the UDTs from Apiaceae (PcPT1, PsPT1, PsPT2) and Moraceae (FcPT1), and the results strongly suggest that the moraceous UDT evolved from a different ancestor than the three apiaceous UDTs. The UDT activity must thus have independently emerged in Moraceae and Apiaceae (Munakata et al. 2020).

In addition, the hydroxylation of xanthotoxin is catalysed by P450s from the CYP71 family in Apiaceae, but by a P450 from the CYP82 family in Rutaceae. Even though no precise gene-family phylogenetic analyses were conducted on these enzymes, this information also suggests an independent emergence of the enzymes in the Apiaceae and the Rutaceae families.

To conclude, if many genes of the pathway have been characterised so far, some are still to be discovered. In particular, finding a gene coding for a marmesin synthase would be required to complete the set of enzymes that allow the conversion of *p*-coumaroyl CoA (a ubiquitous plant metabolite) into psoralen, the first toxic linear furanocoumarin.

In *Ammi majus*, the marmesin synthase may be a P450. Many other steps of the furanocoumarin pathway are also catalysed by P450s. Therefore, to better understand the furanocoumarin production at a molecular level, it is important to investigate the enzymes involved in this production – and the P450s are definitely a fascinating enzyme superfamily to focus on.

II.

ENDLESS P450S MOST BEAUTIFUL

~

History, nature, function and evolution of an essential enzyme superfamily

A. INTRODUCTION: ALL YOU NEED IS A P450

Cytochrome P450s (or P450s) are a ubiquitous superfamily of heme-binding enzymes that exists in (almost) all organisms, including bacteria and eukaryotes as well as viruses and archeae (Nelson 2018; Lamb et al. 2019). P450s catalyse various reactions such as oxidations, dehydratations, C-C cleavages and decarboxylations. The "classical" function of a P450 is a monooxygenation that consists in the reduction of dioxygen into water and the insertion of one of its atoms of oxygen into an organic substrate. This reaction requires a transfer of electrons that can be supported by various redox partner systems (Werck-Reichhart and Feyereisen 2000; Werck-Reichhart et al. 2002; Mclean et al. 2011); it can be written as follow: $RH + O_2 + 2e^- + 2H^+ \rightarrow ROH + H_2O$.

The term of P450 stands for **P**igment absorbing at **450**nm, which comes from their characteristic absorption spectral features in presence of carbon monoxide. Indeed, instead of binding O_2, the heme of P450s can interact with CO with a high affinity, inhibiting the enzymatic activity. This results in a shift in their absorption spectrum, with a peak of maximum absorbance at 450 nm. The binding of CO can also be reversed by light, with a maximum efficiency at 450 nm (Omura and Sato 1964a; Werck-Reichhart et al. 2002; Mclean et al. 2011; Poulos and Johnson 2015).

Different P450s have variable features such as the nature of their electron transfer partner, their subcellular location (soluble or membrane-bounded), the nature of the reaction, the nature of their substrates (from small molecules to big biopolymers) and their specificity (substrate specificity, regio- and stereoselectivity) (Werck-Reichhart and Feyereisen 2000; Werck-Reichhart et al. 2002; Rupasinghe et al. 2003; Poulos and Johnson 2015; Yonekura-Sakakibara et al. 2019). A "typical" plant P450 catalyses a monooxygenation by using atmospheric oxygen and electrons transferred from NADPH via a cytochrome P450 reductase. They are usually anchored in the endoplasmic reticulum membrane, with their catalytic domain in the cytosol. Their molecular weight generally ranges between 45 and 60 kDa, for a coding sequence of about 1500 bp. But of course, there are exceptions; for instance, some chloroplast-localised plant P450s are soluble (Werck-Reichhart and Feyereisen 2000; Baudry et al. 2006; Mizutani and Ohta 2010; Schuler and Rupasinghe 2011; Nelson and Werck-Reichhart 2011).

From a biochemical point of view, the precise incorporation of oxygen into specific molecules is essential to many biosynthetic and catabolic activities. The ability of P450s to catalyse regio- and stereoselective oxygenations in a wide range of substrates make them very interesting and useful enzymes. Therefore, it is not surprising to find out they are involved in various primary and specialised metabolic processes, and also highly prised in biotechnology (Mclean et al. 2011; Ilc et al. 2018).

B. A BRIEF HISTORY OF P450S: RESEARCHES AND DISCOVERIES OVER TIME

The study of P450s is an old story that started in 1955, when Brodie *et al.* described for the first time the hydroxylation of xenobiotic compounds catalysed by some enzyme systems in rabbit liver microsomes (Brodie et al. 1955; Plettner 2018). The P450 characteristic CO absorption spectrum at 450 nm was reported by Klingenberg three years later (Klingenberg 1958), and it was in 1964 that Omura and Sato suggested that P450s were membrane-bound hemoproteins (Omura and Sato 1964a, b); they also introduced the term of "cytochrome P-450". A few years later, some bacterial P450s and there redox partners were reported, chemically characterised, and first amino acid analyses were performed (Dus et al. 1970; Hare and Fulco 1975). Yet, it was in 1979 that Botelho *et al.* determined the first partial amino acid sequence of a P450 from rat liver, using chemical methods (Botelho et al. 1979). Three years later, the first known nucleotide coding sequences of P450s were published. They were the sequences of CYP2B1 and CYP2B2, two phenobarbital-inducible P450s from rat liver (Waxman and Walsh 1982; Fujii-Kuriyama et al. 1982). In plants, the first known P450 was CYP71A1: it was cloned from avocado in 1990, heterologously expressed in yeast, and associated with monoterpenoids metabolism (Bozak et al. 1990, 1992; Nelson and Werck-Reichhart 2011).

Over the years, the interest and fascination for P450s have been endlessly growing and the technological progresses allowed for the discovery and characterisation of more and more genes. In March 2010, a search of the term "cytochrome P450" on PubMed resulted in almost 64,000 references (Nelson 2011). Exactly 10 years later, in 2020, a similar search resulted in about 100,000 references. In the same time, the number of known P450s has been considerably increased: in 2010, a search of "cytochrome P450" on GenBank gave about 36,000 results in the nucleotide section (Nelson 2011). In 2020, it gave almost 500,000 results in the nucleotide section, more than 100,000 results in the gene section and about 855,000 results in the protein section. A significant part of these P450s were from plants (**Table 1**). These increasing numbers reflect all the work that have been done to understand the many roles of this enzyme superfamily, but even now, the great majority of P450s have still not been functionally characterised.

Table 1 Number of results for the search of "cytochrome P450" on GenBank (March 2020).

Section	Number of results	
	All organisms	Plants only
Nucleotide	499,498	84,151
Gene	104,695	35,364
Protein	855,432	88,378

Nevertheless, the number of P450s found in public databases is an underestimation of the number of P450s that are actually known, because many sequences are treated as confidential until published. Therefore, Pr. David Nelson, who contributed to develop the P450 nomenclature and is responsible for assigning names to all P450s, possesses a substantially larger collection: in 2018, he reported that nearly 350,000 P450s had been identified. Most of them were from plants (>184,000), fungi (>85,000), bacteria (>62,000) and animals (>14,000), while less than 1,000 were from protozoa, archaea and viruses together. Most of these P450s were still unnamed: as an example, in plants, only 16,000 out of the 184,000 sequences had a name (Nelson 2018). However, to stick to the example of plants, this number of 184,000 P450s was already way larger in 2018 than what can be found in public databases in 2020. Last but not least, Nelson expected to reach one million P450 sequences by 2020 (Nelson 2018).

C. ONE CLASSIFICATION TO NAME THEM ALL

In order to compare P450 genes belonging to different species and/or identified in different studies, it has been primordial to establish a standardised nomenclature. A classification and nomenclature system has therefore been designed to reflect evolutionary relationships between P450 sequences.

The names of new P450s are assigned by the P450 Nomenclature Committee, based on homology and phylogenetic criteria (Nelson et al. 1996, 2004; Nelson 1999). These names start by "CYP" (or "*Cyp*" for mouse and Drosophila), a root symbol representing "cytochrome P450". It is followed by an Arabic number indicating the family, a letter denoting the subfamily, and another number designating the given gene in its subfamily (Nelson et al. 1996). As an example, CYP73A1 (*Helianthus tuberosus*) and CYP73A5 (*Arabidopsis thaliana*) are two P450s that belong to the CYP73 family, subfamily A. CYP73A1 was the first CYP73A to be named, CYP73A5 was the fifth. It should be noted that these standardised names do not take the function of the P450 into account: for instance, CYP73A1 and CYP73A5 are two orthologous cinnamate hydroxylases (Nelson et al. 1996; Nelson and Werck-Reichhart 2011).

According to the original definitions, the P450s sharing at least 40% amino acid identity are grouped into the same family, while more than 55% identity places these proteins into the same subfamily. These definitions of (sub)family were originally arbitrary, but quite surprisingly, they turned out to be very handy and useful. (Nelson et al. 1996; Nelson 2018). However, as more and more P450s have been added, the "40% boundary" have been stretched: when a new P450 is discovered and added to the phylogenetic trees, it sometimes falls between two adjacent but clearly separated families, is obviously closest to one family than the other, but happens to share just a bit less than the required 40% amino acid identity. In that case, instead of creating a new family for this unique sequence, this new P450 is included in the closest existing family, which "40% boundary" has to be slightly lowered. This phenomenon is called family creep. Because of it, large families can sometimes merge as sequences are being absorbed into an adjacent large family. When families merge, the "small" family can be renamed as a subfamily of the large family that absorbed it. Yet, for historical and practical reasons, it can be better to keep the original name and make a footnote describing the relationship between the two merged families. As an example, it happened to the CYP89 family that merged into the CYP71 family, but kept its original name (Nelson et al. 2008; Nelson and Werck-Reichhart 2011).

Furthermore, as more and more P450 sequences were identified and named, an additional higher level of classification has been created to show the relationship between the families: the clans (Nelson 1998, 1999; Nelson et al. 2004). A clan is composed of related families that derive from a single gene ancestor which is recent enough to see the families cluster together on a phylogenetic tree. Clans are therefore "the deepest branching clade of P450s followed by CYP families" (Nelson 2018). Plant clans are named by a representative family they contain (Nelson 1999).

In summary, the concepts of clans, families and subfamilies mimic the clades in which P450 genes evolved from a common ancestral sequence. The name of a P450 reflects its family and subfamily, but neither its clan, nor its function, nor its orthology.

D. IF YOU PLEASE DRAW ME A FUNCTIONAL P450: FROM STRUCTURE TO ACTIVITY

P450s constitute a superfamily of enzymes which amino acid sequence is highly variable between families and clans. So, what *are* P450s? What does they look like and what are the characteristics that allow them to catalyse similar reactions despite a poorly conserved primary structure?

D.1. P450s: from the primary to the secondary and tertiary structures

D.1.a. Discovery of the P450 fold

If prokaryotic P450s are soluble, most eukaryotic P450s are membrane-bound-proteins anchored in the endoplasmic reticulum or the mitochondrial inner membranes. Because of this difference, it has been more difficult to study the structures of eukaryotic P450s. The first P450 crystal structure was determined in 1985: it was the structure of CYP101A1, the extensively studied P450$_{cam}$ from *Pseudomonas putida* (Poulos et al. 1985, 1987). The second crystal structure was from the bacterial CYP102A1 (Ravichandran et al. 1993). These two pioneer studies largely contributed to the structural elucidation of P450s, but it was only in 2000 that the first crystal structure of a membrane-bound eukaryotic P450 was resolved. To successfully crystallise CYP2C5, the rabbit progesterone hydroxylase, the membrane anchor region of the protein was removed, which gave a truncated but soluble P450 (Williams et al. 2000). Since then, various crystal structures of both prokaryotic and eukaryotic P450s have been published. To date, it seems that only four plant P450s have been crystallised. The first ones were two allene oxide synthases: CYP74A1 from *A. thaliana,* and CYP74A2, its ortholog from *Parthenium argentatum* (Lee et al. 2008; Li et al. 2008b). However, these P450s are very atypical and they will not be discussed here in a first time. So, the only crystal structures of "typical plant P450s" are the structures of the third and fourth ones: CYP76AH1 (*Salvia miltiorrhiza*) and CYP90B1 (*A. thaliana*). Their recent crystallisations have been done, respectively, by Gu *et al.* (2019) and Fujiyama *et al*. (2019), who also had to engineer these enzymes to make them soluble.

The comparison of the various P450 crystal structures shows that, despite a high diversity in their primary structures, P450s share a highly conserved secondary and tertiary structure. For instance, even with an amino-acid sequence identity as low as 13%, P450s harbour a similar structural fold that remained the same throughout evolution. Moreover, enough crystal structures are now available to confidently state that all P450s share a conserved overall architecture – a typical "P450 fold" – which seems to be unique and only found in P450s (Werck-Reichhart and Feyereisen 2000; Rupasinghe et al. 2003; Mclean et al. 2011; Schuler and Rupasinghe 2011; Poulos and Johnson 2015).

D.1.b. The "P450 fold": overall architecture

Many articles describe and review the structure of P450s (Werck-Reichhart and Feyereisen 2000; Werck-Reichhart et al. 2002; Rupasinghe et al. 2003; Baudry et al. 2006; Mclean et al. 2011; Schuler and Rupasinghe 2011; Poulos and Johnson 2015; Fujiyama et al. 2019). The findings of these authors can be synthesised as follow: P450s are heme-thiolate proteins that fold in a triangular prism-shaped configuration. The structural core of a typical P450 is made of two domains: the α-domain, that contains many α-helices, and the β-domains, that contains a substantial number of β-sheets. A heme
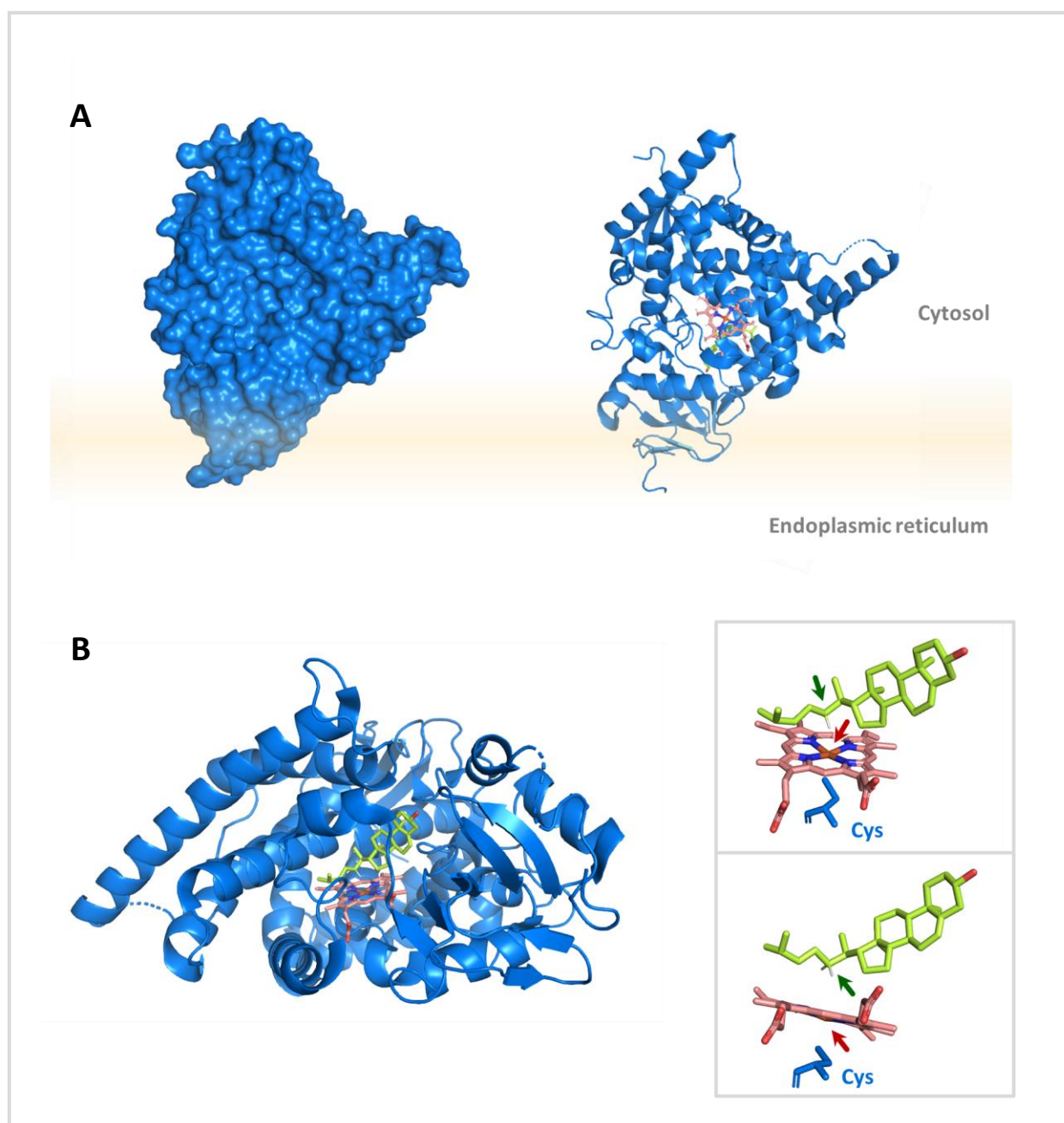
Figure 14 Overall architecture and catalytic centre of a typical membrane-bound plant P450.

CYP90B1 is one of the two typical plant P450s whose crystal structure is known. It has been described by Fujiyama *et al.* (2019) and its structure, complexed with cholesterol, is available on the Protein Data Bank (PDB) under accession code 6A15. Note that CYP90B1 has been engineered prior to crystallisation. **A** – Triangular prism-shaped configuration of CYP90B1, showed as "surface" or in a "cartoon" style representing the α-helices and β-domains. The P450 is in blue, it contains a pink heme, a green substrate (cholesterol), and it is anchored in the endoplasmic reticulum membrane, whose probable location is in yellow. The orientation of the P450 in regard to the membrane have been deduced from those reported in non-plant P450s in Werck-Reichhart and Feyereisen (2000), Werck-Reichhart *et al.* (2000) and Williams *et al.* (2000). **B** – Focus on the catalytic centre. CYP90B1 have been reoriented and zooms have been done to better show the catalytic centre, that consists in the heme iron (red arrow) sandwiched between the conserved cysteine residue (blue Cys) and the substrate. In the substrate, the carbon that will be hydroxylated is positioned just above the iron (green arrow).

cofactor is sandwiched between these two domains, in a relatively hydrophobic cavity. P450s all share a common catalytic centre, buried in the protein, that includes the heme and a cysteine residue which is highly conserved in every P450s. The substrate-binding pocket of a typical P450 is located near the heme, on the other side of the conserved cysteine; it is accessible from the exterior through a wide channel. Compared to soluble P450s, membrane-bound P450s possess a N-terminal signal sequence that is anchored in the membrane and extends across it. From now on, this study will focus on typical plant P450s anchored in the endoplasmic reticulum (**Figure 14**).

Despite a highly conserved architecture, various structural features differ from one P450 to another: usually, the most conserved features are related to the necessary heme binding and the common catalytic properties of P450s. On the contrary, the most variable regions control substrate specificity.

D.1.c. Structurally conserved regions

The structurally conserved regions of a typical P450 are summarised in **Figure 15**.

A typical P450 contains α-helices named α-A to α-L and β-domains named β-1 to β-5; these regions might vary a lot in their amino acid sequence but their secondary structures are well conserved (Gotoh 1992; Werck-Reichhart and Feyereisen 2000; Rupasinghe et al. 2003). Among these helices, α-D, α-E, α-I, α-L, α-J and α-K form a structural core involved in the heme binding. The helices α-I and α-L make contact with the heme: in the α-I helix, the (A/G)Gx(D/E)T(T/S) amino acid sequence is well conserved and mediates dioxygen activation during the reaction. On the α-K helix, other amino acids form the ExxR sequence, located on the proximal face of the heme. Another well conserved sequence is the "PERF" consensus. The E and the R from the ExxR sequence, together with the R of the "PERF" consensus, form the E-R-R triad: this motif is assumed to be involved in locking the heme into position, and to stabilise the conserved core structure (Atkins and Sligar 1988; Aikens and Sligar 1994; Hasemann et al. 1995; Werck-Reichhart et al. 2002; Rupasinghe et al. 2003; Mestres 2004; Mclean et al. 2011).

A second structurally conserved region is the β-bulge section, known as the Cys-pocket, which refers to the absolutely conserved cysteine it contains. This cysteine is bound to the heme and plays an important role during the reaction: indeed, the thiolate of the conserved cysteine is coordinated with the iron contained in the heme, serving as its fifth ligand. The amino acids that surround the cysteine are also well conserved and form a rigid structure that holds the
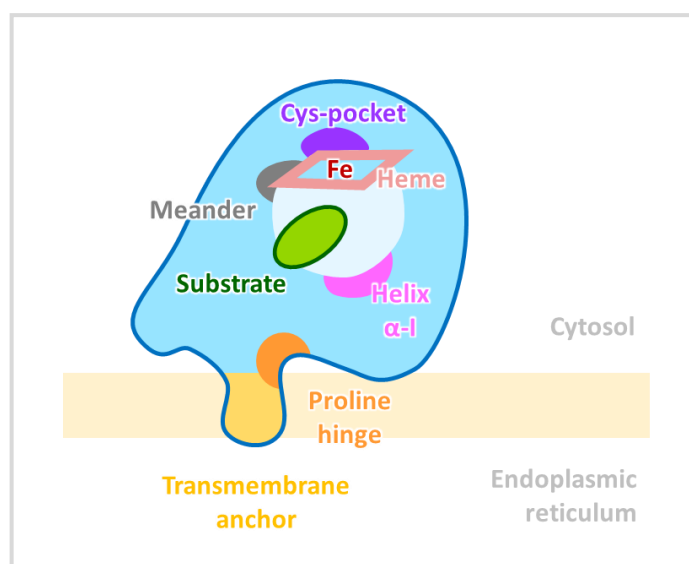
Figure 15 Structurally conserved region of a typical plant P450 anchored in the endoplasmic reticulum. The P450 is in blue, its conserved regions are highlighted in various colours, the substrate is in green, and the membrane in pale yellow. Adapted from Werck-Reichhart *et al.* (2000).

cysteine into place, allowing it to accept hydrogen bonds during the reaction. The cysteine and the amino acids that surround it form the FxxGxRxCxG motif, located on the proximal face of the heme (Atkins and Sligar 1988; Hasemann et al. 1995; Werck-Reichhart et al. 2002; Rupasinghe et al. 2003; Mestres 2004; Mclean et al. 2011).

The "meander loop" is another conserved region located between the α-K helix (structural core) and the Cys-pocket. This loop is assumed to be involved in heme binding and to stabilise the tertiary structure of P450s (Hasemann et al. 1995; Rupasinghe et al. 2003; Mestres 2004; Sirim et al. 2010).

The last region that is well conserved in membrane-bound P450s is the proline hinge: located above the transmembrane anchor, this proline rich region defines a hinge that places the enzyme against the membrane. Also linked to this membrane-bound characteristic, some hydrophobic residues are associated with the cytosolic side of the membrane (Werck-Reichhart et al. 2000; Baudry et al. 2006).

Finally, it should be emphasised that, except for the proline hinge, the conserved domains are mainly located in the C-terminal half of the primary sequence. Among these domains, the only amino acids conserved in all plant P450s are the E-R-R triad and the heme-binding cysteine (Gotoh 1992; Werck-Reichhart et al. 2002; Rupasinghe and Schuler 2006; Schuler and Rupasinghe 2011).

D.1.d. Substrate Recognition Sites

Buried in the conserved core structure of the P450s, six small regions control the substrate access, recognition, positioning and binding in the active site. These regions, called Substrate Recognition Sites (SRSs) and numbered from SRS1 to SRS6, are thus essential to regulate the substrate specificity.

D.1.d.1. Discovery and description of the SRSs

The SRSs have been first identified and defined by the seminal work of Gotoh (1992), who used Xray crystallography and the group-to-group alignment of many P450 sequences to infer their position. His results have later been largely confirmed by mutagenesis approaches (Schoch et al. 2003; Rupasinghe and Schuler 2006; Schuler and Rupasinghe 2011). SRSs are small regions that represent around 16% of all amino acids and are dispersed along the primary sequence (**Figure 16**). If they vary in both sequence and structure, their position in the primary structure is well conserved. So, as the sequences that surround them are well conserved, the SRSs of a given P450 can easily be identified by alignment with another P450 whose SRSs are known. In the typical P450 fold, SRSs are positioned in or near the catalytic site (Gotoh 1992; Sirim et al. 2010; Schuler and Rupasinghe 2011).
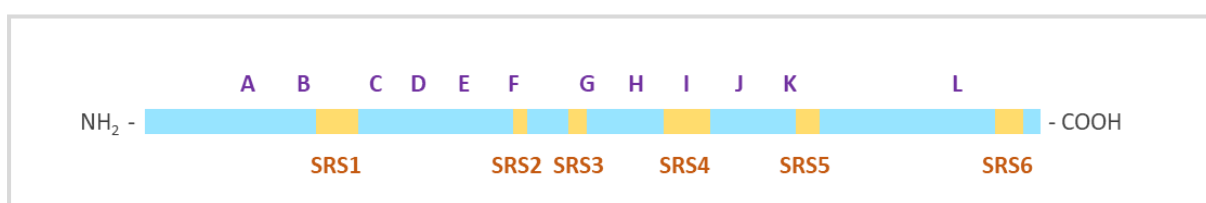
Figure 16 Schematic location of the SRSs in the primary structure of a typical plant P450. The SRSs are in yellow/orange, the purple letters indicate the location of the α-helices.

SRS1 corresponds to the loop located between the helices α-B and α-C (also called the BC-loop – **Figure 16**). It is positioned over the heme and, despite being highly variable, it always houses one residue that points toward the heme and that remains rigid during the substrate binding. Some of its residues are also predicted to contact the distal end of the substrate. SRS2 starts at the C-terminal end of the α-F helix, and extends into the FG-loop. It is closely followed by SRS3, that starts in the FG-loop and extends into the α-G helix. SRS2 and SRS3 contribute to form the access channel for the substrate to the active site. Some of their residues are also predicted to contact the distal end of the substrate. SRS4 is located in the middle of the α-I helix that extends over the heme and is involved in dioxygen activation. In the case of aromatic substrates, some of the residues in the C-terminal portion of SRS4 are predicted to contact the aromatic ring of the substrate. On the contrary, other residues from the N-terminal portion of SRS4 are predicted to contact the distal end of the substrate. It has also been reported that the contact between the substrate and SRS4 is more extended with larger than smaller substrates. SRS5 is located in the loop between the α-K helix and the β-1 sheet, strand 4. SRS6 is housed in the loop between the two strands of the β-4 sheet. SRS5 and SRS6 both protrude into the catalytic site and some of their residues are predicted to contact the aromatic ring of aromatic substrates (Rupasinghe et al. 2003; Larbat et al. 2007; Sirim et al. 2010; Schuler and Rupasinghe 2011; Dueholm et al. 2015).

D.1.d.2. SRSs, a hot spot for site-directed mutagenesis

The SRSs are highly variable and their amino acid sequences are more diversified than the rest of the P450 primary sequence; the most variable ones being SRS1-3 and SRS6. This fact is consistent with the idea that the SRSs are involved in the substrate specificity: their variations contribute to the diversification of the reactions catalysed by P450s.

As a consequence, many mutations affecting the SRSs significantly impact the substrate specificity, and the SRS have thus become a hotspot for site-directed mutagenesis. Indeed, in many studies, site-directed mutagenesis has been performed on a few residues from the SRSs of various P450s, and resulted in changes in the activity, the substrate range, specificity, regio- and stereoselectivity of the enzyme – even with single amino acid substitutions. A quite common approach consists in modelling a P450 and performing docking experiments to identify key residues that might change its activity. The models are subsequently tested by site-directed mutagenesis and might result in new enzymatic activities and performances, which is of high interest for biotechnology applications and the production of new bioactive compounds. However, even though less common, a few mutations outside the SRSs have sometimes been reported to also affect the activity and specificity of P450s (Gotoh 1992; Urlacher 2006; Baudry et al. 2006; Rupasinghe and Schuler 2006; Li et al. 2008a; Nelson et al. 2008; Seifert et al. 2009; Sirim et al. 2010; Schuler and Rupasinghe 2011; Uno et al. 2011; Vazquez-Albacete et al. 2017; Kuhlman and Bradley 2019).

> In brief, P450s have highly diverse amino acid sequences but fold into a highly conserved architecture. The regions that are responsible for substrate specificity are highly variable, allowing P450s to metabolise various compounds. On the contrary, the regions responsible for P450s common features are structurally conserved. P450s conserved regions include the catalytic centre, which allows to maintain a similar activity, but to perform it on various substrates.

D.2. Catalytic activity of P450s

As mentioned in the introduction, a typical P450 catalyses the addition of an atom of oxygen into a substrate, according to the following equation: **$RH + O_2 + 2\ e^- + 2H^+ \rightarrow ROH + H_2O$**.

D.2.a. P450 redox partner systems

The reduction of the molecular O_2 requires two electrons that usually come from reduced cofactors such as NAD(P)H. However, if the NAD(P)H cofactors are two electrons donors, the P450s need the electrons one after the other. P450s therefore require the help of redox partner systems to successively delivers two individual electrons to their active site (Werck-Reichhart and Feyereisen 2000; Werck-Reichhart et al. 2002; Mclean et al. 2011; Plettner 2018).

Depending on their redox partner, P450s can be categorised into four classes, numbered I to IV. Extensive description of these four classes can be found in Werck-Reichhart and Feyereisen (2000), Werck-Reichhart *et al.* (2002), Rupasinghe *et al.* (2003) and Mclean *et al.* (2011). As "typical" plant P450s are class II P450s, they are the only ones that will be mentioned here.

Briefly, class II P450s are present in both eukaryotes (anchored in the endoplasmic reticulum) and prokaryotes (soluble). Their redox partner system is a FAD/FMN-containing P450 reductase. In eukaryote, this reductase can be a NADPH-dependent P450 reductase (**Figure 17**) or a NADH-dependent cytochrome b5 reductase associated to a cytochrome b5. In both cases, two electrons are removed from NAD(P)H to reduce the FAD cofactor, which successively conveys them to the catalytic site of the P450 (Werck-Reichhart and Feyereisen 2000; Rupasinghe et al. 2003; Mclean et al. 2011).
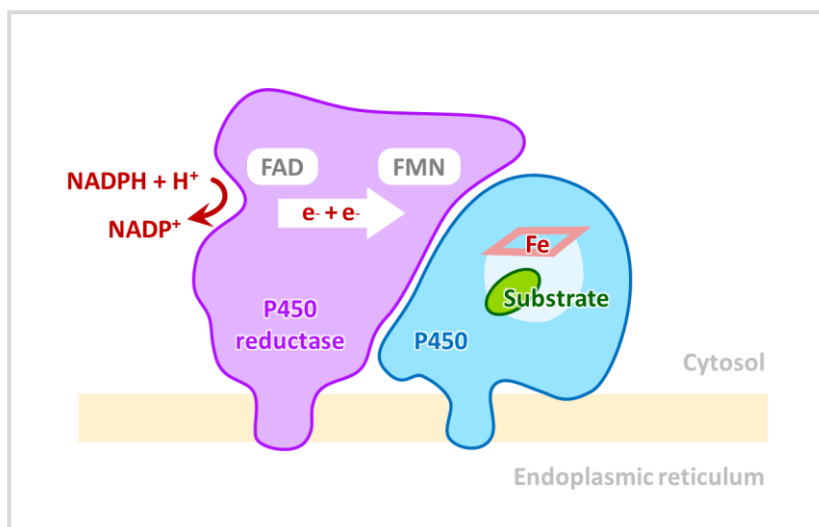
Figure 17 **Typical class II plant P450 and its redox partner.** Adapted from Werck-Reichhart *et al.* (2000).

D.2.b. Catalytic cycle of a typical class II P450

As a reminder, the catalytic centre of all P450s consists in a heme and the absolutely conserved cysteine. The catalytic cycle of a typical P450-mediated hydroxylation is illustrated in **Figure 18**.

The cycle starts with a P450 in its substrate-free form: the iron of the heme, linked to a water molecule, is ferric, six-ligated and in a low-spin state. When a substrate (RH) enters the active site of the P450, it

displaces the sixth water ligand, which shifts the spin of the iron to a high-spin form and results in a higher redox potential. Then, the redox partner delivers a single electron that reduces the heme iron and allows the binding of a dioxygen molecule. This forms a superoxide complex that is subsequently reduced by a second electron delivered by the redox partner. Then, two protons are successively added to the distal oxygen atom, which leads to the heterolysis of the O-O bond: one oxygen atom is released as a water molecule while the other oxygen is still linked to the iron, forming a reactive and transient ferryl-oxo complex. This activation of the oxygen is probably the most important step of the cycle. Then, the iron-bound activated oxygen atom is transferred to the substrate molecule, forming a hydroxylated product (ROH). Finally, the product is released and the P450 comes back to its initial state (Werck-Reichhart and Feyereisen 2000; Zangar et al. 2004; Mclean et al. 2011).
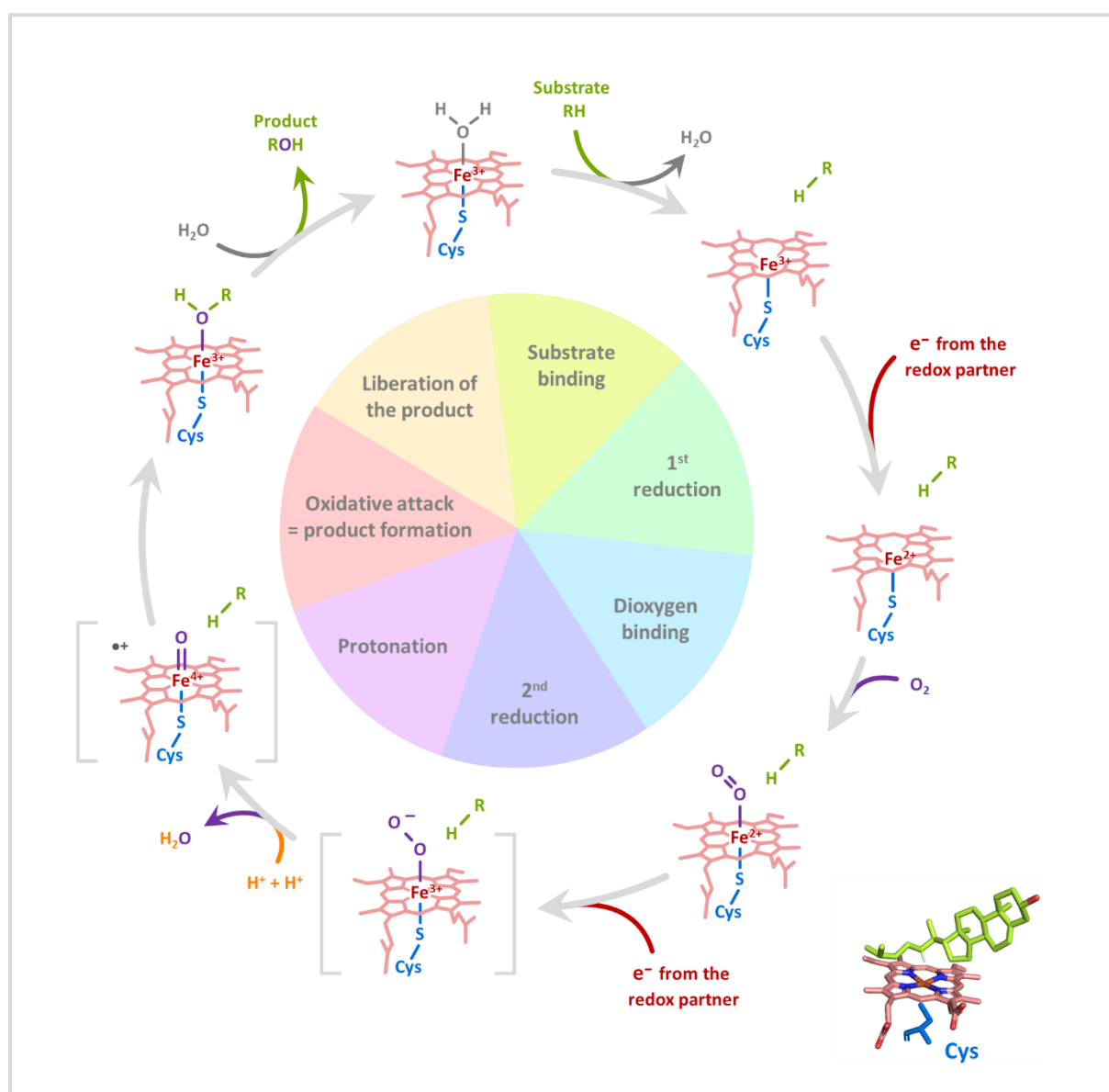
Figure 18 Catalytic cycle of a typical P450-mediated hydroxylation, drawn according to (Werck-Reichhart and Feyereisen 2000; Zangar et al. 2004; Mclean et al. 2011). The conserved cysteine is in blue, its thiolate is designated by "S", the heme is in pink and the substrate in green. The catalytic center of CYP90B1, from **Figure 14** have been added as a reminder.

Finally, it should be noted that this catalytic cycle presents alternative pathways such as the "peroxide shunt": some of these alternatives have been extensively described but, as they are beyond the scope of this project, they will not be presented here (Zangar et al. 2004; Mclean et al. 2011).

D.3. Atypical P450s and unusual P450-mediated reactions

As mentioned earlier, P450s catalyse a large variety of monooxygenations/hydroxylations. However, if these reactions are the most common, some unusual reactions also have to be mentioned. The most atypical plant P450s are probably from the CYP74 family, that contains the allene oxide synthase. As CYP74A1 and CYP74A2 have been crystallised (Lee et al. 2008; Li et al. 2008b), the CYP74 family is well known. Briefly, CYP74s seem to exist in all plant species, but only in a low number of enzymes. They are localised in the chloroplasts and catalyse an atypical reaction that do not require the activation of a molecular dioxygen: instead, they activate a hydroperoxide that is directly provided by their substrate, which short-circuits the standard catalytic cycle. As should be expected from this atypical reaction, the primary sequence of CYP74s is quite different from the one of typical P450s: for instance, the structures involved in oxygen activation and electron transfer are altered (Werck-Reichhart et al. 2002; Nelson et al. 2008; Schuler and Rupasinghe 2011).

Even if they are not as atypical as the CYP74 family members, many other plant P450s catalyse unusual reactions that are involved in various specialised metabolic pathways. In their review, Mizutani and Sato (2011) provide a nice summary of these reactions and describe their proposed mechanisms. Among them, we can quickly list the formation of methylenedioxy-bridges (CYP719A, CYP80Q1), intramolecular C-O or C-C phenol coupling (CYP80A, CYP80G, CYP719B), N-oxidation (CYP79), sterol desaturation (CYP710A), rearrangement of the carbon skeleton (CYP93C2, CYP88A), C-C bond cleavage (CYP72A1), Baeyer-Villiger oxidation (CYP85A2), or the hydroxylation followed by intramolecular cyclisation performed by the menthofuran synthase (CYP71A32 – **Figure 19**) (Mizutani and Sato 2011).
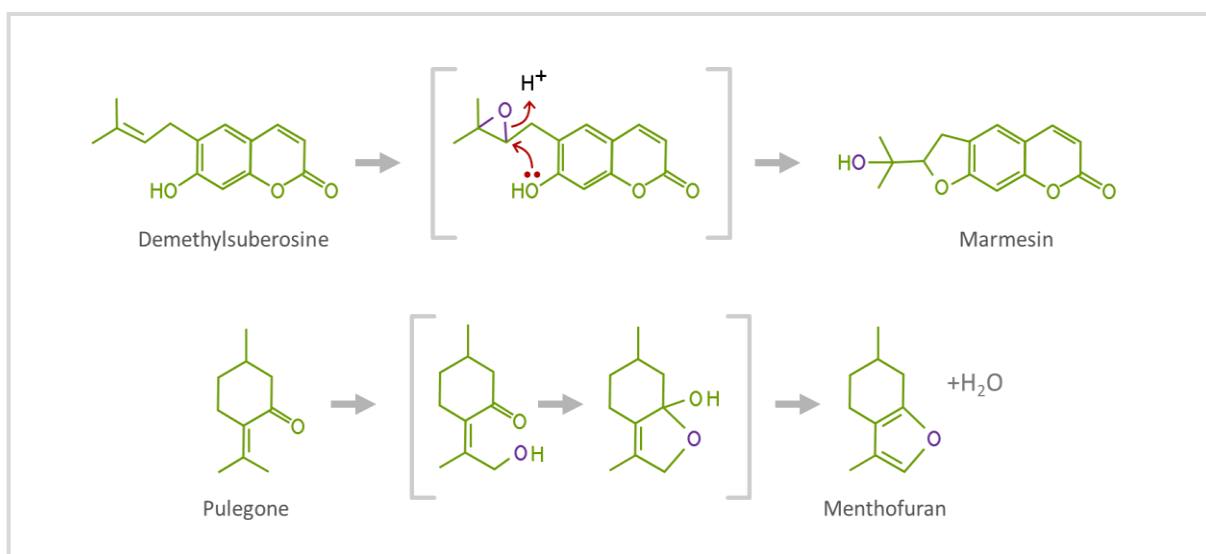
Figure 19 Proposed mechanisms for two oxidations followed by cyclisation. Hypothetical P450-mediated conversion of the DMS into marmesin and conversion of pulegone into menthofuran by CYP71A32. Adapted from Mizutani and Sato (2011).

In the furanocoumarin biosynthesis pathway, the synthesis of marmesin, psoralen and angelicin are described as unusual reactions. Indeed, the conversion of marmesin into psoralen (CYP71AJ3), as well at its angular equivalent, the conversion of columbianetin into angelicin (CYP71AJ4), are C-C bond cleavages that both release an acetone molecule. The hypothetical P450-mediated conversion of DMS into marmesin would be an oxidation followed by a cyclisation. However, the mechanism that have been proposed for this reaction is quite different from the mechanisms of the menthofuran synthesis, which also consists in an oxidation followed by cyclisation (**Figure 19**) (Mizutani and Sato 2011).

In summary, most P450s share a common catalytic mechanism that allows them to perform hydroxylations on various substrates. However, this typical activity is not a strict rule and many unusual reactions have been described, which contributes to increase the range of compounds that can be metabolised and formed by P450s.

E. HIGHWAY TO PHYTOCHEMISTRY: P450S' IMPORTANCE IN THE PLANT KINGDOM

Now that we have seen what P450s are and how they work, let us get interested in their repartition and importance across the tree of life, and especially in the plant kingdom.

E.1. In plants, P450s constitute one of the largest gene superfamily

As mentioned previously, the majority of P450s that have been identified and annotated are from plants: actually, this is not surprising given that plants contain more P450s than any other taxon.

In yeasts, only 2 and 3 P450 sequences have been found in the genomes of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*. The genomes of various animals such as humans, flies, sea squirts, or nematodes contain about 50 to 100 P450 genes, which represents 0.1 to 0.5% of their total genes. With 39 P450 genes found in the genome of *Chlamydomonas reinhardtii,* green algae also seem to contains a relatively small amount of P450s. However, this gene number increases to 71 in *Physcomitrella patens* (moss) and to 225 in *Selaginella moellendorffii* (lycophyte), while the genome of a typical angiosperm is expected to contain about 300 P450 genes. For instance, 245 P450 genes have been found in *Arabidopsis thaliana*, 270 in *Vitis vinifera*, 272 in *Solanum lycopersicum*, 309 in *Oryza sativa*, 312 in *Populus trichocarpa*, 332 in *Glycine max* and 372 in *Sorghum bicolor*. These numbers seem to vary quite a bit, depending on how the genes are counted and exclude putative pseudogenes. To date, the largest set of P450s annotated in a single genome has been found in *Saccharum spontaneum*, the sugarcane, that contains 394 P450 genes. All these sequences represent about 1% of each plant genes (1.3% in the case of sugarcane) – which is 5 to 10 times higher than in animals, and makes P450s the third largest gene superfamily found in plant genomes (Paquette et al. 2000; Werck-Reichhart et al. 2002; Schuler and Werck-Reichhart 2003; Nelson 2004, 2006, 2019; Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011; Mizutani 2012).

E.2. The diversification of P450s is at the heart of plant chemical diversity

Not only plants possess a huge number of P450 genes, but their P450s are also quite diversified. Indeed, in plants, a total of 127 P450 families have been identified so far, and a typical angiosperm is expected to possess around 50 P450 families. The diversification of plant P450s mirrors plant biochemical diversity: as the production of new chemicals requires new enzymes, and as oxygen is very useful to build complex molecules, P450s have been recruited in many metabolic networks, in which they contribute to produce new molecules (Nelson 2011; Nelson and Werck-Reichhart 2011). As a result, and even though most of them are still to be characterised, we already know that plant P450s are involved in the production of a large diversity of both primary and specialised metabolites, that are essential for plant growth, development, and adaptation to their environment (Pichersky and Gang 2000; Nelson et al. 2008; Nelson and Werck-Reichhart 2011; Mizutani 2012; Ilc et al. 2018).

Plant P450s can be grouped into 11 clans. The CYP51, CYP74, CYP97, CYP710, CYP711 and CYP746 clans are single-family clans, and most of them seem to fulfil conserved essential functions such as sterol biosynthesis (CYP51, CYP710), oxylipin biosynthesis (CYP74) or carotenoid biosynthesis (CYP97). On the contrary, the CYP71, CYP72, CYP85, CYP86 and CYP727 clans are multiple family clans involved in diverse functions – except maybe for the CYP727 clan that does not have any known function yet. The CYP86 clan is involved in the metabolism of fatty acids, fatty alcohols and alkanes. The CYP85 clan is mainly involved into isoprenoid metabolism. The CYP72 clan metabolises compounds such as fatty acids, isoprenoids, hormones and cytokinins. Last but not least, the CYP71 clan, which contains more than half of all plant P450s, is involved in the metabolism of a great variety of compounds ranging from phenylpropanoids, terpenoids, alkaloids, fatty acids and hormones (Nelson 2004; Nelson et al. 2008; Nelson and Werck-Reichhart 2011).

Finally, the numerous and highly-diversified P450s found in plants are considered to be one of the dominant driving forces that allowed plants to build a highly complex metabolic network, leading to a huge diversity of metabolites that conferred them survival advantages. This makes P450s the largest enzyme family that supports plant metabolism, and a good mirror of plant evolution (Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011; Bathe and Tissier 2019).

For this reason, rather than to simply describe their functions, it might be better to explain the diversification of P450s by telling their story from an evolutionary perspective.

F. ON THE ORIGIN OF P450 GENES: EVOLUTIVE STORY OF P450S IN LAND PLANTS

P450s constitute a very old genes superfamily: they can be found in all kingdoms of life and the first ancestral P450 gene must have existed before the prokaryote / eukaryote divergence (Nelson et al. 1996; Nelson 2011). The firsts primordial eukaryotic P450s might have been some CYP51s, quickly followed by CYP61s/CYP710s. Other primordial P450s might also have existed, but got extinct or evolved into new families that cannot be traced back to the earliest Eukaryotes. Over time, evolution

of this (these) primordial P450(s) gave rise to all current eukaryotic P450s (Nelson 2018). As evolution includes expansion but also gene loss, specific P450s have been gained in some lineages while others were sometimes lost, which now results in P450 compositions (number of genes from every clans, families and subfamilies) that can be highly variable between species (Uno et al. 2011; Nelson 2018). Given its early appearance, the CYP51 clan is expected to be shared by all Eukaryotes – with the exception of a few groups that lost it. The other clans might have appeared after the animal and fungi kingdoms separated, about 1 billion years ago (Nelson 2011, 2018).

F.1. Early emergence of all plant P450 clans

Depending of their emergence during evolution, the 11 P450 clans found in plants can be divided into three categories (**Figure 20**). Firstly, the CYP51, CYP97, CYP710 and CYP746 clans are ancient clans found in land plants as well as in green algae, that predate the transition from water to land. Secondly, the CYP72, CYP85 and CYP711 clans predate land plants, but not their families – which means that land plants and green algae both contain families belonging to these clans, but the families found in plants do not exist in algae, and vice-versa. As an example, the CYP711 clan contains the CYP743 and CYP744 families which are specific to green algae, but also the CYP711 family that only exists in land plants.
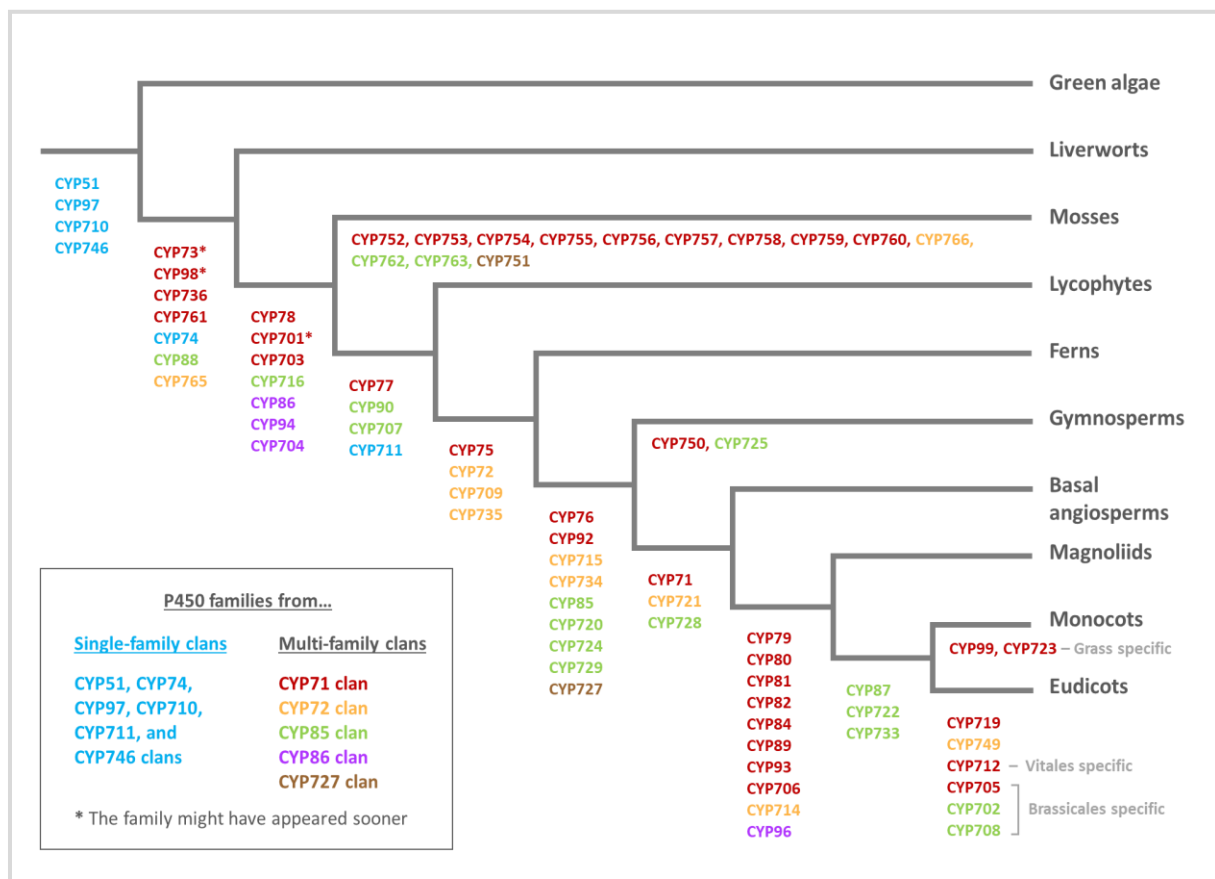
Figure 20 Emergence of plant P450 families, as it can be deduced from their currently known repartition in plants, based on Nelson and Werck-Reichhart, 2011. The families that merged into another are not all represented. As it can be difficult to demonstrate the loss of a P450 family in a given taxon, losses are not shown. Last, note that the CYP72, CYP85 and CYP711 clan also predate land plants but that green algae specific families are not represented.

Thirdly, the CYP71, CYP74, CYP86 and CYP727 clans are absent from green algae but are present in land plants, including moss. These clans must thus have emerged early after the transition from water to land. It should be noted that, if the CYP71, CYP86 and CYP727 clans are specific to land plants, the atypical CYP74 clan also exists in bacteria and animals: it is assumed that the presence of the CYP74 clan in these three kingdoms results from a lateral transfer rather than a common ancestral sequence (Nelson 2006, 2018; Lee et al. 2008; Nelson and Werck-Reichhart 2011).

In brief, all the clans emerged early during evolution: this means that, in P450s, the main sequence innovations already existed when mosses branched from vascular plants. Moreover, a comparison of the P450s from liverworts and angiosperms reveals a highly similar distribution of P450s among clans: thus, the composition of P450 clans has been set early during the emergence on land and it scarcely changed after it (Nelson 2006, 2018; Nelson et al. 2008; Nelson and Werck-Reichhart 2011).

F.2. Subsequent diversification of plant P450 families and subfamilies

If all P450 plant clans existed by the time moss evolved, the same cannot be said for the families and subfamilies that continued to multiply and are still evolving at a small-time scale level. In other words, after plants emerged on land, the sequence innovations have been primary concentrated at the family and subfamily level (Nelson and Werck-Reichhart 2011).

For instance, 31 P450 families have even been found in angiosperms only, suggesting they appeared after the angiosperm-gymnosperm divergence (this number does not include the families that are limited to one specific taxon). Among these 31 families, some appeared before the monocot-dicot divergence while others emerged after it. Some are even restricted to small taxa such as rosid, or are specific to some plant families. See **Figure 20** for more details about the emergence of P450 families over time (Nelson 2006, 2018; Nelson et al. 2008; Nelson and Werck-Reichhart 2011).

Subfamilies are even more diverse and less conserved than families. As an example, out of the 152 subfamilies found in rice and *A. thaliana*, only 31 were present in both species. The overlapping subfamilies include some highly conserved subfamilies involved in fundamental reactions, but also single gene subfamilies. The non-conserved subfamilies might have resulted from one or a small number of ancestral sequences present in the common ancestor of monocot and dicot, that later underwent duplication and divergence (Nelson 2004). Moreover, many subfamilies emerged or expanded specifically in restricted plant families, genus or even species. For instance, the CYP76 family exists in many plants but its CYP76C subfamily seems to be restricted to Brassicaceae, in which it might have appeared and diversified by repeated gene duplications (Höfer et al. 2014).

F.3. The diversification of P450s families: different functions, different patterns

The four ancient P450 clans that predate land plants are single-family clans. The CYP51G, CYP710A and CYP97 are highly conserved and often found in single or low copies in the genomes, which can be explained by a high purifying selection linked to their important functions. Indeed, these three clans seem to be devoted to the biosynthesis of essential compounds such as sterol (CYP51G, CYP710A) and xanthophyll (CYP97). The fourth ancient clan, CYP746, is found in moss and algae but it has been lost

in angiosperm and its function is still unknown (Nelson 2006, 2018; Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011; Mizutani 2012).

In the other clans, P450 families are more recent, diversified, and they are often linked to the adaptation of plants to their environment. On the one hand, a first diversification of P450s happened when plants started to colonise the terrestrial environment, giving rise to plant-specific P450 families that were crucial for the adaptation to life on land. On the other hand, some more recent diversifications, which were often plant-lineage-specific, gave rise to many new families and subfamilies of P450s that can be linked to the high chemical diversity that is now found in every plant lineage (Nelson and Werck-Reichhart 2011).

F.3.a. Plant-specific P450s and the conquest of land

When plants started to colonise land, they have been submitted to new constraints that can be divided into abiotic stresses (drought, UV radiations, gravity, oxidative stress, temperature fluctuations…) and biotic stresses (pests and pathogens). These new stresses have driven many evolutionary adaptations such as the establishment of water barrier, vascular system, structural support, long range signalling, defensive compounds, but also colour and fragrance to attract pollinators. All these innovations have required the production of new biomolecules and the establishment of novel biosynthesis pathways – in which P450s have often been recruited. Moreover, it is assumed that a significant expansion of P450s occurred when plants started to colonise the terrestrial environment, mirroring the need to produce these new compounds. In other words, a part of land plant evolution can be tied to the diversification of P450s; but the early evolutive story of plant P450s is also closely linked to the conquest of land (Paquette et al. 2003; Nelson 2006; Mizutani and Ohta 2010; Alber et al. 2019). Moreover, because of their crucial role to survive in this new environment, these P450s have been well conserved among all plant taxa (Nelson et al. 2008; Mizutani and Ohta 2010; Mizutani 2012).

For instance, lignins are structural components that derive from the phenylpropanoid pathway: some of the core reactions leading to their biosynthesis are catalysed by members of the CYP73A, CYP98A and CYP84A families, from the CYP71 clan (**Figure 21**). The CYP73A family contains the cinnamate 4-hydroxylases (C4H) that converts *trans*-cinnamic acid into *p*-coumaric acid, paving the way to the formation of *p*-hydroxy lignin and many other compounds (Mizutani et al. 1993; Teutsch et al. 1993; Mizutani and Ohta 2010). The CYP98A family includes enzymes that catalyse the meta-hydroxylation of coumaroyl esters, leading to guaiacyl lignin (Franke et al. 2002; Schoch et al. 2003; Mizutani and Ohta 2010). Last, the CYP84A family contains enzymes such as coniferaldehyde and coniferyl alcohol 5-hydroxylases that open the door to syringyl lignin (Meyer et al. 1996; Osakabe et al. 1999; Humphreys et al. 1999; Mizutani and Ohta 2010). To date, the CYP73A and CYP98A families have been found in all studied land plants, suggesting an early appearance and a strong selective pressure to conserve their activity. These families might even predate land plants, because some lignin-like material have been found in algae. On the contrary, the CYP84A family appeared later, in angiosperm, which explains some variations in lignin composition between gymnosperms and angiosperms (Weng et al. 2008; Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011).
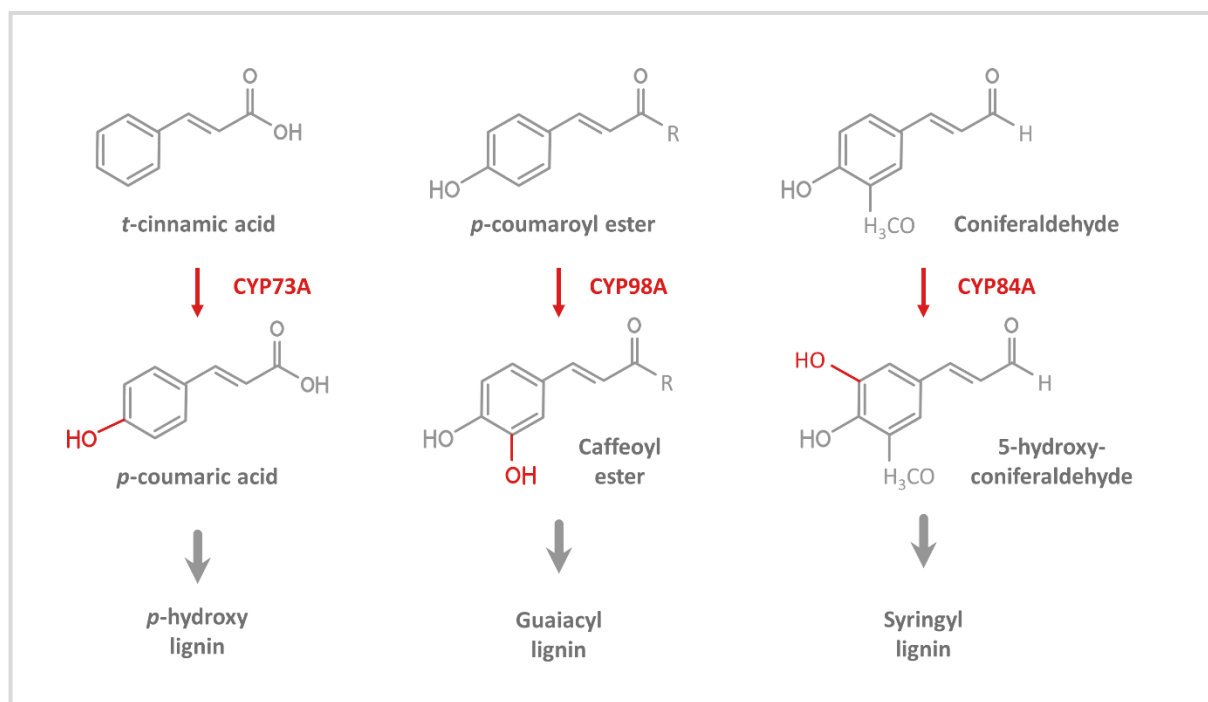
Figure 21 **Hydroxylation of the aromatic ring of cinnamates leading to the formation of lignins**. Adapted from Mizutani and Ohta (2010).

As another example, some P450s contribute to regulate the endogenous level of phytohormones by being involved in their biosynthesis or catabolism. And once again, these P450s – absent in algae – might have emerged early during land plant evolution (Mizutani and Ohta 2010). Plant phytohormones include cytokinins, abscisic acid, jasmonic acid, gibberellins, and brassinosteroids, which biosynthesis and/or catabolism involve many conserved plant-specific P450 families. For instance, the biosynthesis of cytokinins has been linked to members of the CYP735A family (CYP72 clan) (Takei et al. 2004; Nelson et al. 2008). The catabolism of abscisic acid is linked to enzymes from the CYP707A family (CYP85 clan) (Knight et al. 1995; Umezawa et al. 2006; Okamoto et al. 2006). The biosynthesis of jasmonic acid involves the CYP74A subfamily (CYP74 clan), while the CYP74B subfamily contributes to the biosynthesis of green leaf volatiles (Song et al. 1993; Matsui et al. 1996; Mizutani and Ohta 2010). As a last example, members of the CYP85A, CYP90B, CYP90C, CYP90D and CYP724B families contribute to the successive oxidations of campestrol in the brassinosteroid biosynthesis: all these families belong to the CYP85 clan and are conserved in both angiosperms and gymnosperms, which implies they all evolved from a common ancestral gene that have been subjected to duplication and divergence (Ohnishi et al. 2009; Mizutani and Ohta 2010).

F.3.b. P450s involved in plant specialised metabolism

If it is true that a significant number of highly conserved P450s are involved in core pathways and in the adaptation to life on land, these P450s only represent a small part of all plant P450s. Indeed, it is assumed that the majority of plant P450s are in fact involved in specialised metabolism and in the adaptation to various and fluctuating environmental constraints (Nelson 2006). Moreover, in many cases, plants that produce taxon-specific metabolites (*e.g.* restricted to families, genera or even species) also contain species-specific P450 (sub)families associated to this production. As an example,

the CYP726 family (which might, in fact, be a CYP71 subfamily) has only been found in some Euphorbiaceae where it contributes to the biosynthesis of vernolic acid, an epoxy fatty acid (Cahoon et al. 2002; Nelson et al. 2008). However, it seems that there is no unique pattern that can explain which P450(s) family is recruited in a given specialised metabolic pathway.

F.3.b.1. Evolutionary patterns

In general, after the successful duplication of an ancestral gene and the functional divergence of one of the duplicates, the resulting gene often codes for an enzyme which metabolises compound(s) that are similar to its ancestral substrate(s). Therefore, P450s belonging to a single (sub)family are usually involved in the metabolization of similar compounds (Nelson and Werck-Reichhart 2011). This can be exemplified with CYP86A1 and CYP86B1, that are both involved in the biosynthesis pathway of suberin, in which they hydroxylate the terminal methyl of acyl chains. The only difference between these two enzymes seems to be their chain length preference (Pinot and Beisson 2011). Such evolution in a given P450 (sub)family can also result in the mediation of successive reactions, the product of a first P450 becoming the substrate of another one. As an example, in maize, four P450s from the CYP71C subfamily catalyse four successive steps that lead to the conversion of indole into DIBOA (2.4-dihydroxy-2*H*-1,4-benzoxazin-3(4*H*)-one) in the benzoxazinoid biosynthesis pathway (**Figure 22**) (Frey et al. 1997; Gierl and Frey 2001; Nelson and Werck-Reichhart 2011).
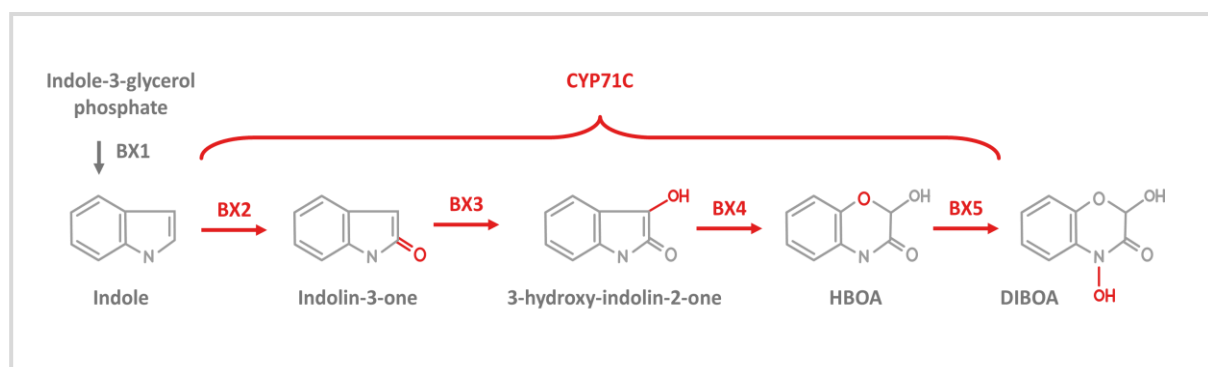
Figure 22 Beginning of the benzoxazinoid biosynthesis pathway until the production of DIBOA (2.4-dihydroxy-2H-1,4-benzoxazin-3(4H)-one). The four enzymes named BX2-5 are P450s belonging to the CYP71C subfamily. Adapted from Frey *et al.* (1997) and Gierl and Frey (2001).

Yet, a single biosynthesis pathway can also involve P450s belonging to different (sub)families, which can be exemplified by the production of benzylisoquinoline alkaloids, a group of defensive metabolites found in basal angiosperms such as the Ranunculales and the eumagnoliids. The biosynthesis of these compounds involves several P450s from the CYP80A, CYP80B and CYP80G subfamilies (CYP71 clan), together with the CYP719 family (CYP71 clan). To date, the CYP719 family have only been found in Ranunculales, while the CYP80A, CYP80B and CYP80G subfamilies might be specific to alkaloid producing plants (Liscombe and Facchini 2008; Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011). In short, a biosynthesis pathway can contain several P450s resulting from the expansion of one or several (sub)families.

In other cases, evolution found multiple solutions to produce the same compounds. Indeed, given that many specialised metabolites have arisen multiple times in different plant lineages by convergent evolution, one should not be surprised to find out that some plant species have independently recruited different P450s to catalyse the same reactions. As an example of repeated evolution, in the monocot *Paris polyphylla* and the eudicot *Trigonella foenum-graecum,* the oxidative 5,6-spiroketalisation of cholesterol can be catalysed by two enzymes from different CYP90 subfamilies (respectively, CYP90G4 and CYP90B50) that evolved independently (Pichersky and Gang 2000; Mizutani and Ohta 2010; Christ et al. 2019).

In addition, if P450s from different (sub)families can metabolise similar compounds, P450s from the same (sub)families can also metabolise various substrates. As an example, many P450s belonging to the CYP76 family are known for being involved in the metabolism of various specialised diterpenes, but some members of the CYP76C subfamily can also metabolise herbicides belonging to the class of phenylurea (Robineau et al. 1998; Swaminathan et al. 2009; Hofer et al. 2014; Bathe and Tissier 2019). So, in specialised metabolites, there is little discernible correlation between enzyme sequence similarity and substrate preference similarity. It can therefore be difficult to predict the substrate preference of a P450 by sequence membership, especially in the CYP71, CYP72 and CYP85 clans, which are the most expanded (Pichersky and Gang 2000; Nelson and Werck-Reichhart 2011).

Lastly, it is assumed that the most recently evolved gene families are more versatile and more likely to accumulate the mutations that will lead to new reactions or substrate specificity. But once again, there are exceptions. For instance, the ancient and highly conserved CYP51 clan, usually found in single or low copy in a given genome, has expanded in grasses and gave the CYP51H subfamily which is involved in the biosynthesis of avenacin, an antimicrobial triterpene glycoside. In all kingdoms, this is the only known example of the CYP51 clan that branched out from its ancestral function (Pichersky and Gang 2000; Nelson et al. 2008; Nelson and Werck-Reichhart 2011).

F.3.b.2. Two examples of lineage-specific evolution

In the CYP85 clan, some conifer-specific families are involved in the biosynthesis of conifer-specific specialised metabolites. For instance, taxoids such as taxol are defensive terpenoids found in the *Taxus* genus. Their biosynthesis requires many steps and, among other enzymatic reactions, involves at least eight P450-mediated reactions. To date, 7 P450s that hydroxylates taxoids have been identified; they all belong to the CYP725 family (CYP85 clan) which is specific to the *Taxus* genus. As the defensive properties of taxoids might confer a survival advantage, it may have been a driving force leading to the specific expansion of the CYP725 family in *Taxus* (Jennewein et al. 2004; Kaspera and Croteau 2006; Hamberger and Bohlmann 2006; Mizutani and Ohta 2010; Mizutani 2012). As another example, conifers also produce diterpenoid resin acids, which are important defensive compounds against herbivores and pathogens. This time, it is the CYP720B subfamily from the CYP85 clan, which is also conifer-specific, that contains many enzymes involved in the biosynthesis of these compounds (Ro et al. 2005; Keeling and Bohlmann 2006; Hamberger and Bohlmann 2006).

CYP93 is a family from the CYP71 clan that have only been found in angiosperms. To date, 10 CYP93 subfamilies have been identified: the CYP93A subfamily exists in both monocots and dicots and constitutes the ancestral group that underwent multiple gene duplications and divergence, yielding to

the 9 other families – CYP93B-CYP93K. Some of these subfamilies (CYP93F, CYP93G, and CYP93J) are found in monocots while the others (CYP93B, CYP93C, CYP93D, CYP93E, CYP93H, and CYP93K) are only found in eudicots. Some of these subfamilies even seem to be taxon-specific. For instance, the CYP93E subfamily has only been found in legumes, and the CYP93Fs only in grasses (Du et al. 2016; Yonekura-Sakakibara et al. 2019). The CYP93 family seems to be mainly involved in flavonoids metabolism: the ancestral subfamily, CYP93A, is involved in the biosynthesis of pterocarpanoids, which are derivative of isoflavonoids. In eudicot, the CYP93Bs catalyse the conversion of flavanones to flavones while the CYP93C subfamily is involved in the biosynthesis of isoflavones. In dicots, the CYP93Gs can also convert flavanones into flavones, which suggests that the monocot- and eudicot-specific subfamilies might have evolved a similar role in flavone metabolism. However, the CYP93Es are involved in the biosynthesis of triterpenoid saponins. The function of the remaining subfamilies, CYP93D, CYP93F, CYP93H, CYP93J, and CYP93K are still unknown (Schopfer et al. 1998; Moses et al. 2014; Du et al. 2016; Yonekura-Sakakibara et al. 2019). So, the CYP93 family started from a single subfamily that expanded and gave rise to different subfamilies restricted to limited plant taxa. Most of these new subfamilies take various flavonoids as substrates, but one got recruited in a different pathway and now mediate the biosynthesis of some triterpenes.

F.6. Summary of plant P450 evolution

The oldest P450 single-family clans that predate the emergence on land are devoted to the biosynthesis of essential compounds such as sterols and carotenoids. They are highly conserved and often found in single or low copies in the genomes.

In the other P450 clans, the families are more recent and specific to plants. A first P450 diversification happened during land colonisation and gave rise to new P450 families that mediate crucial adaptation for life on the terrestrial environment. Therefore, these P450 families are also well conserved among land plants.

Since then, new P450 families and subfamilies expanded specifically in some plant lineage(s) and got recruited in more diverse functions linked to the adaptation of plants to their environment. These new P450s are involved in the biosynthesis of pigments, odours, flavours and specialised metabolites that are often limited to restricted plant taxa.

All these P450s contribute to the characteristic high diversity of plant chemicals.

III.

OBJECTIVE AND APPROACH OF THE PHD

~

What, why and how?

Through evolution, plants have evolved many defence mechanisms to cope with pests and pathogens; and their chemical defences mainly rely on the biosynthesis of a vast array of specialised metabolites such as furanocoumarins. As crop species have been mostly bred to increase agronomical traits such as crop yield and quality, they have partially lost some defensive traits, which has been counterbalanced by an increased use of pesticides. But today, the use of pesticides is being more and more restricted: we thus need to find new alternatives to protect our crops, and to do so, the age-old plant defence mechanisms represent an interesting source of inspiration. However, the establishment of defences is costly and can negatively impact plant growth and reproduction, which means it would impact crop yield. In this context, improving our understanding of plant defences and the mechanisms underlying their production becomes essential to rethink our crop management strategies.

My PhD aimed to improve our knowledge of plant defence through the example of the furanocoumarin biosynthesis. This study was articulated around two distinct axes:

My first objective was to pursue the molecular elucidation of the furanocoumarin biosynthesis pathway in *Ficus carica*. Indeed, some enzymatic steps of the pathway are still unresolved. In particular, the conversion of DMS into marmesin is of great interest because it is the last unresolved step leading to the biosynthesis of psoralen in plants. Therefore, I have concentrated my efforts in searching, isolating, characterising and analysing new genes involved in the biosynthesis pathway of furanocoumarins, by using transcriptomic, metabolomic, phylogenetic and modelling approaches. I focused on the genes coding for P450s, an enzyme superfamily to which the marmesin synthase might belong. As for the plant material, I used *Ficus carica*, the fig tree (Moraceae), which produces linear furanocoumarins but which P450s have never been studied for furanocoumarin biosynthesis. In fact, no P450 involved in furanocoumarin production has been characterised so far in the Moraceae family. Identifying and characterising P450s from the fig tree would thus provide new evolutionary insight of the emergence and evolution of the furanocoumarin biosynthesis pathway in higher plants.

The second objective of my PhD consisted in assessing the metabolic cost of furanocoumarin production by inserting the beginning of the furanocoumarin pathway in the genome of tomato (*Solanum lycopersicum*), a plant that does not naturally produce these molecules. To do so, I used a multi-gene cloning technology, transgenesis and *in vitro* culture methods to introduce several genes of the furanocoumarin pathway into the genome of tomato. The idea was thus to generate furanocoumarin producing tomatoes which growth and defence might be compared to the ones of wild-type tomatoes.

Chapter II
Whole new genes
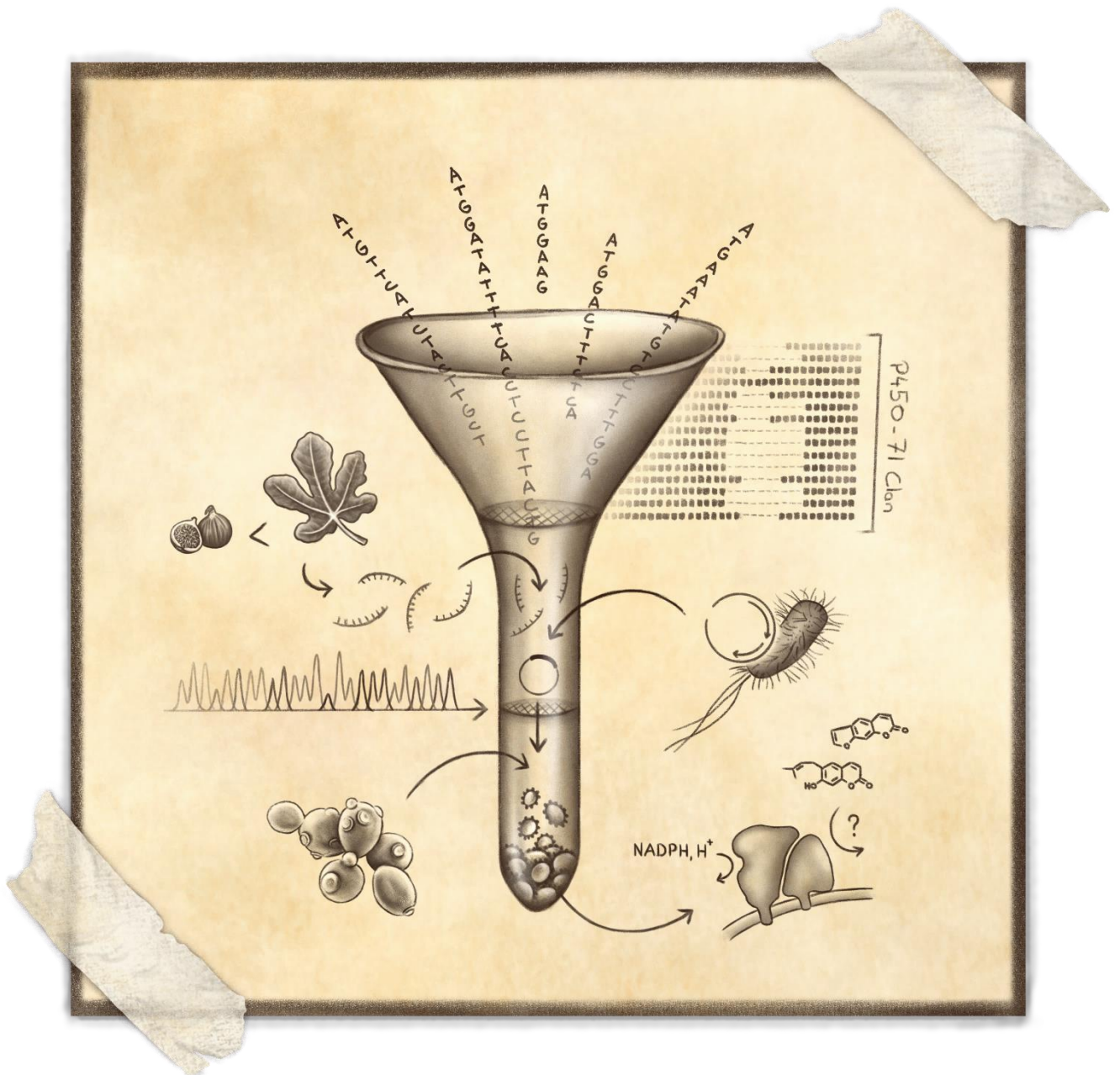
CHAPTER II.

WHOLE NEW GENES

~

Search, identification and functional characterisation
of candidate genes involved in the biosynthesis of furanocoumarins

A. INTRODUCTION AND STRATEGY: IN SEARCH OF THE LOST GENE

As depicted in **Chapter I** (**I.E.1**), the furanocoumarin biosynthesis pathway has been well studied over the last 30 years. First, the successive intermediates of the pathway were described in the late 1980's. Subsequently, many genes involved in this pathway were identified and functionally characterised – especially at the Laboratory of Agronomy and Environment (LAE – Nancy – France). These genes encode enzymes belonging to 4 different families: dioxygenases, prenyltransferases, P450s and methyltransferases (Hehmann et al. 2004; Larbat et al. 2007, 2009; Vialart et al. 2012; Karamat et al. 2014; Dueholm et al. 2015; Munakata et al. 2016, 2020; Roselli et al. 2017; Krieger et al. 2018; Limones-Mendez et al. 2020). Despite these important discoveries, some key-steps of the pathway are still unresolved at a molecular level. Among them, the most peculiar concerns the conversion of demethylsuberosin (DMS) into marmesin. Previous studies suggested that, in *Ammi majus*, the marmesin synthase should be a cytochrome P450 (Hamerski and Matern 1988a), which is quite original since cyclisation reactions are not very common for such enzymes (Mizutani and Sato 2011). Finding a marmesin synthase would also complete the set of enzymes that allows the production of psoralen, a toxic furanocoumarin, from the ubiquitous plant compound coumaric acid.

> **Therefore, the first objective of my PhD project was to identify new genes involved in the biosynthesis of furanocoumarins. Finding a marmesin synthase was my primary goal.** For this purpose, I combined transcriptomic and metabolomic approaches. First, I used a differential RNA-seq database to identify candidate genes from *Ficus carica*. Then, I cloned the coding sequences of the candidates and heterologously expressed them. Finally, I performed enzymatic assays to test their activity and characterise them.

B. CANDIDATE GENES AND HOW TO FIND THEM

B.1. Approach: where to search, what to search?

In this study, I focused on *Ficus carica*, the common fig tree, which belongs to the Moraceae family and produces linear furanocoumarins. This choice relies on three main reasons. Firstly, thanks to a

collaboration established between the LAE and S. Kitajima (Kyoto University, Japan), we had access to a differential RNA-seq library constructed from the latexes of fruits, petioles, and lignified trunks of *F. carica*. The latexes from these three tissues all contain different furanocoumarin concentrations: indeed, in *F. carica*, the latex from leaves contain higher concentration of furanocoumarins than the latexes from trunks and fruits (Kitajima et al. 2018). It can thus be hypothesised that the genes involved in furanocoumarin biosynthesis have differential expression in the different latexes. The second reason is that only one *Ficus* gene has been described thus far in the furanocoumarin pathway. This gene, recently characterised at the LAE by R. Munakata, encodes FcPT1, a prenyltransferase that converts umbelliferone into DMS (Munakata et al. 2020). The third reason is that the expression pattern of FcPT1 was consistent with the furanocoumarin contents in the three latexes (Munakata et al. 2020).

Therefore, my strategy was to screen the *F. carica* RNA-seq library to identify sequences that would be preferentially expressed in the latex from leaves rather than in those from trunks and fruits.

As my priority was to identify a marmesin synthase, I focused on the identification of P450 sequences. Indeed, in apiaceous species, the marmesin synthase is assumed to be a P450 (Hamerski and Matern 1988a). By analogy, it could be hypothesised that the fig tree marmesin synthase was also a P450. Yet, P450s constitute a very large superfamily composed of many clans, families and subfamilies. Under these conditions, it was not reasonable to investigate all the P450s found in the fig tree – demanding that I narrow my search. The P450s involved in the furanocoumarin pathway that have been characterised so far all belong to the CYP71 and CYP82 families, which are both from the CYP71 clan (**Chapter I, Figure 20**). Thus, the CYP71 clan was my initial target of interest.

The CYP71 clan contains more than half of all P450s found in higher plants, and it is the main clan involved in the phenylpropanoid pathway (Nelson and Werck-Reichhart 2011; Hamberger and Bak 2013). In fact, this clan is involved in both primary and specialised metabolism: it contains some ancient families that played a crucial role in the conquest of land, but most of its members seem to be associated to specialised metabolism (*e.g.* terpenoids, alkaloids, glucosinolates, phenylpropanoids, *etc.*). A significant number of its families and subfamilies have also expanded into lineage-specific plant families (Nelson 2004; Nelson et al. 2008; Nelson and Werck-Reichhart 2011; Hamberger and Bak 2013; Yonekura-Sakakibara et al. 2019). All of this made the CYP71 clan a candidate full of potential for further elucidation of the furanocoumarin biosynthesis pathway.

B.2. *In silico* screening of the *F. carica* RNA-seq library

To facilitate the understanding of my screening strategy of the *F. carica* RNA-seq library, I have summarised my approach in **Figure 23**. More information is available in the materials and methods section (**Chapter VII**, **B.6.a**). First, I used a similarity approach to identify as many contigs of interest as possible. For this purpose, I used *CYP71AJ3*, the parsnip gene encoding for a psoralen synthase (Larbat et al. 2009), to perform my initial BLAST searches on the *F. carica* RNA-seq library. This led me to the identification of 50 contigs of various length corresponding to partial or complete gene sequences.

In a second step, I sorted the contigs to eliminate duplicates and I assembled the sequences that could be merged in order to reconstruct "complete coding sequences" from partial sequences. The contigs from the RNA-seq library were not always sufficient to obtain full-length gene sequences. Yet, in some

cases, I could complete these contigs by using the data found in another database. In particular, some cDNAs from the related *Ficus religiosa* were available on 1KP (https://db.cngb.org/onekp/ – One Thousand Plant Transcriptomes Initiative 2019; Carpenter et al. 2019), which included a few sequences that were highly similar to some of the contigs of interest. Therefore, I sometimes used these homologous *F. religiosa* sequences as a frame to assemble different contigs from *F. carica* with limited overlap, or to complete the contigs by filling the gaps between them. This allowed me to assemble a larger number of complete coding sequences. However, it should be kept in mind that this reconstruction was artificial and might have created chimeric sequences (merging of contigs from different genes, *F. carica* sequences completed by *F. religiosa* sequences). Partial genomic sequences from *F. carica* were also available on GenBank (https://www.ncbi.nlm.nih.gov/genbank/ – Benson et al. 2012), but as they did not provide additional useful information, they were not used. In the end, I obtained 22 P450 sequences, among which only 13 could be considered as "complete coding sequences". The sequences that were still too short to correspond to entire coding sequences have been set aside; the work only continued with the putative P450 complete coding sequences.
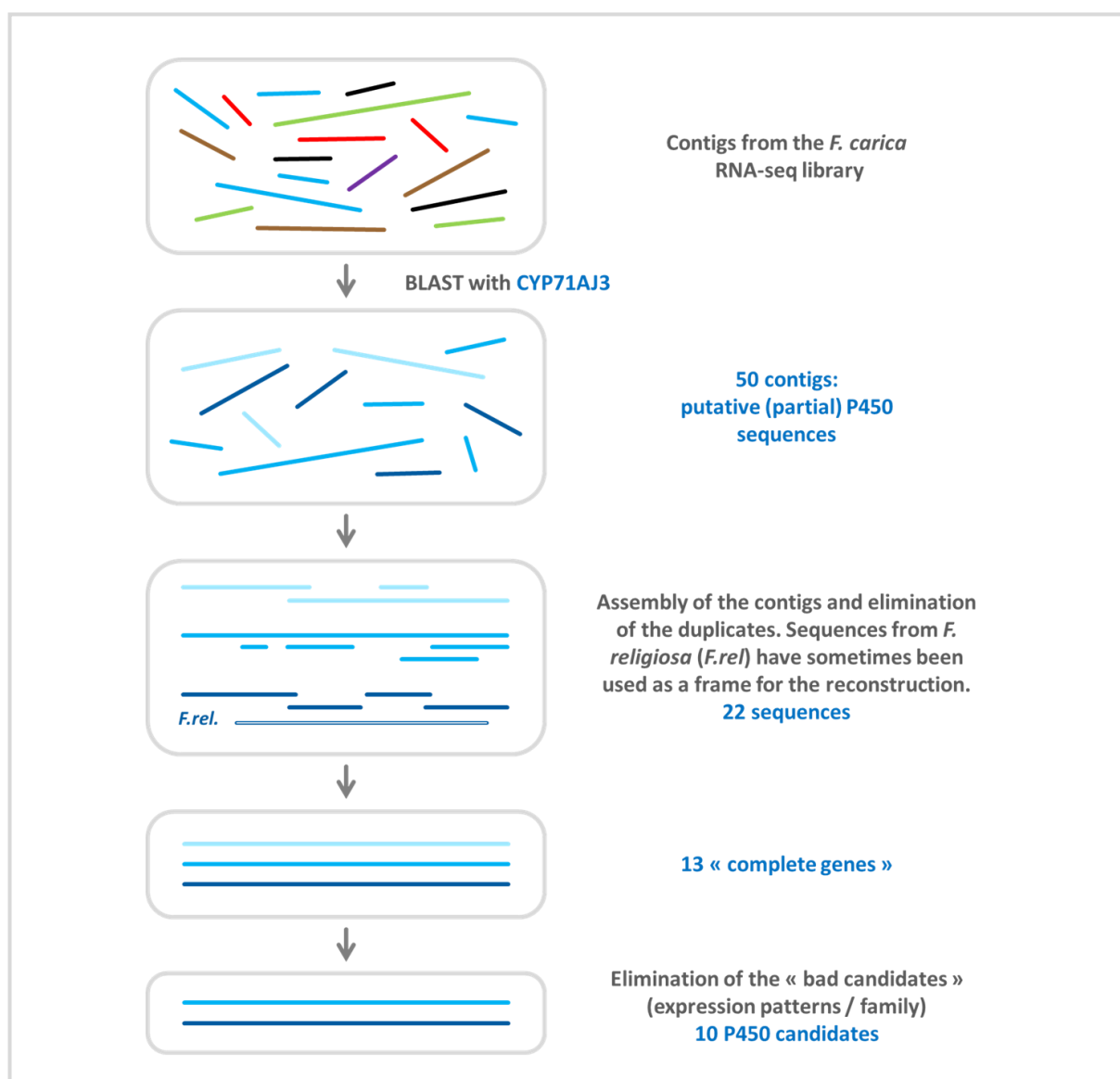
Figure 23 Identification of the P450 candidates from the *F. carica* library.

In a last step, I analysed the expression level of the candidate P450s from the three latexes, and kept only those which were more expressed in the latex from the leaves. In addition, in order to eliminate some unlikely candidates, I performed a BLAST search with every candidate on a set of known P450s belonging to different families from the CYP71 clan. This allowed me to get a first impression of the P450 family to which every candidate belonged. Then, I eliminated the candidates from the families that had a very low chance to be involved in the biosynthesis of furanocoumarins – for instance, the ancient and/or very conserved families involved in core pathways, such as the CYP73 family. Thus, I limited the P450 candidates to 10 sequences.

The main characteristics of these 10 candidates are summarised in **Table 2**. Additional details can be found in **Supp. Table 1**. It should be noted that only three and one candidates from the CYP71 and CYP82 families were identified: these P450 families are the ones that had previously been identified in the furanocoumarin biosynthesis pathway in Apiaceae and Rutaceae, respectively. However, at this point, the candidates only corresponded to *in silico* reconstructed genes – that is, to putative sequences. The next step, which consisted in amplifying and sequencing these candidates from *F. carica* material, was therefore essential to confirm the existence of the candidates in the fig tree, and to verify their complete sequences. This was all the more important for the candidates which have been reconstructed from several contigs and/or assembled using *F. religiosa* sequence data.

Table 2 Summary of the P450 candidate genes. The relative expression levels of every candidate correspond to the expression patterns of the contigs from the RNA-seq library (Kitajima et al. 2018). The white to green shades reflect the low to high expression levels in the three latexes. The candidates have been arbitrarily numbered, and coloured according to their P450 family (determined with a BLAST search performed on a set of P450s from the CYP71 clan). More information such as the accession numbers and expression patterns can be found in **Supp. Table 1**.

Candidate number	Relative expression levels in the three latexes			Length of the amino acid sequence	Closest P450 family
	Petiole	Trunk	Fruit		
1	83,7%	15,6%	0,7%	507	CYP71
2	80,4%	18,8%	0,9%	504	CYP76
3	62,0%	22,2%	15,9%	497	CYP81
4	84,6%	6,0%	9,4%	497	CYP81
5	89,5%	10,5%	0,0%	506	CYP76
6	44,7%	33,7%	21,6%	502	CYP81
7	89,1%	10,9%	0,0%	504	CYP76
8	85,9%	13,4%	0,6%	506	CYP71
9	40,7%	39,1%	20,1%	520	CYP82
10	44,3%	24,6%	31,0%	499	CYP71

In summary, I identified 10 candidate genes belonging to the P450 superfamily. The next steps consisted in amplifying, sequencing and expressing them.

The same screening approach was also done for the dioxygenase and methyltransferase families, also involved in furanocoumarin biosynthesis. This led to the identification of 7 candidate genes for both families (**Supp. Table 1**, **Supp. Table 2**). The subsequent results will be restricted to P450s; the other enzymes will be discussed in the perspectives (**Chapter VI**).

C. I'LL MAKE AN ENZYME OUT OF YOU: CLONING AND EXPRESSION OF THE P450S

C.1. Amplification, cloning and sequencing of the P450 candidates

To amplify the 10 P450 candidates, the mRNAs from a *F. carica* leaf were extracted and retrotranscribed into cDNA. PCR amplification was then performed on the cDNA using pairs of primers designed for each candidate (**Supp. Table 3**). A fragment of about 1500 bp was specifically amplified for each candidate, except for candidate n°8 (as numbered in **Table 2**), which I was never able to amplify. The result obtained for candidate n°8 was not surprising: indeed, to reconstruct its sequence, two contigs from the *F. carica* RNA-seq library were assembled and completed with the genetic data from *F. religiosa*. Thus, candidate n°8 was likely a false chimeric sequence. I put it aside and proceeded only with the remaining successfully amplified 9 candidates.

The fragments amplified for each candidate were then inserted into the cloning vector pCR™8/GW/TOPO™ and sequenced. To avoid potential errors of the polymerase during the initial amplification from the cDNA, each candidate was sequenced from two independent PCRs. For 8 of the 9 remaining candidates, the sequences amplified from *F. carica* cDNAs shared 99 to 100% identity with the putative sequences determined from the RNA-seq database – which included a few silent and missense mutations, but no nonsense nor frameshift mutations. These 8 sequences therefore corresponded to coding sequences of about 1500 bp. However, the fragment amplified with the primers corresponding to the candidate n°1 shared only 84% identity with the expected sequence and contained stop codons. Thus, candidate n°1, even though it came from one single contig, might have been a chimeric sequence. It is also possible that the amplification was not specific enough and I did not amplify the right sequence. Anyway, this candidate was eliminated from future experiments.

The sequences of the 8 remaining candidates were sent to Pr. David Nelson (P450 nomenclature comity, https://drnelson.uthsc.edu/CytochromeP450.html), who named them according to the standardised P450 classification and nomenclature. Henceforth, I will use these "official" names, which are reported in **Table 3**. The sequences of the candidates are available in **Supp. Figure 1**. It should be emphasised that the 8 remaining candidates did not include any P450 from the CYP71 family, which is the one involved in the synthesis of furanocoumarins in Apiaceae. Finally, the candidate n°10 did not belong to the CYP71 clan but to the CYP88 family, from the CYP85 clan. This made it an unlikely candidate, but as it was already cloned, CYP88A103 has been kept in the following steps.

Table 3 **Amplification, sequencing and heterologous expression of the P450 candidates.** The colours refer to the P450 family of the candidates; eliminated candidates are in orange.

Candidate	Amplifi-cation from the cDNA?	Identity b/w the amplified and expected sequences	Standardized name	Length of the protein	Weight of the protein	Heterologous expression (Figure 24)
1	Yes	84% – stop codons – elimination of the candidate				
2	Yes	100%	CYP76F110	504 AA	56.68 kDa	Correct
3	Yes	99%	CYP81B114	497 AA	56.72 kDa	Correct
4	Yes	99%	CYP81BN4	497 AA	56.10 kDa	Low
5	Yes	99%	CYP76F111	506 AA	56.62 kDa	Correct
6	Yes	100%	CYP81CA1	502 AA	57.39 kDa	Very low
7	Yes	100%	CYP76F112	504 AA	56.95 kDa	High
8	No: elimination of the candidate					
9	Yes	99%	CYP82J18	520 AA	58.78 kDa	Correct
10	Yes	99%	CYP88A103	497 AA	56.95 kDa	Very low

C.2. Heterologous expression of the P450 candidates

Identifying the function of new enzymes is a major challenge. For this purpose, proteins need to be heterologously expressed and then incubated in the presence of potential substrates. Things are clear when there is a metabolization of the substrate, but the absence of enzymatic reaction could be due to three different reasons. First, the tested molecule might not be a substrate of the enzyme. Second, the enzyme might not be functional in the conditions of the reaction, or impaired. Third, the protein might not have been expressed. To ensure this is not the case, tagging strategies can provide us the possibility to highlight the expression of the proteins using an immunodetection approach.

Therefore, prior to expressing the candidate genes, a His-tag composed of 6 CAT codons was added right before their stop codon by reamplifying the candidates with modified primers (**Supp. Table 3**). The His-tagged candidate genes were inserted by recombination into the expression vector pYeDP60_GW and transformed into the WAT21 yeast strain. Finally, each candidate was heterologously expressed in yeast and the produced proteins were collected by preparing yeast microsomes, as explained in the materials and methods section (**Chapter VII**, **B.2.c**).

The expression of the candidates has been evaluated by immunodetection, using a serum directed against the His-tag. The results, presented in **Figure 24**, confirmed the successful expression of the apoprotein part of all eight candidates. Indeed, some signals were detected around 55 kDa, which corresponds to their theoretical molecular weight. Such signals were absent in the negative control. However, some candidates were more highly expressed than others. In particular, CYP76F112 was highly expressed while CYP81CA1 and CYP88A103 were barely detectable (**Figure 24**). These results are also summarised in the above **Table 3**.
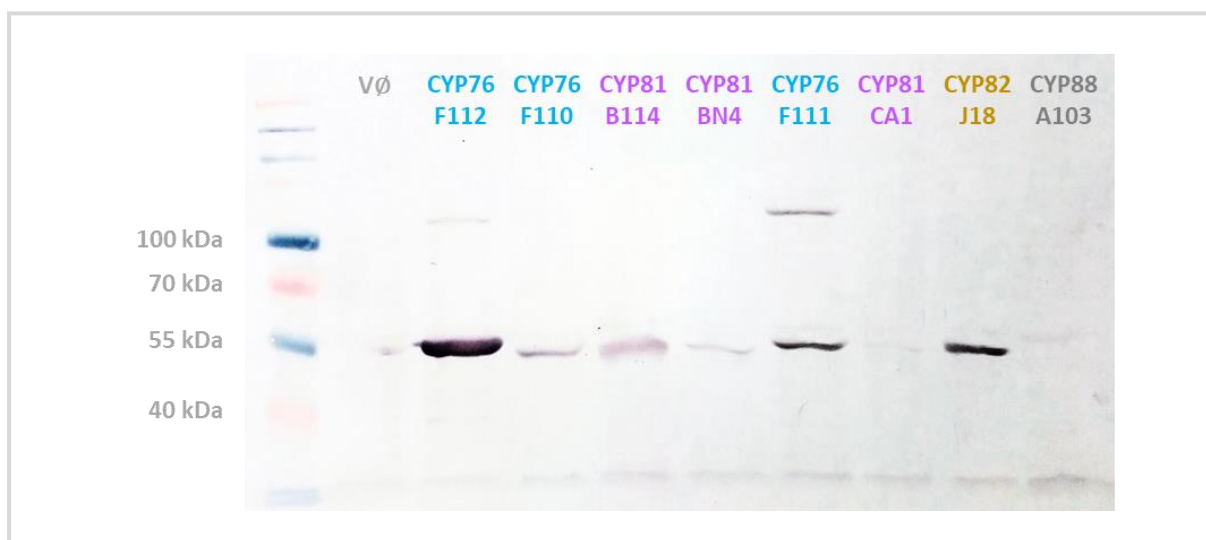
Figure 24 **Evaluation of the expression of the P450 candidates by immunodetection.** Western Blot performed on the microsomes collected after the expression of the P450 candidates in the yeast WAT21, directed against the His-tag added in the sequence of every candidate. VØ refers to the microsomes from the yeasts transformed with the empty pYeDP60_GW vector (negative control). The candidates' names are coloured according to their P450 family. The molecular weight marker (first well) is described in the materials and methods section (**Chapter VII**, **B.2.d**).

In brief, I successfully amplified 8 of the P450 candidates, which included 3 P450s from the CYP76 family, 3 from the CYP81 family, 1 from the CYP82 family and 1 from the CYP88 family. These candidates have all been heterologously expressed, with varying degrees of success.

D. CONVERT THIS AND I'LL LOVE YOU: ENZYME ASSAYS AND CHARACTERISATION

D.1. Enzyme assays and functional screening

In order to determine the function of the P450 candidates, I performed enzyme assays with a large range of 40 substrates that included coumarins and linear / angular furanocoumarins (**Table 4**). The most interesting substrates were obviously the DMS and the linear furanocoumarins. Yet, as coumarins and angular furanocoumarins have similar structure, and as some of them are produced (even if in low concentration) in the fig tree, they were also included.

To perform the functional screening, the microsomes prepared from the yeasts were incubated in the presence of NADPH and one substrate at a time. It should be noted that the microsomes also contained a NADPH reductase from *A. thaliana* (ATR2), produced by the yeast strain WAT21. The presence of this redox partner is essential for P450-mediated reactions (**Chapter I**, **D.2.a**). As it is also well known that P450s are rather unstable proteins, I performed all the incubations with fresh microsomes produced within the day. After the incubations, the reaction products were analysed by UHPLC-MS, by monitoring either the appearance of new peaks or the decrease of substrate peak.

Table 4 **List of the potential substrates used for the functional screening.** The substrates highlighted in red were of primary interest: they are intermediates from the main linear furanocoumarin pathways, and they might all be hydroxylated by P450s. The three substrates highlighted in blue are additional substrates that have been tested on CYP76F110, CYP76F111 and CYP76F112.

Linear furanocoumarins	Coumarins	Angular furanocoumarins
Bergamottin	Aurapten	Angelicin
Bergapten	Daphnetin	Cnidicin
Byakangelicine	Daphnetin dimethyl ether	Columbianadin
Byakangelicol	Demethylsuberosin (DMS)	Columbianetin
Cnidilin	Esculetin	Pimpinellin
Heraclenin	Fraxetin	Sphondinol
Heraclenol	Herniarin	Osthenol
Isopimpinellin	Limettin	
Marmesin	Osthol	**Others**
Psoralen	Scoparone	Cinnamic acid
Xanthotoxin	Scopoletin (Scopoletol)	*p*-coumaric acid
Xanthotoxol	Suberosin	O-coumarate
5-hydroxyxanthotoxin	Umbelliferone	
8-geranopsoralen	4-hydroxyherniarin	**Herbicides**
8-hydroxybergapten	4-methylumbelliferone	Chlortoluron
	7,8-dihydroxy-6-methoxycoumarin	Isoproturon

In the case of CYP76F110, CYP76F111, CYP81B114, CYP81CA1 and CYP88A103, no new metabolite was formed, no matter what the substrate.

On the contrary, the analyses revealed the appearance of new products when CYP76F112, CYP82J18, and CYP81BN4 were respectively incubated with DMS, auraptene, and cnidilin.

D.1.a. CYP76F112, a marmesin synthase

The incubation of CYP76F112 in the presence of DMS (230 g/mol) and NADPH yielded an enzymatic reaction product (246 g/mol) concomitant with the consumption of DMS. This product was identified as marmesin by direct comparison of its retention time and mass spectrometry spectrum with those of the standard molecule. In addition, the formation of this product was NADPH-dependent and it did not appear in the control incubations done with microsomes from yeasts transformed with the empty vector (**Figure 25**). Thus, this product was indeed the result of a specific P450-mediated reaction.

In other words, CYP76F112 is responsible for the conversion of DMS into marmesin, which makes it a *F. carica* marmesin synthase.
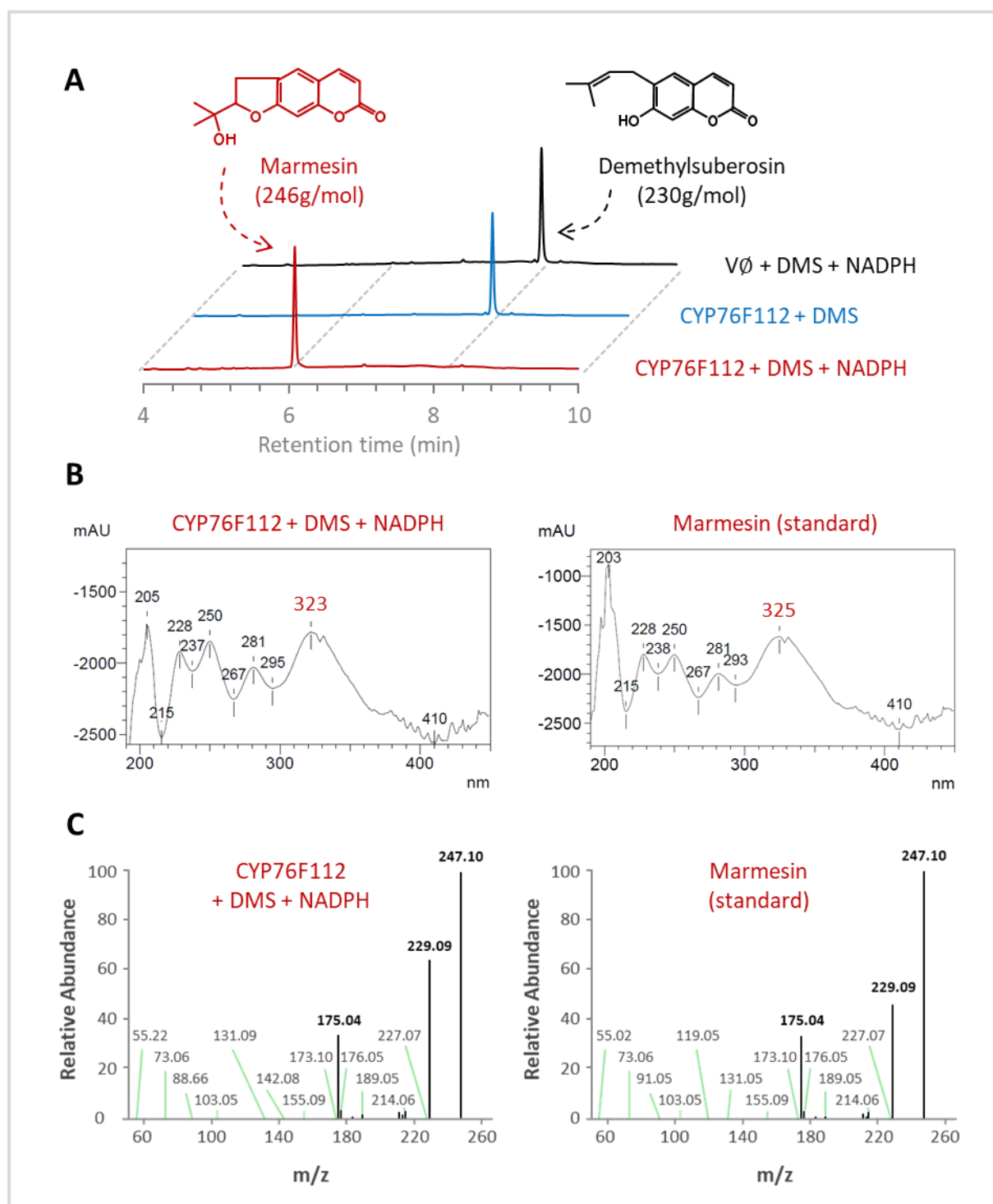
Figure 25 Conversion of DMS into marmesin, catalysed by CYP76F112. Analyses of the reaction products generated by microsomes incubated in the presence of DMS, with or without NADPH. The microsomes were collected from yeast producing CYP76F112 or the empty pYeDP60_GW vector (VØ, negative control). **A** UHPLC separation profile: the reaction products were extracted and separated by UHPLC monitored at 320 nm. The retention times of 6 and 8 minutes correspond, respectively, to the retention times of the standards of marmesin and DMS. **B** UV spectrum of the compound produced by CYP76F112, in comparison with a standard of marmesin. **C** Mass spectrometry fragmentation pattern (MSMS analysis) of the reaction product generated by CYP76F112, in comparison with a standard of marmesin.

D.1.b. CYP82J18, a P450 that hydroxylates auraptene

Similarly, the incubation of CYP82J18 in the presence of auraptene (a simple coumarin – 298 g/mol) and NADPH led to the formation of four reaction products that did not appear in the controls (A, B, C, D – **Figure 26**). With a molecular mass of 314 g/mol (C/D), 330 g/mol (B) and 332 g/mol (A), these products might respectively correspond to auraptene derivatives that would were hydroxylated (+16), dihydroxylated (+32), and dihydroxylated and desaturated (+34) (**Figure 26**). However, many hydroxylated derivatives of auraptene might correspond to these products. In the absence of standards to compare their retention times and mass spectrometry spectrums, they have not been precisely identified. Thus, CYP82J18 is responsible for the hydroxylation of auraptene into four unidentified products, that should now be formally identified using LC-NMR approaches.

Compound	Ret. time	Mass	Corresponding reaction?
Auraptene	10.0 min	298 g/mol	---
A	7.4 min	332 g/mol	Dihydroxylation + insaturation
B	7.5 min	330 g/mol	Dihydroxylation
C	8.2 min	314 g/mol	Hydroxylation
D	8.3 min	314 g/mol	Hydroxylation
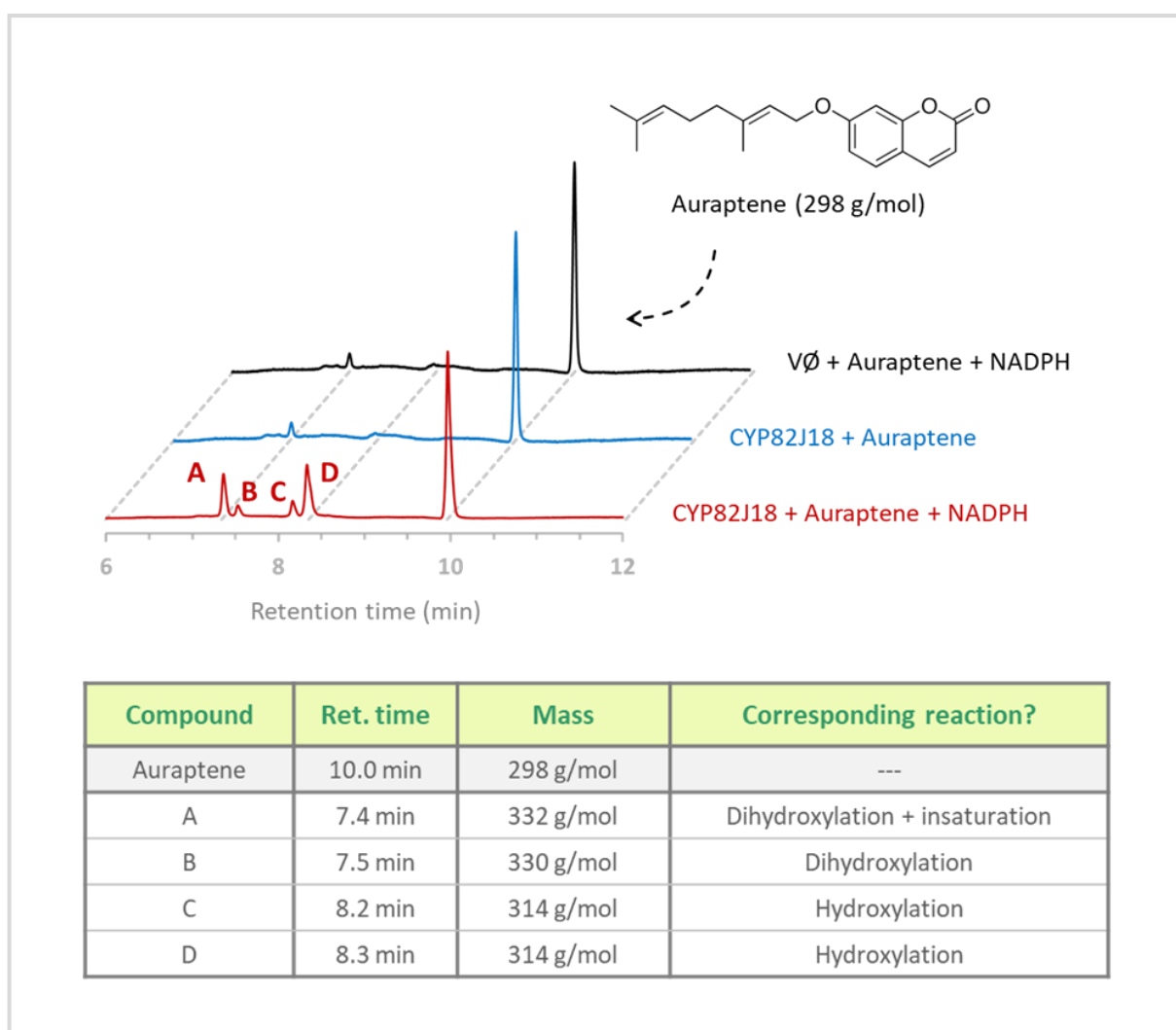
Figure 26 Metabolization of auraptene by CYP82J18. UHPLC-MS analyses of the reaction products generated by microsomes incubated in the presence of auraptene, with or without NADPH. The microsomes were collected from yeasts producing CYP82J18 or the empty pYeDP60_GW vector (VØ, negative control). In the UHPLC separation profile, the retention time of 10 minutes corresponds to the auraptene standard. The mass of the products A, B, C and D were determined by mass spectrometry, but these products have not been precisely identified.

D.1.c. CYP81BN4, a P450 that hydroxylates cnidilin

Lastly, the incubation of CYP81BN4 in the presence of cnidilin (a prenylated derivative of psoralen – 300 g/mol) and NADPH yielded three enzymatic reaction products that did not appear in the controls (**Figure 27** – E, F, G). Their molecular masses were 316 g/mol (G), 332 g/mol (F) and 334 g/mol (E). These reaction products might thus correspond to cnidilin derivatives that would were respectively hydroxylated (+16), dihydroxylated (+32), and dihydroxylated and desaturated (+34) (**Figure 27**). Once again, several hydroxylated derivatives of cnidilin could correspond to these molecular masses and, in the absence of standards to compare the retention times and mass spectrometry spectrums of the reaction products, they have not been precisely identified. CYP81BN4 is thus responsible for the hydroxylation of cnidilin into three unidentified products. As for CYP82J18, the three reaction products of CYP81BN4 still have to be identified using LC-NMR approaches.

Compound	Ret. time	Mass	Corresponding reaction?
Cnidilin	7.4 min	300 g/mol	---
E	4.3 min	334 g/mol	Dihydroxylation + insaturation
F	5.1 min	332 g/mol	Dihydroxylation
G	6.2 min	316 g/mol	Hydroxylation
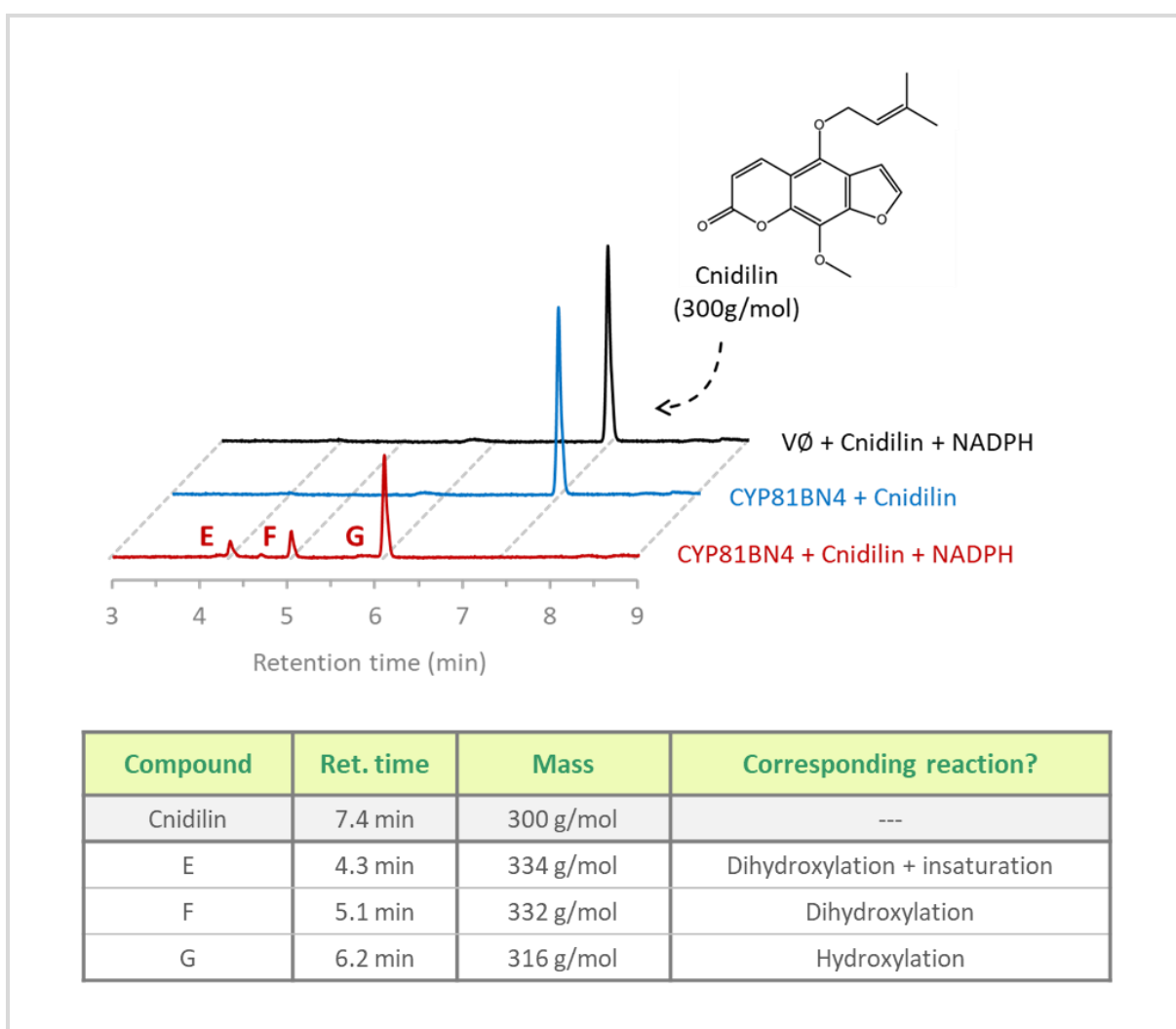
Figure 27 Metabolization of cnidilin by CYP81BN4. UHPLC-MS analyses of the reaction products generated by microsomes incubated in the presence of cnidilin, with or without NADPH. The microsomes were collected from yeasts producing CYP81BN4 or the empty pYeDP60_GW vector (VØ, negative control). In the UHPLC separation profile, the retention time of 7 minutes corresponds to the cnidilin standard. The mass of the products E, F and G were determined by mass spectrometry, but these products have not been precisely identified.

To summarise, the major result of this study was the identification of CYP76F112, a P450 from *F. carica* with a marmesin synthase activity.

I also identified CYP82J18 and CYP81BN4, which respectively metabolise auraptene and cnidilin, leading to the formation of several hydroxylated products that have not yet been chemically characterised.

Compared to CYP76F112 and its marmesin synthase activity, CYP82J18 and CYP81BN4 were of lesser interest for the study of core furanocoumarin biosynthesis. Indeed, auraptene is a simple coumarin that is not involved in the furanocoumarin pathway, while cnidilin is a derivative of psoralen formed downstream to the main steps of the furanocoumarin pathway. In addition, as far as I know, auraptene and cnidilin have never been described in the fig tree and their toxicity has never been specifically tested (discussed later). For these reasons, the subsequent analyses are more detailed for CYP76F112 rather than for CYP82J18 and CYP81BN4.

D.2. Functional characterisation of CYP76F112, CYP82J18 and CYP81BN4

D.2.a. CYP76F112: an enzyme with a strong specificity and affinity

A functional characterisation of the marmesin synthase has been realised to determine its substrate specificity, optimal pH and temperature conditions, and some kinetic parameters such as the apparent K_m and K_{cat} values.

D.2.a.1. Substrate specificity

First, among the 40 substrates tested during the functional screening (**Table 4**) CYP76F112 was only able to metabolise DMS. Even molecules such as marmesin, which are structurally close to DMS (**Figure 28**), were not accepted as substrate. Additional incubations were performed with three new substrates: osthenol, chlortoluron and isoproturon (**Figure 28**). Osthenol is the angular isomer of DMS, which was not included in the initial screening. Chlortoluron and isoproturon are two herbicides that can be metabolised by several CYP76s, such as CYP76B1, CYP76C1, CYP76C2 and CYP76C4 (Robineau et al. 1998; Höfer et al. 2014). However, none of these additional substrates was metabolised by CYP76F112.

CYP76F110 and CYP76F111 (**Table 3**), two candidates belonging to the same subfamily than CYP76F112 and sharing about 70% nucleotide identity, were also incubated with osthenol, chlortoluron and isoproturon. Here again, no metabolization was observed.

Thus, out of the 43 substrates that were incubated with CYP76F112, only DMS was metabolised, leading to the formation of a unique product: marmesin. CYP76F112 therefore seems to be a highly specific enzyme converting a single substrate into a single product.
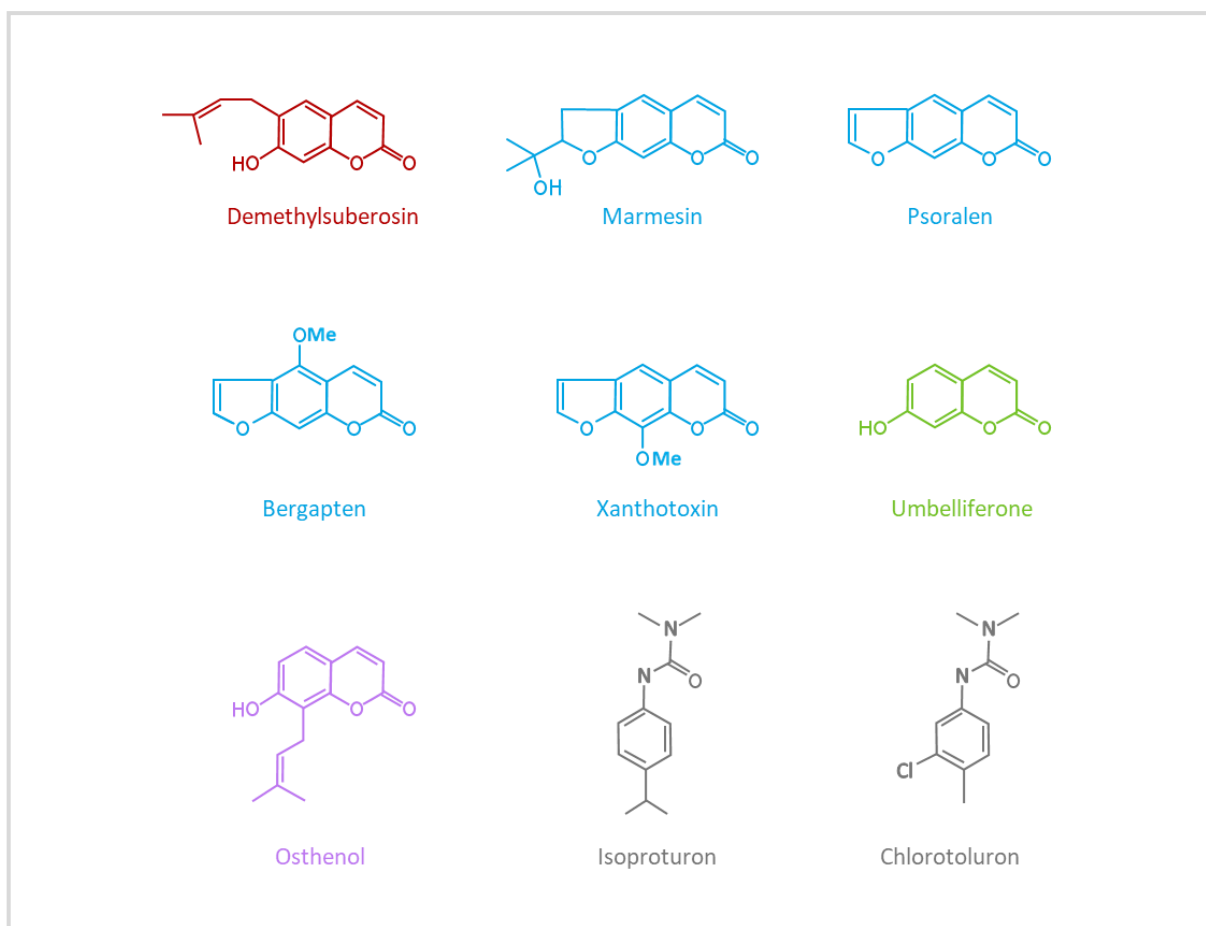
Figure 28 Structure of some of the potential precursor compounds incubated with CYP76F112. DMS is in red, some of the tested linear furanocoumarins are in blue, and an example of tested coumarin is in green. Osthenol, the angular isomer of DMS, is in purple. The two tested herbicides are in grey. Only DMS was successfully metabolised by CYP76F112.

D.2.a.2. Optimal enzymatic conditions

To determine CYP76F112 optimal reaction temperature, I tested the efficiency of marmesin synthesis with incubation temperatures ranging from 20°C to 45°C. The maximal activity was measured for a temperature of 27°C, but CYP76F112 presented a low thermo-specificity since it kept more than 50% activity between 20°C to 35°C (**Figure 29**).

The optimal pH was determined by incubating the microsomes in buffer solutions ranging from pH=4 to pH=10. The maximal activity was measured for a pH of 7; and CYP76F112 seems to be mainly active in a range of pH buffer restricted to 6.5-8 (**Figure 29**).

The optimal temperature and pH conditions of CYP76F112 are consistent with those of other P450s such as CYP71AJ1-4, the apiaceous psoralen and angelicin synthases (Larbat et al. 2007, 2009). However, contrary to CYP71AJ1-4, CYP76F112 has proven to be very resistant to freezing, making it a very convenient enzyme to study. Indeed, after more than a year of storage at -20°C, the microsomes containing CYP76F112 were still able to convert DMS into marmesin, even though the activity of the defrost microsomes was a bit lower than that of freshly prepared microsomes (not quantified).
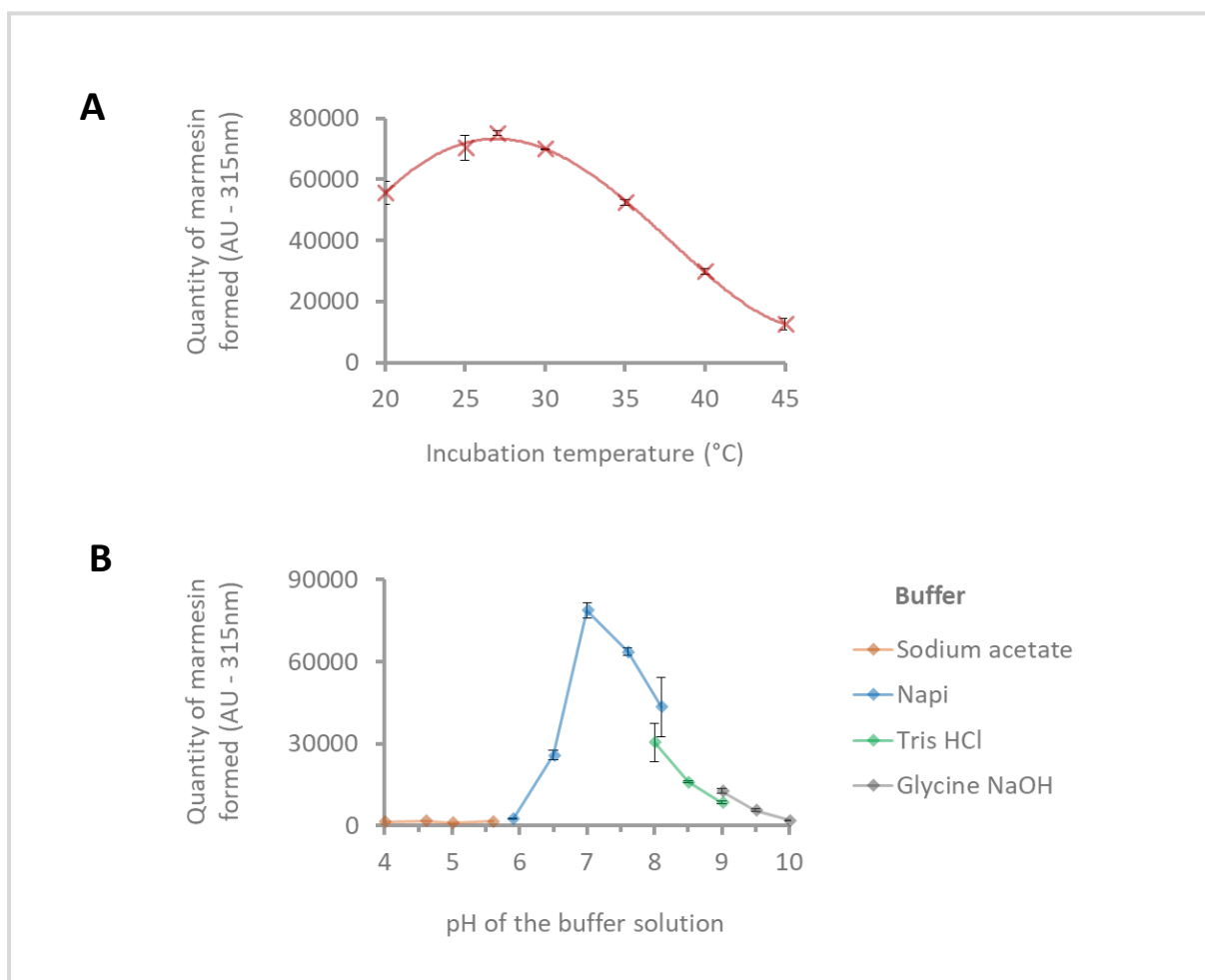
Figure 29 Determination of the optimal temperature and pH for the activity of CYP76F112, incubated in presence of DMS and NADPH. All incubations were repeated in triplicates and analysed by UHPLC-MS. The peak areas corresponding to the formation of marmesin were measured at 315nm. **A** The incubations were performed at pH=7, with variable temperatures. B. The incubations were performed at 27°C, in buffer solutions of variable pH.

D.2.a.3. Kinetic parameters: affinity, catalytic constant and catalytic efficiency

The kinetic parameters that can be measured to characterise an enzyme include the Michaelis constant K_m, the catalytic constant K_{cat} and the catalytic efficiency K_{cat}/K_m. The K_m represents the affinity of the substrate toward the enzyme. The K_{cat} represents the maximum number of molecules of substrate converted per unit of time and per molecule of enzyme. The K_{cat}/K_m is a constant that measures how efficiently an enzyme is able to convert its substrate into product.

Determining the kinetic parameters associated to the conversion of DMS into marmesin by CYP76F112 has proved to be a challenging task. Indeed, some preliminary experiments showed that the K_m value associated to the conversion of DMS by CYP76F112 was lower than 0.1 µM, which was below the detection limit of the UHPLC-MS I could use. Therefore, I had to develop a specific protocol to overcome this technical limit. This protocol is detailed in the materials and methods section (**Chapter VII**, **B.2.f.3**). In brief, because of the need to maintain the incubations at a temperature of 27°C and

agitated, I had to perform the reactions in 2 mL tubes, and was limited to 1 mL of reaction mix at a time. Consequently, for every modality, I performed 5 to 15 incubations of 1 mL, pooled them together and concentrated them in a final volume of 100 µL (for a concentration factor of 50 to 150). The concentrated solutions were then analysed by UHPLC-MS. This adapted protocol allowed me to test DMS concentrations as low as 10 nM. To make sure that the conversion of the substrate did not exceed 50%, even for the smaller concentrations, I also had to dilute the microsomes and to reduce the incubation time – which required many preliminary experiments (not shown).

In addition, in order to calculate the K_{cat}, I also needed to know the amount of active CYP76F112 added in every incubation mix. For this purpose, I dosed the P450s present in the CYP76F112 microsomal solution by using the differential CO spectrum method. The spectrum I obtained showed a characteristic peak around 450nm, which indicates the presence of functional P450s in the solution (**Figure 30**). The height of this peak allowed me to calculate the concentration of functional CYP76F112 present in the microsomal solution used for the kinetic assays and therefore the quantity of active enzyme added in every incubation (**Figure 30**). The details of the calculation can be found in the materials and methods section (**Chapter VII**, **B.2.e**).

Microsomal solution	ΔDO	Dilution factor D	Concentration of functional P450s C=ΔDO*D/ε (mM)
CYP76F112	2,46E-03	40	1,08E-03
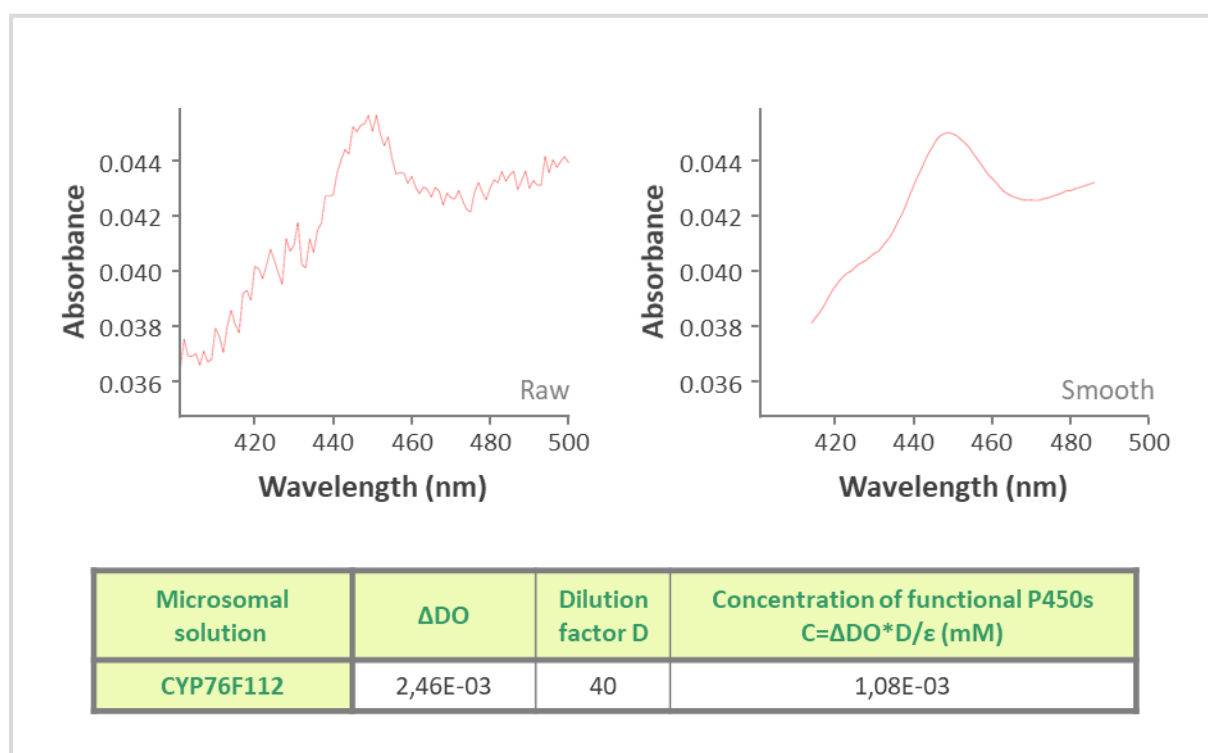
Figure 30 Differential CO spectrum recorded with the microsomes collected from the yeast producing CYP76F112. The raw spectrum has been smoothed prior to measuring the peak height, which corresponds to the ΔDO used to calculate the concentration of functional P450s in the microsomal solution. The dilution factor D corresponds to the dilution of the initial microsomal solution used to record the CO spectrum. ε = 91mM^{-1}.cm^{-1} (see **Chapter VII**, **B.2.e**).

Then, the kinetic parameters were determined. Briefly, for every 1 mL incubation, 0.54 pmol of functional CYP76F112 were incubated in presence of 10 to 300 nM of DMS for 1min, before being stopped by the addition of concentrated HCl. The reaction mixtures were concentrated, and the

quantity of marmesin was quantified via UHPLC-MS analyses. The results are given as a quantity of marmesin (mol) formed by minute by mol of functional P450s – which corresponds to the specific activity of CYP76F112 (**Supp. Table 4**). The whole experiments could not be repeated in triplicates, but was performed in duplicate.

Finally, a Michaelis-Menten saturation curve was plotted. It permitted to determinate the kinetic parameters associated to the conversion of DMS into marmesin by CYP76F112 (**Figure 31**). The apparent K_m value of CYP76F112 for DMS was of 32.2 ± 3.9 nM, which indicated a very high affinity of the substrate for the enzyme. The apparent K_{cat} associated to this conversion was of 22.1 ± 0.9 min^{-1}. The catalytic efficiency K_{cat}/K_m was therefore of 687 µM^{-1}.min^{-1} (or 11.5 µM^{-1}.s^{-1}).

Enzyme	K_m (nm)	K_{cat} (min^{-1})	K_{cat}/K_m (µM^{-1}.min^{-1})
CYP76F112	32.2 ± 3.9	22.1 ± 0.9	687
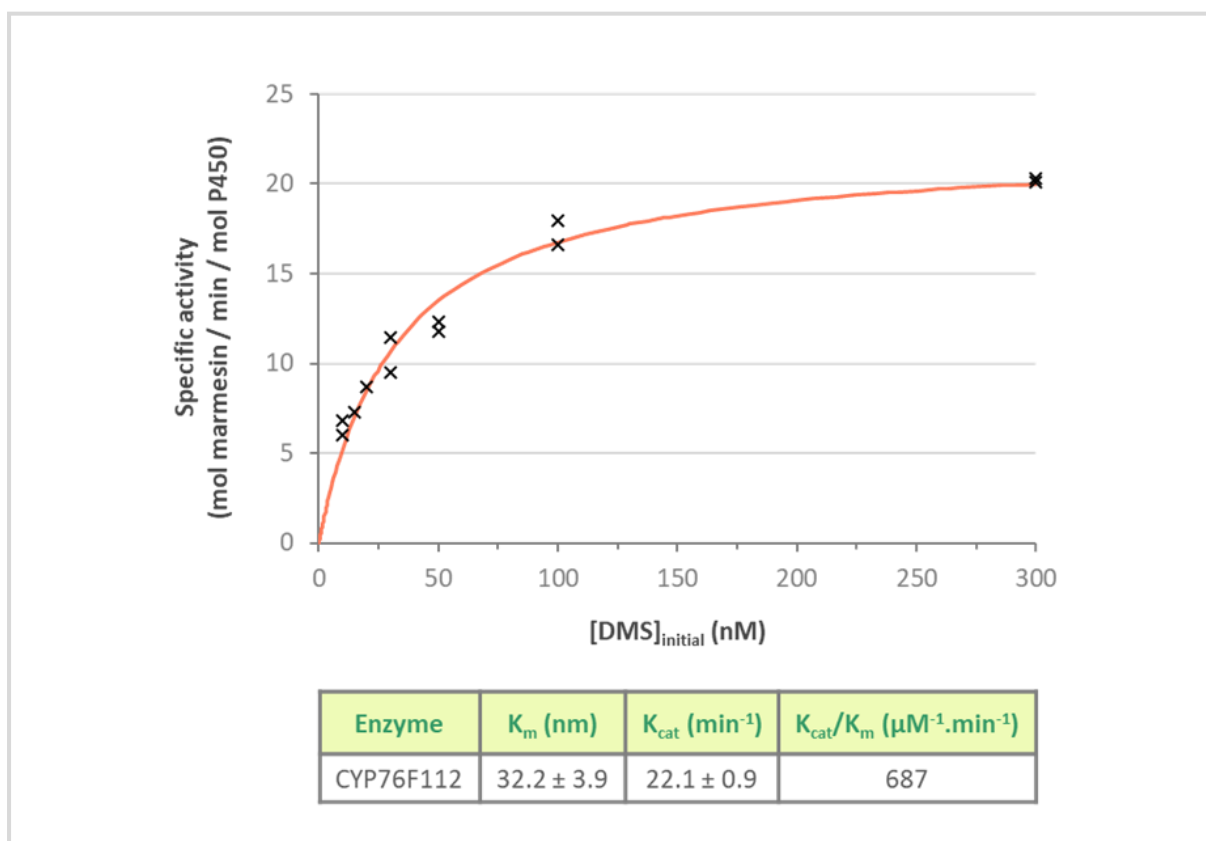
Figure 31 Determination of the kinetic parameters associated to the conversion of DMS into marmesin by CYP76F112. Specific activity of CYP76F112 in the presence of various initial substrate concentrations. The experimental points are marked by black crosses (**Supp. Table 4**) The Michaelis-Menten model curve, plotted on SigmaPlot to fit the experimental data, is in light red. The apparent K_m and K_{cat} values were determined in SigmaPlot, by using the Michaelis-Menten equation fitting the experimental data.

In brief, CYP76F112 is a P450 which is characterised by a high specificity of substrate and reaction, and interact with a very high affinity with DMS, its presumed unique substrate.

D.2.b. The cases of CYP82J18 and CYP81BN4

Among the 40 substrates tested during the functional screening, CYP82J18 and CYP81BN4 were only able to metabolise auraptene and cnidilin, respectively. Yet, both enzymes led to the production of several products, which indicates that the reactions they catalysed were not as specific as the one catalysed by CYP76F112. In addition, the reaction products obtained with CYP82J18 and CYP81BN4 seem to correspond, respectively, to auraptene and cnidilin derivatives that have been hydroxylated, dihydroxylated, dihydroxylated and desaturated. It is thus highly possible that the products corresponding to the simple hydroxylation of auraptene and cnidilin can be accepted again as substrates by CYP82J18 and CYP81BN4, respectively. Therefore, CYP82J18 and CYP81BN4 seem to be enzymes with an only moderate specificity.

The optimal pH and temperature were determined for CYP82J18. The different metabolization products showed different patterns when the temperature and especially the pH varied. Yet, the maximal activity was measured for a temperature of 22°C, and the optimal pH ranged around 7 to 9 (**Supp. Figure 2**).

The optimal conditions have not been determined for CYP81BN4, nor the kinetic parameters of both CYP82J18 and CYP81BN4. Indeed, as these two candidates were less interesting than CYP76F112 for the study of the furanocoumarin pathway, I preferred to concentrate my efforts on the analysis of the marmesin synthase. However, it can still be noted that, as for CYP76F112, CYP82J18 and CYP81BN4 can be frozen and defrost without losing their activity.

E. Discussion: another P450 in the pathway

E.1. An approach that proved itself effective

In this chapter, my first objective was to identify new *F. carica* genes involved in the furanocoumarin biosynthesis pathway. As furanocoumarins are preferentially found in the leaves of *F. carica*, it has been hypothesised that the genes involved in their biosynthesis might also be preferentially expressed in the leaves. By using a differential RNA-seq library, I identified 10 putative P450s that were preferentially expressed in the latex of *F. carica* petioles.

Such a differential approach has already been proven efficient for the identification of new genes involved in furanocoumarin biosynthesis. For instance, previous studies based on *Ammi majus* cell cultures induced with *Phytophtora sojae* extracts led to the identification of CYP71AJ1, a psoralen synthase. In this older study, the method was based on a differential display RT-PCR (Larbat et al. 2007). More recently, an approach identical to that of the present study was successfully used for the identification of FcPT1, which was preferentially expressed in *F. carica* leaves (Munakata et al. 2020).

Among the 10 P450 candidates, 3 new enzymatic activities were identified (**Table 5**). CYP76F112 was able to convert DMS into marmesin, while CYP82J18 and CYP81BN4 respectively metabolised auraptene and cnidilin into multiple products. Two of these enzymes seem to be preferentially

expressed in the *F. carica* tissues that are rich in furanocoumarins: indeed, according to the RNA-seq library, CYP76F112 is expressed at 89% in the latex of petioles (**Table 5**), compared to the latexes of trunks and fruits. This is consistent with the higher levels of furanocoumarins found if the latex of leaves (Kitajima et al. 2018) and the fact that the conversion of DMS into marmesin represents an essential step leading to the production of furanocoumarins. Similarly, CYP81BN4 is expressed at 85% in the latex of petioles, compared to those of trunks and fruits (**Table 5**): this is also consistent with the fact that it can hydroxylate cnidilin, a linear furanocoumarin. On the contrary, CYP82J18 might be "only" expressed at 41% in petiole latex (**Table 5**). Yet, this enzyme was shown to metabolise auraptene, a simple coumarin that is not an intermediate of the furanocoumarin pathway and that might have other repartition patterns in the fig tree.

Table 5 **Summary of all P450 candidates.** The colours refer to the P450s families.

Candidate	Relative expression in petiole latex	Amplifi-cation from the cDNA?	Identity b/w the amplified and expected sequences	Standardized name	Length of the protein	Weight of the protein	Heterologous expression	Metabolized substrate
1	83,7%	Yes	84% – stop codons – elimination of the candidate					
2	80,4%	Yes	100%	CYP76F110	504 AA	56.68 kDa	Correct	
3	62,0%	Yes	99%	CYP81B114	497 AA	56.72 kDa	Correct	---
4	84,6%	Yes	99%	CYP81BN4	497 AA	56.10 kDa	Low	Cnidilin
5	89,5%	Yes	99%	CYP76F111	506 AA	56.62 kDa	Correct	---
6	44,7%	Yes	100%	CYP81CA1	502 AA	57.39 kDa	Very low	---
7	89,1%	Yes	100%	CYP76F112	504 AA	56.95 kDa	High	DMS
8	85,9%	No: elimination of the candidate						
9	40,7%	Yes	99%	CYP82J18	520 AA	58.78 kDa	Correct	Auraptene
10	44,3%	Yes	99%	CYP88A103	497 AA	56.95 kDa	Very low	---

E.2. Physiological or promiscuous activities?

Many P450s are related to the adaptation of plants to their environment, by being involved in the production of specialised metabolites, or sometimes, by being able to detoxify xenobiotic compounds such as herbicides. This brings us to another essential question: are DMS, auraptene and cnidilin really present in the fig tree, and can they be the physiological substrates of CYP76F112, CYP82J18 and CYP81BN4? In other words, the activities described here are they the primary activities of the P450s or are they only opportunistic activities?

E.2.a. CYP76F112: a physiological marmesin synthase activity?

The simplest case is that of CYP76F112. To the best of my knowledge, DMS has never been reported in phytochemical analysis of any species from the *Ficus* genus. Yet, marmesin and other furanocoumarins such as psoralen and bergapten have been found in *F. carica* leaves (Innocenti et al.

1982; Oliveira et al. 2012). So, if these compounds exist in the fig tree, the DMS – which is their direct precursor – might also be present. In addition, *F. carica* possesses and produces FcPT1, an enzyme that can convert umbelliferone into DMS (Munakata et al. 2020). So, it is more than likely that DMS does exist in *F. carica* leaves, which is where CYP76F112 is preferentially expressed: this makes DMS a realistic physiological substrate for CYP76F112.

Moreover, CYP76F112 is a very selective and efficient enzyme that has a high specificity and affinity for the DMS, which is also in favour of a physiological function. It should be highlighted that, thanks to its apparent K_m value of 32.2 ± 3.9 nM, CYP76F112 can stay relatively active under very low concentrations, at least under *in vitro* conditions. Thus, even if the actual concentration of DMS in the fig tree is very low – which would explain why it has never been detected so far – it might still be efficiently converted into marmesin by CYP76F112.

Compared to the other P450s that have already been described in the furanocoumarin pathway, CYP76F112 is by far the one with the higher affinity for its substrate: for instance, the apparent K_m values of CY71AJ1-4, CYP71AZ1,4,6 and CYP82D64 for their respective substrates ranged from 0.54 µM (CYP71AJ2) to 13.1 µM (CYP71AZ1), which is 10 to more than 100 times higher than the apparent K_m value of CYP76F112 (Larbat et al. 2007, 2009; Dugrand-Judek 2015; Krieger et al. 2018; Limones-Mendez et al. 2020). Similarly, FcPT1, the *F. carica* prenyltransferase that converts umbelliferone into DMS, is associated to an even higher apparent K_m value of 35 µM (Munakata et al. 2020).

On the contrary, the specific activity of CYP76F112 (22.1 ± 0.9 min^{-1}) was approximatively 10 times lower than that of CYP71AJ1 (340 min^{-1}) and CYP71AJ4 (112 min^{-1}) (Larbat et al. 2007, 2009). The value of this constant was not determined for the other P450s involved in the furanocoumarin biosynthesis pathway. Yet, because of its very high affinity, the catalytic efficiency of CYP76F112 (K_{cat}/K_m = 687 µM^{-1}.min^{-1}) was still higher than that of CYP71AJ1 and CYP71AJ4 (226 and 55 µM^{-1}.min^{-1}, respectively). This makes CYP76F112 the most efficient P450 of the furanocoumarin biosynthesis pathway described so far. The comparison of CYP76F112 with other P450s which are not involved in the furanocoumarin biosynthesis pathway is discussed later.

As a result, in the fig tree, DMS may be converted into marmesin by CYP76F112 as soon as it is produced by FcPT1, without being accumulated. This would make DMS a fleeting intermediate, whose production from umbelliferone might be a bottleneck for the biosynthesis of downstream toxic furanocoumarins.

For all these reasons, we can confidently assume that the conversion of DMS into marmesin corresponds to the physiological activity of CYP76F112. Yet, to get the evidences required to state it with certainty, additional experiments would have to be performed. For instance, it would be possible to study the consequences of the silencing of CYP76F112 on the accumulation of furanocoumarins in *F. carica*, by using techniques such as stable transformations or virus-induced gene silencing (VIGS).

E.2.b. CYP81BN4 and CYP82J18: promiscuous activities?

As far as I know, auraptene and cnidilin have never been reported in the *Ficus* genus. But contrary to DMS, their closest precursors and potential derivatives have not been identified in *Ficus* species either,

which makes them unlikely physiological substrates for CYP82J18 and CYP81BN4. Nonetheless, as cnidilin has been described in the *Dorstenia* genus, which is also from the Moraceae family (Franke et al. 2001; Ngadjui and Abegaz 2003), it is possible that it does exist in the fig tree but has not been identified or detected so far. In any cases, additional data would be required to go further in the discussion: for instance, detecting auraptene or cnidilin in the *Ficus* genus would be in favour of a physiological activity. In addition, the chemical characterisation of the metabolization products of CYP81BN4 and CYP82J18 might help to identify putative derivatives that might be present in the fig tree. And without this kind of information, the expression patterns of CYP81BN4 and CYP82J18 cannot be linked to the repartition of their respective substrates in *F. carica*.

Among all the substrates that were tested, CYP81BN4 only metabolised cnidilin while CYP82J18 only metabolised auraptene. Yet, these reactions were not very specific since they led to the formation of multiple products. Therefore, we do not have any evidence which would suggest that the metabolizations of auraptene and cnidilin by CYP82J18 and CYP81BN4 correspond to the physiological activities of these enzymes. These reactions might thus only be opportunistic. However, these are only assumptions that cannot be confirm without additional data or analyses.

E.3. New P450 families involved in the furanocoumarin pathway

Thus far, the P450s identified in the biosynthesis of furanocoumarins belong to the CYP71 family in the Apiaceae and to the CYP82 family in the Rutaceae. The marmesin synthase I identified belongs to the CYP76 family, which makes it the first known CYP76 of the pathway. Similarly, and even if its activity may only be promiscuous, CYP81BN4 is the first known CYP81 that can accept a furanocoumarin as a substrate. On the contrary, CYP82J18 belongs to the CYP82 family that includes CYP82D64, a xanthotoxin hydroxylase (Dugrand-Judek 2015; Limones-Mendez et al. 2020). Yet, CYP82J18 has only been shown to metabolise a simple coumarin, which might not even be its physiological substrate. These three enzymes therefore offer some promising evolutionary perspectives for the emergence of the furanocoumarin pathway.

E.3.a. The CYP76 family: from terpenoids to furanocoumarins?

E.3.a.1. General overview of the CYP76 family

Several enzymes from the CYP76 family have already been described. Many of them are involved in the terpenoid metabolism, and they can have unique or multiple functions. In addition, the diversification and expansion of some CYP76 subfamilies seem to be species-specific. For instance, in Lamiaceae, promiscuous enzymes from the CYP76AH and CYP76AK subfamilies lead to the biosynthesis of various labdane diterpenes (Bathe and Tissier 2019). Similarly, the CYP76C subfamily seems to be specific to Brassicaceae, and two of its members have been associated with monoterpenol metabolism (Höfer et al. 2014). The diversity of substrates metabolised by the CYP76 family is shown in **Figure 32**.

The various terpenoids that can be oxidised by members of the CYP76 family include monoterpenoids such as linalool (CYP76C1 – *A. thaliana*) (Boachon et al. 2015), diterpenes such as 13*R*-manoyl oxide (CYP76AH15 – Lamiaceae) (Pateraki et al. 2017; Forman et al. 2018), or sesquiterpenes such as santalene and bergamotene (CYP76F36-42 – *Santalum album*) (Diaz-Chavez et al. 2013). Among
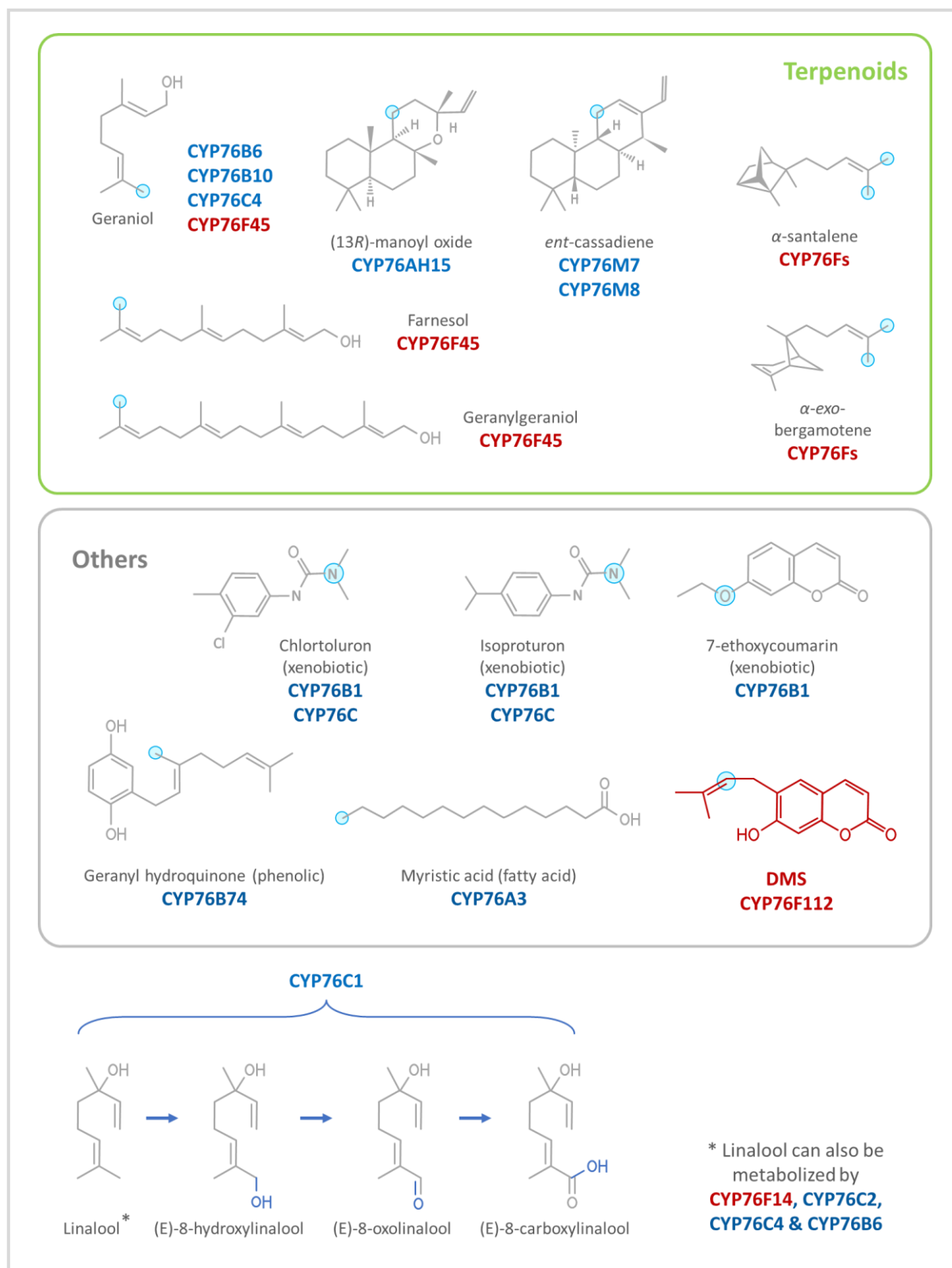
Figure 32 Some substrates metabolised by enzymes from the CYP76 family. The blue circles show the position of the P450-mediated hydroxylations / reactions. The members of the CYP76F subfamily and the DMS are highlighted in red. The successive oxidations catalysed by CYP76C1 are detailed. Based on (Batard et al. 1998; Collu et al. 2001; Imaishi and Ishitobi 2008; Swaminathan et al. 2009; Wang et al. 2010, 2012, 2018; Sung et al. 2011; Diaz-Chavez et al. 2013; Höfer et al. 2013; Hofer et al. 2014; Sintupachee et al. 2015; Boachon et al. 2015; Pateraki et al. 2017; Bathe and Tissier 2019).

terpenoids, one well known example is that of geraniol, which can be hydroxylated by CYP76B6 (*Catharanthus roseus*), CYP76B10 (*Swertia mussotii*), CYP76C4 (*A. thaliana*) and CYP76F45 (*Croton stellatopilosus*) (Collu et al. 2001; Wang et al. 2010; Sung et al. 2011; Höfer et al. 2013; Sintupachee et al. 2015). Three CYP76 subfamilies have therefore been associated to the same reaction in different plant species. Some of these enzymes also have multiple functions: as an example, in addition to geraniol, CYP76F45 hydroxylates farnesol and geranylgeraniol (Sintupachee et al. 2015). Among the other CYP76s with multiple functions, CYP76C1 (*A. thaliana*) catalyses three sequential oxidations on linalool and its successive oxides (Boachon et al. 2015).

While the CYP76 family has mainly been associated to terpenoids, other substrates have also been described. For instance, some CYP76s can catalyse the ω-hydroxylation of fatty acids such as myristic acid (CYP76A3 – *Petunia hybrida*), capric or lauric acids (CYP76B9 – *Petunia hybrida*) (Imaishi and Petkova-Andonova 2007; Imaishi and Ishitobi 2008). On its side, CYP76B74 (*Arnebia euchroma*) is involved in shikonin biosynthesis (phenolic compounds) by catalysing the 3''-hydroxylation of geranyl hydroquinone (Wang et al. 2018).

Finally, some CYP76s are known to metabolise xenobiotics. For example, CYP76B1 (*Helianthus tuberosus*) metabolises a large range of xenobiotics such as alkoxyresorufins, alkoxycoumarins, and some herbicides of the class of phenylureas. In particular, it can dealkylate 7-ethoxycoumarin (Batard et al. 1998; Robineau et al. 1998), a synthetic molecule that is similar to natural coumarin and which has a structure similar to DMS (**Figure 32**). Another example is that of the CYP76Cs, which seem to be Brassicaceae specific: CYP76C1, CYP76C2 and CYP76C4 have been reported to metabolise a large subset of herbicides belonging to the class of phenylurea such as isoproturon and chlortoluron, and leading to the formation of multiple products. Yet, these CYP76Cs can also metabolise monoterpenols such as geraniol (CYP76C4), nerol (CYP76C2, CYP76C4) or linalool (CYP76C1, CYP76C2, CYP76C4) (**Figure 32**). These CYP76Cs are therefore enzymes with variable degrees of promiscuity that can metabolise both xenobiotics and natural compounds (Höfer et al. 2014).

E.3.a.2. Focus on the CYP76F subfamily and the marmesin synthase

If 7-ethoxycoumarin is similar to DMS, it is metabolised by a P450 from the CYP76B subfamily, not from the CYP76F subfamily. Moreover, to the best of my knowledge, the only members of the CYP76F subfamilies described so far are involved in terpenoid metabolism (all the CYP76F substrates I found in the literature are shown in **Figure 32**). The substrate metabolised by CYP76F112 is therefore quite different from those of others CYP76Fs. Yet, like the other CYP76F substrates, DMS is a prenylated compound, and it is oxidised by CYP76F112 on its prenyl group (**Figure 33**). So, the enzymes belonging to the CYP76F subfamily described so far seem to be characterised by their ability to oxidise prenyl groups – even though we would need to functionally characterise more CYP76Fs to confirm this. It is even possible that the CYP76F family had been recruited for the conversion of DMS into marmesin thanks to its ability to metabolise prenylated compounds.

Lastly, it can also be noted that the apparent K_m values measured for other CYP76Fs ranged from 0.066 μM (CYP76F45) to 20-150 μM (CYP76F36-42) (Diaz-Chavez et al. 2013; Sintupachee et al. 2015). Consequently, if the very high affinity of the marmesin synthase is still unique in the furanocoumarin biosynthesis pathway, it is not the only CYP76F that can function with very low concentrations.
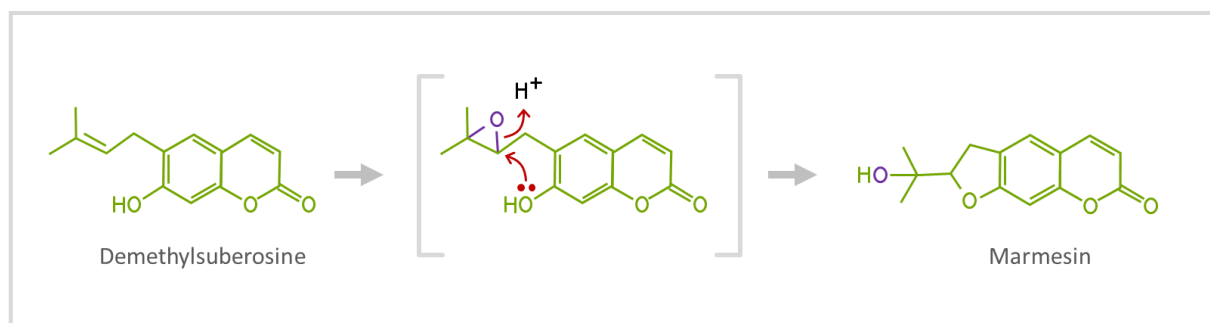
Figure 33 Proposed mechanisms for the conversion of DMS into marmesin. Adapted from Mizutani and Sato (2011). Also see **Figure 19** from **Chapter I**.

Ultimately, CYP76F112 is the first known marmesin synthase, as well as the first CYP76 described in the furanocoumarin biosynthesis pathway. This could suggest that other marmesin synthases from other plant species may also belong to the CYP76 family. Yet, CYP76F112 is also the first moraceous P450 involved in furanocoumarin biosynthesis, which could also suggest that, in *Ficus*, the P450s involved in the furanocoumarin pathway may have all been recruited in the CYP76 family. This raises an interesting question: is CYP76F112 a CYP76 because of its marmesin synthase activity, or because it has been identified in the *Ficus* genus? Here, two hypotheses linked to the emergence of CYP76F112 and furanocoumarins in higher plants can be advanced.

The first hypothesis would be supported by a unique emergence of the furanocoumarin pathway in higher plants, followed by multiple losses: if this pathway emerged only once, an ancestral marmesin synthase would have been present in the most recent common ancestor of all furanocoumarin producing plants. Under this scenario, it may thus be hypothesised that, in other plant families, the marmesin synthase would also belong to the CYP76 family. Such scenario has already been described for some ancient pathways that are well conserved, even between phylogenetically distant species. For instance, in the biosynthesis pathway of *p*-hydroxy lignin in higher plants, the conversion of *trans*-cinnamic acid into *p*-coumaric acid is catalysed by cinnamate 4-hydroxylases (C4H) that belongs to the CYP73A family (Mizutani et al. 1993; Teutsch et al. 1993; Mizutani and Ohta 2010). Yet, contrary to the CYP73As that are well conserved in higher plants (Nelson and Werck-Reichhart 2011), studies have shown that the diversification and expansion of the CYP76 subfamilies are often species-specific (Höfer et al. 2014). Moreover, I performed a BLAST search on the partial databases available for some furanocoumarin-producing plants from the Rutaceae and Apiaceae families, but it did not reveal the presence of any CYP76F that would be homologous to CYP76F112. This makes this first hypothesis of a unique emergence of furanocoumarins very unlikely, and brings us to the second one.

The second hypothesis would be in favour of a multiple emergence of the furanocoumarin pathway in higher plants. The repeated but independent emergence of enzymes with the same function in distant taxa is quite common, and usually referred to as "convergent evolution" (Pichersky and Gang 2000; Weng 2014). As an example, the release of hydrogen cyanide (a defensive compound) from cyanogenic glycosides can be catalysed by hydroxynitrile lyases which have emerged multiple times in different plant taxa (Hickel et al. 1996). Similarly, it could be assumed that CYP76F112 results of a lineage-specific diversification, and is therefore specific of the *Ficus* genus or the Moraceae family. In this scenario, CYP76F112 would have no phylogenetic link with the marmesin synthases from other plant

families – which may thus belong to other P450 (sub)families or clans. In addition, if CYP76F112 originated from a more recent diversification of CYP76F, there would be a chance that other *Ficus* P450s involved in the furanocoumarin pathway also belong to the CYP76F subfamily. However, to provide further information in favour of one hypothesis or the other, a phylogenetic analysis of CYP76F112 would be required. This will be done in the next chapter.

E.3.b. The cases of CYP81BN4 and CYP82J18

Given their relatively limited interest in the context of this study, CYP81BN4 and CYP82J18 will be discussed more briefly.

It seems that the CYP81 family members characterised so far have been mostly associated to flavonoid metabolism (Akashi et al. 1998; Shimada et al. 2000; Liu et al. 2003) and the detoxification of xenobiotics (Cabello-Hurtado et al. 1998; Yamada et al. 2000; Pan et al. 2006; Zhang et al. 2006). Occasionally, they have also been associated with other stresses such as salinity tolerance (Wang et al. 2020), or the biosynthesis of other compounds such as (+)-sesamin (Ono et al. 2006) or fatty acids (Cabello-Hurtado et al. 1998). But as far as I know, no P450 from the CYP81BN subfamily had been characterised before, which makes CYP81BN4 the first one to be described. As long as the physiological function of CYP81BN4 is unsure, it is prudent not to rely on this enzyme to discuss the emergence of the furanocoumarin pathway in higher plants. Yet, as the CYP81 family have been associated to the biosynthesis of numerous specialised compounds, to stress responses, and to the detoxification of many xenobiotics, it would not be particularly surprising to find out that the CYP81BN are indeed involved in the biosynthesis of defensive furanocoumarins. This might be deeper investigated elsewhere.

Enzymes from various CYP82 subfamilies have been described to metabolise alkaloids (Siminszky et al. 2005; Beaudoin and Facchini 2013; Dang and Facchini 2014), terpenoids (Lee et al. 2010) and phenolic compounds (Berim and Gang 2013), including (furano)coumarins. Indeed, CYP82C4 (*A. thaliana*) has been described to hydroxylate fraxetin, a simple coumarin (Rajniak et al. 2018), while CYP82C2 (*A. thaliana*) and CYP82D64 (*Citrus paradisi*) can both hydroxylate xanthotoxin, a linear furanocoumarin (Kruse et al. 2008; Dugrand-Judek 2015; Limones-Mendez et al. 2020). But here again, to the best of my knowledge, no P450 from the CYP82J subfamily has been described so far, making CYP82J18 the first one. As for CYP81BN4, in absence of additional information that would confirm its physiological activity, the emergence of the furanocoumarin pathway cannot be discussed with CYP82J18.

E.4. Conclusion and future perspectives

To summarise, the differential approach used in this study proved itself very useful since it allowed me to identify three enzymes with original activities. CYP76F112, CYP81BN4 and CYP82J18 constitute the first P450s from *F. carica*, and even from the Moraceae family, that can metabolise coumarins and linear furanocoumarins. CYP81BN4 and CYP82J18, which were respectively shown to metabolise cnidilin and auraptene, are the first enzymes from their respective subfamilies to be characterised. However, we do not have enough evidences to state if these reactions correspond to the physiological or opportunistic activities of the two enzymes. On the contrary, CYP76F112 was shown to have a marmesin synthase activity, which is likely its *in planta* physiological function. This enzyme is highly

specific: it interacts with a high affinity with its assumed unique substrate (DMS) and converts it into a single product (marmesin). This marmesin synthase therefore complete the set of four enzymes that allow to convert the common coumaric acid into toxic psoralen (**Figure 34**). With the discovery of a marmesin synthase, my first objective was thus achieved, and it allowed for new uses of the furanocoumarin biosynthesis pathway, as it will be shown in **Chapter V**.
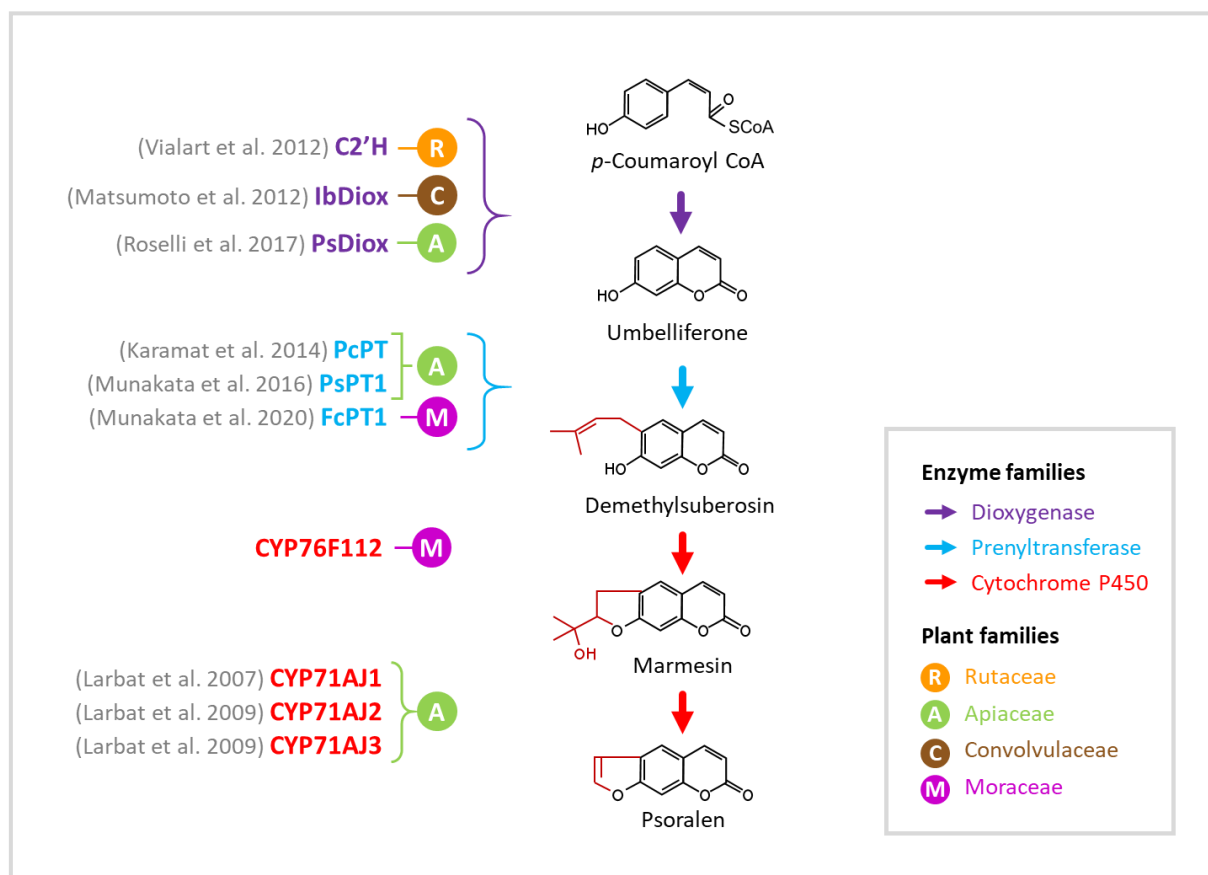
Figure 34 Conversion of coumaric acid into psoralen, a toxic furanocoumarin

Moreover, CYP76F112 is the first CYP76 known to be involved in the furanocoumarin biosynthesis, which could provide new information in favour of a unique or multiple emergence(s) of furanocoumarins in higher plants. For instance, if *CYP76F112* – which is a moraceous gene – has an ancient enough origin, it might possess homologues in other furanocoumarin-producing plant families, which would be in favour of a unique emergence of the furanocoumarin pathway. On the contrary, if it is a more recent gene that is specific of a restricted taxon, this would be in favour a multiple emergence of the furanocoumarin pathway. Yet, further phylogenetic analyses were required to confirm one hypothesis or the other: this will therefore be the objective of **Chapter III**. Phylogenetics may also allow us to identify new candidate genes that might be involved in furanocoumarin biosynthesis. Indeed, we already know that *F. carica* possesses two other CYP76Fs: CYP76F110 and CYP76F111. As I could not find any activity for any of them, they may be involved in other pathways, or they were not active in the conditions of the incubations. Yet, the fig tree or other plants might possess other CYP76Fs that may be interesting candidates to find other enzymes involved in furanocoumarin biosynthesis.

Chapter III
Once upon a P450

CHAPTER III.

ONCE UPON A P450

~

Evolutive stories of CYP76F112,
CYP81BN4 and CYP82J18

A. BACK TO THE PAST: INTRODUCTION AND STRATEGY

A.1. Objective and strategy

The identification of CYP76F112 as the fig-tree marmesin synthase suggests new hypotheses that can be tested about the emergence of furanocoumarins in higher plants. A way to go deeper into details would be to perform a detailed gene-family phylogenetic analysis to help to answer questions such as: when did this gene appear? Which genes did it evolve from? In which taxon? How did its primary sequence, and particularly its substrate recognition sites, evolve to allow the emergence of the marmesin synthase activity? In other words, what is its evolutionary story? A similar analysis performed on CYP81BN4 and CYP82J18 might also help to elucidate their respective physiological functions. For instance, if the physiological function of CYP81BN4 is really to oxidise cnidilin, its evolutionary history may mirror the one of CYP76F112. Another possibility could be that *CYP81BN4* possesses homologues in many plants that do not produce furanocoumarins, which would be in favour of the opportunistic metabolization of cnidilin.

In this chapter, my objective was to reconstruct the evolutionary history of CYP76F112, CYP81BN4 and CYP82J18 in order to determine how, where and when they phylogenetically appeared. First, I performed an extensive data mining to identify sequences that are similar to those of the three genes of interest. Second, I aligned these sequences and generated phylogenetic gene trees.

To ensure an easy understanding of the following evolutive stories, a quick introduction to gene and enzyme evolution might be appreciated. Therefore, before starting the analysis of my favourite genes, let us make a short state of the art on the various evolution possibilities.

A.2. Prerequisite: how do genes evolve and diversify?

As the biosynthesis of plant metabolites requires multistep metabolic pathways, the diversification of phytochemicals implies the emergence of new enzymes. In general, such diversification does not require the creation of completely new protein folds. Instead, new functions often arise via the recruitment of pre-existing enzymes involved in other pathways (Weng 2014).

A.2.a. Gene duplication

The main evolutionary mechanism allowing the emergence of new enzymatic activities is the creation and recruitment of new genetic variation. For example, the functional divergence of single-copy genes or the duplication of pre-existing genes followed by their divergence. It has been hypothesised that gene duplication is a major force driving the recruitment of new genes in specialised metabolic pathways. Duplication generates two copies of an ancestral gene, thus allowing one copy to retain the ancestral function while the other might diversify and, sometimes, give rise to a new catalytic activity. In the end, selection and/or genetic drift operate to fix (or not) the gene in the population (Prince and Pickett 2002; Ober 2005; Weng and Noel 2012; Weng et al. 2012; Weng 2014; Moghe and Last 2015).

Genes can be duplicated via various mechanisms such as retroduplication or mechanisms based on transposable elements. In plants, the majority of duplicates emerge via whole-genome duplications or tandem duplications, illustrated in **Figure 35** (Zhang 2003; Panchy et al. 2016). As its name implies, a whole-genome duplication event results from polyploidisation through the doubling of a complete genome either from the same species (autopolyploidy) or by the hybridisation of two species (allopolyploidy). The primary mechanism for polyploidisation is thought to be the production of unreduced gametes. Yet, in rare cases, a somatic doubling of chromosomes can occur in the zygote. Whole-genome duplications are uncommon in animals, but frequent in plants (Prince and Pickett 2002; Moghe and Shiu 2014; Panchy et al. 2016). On its side, tandem duplication is a local event that can duplicate a part of a gene, an entire gene, or a bigger fragment. It arises from unequal crossing-over events that happen during meiosis, from multiple repeats during DNA repair, or from intrachromosomal rearrangements. These duplications are termed "tandem" because the duplicated sequences are usually arranged one after the other on the chromosome, in a head-to-tail orientation. The duplicates resulting from intrachromosomal rearrangements might also be organised with an opposite head-to-head orientation (Meyers et al. 2003; Zhang 2003; Panchy et al. 2016). One of the consequences of tandem duplication is the formation of physical gene clusters: indeed, tandem duplicates are usually found at adjacent loci, and repeated tandem duplications can generate clusters of numerous paralogues. Such clustering events seem to be a common feature of the expansion of specialised metabolic genes (Zhang 2003; Ober 2005; Weng and Noel 2012; Panchy et al. 2016).

A.2.b. The fate of duplicated genes

When a gene is duplicated, the two duplicates first experience a "brief" period of functional redundancy and relaxed selection (several generations). Then, various fates await them. In most cases, one copy of the duplicated gene experiences a strong purifying selection which retains the ancestral function, while the other copy is silenced or lost within a few million years, via pseudogenisation and/or deletion from the genome (chromosomal remodelling, locus deletion) (Lynch 2000; Prince and Pickett 2002; Ober 2005; Weng et al. 2012; Moghe and Shiu 2014; Panchy et al. 2016).

In rare cases, duplicated genes can both be retained if they provide a beneficial diversification: for instance, through the evolution of the coding sequence, gene expression divergence, or network and epistatic interactions (Ober 2005; Weng 2014; Moghe and Shiu 2014; Moghe and Last 2015). Here, I will mostly focus on coding sequence variations that lead to the evolution of an ancestral enzymatic activity, through the sub-functionalisation and neo-functionalisation mechanisms (**Figure 35**). In the

sub-functionalisation process, the ancestral activity is partitioned between the two copies of the gene: each copy acquires independent mutations and loses a complementary part of the ancestral activity, until both copies are required to maintain the original function. In the neo-functionalisation process, one copy retains the ancestral activity while the other acquires a new function through an alteration in the coding or the regulatory sequence. If the new function of the duplicate is beneficial, it might be fixed by positive selection (Lynch 2000; Prince and Pickett 2002; Moghe and Last 2015).
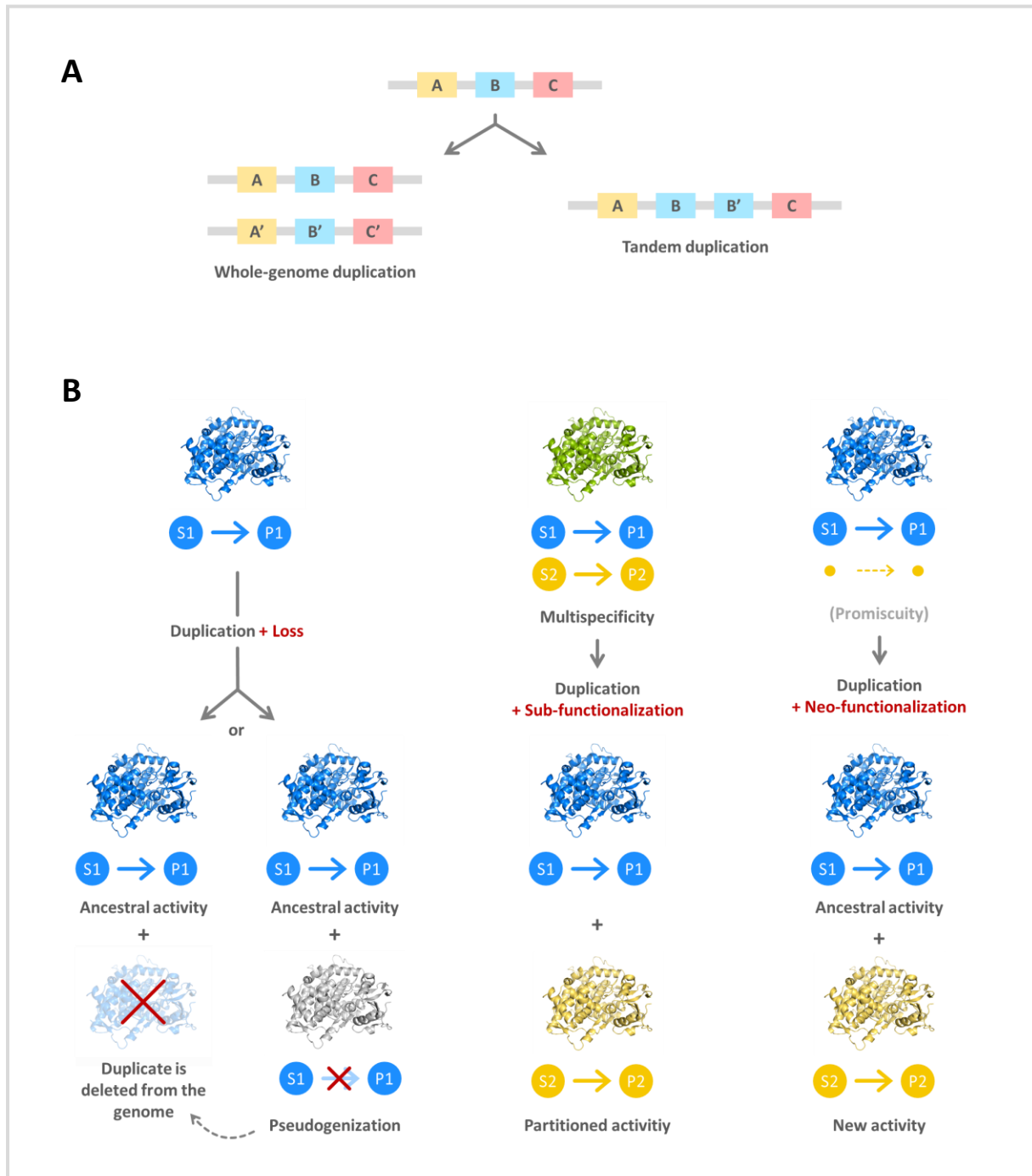
Figure 35 Duplication and diversification. A Main event contributing to duplicate plant genomic sequences. **B** Various fates of duplicated genes. The most common fate is loss, but a change in the coding or the regulatory sequence can also lead to sub- or neo-functionalisation.

It should be noted that the number of mutations required to alter the activity of an enzyme is variable. Indeed, enzymes can accumulate neutral changes, but one or a few mutations might also be enough to change the substrate specificity of an enzyme, its structure, or its catalytic mechanism (Pichersky and Gang 2000; Weng and Noel 2012; Moghe and Last 2015).

A.2.c. All duplicates are not created equal

The successful recruitment of a newly duplicated gene could be rare. However, factors such as the duplication mechanism can influence the chance of such a recruitment event occurring. For instance, all the duplicates created by whole-genome duplications tend to return to single-copy status – which might be partially explained by the need to maintain stoichiometry, and other factors such as expression level or network connectivity (Chae et al. 2014; Moghe and Last 2015). On the contrary, the fate of a tandem duplicates is partly linked to the metabolism to which it is associated: primary metabolic genes tend to revert back to single-copy status, while specialised metabolic genes are more prone to be retained as duplicates (Weng et al. 2012; Chae et al. 2014; Moghe and Last 2015). Indeed, as primary metabolism is associated to essential and highly conserved functions, the corresponding genes are subjected to a constant and strong selective pressure (Weng and Noel 2012; Weng et al. 2012; Weng 2014; Moghe and Last 2015). Conversely, in specialised metabolism, some enzymes might be more or less essential depending on environmental fluctuations: for instance, a gene involved in the biosynthesis of a no-longer effective defensive chemical can evolve more easily to allow the synthesis of a more effective defensive compound. So, compared to their primary metabolic counterparts, the genes involved in specialised metabolism seem to have a higher potential for evolution (Weng and Noel 2012; Weng et al. 2012; Weng 2014; Chae et al. 2014; Leong and Last 2017).

In addition, some enzymatic features such as promiscuity might influence the retention rate of duplicate genes, and therefore play a key role in enzyme evolution. Promiscuity has been defined as the ability of an enzyme to catalyse (an)other reaction(s) than its primary one. It was reported that enzymes contain various degrees of promiscuity, allowed by variable levels of dynamic flexibility in the active site and the ligand-binding pocket. Enzymatic promiscuity can lead to the production of low-level metabolites, which results in a metabolic noise: usually, these low-level metabolites are selectively neutral, but if the environment changes, they can gain selective advantages that leads to their spreading in the population. Moreover, as some genetic mutations might significantly impact the promiscuous function(s) of an enzyme, without necessarily impacting its primary activity, enzymes might evolve new functions without losing their initial activity – at least in a first time. And indeed, it seems that the mutations that affect promiscuous rather than native functions are a good starting point that leads to the emergence of new activities. Consequently, promiscuity is generally considered as a key stone for the evolution of specialised metabolic enzymes (Aharoni et al. 2005; Tokuriki and Tawfik 2009; Weng and Noel 2012; Weng 2014; Moghe and Last 2015; Leong and Last 2017).

> This bibliographic survey highlights that new enzymatic activities are more likely to occur from tandem duplications followed by functional divergence. As duplicates are generally silenced, successful recruitment events are quite rare. Yet, specialised metabolism genes are usually more prone to be retained and to evolve new functions than their primary metabolic counterparts.

B. Data mining: the P450s coming out of the Nitrogen Fixing Clade

The first step of the phylogenetic gene-family analysis consisted in generating the datasets that have subsequently been used to produce phylogenetic gene trees. To produce the best trees possible, I only worked with nucleotide sequences, which provide more information than their amino acid counterparts. Indeed, as a single amino acid can result from several codons, mutations on the nucleotide sequence do not always impact the amino acid sequence, which means that the evolution of the amino acid sequence does not faithfully reflect the evolution of the nucleotide sequence.

The idea was therefore to perform similarity searches using various genetic databases in order to identify many nucleotide sequences that belong to the same P450 subfamilies than the three enzymes of interest. For this purpose, I successively used *CYP76F112*, *CYP81BN4* and *CYP82J18* as a query to perform BLAST analyses on the genetic resources from various plants found in 1KP (One Thousand Plant Transcriptomes Initiative 2019; Carpenter et al. 2019) and GenBank (Benson et al. 2012).

B.1. Approach: from *Ficus* to the Nitrogen Fixing Clade

CYP76F112, *CYP81BN4* and *CYP82J18* were identified in *Ficus carica*, which belongs to the Moraceae family. The plants that were the most likely to possess homologous sequences of the three genes of interest were thus the plant species that are phylogenetically close to *F. carica*. Consequently, to infer the evolutionary stories of *CYP76F112*, *CYP81BN4* and *CYP82J18*, I started the data mining in other *Ficus* and moraceous species. Then, I progressively widened my search to more distant species: first, I extended it to the Urticaceae and Cannabaceae, which are Moraceae's closest families (**Figure 36**). Then, I included other families from the Rosales order (to which belongs the Moraceae family), and from the three neighbouring orders: the Cucurbitales, Fagales and Fabales. Together, these four orders constitute the Nitrogen Fixing Clade (NFC), which is the clade in which all nodulating plant species occur (van Velzen et al. 2019). The NFC is the largest clade in which I performed the data mining (**Figure 36**). It should be noted that the Fabales order contains furanocoumarin-producing species such as *Psoralea cinerea*: therefore, if the furanocoumarin pathway emerged only once in plants, it may have been possible to trace the marmesin synthase back to the root of the NFC, and even further before. All the species belonging to the NFC were not included in the screening, because genetic data were not available for all of them, but also because it was not absolutely necessary to do so.

Practically, for the Moraceae, Urticaceae and Cannabaceae families, I used all the species for which genetic resources were available in 1KP and GenBank, whether they were partial cDNA sequences or complete genomes (**Table 6**). Yet, I could not find any database related to the Ulmaceae family, which is the closest relative of the three other families (**Figure 36**). Among the species I used, those whose genomes have been entirely sequenced, such as *Ficus erecta*, *Morus notabilis*, and *Cannabis sativa* (Ma et al. 2014; Grassa et al. 2018; Shirasawa et al. 2019; Jenkins and Orsburn 2019) were of particular interest. The access to complete genomes allows the identification of many sequences, and contrary to partial or transcriptomic resources, it also allows to take into account the absence of homologous genes. Indeed, if a BLAST on a complete genome does not lead to the identification of any sequence that is homologous to the query, we can be sure that such sequence does not exist in the corresponding plant.
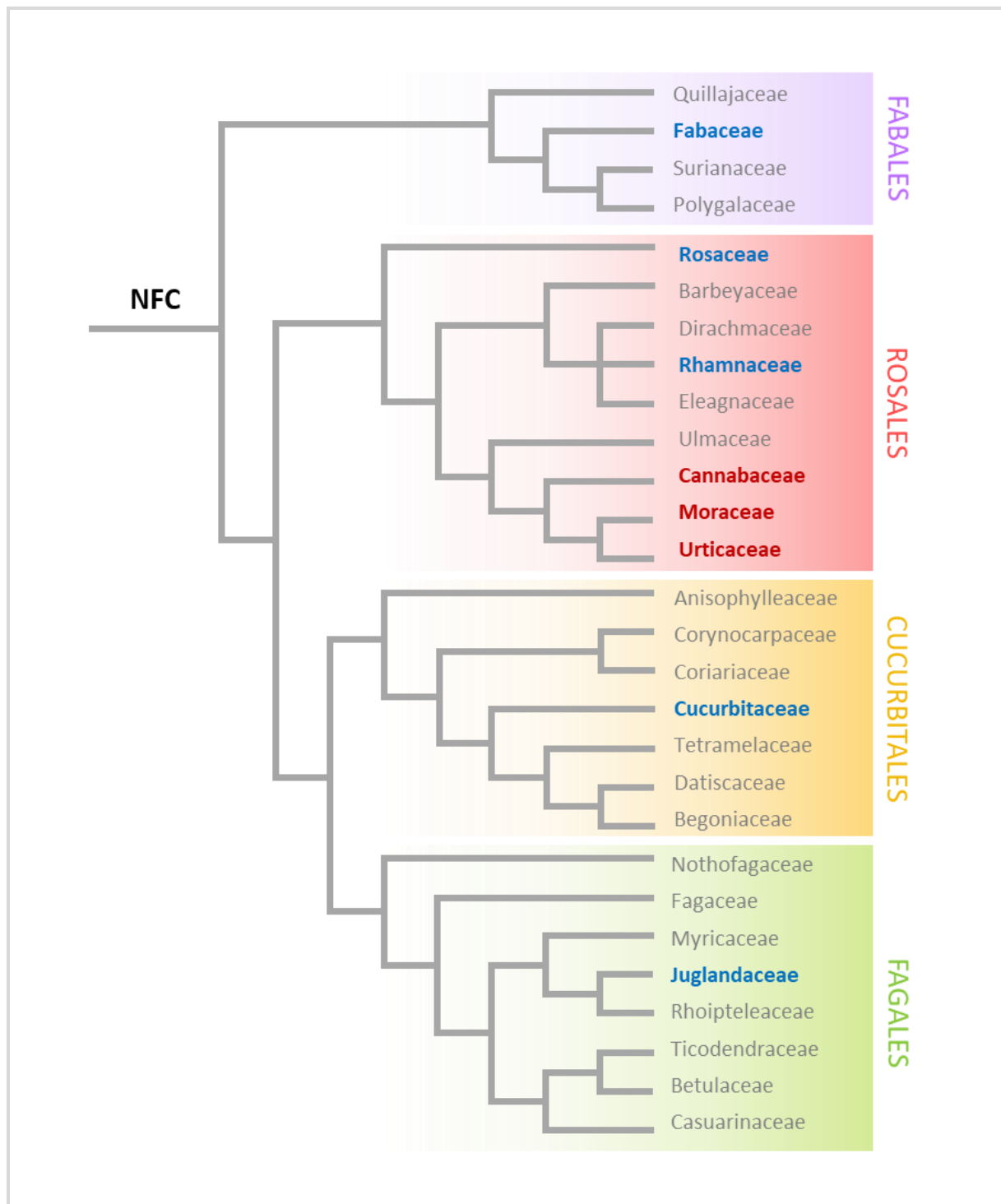
Figure 36 Plant families of the Nitrogen Fixing clade included in the data mining. The families in red are the families in which an exhaustive data mining has been performed in all currently available genetic resources. The families in blue are the ones in which the search was restricted to some "representative" species. The families in grey are the ones in which no genetic resource was available, or the ones that were not used. The topology of the tree has been drawn according to the Angiosperm Phylogeny Group IV (Stevens 2001; The Angiosperm Phylogeny Group 2016).

In the other Rosales families, genetic data was available for many plants. Yet, whenever possible, I limited the search to a few species with completely sequenced genomes. For instance, among the many Rosaceae which complete genomes are published, I only selected five species and considered them as "representative" (**Figure 36**, **Table 6**). Indeed, if CYP76F112 (or CYP81BN4, or CYP82J18) has an ancient enough origin to possess homologues in the Rosaceae family, it is very likely that the analysis of a complete rosaceous genome would reveal it. Of course, this putative orthologue may have been lost in one or a few species/taxa, but if none were found in the five "representative" genomes, it would be reasonable to assume it does not exist in the Rosaceae family.

The same logic was applied to the orders of Cucurbitales, Fagales and Fabales, which were used to confirm the patterns observed in *Ficus* closest relatives. If CYP76F112 (or CYP81BN4, or CYP82J18) were found in the Moraceae family but were absent from the other Rosales families, it would be very unlikely to find it in plant species from the neighbouring orders. On the contrary, if this gene had an ancient origin and were highly conserved in the entire Rosales order, a single genome from the neighbouring orders would certainly be enough to reveal its presence or absence. For this reason, I only selected a few "representative" species from the Cucurbitales, Fagales and Fabales orders (**Figure 36**, **Table 6**). In the Fabales order, no furanocoumarin-producing species (such as *Psoralea cinerea*) was included in the screening: indeed, these species only have scarce genetic data, and a BLAST search with CYP76F112 on these uncomplete resources did not lead to any satisfying result. As this might either be related to the "partial" status of the data or to the real absence of homologues, I set these furanocoumarin-producing species aside, in favour of species with more complete genetic resources.

The final list of species used to reconstruct the respective phylogenetic stories of CYP76F112, CYP82J18 and CYP81BN4 is summarised in **Table 6**. To the best of my knowledge, the only species on this list that produce furanocoumarins are *Ficus* species.

Table 6 Species included in the phylogenetic analysis. The colour code refers to **Figure 36**. The species highlighted with a * are the ones for which the genetic data were only partial. For the other species, (almost) complete genomes were available (scaffold to chromosome assembly level).

Order	Family	Species
Rosales	Moraceae	*Ficus carica*, Ficus religiosa*, Ficus erecta, Morus notabilis*
	Urticaceae	*Boehmeria nivea*
	Cannabaceae	*Cannabis sativa, Parasponia andersonii, Trema orientale, Humulus lupulus*
	Rhamnaceae	*Ziziphus jujuba*
	Rosaceae	*Prunus avium, Fragaria vesca subsp. Vesca, Rosa chinensis, Malus domestica, Pyrus x bretschneideri*
Fabales	Fabaceae	*Cicer arietinum, Cajanus cajan, Medicago truncatula*
Cucurbitales	Cucurbitaceae	*Cucumis sativus*
Fagales	Juglandaceae	*Juglans regia*

B.2. Constitution of the datasets

Three datasets of CYP76F, CYP81BN and CYP82J nucleotide sequences were constituted. The sequences included in these datasets were identified by exhaustive BLASTn searches performed respectively with *CYP76F112*, *CYP81BN* and *CYP82J18* on the genetic resources of every selected species (**Table 6**). To obtain good alignments and produce the best trees possible, I only kept the sequences that looked like complete coding sequences. This means the sequences needed to begin with a start codon, to end with a stop codon, and to be rid of introns. Therefore, I put aside the partial sequences, and I removed the putative intron(s) of the sequences that were identified from poorly or non-annotated genomic resources, as explained in the materials and methods section (**Chapter VII**, **B.6.b.1**). The sequences which contained mutations that created a frameshift or introduced premature stop codons were considered as putative pseudogenes. The frameshift-containing pseudogenes were eliminated because their corresponding amino acid sequences did not align well with those of the other genes and could therefore not be used to construct partitioned phylogenetic trees. The pseudogenes which contained premature stop codons were kept in the analysis because they did not contain any frameshift that would hinder their correct alignment and partitioning.

The similarity search performed with *CYP76F112* led to the identification of 123 sequences, which includes 5 frameshift-containing pseudogenes (excluded), 2 pseudogenes with stop codons (included) and 116 putative complete coding sequences. Based on the identity they shared with CYP76F112 (>55% amino acid identity), the great majority of these sequences seemed to correspond to genes from the CYP76F subfamily. A unique name has been attributed to every sequence, according to the following model: "G03_76F_F.erecta" is the third CYP76F gene (G) identified in *F. erecta*. The main characteristics of these sequences are presented in **Supp. Table 5**. For obvious reasons of space, the nucleotide sequences are not reported in this document, but they can be provided upon request.

Similarly, for *CYP81BN4*, I identified 93 sequences that included 1 frameshift-containing pseudogene (excluded) and 2 pseudogenes with stop codons (included). The majority of these sequences seemed to belong to the CYP81BN subfamily, but some were obviously from other subfamilies (<55% amino acid identity with CYP81BN4; higher similitude with known P450s from other CYP81 subfamilies). As far as I know, no other CYP81BN has been described in any other plant species. So, to help distinguish the sequences belonging to the CYP81BN from the ones belonging to other CYP81 subfamilies, I added four known sequences to the dataset, which were arbitrarily chosen in 4 other subfamilies: *CYP81E8*, *CYP81Q32*, *CYP81D1* and *CYP81B1c* (**Supp. Table 6**).

For *CYP82J18*, and despite a search which was as exhaustive as the previous ones, I was only able to identify 54 sequences that were somehow close enough to CYP82J18 (>50% nucleotide identity) to correspond to genes from the CYP82J subfamily, or to neighbouring subfamilies. These sequences did not include any pseudogenes but, based on the identity they share with CYP82J18, only half of them seemed to belong to the CYP82J subfamily (>55% amino acid identity). I also added extra sequences to help distinguish the different CYP82 subfamilies: *CYP82G1*, *CYP82B1*, *CYP82C1*, *CYP82A4*, *CYP82D3* and *CYP82H1* (**Supp. Table 7**).

Finally, some additional sequences were added in each of the 3 datasets to constitute the outgroups of the 3 respective phylogenetic trees: this included the apiaceous CYP71s described in the

furanocoumarin pathway, the 5-hydroxyxanthotoxin synthases from *A. thaliana*, and all the *F. carica* candidate genes (**Chapter II**) – except for CYP88A103, which is not from the CYP71 clan and did not align well with the other genes.

In summary, I screened various species from the Nitrogen Fixing Clade to identify putative CYP76F, CYP81BN and CYP82J (pseudo)genes that constitute the datasets used to infer the following gene-tree phylogenies.

C. INFERRING PHYLOGENIES: THE REALM OF THE ELDER GENES

For each of the 3 datasets, the nucleotide sequences were aligned by using the translation alignment tool of the Geneious software (https://www.geneious.com/). Then, the alignments were partitioned and used to generate Bayesian inference trees. The computation of the trees was done on the CIPRES Science Gateway (http://www.phylo.org/) with the parameters described in the materials and methods section (**Chapter VII**, **B.6.b.3**). According to the gene trees I obtained for *CYP76F112*, *CYP81BN4* and *CYP82J18*, the 3 respective datasets were refined by keeping only the "most interesting sequences" (detailed below for *CYP76F112*). As they allowed for a better alignment, these restricted datasets have been aligned again and used to generate secondary phylogenetic gene trees.

C.1. Expansion and diversification of CYP76Fs

C.1.a. Gene-family phylogeny of the CYP76Fs in the Nitrogen Fixing Clade

As a first step, I used the complete CYP76F dataset (**Supp. Table 5**) to generate a gene-family phylogenetic tree, presented in **Figure 37**. The general topology of this gene tree reflects the phylogeny of the NFC (**Figure 36**), with taxon-specific gene expansions. For instance, if we focus on the Rosales order, the node highlighted with a red dot on **Figure 37** leads to two clades: one which is Rosaceae-specific, and the second that includes sequences from the Rhamnaceae, Cannabaceae, Moraceae and Urticaceae families. In this second clade, the Rhamnaceae sequences branch before the others. The red node (**Figure 37**) therefore represents a unique ancestral CYP76F gene that underwent independent duplication events in the different taxa. It gave rise to 5-9 sequences in the screened Rosaceae species, 2 sequences in *Ziziphus jujuba* (Rhamnaceae), 1-2 sequences in the Cannabaceae species and 5 sequences in *Boehmeria nivea* (Urticaceae). In the Moraceae family, and particularly in the *Ficus* genus, this putative ancestral gene might have undergone many duplications. Indeed, 4 sequences have been identified in *Morus notabilis*, 4 and 3 in the partial resources of *Ficus religiosa* and *F. carica* (*CYP76F110-112*), and a total of 15 sequences were identified in the genome of *F. erecta*.

The most interesting part of this gene-family phylogeny corresponds to the clade that contains *CYP76F110-112*. To analyse it with more precision, I restricted the dataset to a smaller number of sequences that allowed a better alignment. This reduced dataset corresponds to the clade highlighted
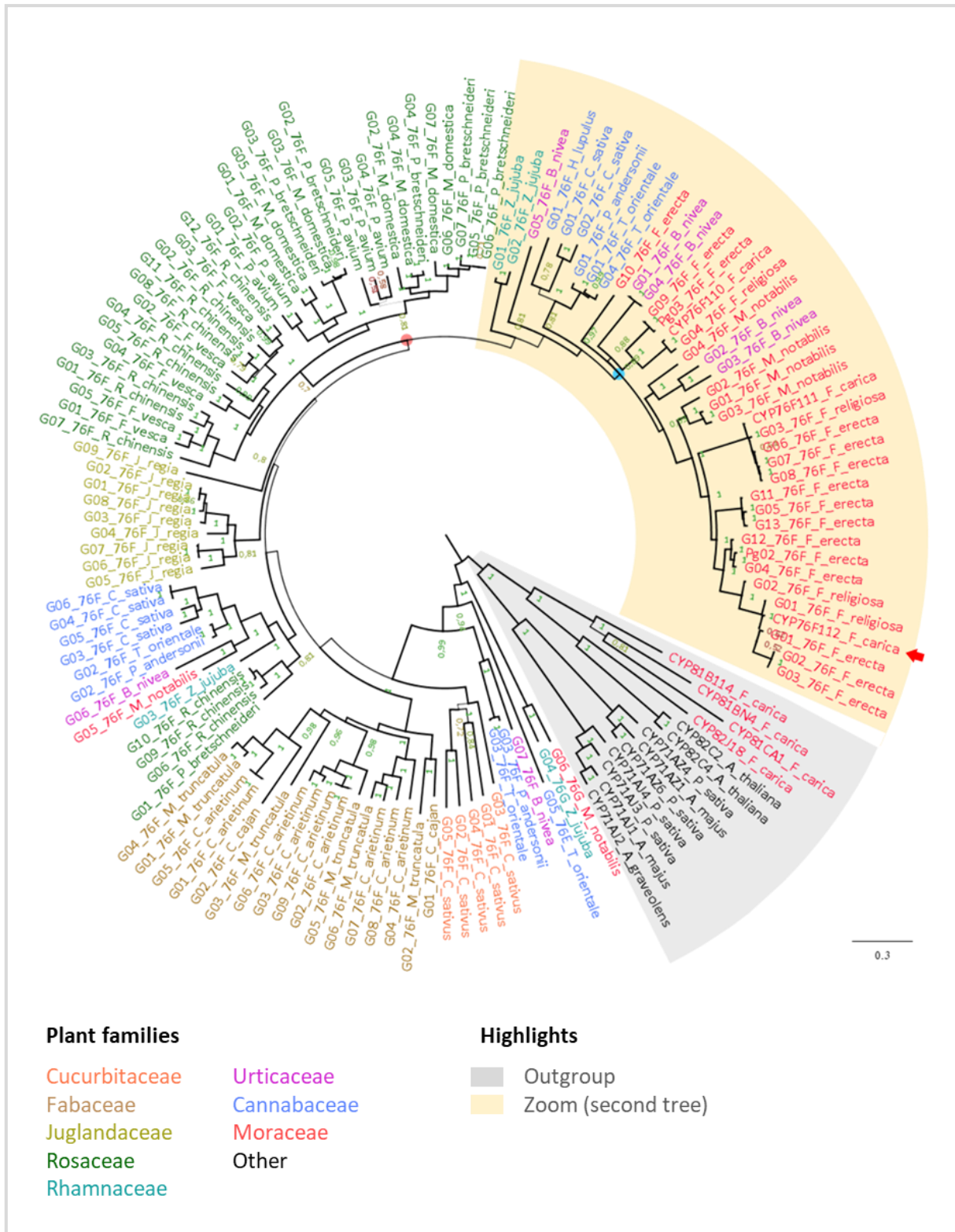
Figure 37 Gene-tree phylogeny of the CYP76F subfamily across the Nitrogen Fixing Clade. The alignment of full-length P450 nucleotide sequences from the CYP76F dataset (**Supp. Table 5**) was used to generate a Bayesian inference tree. The tree was rooted using P450s from the CYP71, CYP81 and CYP82 families. CYP76F112 is highlighted with a red arrow. The outgroup is highlighted in grey. The clade selected to generate a second "zoomed" tree is highlighted in yellow. The node highlighted in red is used as an example in the text. The blue circle shows a polytomy.

in yellow in **Figure 37**. It contains sequences from the Cannabaceae, Moraceae (including *CYP76F110-112*), and Urticaceae families, together with sequences from *Ziziphus jujuba* that were used as the outgroup for subsequent analyses (**Supp. Table 5**). These sequences were realigned and used to generate a second gene tree, presented in **Figure 38**. On this figure, I also reported the genomic position of all the CYP76Fs within the *F. erecta* genome (for clustering analysis, see below). This second gene tree confirmed the general topology seen in the first one, but it also resolved a polytomy (unresolved node – highlighted with a blue dot on **Figure 37**) and displayed higher probabilities.

Based on the gene-family phylogenies inferred on **Figure 37** and **Figure 38**, we can assume that the CYP76F subfamily expanded and diversified independently in the different taxa of the NFC. In particular, the CYP76Fs have dramatically expanded in the Moraceae family and maybe specifically in the *Ficus* genus. Indeed, *CYP76F112* belongs to a lineage-specific clade that is exclusively composed of 17 *Ficus* sequences: the "F112-like clade", highlighted in blue in **Figure 38**. This clade contains *CYP76F111*, *CYP76F112*, 3 sequences from *F. religiosa* and 12 sequences from *F. erecta*. It can be subdivided into 4 subclades, highlighted in green, purple, orange and red in **Figure 38**.

All the CYP76F sequences from *F. erecta* are clustered on the chromosomes 9 and 10. In addition, their genomic locations show repeated patterns that are particularly visible on **Figure 38** with the "red-green" and "orange-purple" repeats – each colour corresponding to a subclade of CYP76F. These patterns strongly suggest that the CYP76Fs from *F. erecta* emerged via repeated tandem duplications. Indeed, many of the duplicates are in a head-to-tail orientation (for instance, the genes G12 and G13 on the chromosome 9, or the genes G06, G07 and G08 on the chromosome 10), but some of them are organised on the chromosomes with the opposite orientation (for instance, the genes G01 and G02 on the chromosome 10), which implies that intrachromosomal rearrangements might have happened. As the complete genomes of *F. carica* and *F. religiosa* were not available at this time, it is possible that every *F. erecta* sequence possesses orthologues in *F. carica* and *F. religiosa* – depending on when the tandem duplications happened. Anyway, two sequences from *F. erecta* and *F. religiosa* share more than 98% identity with *CYP76F112* (**Figure 38**), their phylogenetic neighbour sequence. This makes them putative marmesin synthases that would be orthologues of the *F. carica* one. These results suggest that the *Ficus* marmesin synthases might originate from a bloom of CYP76Fs that happened via repeated duplications in a restricted taxon of the Moraceae family.

Last, it should be noted that *CYP76F110* and *CYP76F111* are quite distant from *CYP76F112*, with which they share only 70% identity (**Figure 38**). Indeed, *CYP76F110* does not belong to the F112-like clade, and it is closer to some *B. nivea* sequences (Urticaceae). The respective ancestors of *CYP76F110* and *CYP76F112* must therefore have diverged early, before the Moraceae-specific bloom. *CYP76F110* may even function in a pathway that is common to the Urticaceae and the Moraceae family, which would explain why it did not metabolise any of the furanocoumarins I tested on it. On its side, *CYP76F111* belongs to the F112-like clade. Yet, *CYP76F111* branched early from the F112-like clade, and belongs to a different subclade than that of *CYP76F112* (respectively, the subclades coloured in green and red in **Figure 38**). It would thus not be surprising to find out that *CYP76F111* functions in a pathway that is common to *Morus* and *Ficus*, and has nothing to do with furanocoumarins – as suggested by the absence of activity described in **Chapter II**. This point will be further discussed in **Chapter IV**, with the use of a docking and a site-directed mutagenesis experiment.
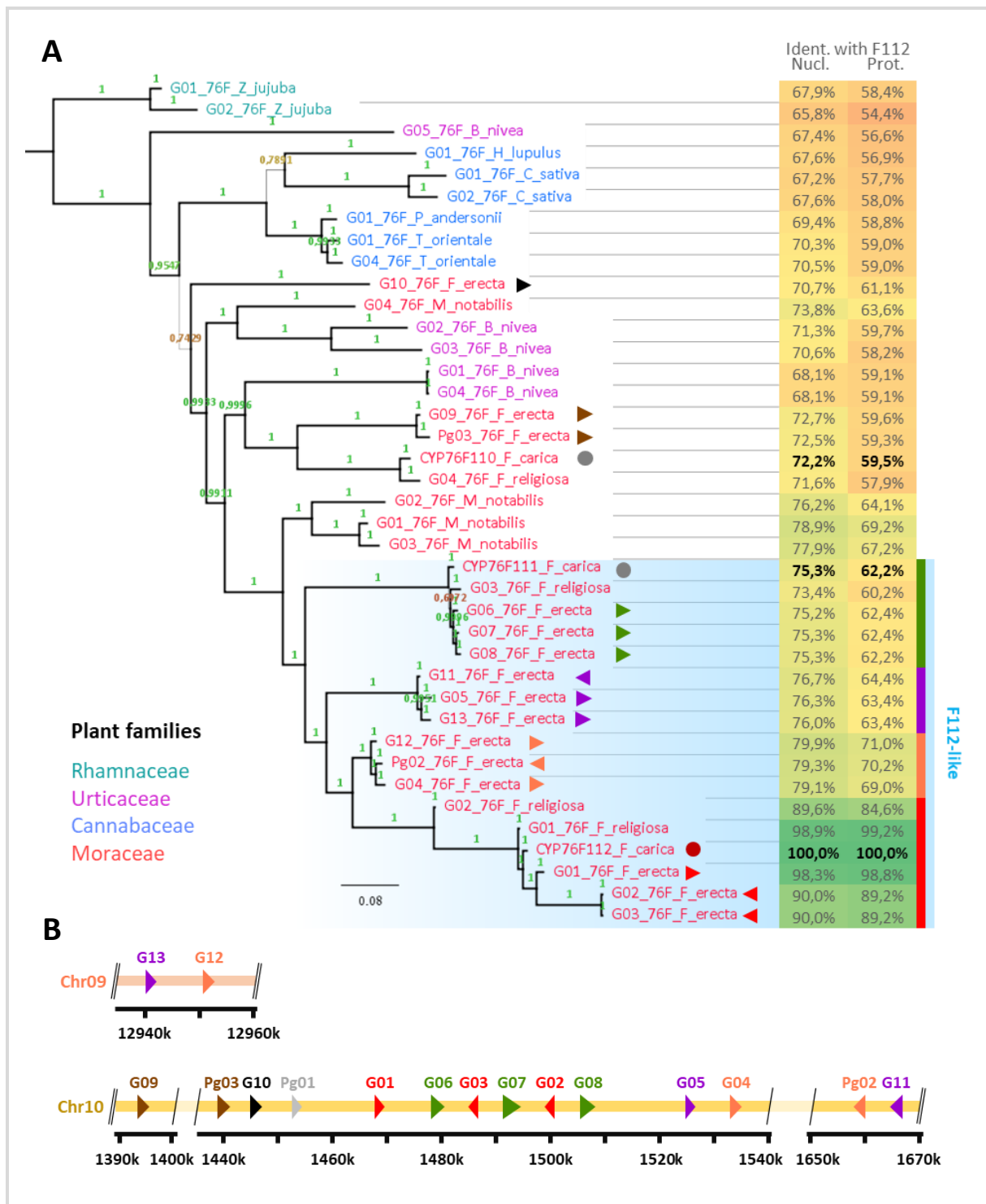
Figure 38 Gene-tree phylogeny of the CYP76Fs from the Cannabaceae, Moraceae and Urticaceae families. A The alignment of full-length P450 nucleotide sequences from the restricted CYP76F dataset (**Supp. Table 5**) was used to generate a Bayesian inference tree. The tree was rooted using CYP76Fs from the Rhamnaceae family. *CYP76F112* is highlighted with a red circle, *CYP76F110-111* with a grey circle. The identity shared with *CYP76F112*, at a nucleotide and protein levels, are detailed. The "F112-like clade" is highlighted in blue, it can be subdivided into 4 subclades that are coloured in green, purple, orange and red. **B** Genomic position of the *F. erecta* sequences, coloured according to their subclades, and reported in the tree with corresponding arrows. The direction of the arrow reflects the orientation of the gene in the genome (Shirasawa et al. 2019).

C.1.b. Evolution of CYP76Fs' SRSs

To further inquire the emergence of the marmesin synthase activity, I studied the substrate recognition sites (SRSs) of CYP76F112 and its closest homologues. As explained in **Chapter I** (**II.D.1.d**), the SRSs play a key role in regulating the substrate specificity of P450 enzymes. Therefore, a single or a limited number of mutations in these regions might be enough to change a P450 substrate specificity (Gotoh 1992; Baudry et al. 2006; Rupasinghe and Schuler 2006; Vazquez-Albacete et al. 2017).

The SRSs of all F112-like sequences were identified by protein alignment with other P450s which SRSs are known. To do this, I used as references the P450s described in Gotoh (1992), and CYP71AJ1 (Larbat 2006). The 6 SRSs of each F112-like and CYP76F110 are reported in **Figure 39**. To ease the comparison, the amino acids have been coloured according to their characteristics (charge, polarity, etc) and their sameness to those of CYP76F112.
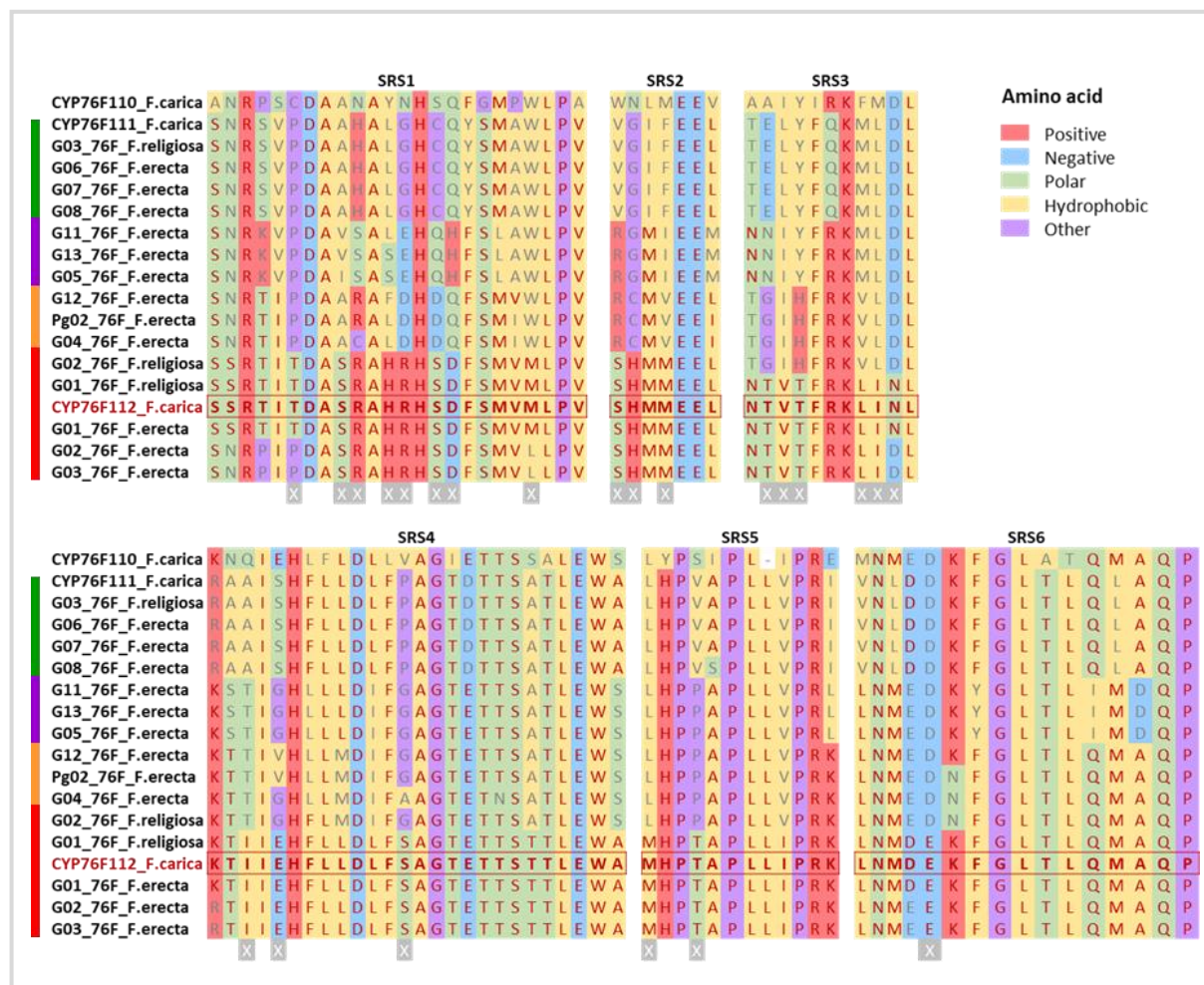
Figure 39 SRSs of the CYP76F112-like and CYP76F110. The SRSs of CYP76F112 are highlighted in bold and framed in red. In the other sequences, the amino acids identical to those of CYP76F112 are in red; the ones that are different are in grey. A background colour has been associated to every amino acid, according to the characteristics of its side chain (charge, polarity, etc). The colours on the left of the sequences' names mirror the subclades defined on **Figure 38**. The amino acids that might be specific to the red subclade are highlighted with a white "X" on grey background.

First, it should be noted that some SRSs are more conserved than others. For instance, SRS6 is well conserved while SRS1 shows variations that mirror the phylogenetic relationship between the F112-like subclades (**Figure 38**). In particular, SRS2 and SRS3, which are assumed to form the substrate access channel (Larbat et al. 2007), are quite variable from one subclade to another. This suggests that the CYP76Fs from the different subclades may convey substrates with different structures to their active site. This assumption is consistent with my results showing that CYP76F112 (red subclade) metabolises DMS while CYP76F111 (green subclade) does not (**Chapter II**).

The 6 sequences from the "red subclade" (including CYP76F112 – **Figure 39**) possess highly similar SRSs and share more than 20 exclusive amino acids, distributed in the 6 SRSs. Among these 6 sequences, the genes G01 from *F. erecta* and from *F. religiosa,* previously highlighted for their high identity with CYP76F112, possess the exact same SRSs than CYP76F112. This strongly suggests that these genes are orthologous marmesin synthases, but this needs to be confirmed. The SRSs of the three other CYP76Fs of the "red subclade" display more mutations compared to CYP76F112. Yet, in absence of more data, it is not possible to know if these mutations are enough or not to modify the marmesin synthase activity, or how they can affect it.

In summary, the phylogenetic analysis strongly suggests that CYP76F112 – and probably the other *Ficus* marmesin synthases – originates from a Moraceae-specific bloom of CYP76Fs, that happened via multiple duplication events. Yet, the general patterns shown by the SRSs are not clear enough to conclude about the emergence of the marmesin synthase activity itself.

C.2. Evolutionary patterns of the CYP81BNs across the Nitrogen Fixing Clade

Similarly to the CYP76F, I used the complete CYP81BN dataset (**Supp. Table 6**) to generate a phylogenetic gene tree, presented in **Figure 40**. Once again, the general topology of this gene tree reflects the phylogeny of the NFC, with lineage-specific diversifications. Yet, one gene duplication might have happened before the divergence of the Cannabaceae and the Moraceae/Urticaceae families because the respective sequences from these plant species are split into two distinct clades, which are themselves split into family-specific subclades. As for CYP76F, the initial CYP81BN dataset was restricted to the yellow clade (**Figure 40**). It contains sequences from the Cannabaceae, Moraceae and Urticaceae families, together with sequences from *Z. jujuba* that were subsequently used as an outgroup (**Supp. Table 6**). The restricted dataset was used to generate a second gene tree (**Figure 41**), which confirms the general topology inferred by the first one, and provides a better support.

According to the gene-family phylogenies inferred on **Figure 40** and **Figure 41**, *CYP81BN4* belongs to a small clade of four sequences, called the "BN4-like", that includes one *F. erecta* and two *F. religiosa* genes. Their closest homologue is another *Ficus* sequence (G02_81_F.erecta), but except for it, the closest sequences belong to the *Morus* genus and then, to the Cannabaceae family. Therefore, we can assume that the last Cannabaceae/Moraceae/Urticaceae ancestor might had two CYP81BN ancestral genes. The first one underwent a few duplications in the Cannabaceae (4 sequences in *C. sativa*, 3 in *T. orientale*), is present in low-copy number in the Moraceae (2 sequences in *M. notabilis*, *F. erecta* and *F. religiosa*, and *CYP81BN4* in *F. carica*), and seems to have disappeared from the Urticaceae.
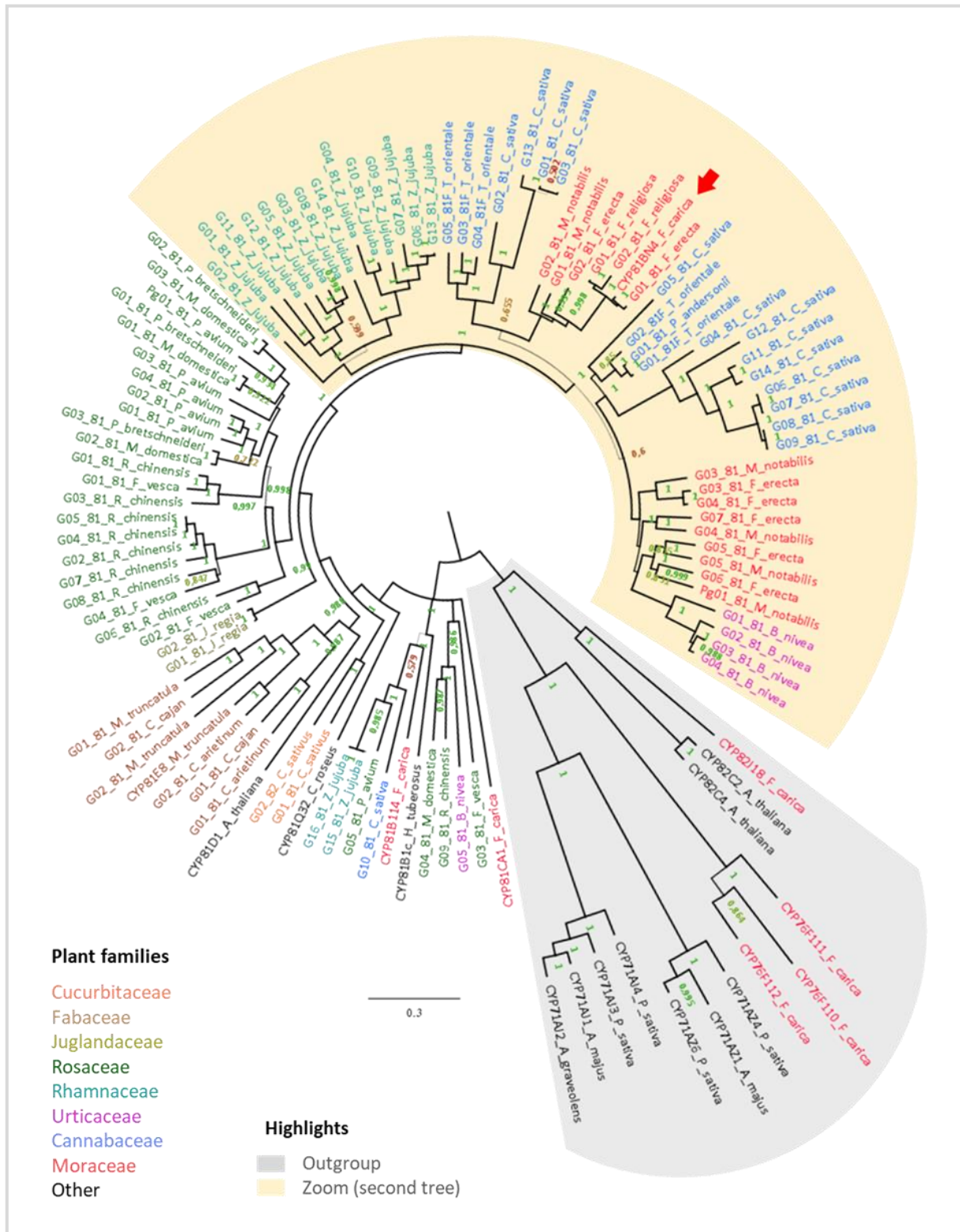
Figure 40 Gene-tree phylogeny of the CYP81BN subfamily across the Nitrogen Fixing Clade. The alignment of full-length P450 nucleotide sequences from the CYP81BN dataset (**Supp. Table 6**) was used to generate a Bayesian inference tree. The tree was rooted using P450s from the CYP71, CYP76, CYP82 families. *CYP81BN4* is highlighted with a red arrow. The outgroup is highlighted in grey. The clade selected to generate a second "zoomed" tree is highlighted in yellow.
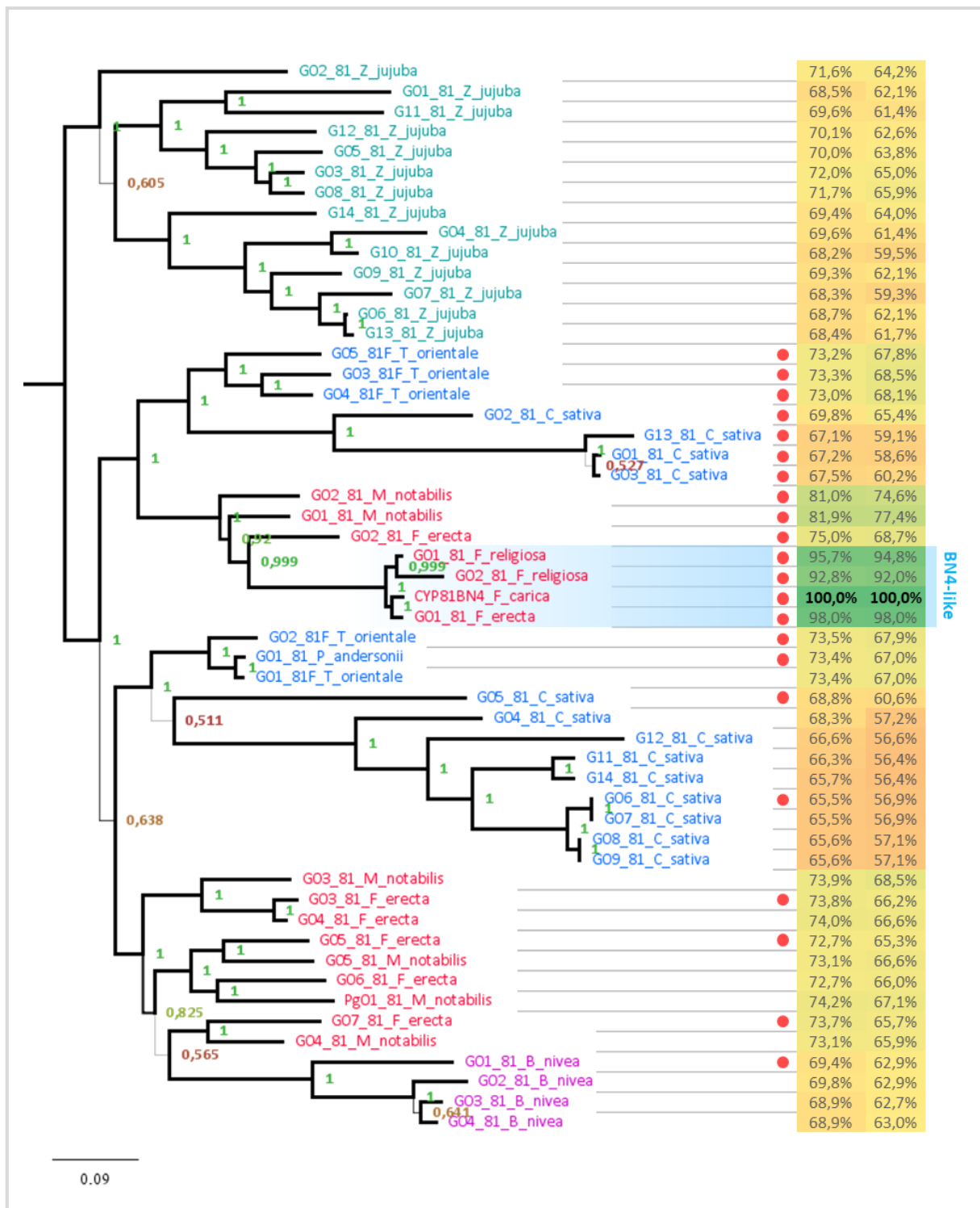
Figure 41 Gene-tree phylogeny of the CYP81BNs from the Cannabaceae, Moraceae and Urticaceae families. The alignment of full-length P450 nucleotide sequences from the restricted CYP81BN dataset (**Supp. Table 6**) was used to generate a Bayesian inference tree. The gene tree was rooted using CYP82s from the Rhamnaceae family. The identity shared with *CYP81BN4*, at a nucleotide and protein levels, are detailed. The CYP81BN4-like are highlighted in blue. The red circles refer to **Supp. Figure 3**.

This analysis shows that, in the Moraceae family, there is no expansion of the CYP81BNs that is comparable to that of the CYP76Fs. This may suggest that the CYP81BN enzymes from *Morus* and *Ficus*, and maybe also from the Cannabaceae, share the same function. If this assumption is right, the physiological activity of CYP81BN4 would not be to oxidise cnidilin since *Morus* and the Cannabaceae have never been reported so far to produce furanocoumarins. However, the possibility of a physiological activity cannot be completely excluded since the BN4-like share only 80% identity with the *Morus* sequences – which could be enough to modify a P450 activity.

To go deeper in the analysis, I also compared the SRSs of the CYP81BNs (**Supp. Figure 3**). As for the CYP76Fs, some SRSs are more variable than others. Several amino acids, which looks randomly distributed, seem to be almost exclusive to the 4 sequences from the BN4-like clade. These specific amino acids might be in favour of an enzymatic activity that would be different in the BN4-like clade than in the others sequences. Yet, as other CYP81BN subclades also possess their own exclusive amino acids, and since the other CYP81BNs are functionally orphans, it is possible that these "exclusive" amino acids are only the reflect of neutral variations that do not affect the activity of the CYP81BNs.

In brief, the evolutionary patterns shown by the CYP81BN subfamily are not pronounced enough to confidently conclude anything about the physiological activity of CYP81BN4. CYP81BN4 does not result from a Moraceae-specific expansion and its homologues tend to remain low-copy. Yet, as the CYP81BN4-like are still different from their closest *Morus* homologues, it is possible that they evolved a new activity.

C.3. Conservation of the CYP82J subfamily across the Nitrogen Fixing Clade

The complete CYP82J dataset (**Supp. Table 7**) was used to generate a first phylogenetic gene tree (**Figure 42**). This tree confirms what has been foreseen during the construction of the CYP82J dataset, since only a small number of the sequences truly belong to the CYP82J subfamily. Indeed, the CYP82J clade, which includes *CYP82J18*, is only composed of 24 sequences that are highlighted in dark yellow on **Figure 42**. The remaining sequences are grouped with CYP82s from other subfamilies, such as the CYP82G subfamily. Therefore, I reduced the initial dataset to the CYP82J subfamily, together with the CYP82Gs that were used as an outgroup (yellow clade on **Figure 42**, **Supp. Table 7**). The sequences from this restricted dataset were used to generate a second phylogenetic gene tree (**Figure 43**).

Both gene trees clearly indicate that the CYP82Js constitute a conserved subfamily which origin is older than the NFC, and which did not undergo any major expansion. Indeed, the CYP82Js have been found in one or two copies (3 in *Juglans regia*) in most species belonging to the Rosales, Fabales and Fagales orders. In addition, the phylogenetic relationships between these sequences perfectly mirror the phylogeny of the NFC (**Figure 36**), which strongly suggests that the last common ancestor of the NFC species possessed a unique CYP82J that has been transmitted to all its descendants. This ancestral CYP82J remained single or low copy in most species, such as *F. erecta*, but it also disappeared from a few taxa. For instance, despite the access to complete genomes, no CYP82J could be found in *Cannabis sativa*, nor in *Cucumis sativus* which was the only Cucurbitales species included in this analysis.
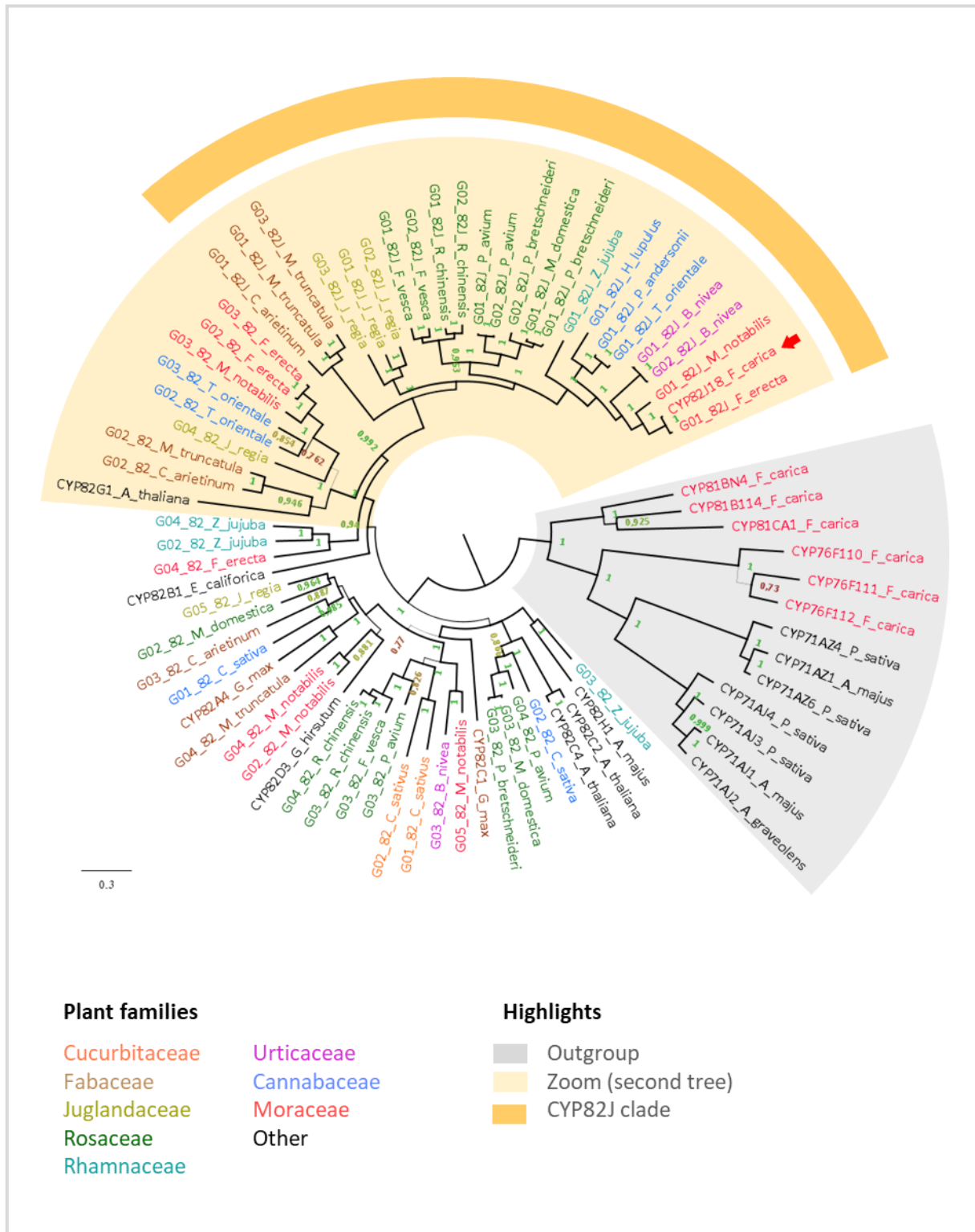
Figure 42 Gene-tree phylogeny of the CYP82J family across the Nitrogen Fixing Clade. The alignment of full-length P450 nucleotide sequences from the CYP82J dataset (**Supp. Table 7**) was used to generate a Bayesian inference tree. The tree was rooted using P450s from the CYP71, CYP76, CYP81 families. *CYP82J18* is highlighted with a red arrow. The outgroup is highlighted in grey. The clade selected to generate a second "zoomed" tree is highlighted in yellow.

Figure 43 Restricted gene-tree phylogeny of the CYP82J subfamily. The alignment of full-length P450 nucleotide sequences from the restricted CYP82J dataset (**Supp. Table 7**) was used to generate a Bayesian inference tree. The tree was rooted using CYP82G sequences. The identity shared with CYP82J, at a nucleotide and protein levels, are detailed.

The conservation of the CYP82J subfamily in the NFC suggests that all the CYP82Js may share the same function. To further inquire this point, I identified and compared the SRSs of the CYP82Js (**Supp. Figure 4**). This time, the SRSs are mostly conserved across all CYP82J sequences – except for SRS1 which is a bit more variable. In addition, no amino acid is *Ficus* specific. These results led us to assume that CYP82J18 may possess the same physiological function than its homologues from *Morus*, from the others families of the Rosales order, and maybe even from the entire NFC.

In brief, we can assume that *CYP82J18* originates from an ancient sequence that predates the NFC. Despite some taxa-specific losses, this ancestral sequence has been well conserved in the NFC where it remained single or low copy, and may still possess its ancestral activity.

D. DISCUSSION: THE STORY O' MY P450S

In this chapter, my objective was to reconstruct the evolutive stories of *CYP76F112*, *CYP81BN4* and *CYP82J18*. By inferring the phylogenies of their respective subfamilies in the Nitrogen Fixing Clade, I found out that the three genes have very different stories: *CYP76F112* may have appeared recently, during an expansion of the CYP76Fs in the Moraceae family. *CYP81BN4* belongs to a subfamily that did not particularly expand nor diversified in the Moraceae. Last, *CYP82J18* belongs to an old and conserved subfamily presents in single or low-copies in the screened plant species. Such evolutionary patterns have already been described in the P450s. Indeed, within the P450s, the families that fulfil important functions are usually conserved and present in low-copy in the genome, which might be explained by a rare retention of the duplicates due to a high purifying pressure. On the contrary, other P450 families and subfamilies are larger, which suggests they underwent multiple duplications and might be involved in a broader range of reactions (Nelson et al. 2008; Ilc et al. 2018).

D.1. Evolutive story of *CYP76F112*, the F. carica marmesin synthase

D.1.a. Evolution of the CYP76 family and emergence of *CYP76F112*

The CYP76s constitute an old family that is assumed to predate the monocot-eudicot divergence. It has extensively diversified and led to (at least) 34 distinct subfamilies (Nelson and Werck-Reichhart 2011; Bathe and Tissier 2019). The CYP76 family and its subfamilies seemed to have bloomed in a lineage-specific way through gene duplications. For instance, it has been reported that the CYP76 family is more expanded in the genome of the grapevine (CYP76A, G, F, T and Y) than in most other genomes (Ilc et al. 2018). As another example, the CYP76M and CYP76C subfamilies respectively bloomed in rice and Brassicaceae via gene duplications. Such expansions are often assumed to be associated with new species-specific adaptative functions: in rice, the expansion of the CYP76M subfamily gave rise to 11 genes and 2 pseudogenes, among which 4 clustered genes (or more) are involved in the biosynthesis of defensive diterpenes (Swaminathan et al. 2009; Wang et al. 2012; Höfer et al. 2014; Ilc et al. 2018).

In this study, the evolutionary patterns observed with the CYP76Fs in the NFC are in perfect accordance with these previous findings: from a single ancestral gene, the CYP76F subfamily expanded specifically in some taxa of the NFC (15 sequences in *F. erecta*) but remained low copies in others (2 in *C. sativa*). In particular, a lineage-specific expansion happened in the Moraceae family and gave rise to many *Ficus* sequences that include the *F. carica* marmesin synthase. Indeed, *CYP76F112* belongs to a clade composed of 17 *Ficus* CYP76Fs – called the F112-like clade (**Figure 38**, **Figure 44**). The closest non-*Ficus* sequences are from *M. notabilis,* that does not produce furanocoumarins. Since the entire genome of *M. notabilis* has been investigated, we can be sure that this species does not contain any CYP76F that would be closer to the *Ficus* ones. Moreover, inside the F112-like clade, the earliest branching gave rise to *CYP76F111*, which did not show any activity related to the synthesis of (furano)coumarins (**Chapter II**). On its side, *CYP76F112* emerged from a subsequent diversification that also gave rise to two putative marmesin synthases from *F. erecta* and *F. religiosa* (**Figure 39**, **Figure 44**). Therefore, it can reasonably be assumed that the marmesin synthase activity is a recent innovation that emerged after the divergence of the *Ficus* and *Morus* genera. Inside the F112-like clade, this activity probably emerged after the initial branching that lead to *CYP76F111*, as highlighted in purple on **Figure 44**.
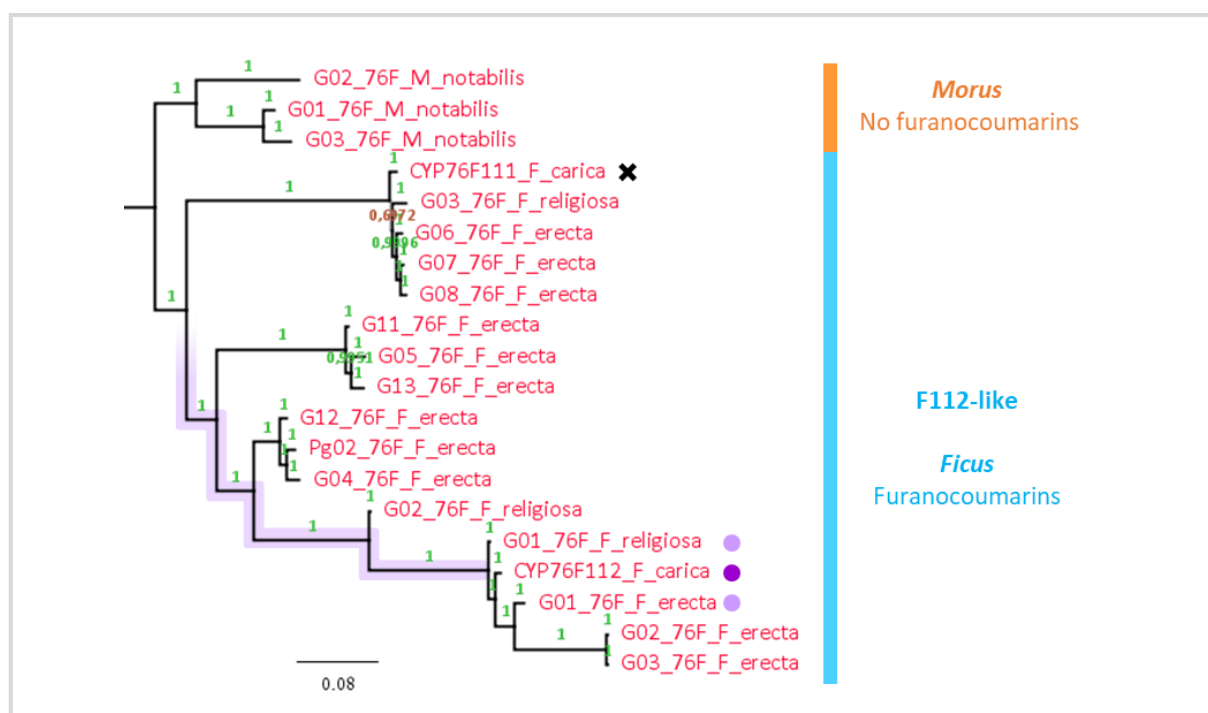
Figure 44 Focus on the F112-like clade. The phylogenetic gene tree is a detail from **Figure 38**. The *F. carica* marmesin synthase is highlighted with a deep purple circle, the putative marmesin synthases from *F. religiosa* and *F. erecta* are highlighted with light purple circles. CYP76F111, that does not metabolise DMS, is highlighted with a black cross. The putative emergence of the marmesin synthase activity is highlighted on the branches with a purple path.

Therefore, as for the rice-specific CYP76M expansion that gave rise to genes involved in the biosynthesis of defensive diterpenes (Swaminathan et al. 2009; Wang et al. 2012), the CYP76F expansion observed in the *Ficus* genus gave rise to at least one new activity involved in a species-specific adaptation: the marmesin synthase activity, which is involved in the biosynthesis of defensive furanocoumarins.

D.1.b. Clustering of the CYP76Fs

The CYP76F expansion observed in *Ficus* probably happened via multiple tandem-duplications. In *F. erecta*, these duplications resulted in the formation of a large cluster on chromosome 10 (**Figure 38**) which contains all the *F. erecta* CYP76Fs, except for two sequences clustered on chromosome 9.

In plants, clusters of homologous P450s resulting from tandem duplications are quite common. For instance, in the grapevine genome, 78% of the identified P450 sequences belong to small or large clusters (Ilc et al. 2018). A close clustering on a genome seems to be a common feature of recently tandem-duplicated P450s and suggests active evolution dynamics that may be in favour of the emergence of new activities (Feyereisen 2011; Liu et al. 2016; Ilc et al. 2018). Indeed, it seems that there is a snowball effect in gene evolution: when a gene is duplicated, it increases its presence in the genome and has thus more chance to be duplicated again. So, when a gene family starts to expand by duplication, it increases the probability that at least one member of the family will be duplicated again, diverge, and evolve a new function. Therefore, it is assumed that the more recently evolved families

are quite versatile and more likely to expand again and generate the diversity required to adapt to environmental fluctuations (Pichersky and Gang 2000; Luscombe et al. 2002; Feyereisen 2011; Nelson and Werck-Reichhart 2011). Consequently, the clustering of the CYP76F sequences on the genome of *F. erecta* is another indication in favour of a recent diversification of the CYP76Fs that would have given rise to the marmesin synthase activity.

As physical clustering often suggests that some of the clustered genes function in the same pathways (Frey et al. 1997; Field and Osbourn 2008; Osbourn and Field 2009; Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011), other *Ficus* CYP76Fs may catalyse other reactional steps of the furanocoumarin pathway. Such a clustering of P450 genes involved in furanocoumarin biosynthesis has already been described in the Apiaceae: for instance, *CYP71AJ3* (psoralen synthase) and *CYP71AJ4* (angelicin synthase) are clustered on the genome of *P. sativa* (Roselli et al. 2017). It has also been hypothesised that *PsPT2* (osthenol synthase) appeared via the duplication and neofunctionalisation of *PsPT1* (DMS synthase) (Munakata et al. 2016). Based on these reports, the many duplicated CYP76Fs clustered on the genome of *F. erecta* appear as significant candidates to identify new *Ficus* enzymes involved in furanocoumarin biosynthesis. Among them, the most interesting are probably those which are phylogenetically closer to the marmesin synthase(s) gene(s) rather than to *CYP76F111* (which does not seem to be involved in the furanocoumarin biosynthesis, **Chapter II**).

Lastly, clusters are not always made of paralogous genes: on the contrary, genes coding for different types of enzyme fold might also be arranged into biosynthetic clusters that form functional units and exhibit integrated biological functions (Ober 2005; Mizutani and Ohta 2010; Nützmann et al. 2016). Numerous metabolic clusters were reported and many of them are involved in species-specific metabolic activities (Frey et al. 1997; Qi et al. 2004; Shimura et al. 2007; Field and Osbourn 2008; Mizutani and Ohta 2010; Nützmann and Osbourn 2014; Nützmann et al. 2016; Bathe and Tissier 2019). For instance, a metabolic gene cluster involved in diterpene metabolism has been found in Euphorbiaceae. It contains various enzymes such as an acyltransferase, some casbene, neocembrene and terpene synthases, but also P450s from the CYP726A and CYP80C subfamilies (Bathe and Tissier 2019). Another example from the furanocoumarin pathway is the clustering of *PsPT1* and *PsDiox* (a prenyltransferase and a dioxygenase producing umbelliferone and DMS, respectively) on the genome of *P. sativa* (Roselli et al. 2017). The processes leading to the physical clustering of non-paralogous genes have not been completely clarified, but clusters seem to provide several advantages. For instance, they might facilitate the transmission of genes involved in the same pathway by reducing the risk of disruption by recombination. They also seem to help regulating the expression of a complete pathway. These features contribute to reduce the risk of producing deleterious intermediates that might be accumulated if one or several genes of the pathway were missing or not expressed (Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011; Nützmann et al. 2016).

For these reasons, some non-P450 genes involved in furanocoumarin biosynthesis might have been clustered with the CYP76Fs. Therefore, I performed some similarity searches on the chromosomes 9 and 10 of *F. erecta*, in order to identify dioxygenases, prenyltransferases or methyltransferases that would have been located in the same genomic region than the CYP76Fs. Yet, I could not identify any. On the contrary, a BLAST analysis performed on the whole genome of *F. erecta* using *FcPT1* and *FcPT2* (the *F. carica* prenyltransferases converting umbelliferone into DMS and osthenol, respectively) as a query revealed the presence of highly similar sequences on chromosome 12. So, we can assume that,

in *F. erecta*, the genes involved in the furanocoumarin biosynthesis pathway do not form a single metabolic cluster. Considering the recent origin of *CYP76F112*, this would not be surprising: indeed, it could be hypothesised that the different type of genes involved in the *Ficus* furanocoumarin biosynthesis pathway – at least *CYP76F112* and the FcPTs – evolved recently from the duplication and divergence of genes localised on different chromosomes, and that they did not have the time (yet) to be rearranged into a single metabolic cluster.

D.1.c. Multiple origin of the furanocoumarin pathway in higher plants

All the points discussed above converge to one strong assumption: *CYP76F112* and its marmesin synthase activity emerged recently within the Moraceae family. Therefore, the *Ficus* marmesin synthases must have emerged independently of the marmesin synthases of Fabaceous, Apiaceous and Rutaceous species – even though none of them are known yet. In a previous study, Munakata *et al.* (2020) suggested that the UDTs from Apiaceae (PcPT1, PsPT1, PsPT2) and Moraceae (FcPT1) evolved from a different ancestral sequence. Moreover, in *F. carica*, FcPT1 and CYP76F112 catalyse successive enzymatic steps that are at the very beginning of the furanocoumarin pathway – which imply they should have been among the first enzymes recruited in this pathway. So, if we combine this suggestion with the independent emergence of the moraceous and apiaceous UDTs, and with the recent emergence of *CYP76F112*, we now have sufficient evidences to propose the following hypothesis: **in Moraceae, the furanocoumarin pathway emerged independently from the furanocoumarin pathway of other furanocoumarin-producing families such as Apiaceae, Rutaceae, and even the Fabaceae family which also belongs to the NFC**.

In plants, the repeated but independent emergence of similar metabolic traits in distant taxa is quite common; which means that new enzymes with the same function independently evolved in separate plant lineages. This independent acquisition of identical catalytic properties is usually termed "convergent evolution". When it happened between enzymes belonging to the same fold family, it can be referred to as "parallel evolution". Similarly, another special form of convergence, very common in specialised metabolism, is "repeated evolution". In this case, a single pool of enzymes with similar functions are independently recruited in separated plant lineages to catalyse the same new reaction (Pichersky and Gang 2000; Weng 2014). The multiple and independent putative emergence of the furanocoumarin pathway implies that the marmesin synthase activity emerged multiple time through convergent evolution. As the apiaceous marmesin synthase is assumed to be a P450 (Hamerski and Matern 1988b), like the moraceous one, this would be a case of parallel evolution. However, this is probably not a case of repeated evolution: indeed, a similarity search on the partial genetic databases of furanocoumarin-producing species that do not belong to the Moraceae family did not lead to the identification of any CYP76F gene that would be close to *CYP76F112*. So, in the Apiaceae, Rutaceae and Fabaceae families, the marmesin synthase might not belong to the CYP76 family – or, at least not to the CYP76F subfamily.

D.1.d. The CYP76Fs: a marmesin synthase activity specific of the *Ficus* genus?

Until now, I described the expansion of the CYP76Fs as specific of a restricted taxon of the Moraceae family. I would like to remind that the only moraceous species included in this study are from the *Morus* and *Ficus* genera, and that the Moraceae family contains many other genera which include

other furanocoumarin producing species. As stated in **Chapter I** (**Figure 10**), the presence of furanocoumarins has been reported in some species belonging to the *Maquira* (Rovinski and Sneden 1984), *Brosimum* (Vieira et al. 1999), *Fatoua* (Chiang et al. 2010), *Dorstenia* (Abegaz et al. 2004; Heinke et al. 2012), and of course, *Ficus* (Späth 1937; Athnasios et al. 1962) genera. Within the phylogeny of the Moraceae (Clement and Weiblen 2009; Zerega and Gardner 2019), these furanocoumarin-producing species are grouped into a single clade that contains 3 tribes: the Ficeae (that includes the *Ficus* genus), the Castilleae (that includes the *Maquira* genus), and the Dorstenieae (that includes the *Brosimum* and *Dorstenia* genera). As the position of the *Fatoua* genus in the Moraceae phylogeny is still conflictuous, it has not been taken into account. On the contrary, the *Morus* genus belongs to another clade that contains the Maclureae, Artocarpeae and Moreae tribes (Clement and Weiblen 2009; Zerega and Gardner 2019), in which no furanocoumarin producers has been reported so far (**Figure 45**).
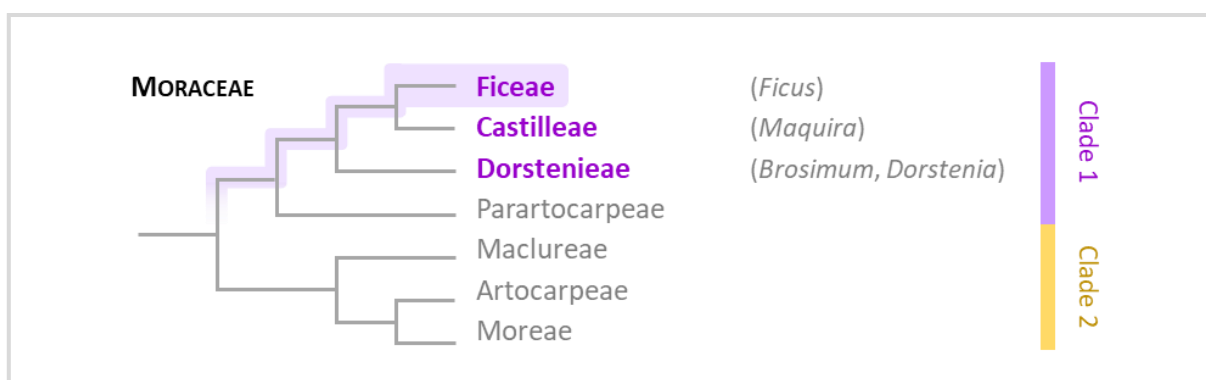
Figure 45 Emergence of the marmesin synthase activity associated to *CYP76F112* in the Moraceae family. The Moraceae tribes which are known to contains furanocoumarin producers are in purple. The phylogeny of the Moraceae family and the positioning of the different furanocoumarin-producing genera have been drawn according to Clement and Weiblen (2009) and Zerega and Gardner (2019) – also see **Chapter I**, **Figure 10**. The putative emergence of the marmesin synthase activity associated to CYP76F112, during the evolution of the Moraceae family is highlighted in purple.

This means that the marmesin synthase activity probably emerged in the Moraceae family after the divergence of the Ficeae / Castilleae / Dorstenieae / Parartocarpeae clade (Clade 1) and the Maclureae / Artocarpeae / Moreae clade (Clade 2 - **Figure 45**). After this initial divergence, several scenarios might be suggested:

First, it can be hypothesised that the marmesin synthase activity of the CYP76Fs emerged early after the divergence of the Clades 1 and 2: according to this scenario, some orthologues of *CYP76F112* should exist in various species of the Castilleae and Dorstenieae tribes. In the absence of additional genetic resources, these putative orthologous marmesin synthases can, however, not be investigated. If the marmesin synthase activity emerged early enough within Clade 1, it is even possible that *CYP76F112* orthologues exist in the Parartocarpeae tribe. The furanocoumarin content was not investigated in this tribe, but it contains poorly studied genera such as *Hullettia* and *Parartocarpus* (Zerega and Gardner 2019). According to this first hypothesis, the marmesin synthase activity would have emerged only once in Clade 1 and, unless it has been lost several times, furanocoumarins might be present in most species of the Castilleae, Dorstenieae (and Parartocarpeae) tribes.

The second hypothesis would be in favour of a multiple emergence of the marmesin synthase activity in Clade 1, probably by repeated evolution. In this scenario, *CYP76F112* might have emerged specifically in the *Ficus* genus, from a very recent expansion of CYP76F. The ancestral substrate(s) of its ancestral CYP76Fs may have been prenylated compounds (see **Chapter II**) similar enough to the DMS to allow the neo-functionalisation into a marmesin synthase. According to this scenario, it would be easy to imagine that the species from the Castilleae and Dorstenieae tribes independently recruited their own marmesin synthases from the same initial pool of CYP76Fs. Compared to the first hypothesis, if the marmesin synthases have been independently recruited multiple times in Clade 1, we could expect a less systematic presence of furanocoumarins in the different taxa.

Genetic resources and data concerning the presence of furanocoumarins in Clade 1 are scarce and not sufficient to promote one hypothesis rather than the other. In other words, the CYP76F subfamily has probably diversified in Clade 1 to give rise to the marmesin synthase activity, but we do not have enough information to go further into details.

> In summary, *CYP76F112* results from an expansion of the CYP76F subfamily that happened in the Moraceae family via multiple tandem duplication events followed by functional divergence. In particular, the marmesin synthase activity probably emerged in the Moraceae family after the divergence of the clades that contains the *Ficus* and the *Morus* genera. This recent evolution is a strong enough evidence to confidently state that the furanocoumarin biosynthesis pathway emerged multiple times in different plant families.

D.2. Evolutive story of *CYP81BN4*

The phylogenetic analysis of *CYP81BN4* – which has been found to metabolise cnidilin (**Chapter II**) – does not really allow to further discuss its physiological function. Indeed, *CYP81BN4* does not result from a Moreaceae-specific expansion, which would be in favour of a functional divergence, as it has been observed for *CYP76F112*. On the contrary, the closest *CYP81BN4* homologues tend to remain low-copy (1-2 copies in the Moraceae species) and they may all have the same activity, or function in the same pathway. If CYP81BN4 has the same function than its closest homologues found in the *Morus* genus, this physiological function would not be to oxidise cnidilin – since *Morus* do not produce furanocoumarins. *CYP81BN4* belongs to a subclade composed of 4 *Ficus* sequences which may have a different function than their closest *Morus* homologues, because the SRS analysis revealed the existence of some *Ficus*-specific amino acids. Such evolution of the SRSs might or might not be enough to change a P450's activity. But since no other activity is described for any other CYP81BN, the physiological function of CYP81BN4 and the emergence of this activity cannot be further clarified. In addition, even if its physiological activity were to oxidise cnidilin, CYP81BN4 could still not be used to find other candidates involved in the furanocoumarin biosynthesis pathway, because it does not possess many homologues in the *Ficus* species (only two homologues were found in the entire *F. erecta* genome).

For all these reasons, CYP81BN4 will not be further studied here.

D.3. Evolutive story of *CYP82J18*

Lastly, we can assume that the CYP82J subfamily is older than the NFC, and quite conserved in it. As the CYP82Js sequences were found in low-copy in the screened species, and all possess very similar SRSs, they might all code for enzymes that fulfil the same function. In **Chapter II**, CYP82J18 have been found to metabolise auraptene. This molecule has already been reported in several genera from the Rutaceae family, which includes many *Citrus* species (Gray and Waterman 1978; Ogawa et al. 2000; Genovese and Epifano 2011), and in some Apiaceae such as *Ferula szowitsiana* (Iranshahi et al. 2007; Soltani et al. 2010). To the best of my knowledge, it has never been described in any species belonging to the NFC. If auraptene is not present in this entire clade, the metabolization observed for CYP82J18 would clearly reflect an opportunistic activity, and the physiological function of the CYP82Js found in the NFC would still have to be elucidated. The "conserved" and "low copy" status of the CYP82J would be in favour of an essential function conserved by purifying selection (Nelson et al. 2008; Ilc et al. 2018). However, I would like to remind that CYP82Js disappeared in several species, such as *C. sativa*. All this taken together leads me to assume that the CYP82Js found in the NFC are probably not involved in a core primary metabolism pathway, but would be more likely involved in the specialised metabolism.

CYP82J18, which (opportunistic) activity is to metabolise a simple coumarin, has no particular interest for the elucidation of the furanocoumarin biosynthesis pathway, and it will not be further studied here.

D.4. Conclusion and future perspectives

As a very brief summary, in this chapter, I wanted to investigate how, where and when the 3 candidate P450s and their respective activities appeared in the course of evolution.

For *CYP81BN4*, I was not able to confidently answer to these questions.

For *CYP82J18*, it can be assumed that an ancestral CYP82J sequence predated the NFC. This putative ancestor may have already possessed the physiological function that might now be shared by all the CYP82Js found in the NFC, including *CYP82J18*.

For *CYP76F112* and its associated marmesin synthase activity, we can confidently conclude that they appeared recently. They emerged from a Moraceae-specific expansion that happened via repeated duplications events followed by diversification. This recent emergence is a significant finding in favour a multiple emergence of the furanocoumarin biosynthesis pathway in higher plants, by repeated evolution. In addition, these multiple duplications gave rise to numerous sequences, among which some might be interesting candidates for a further elucidation of the furanocoumarin biosynthesis pathway in the fig tree. Yet, one question has not been answered: how did the marmesin synthase activity emerged? Indeed, the general analysis of the CYP76Fs SRSs was not enough to identify the amino acids which might be specifically responsible for the acquisition of the marmesin synthase activity. Answering to this question has therefore been the objective of the next chapter.

Chapter IV
In the active site of the marmesin synthases
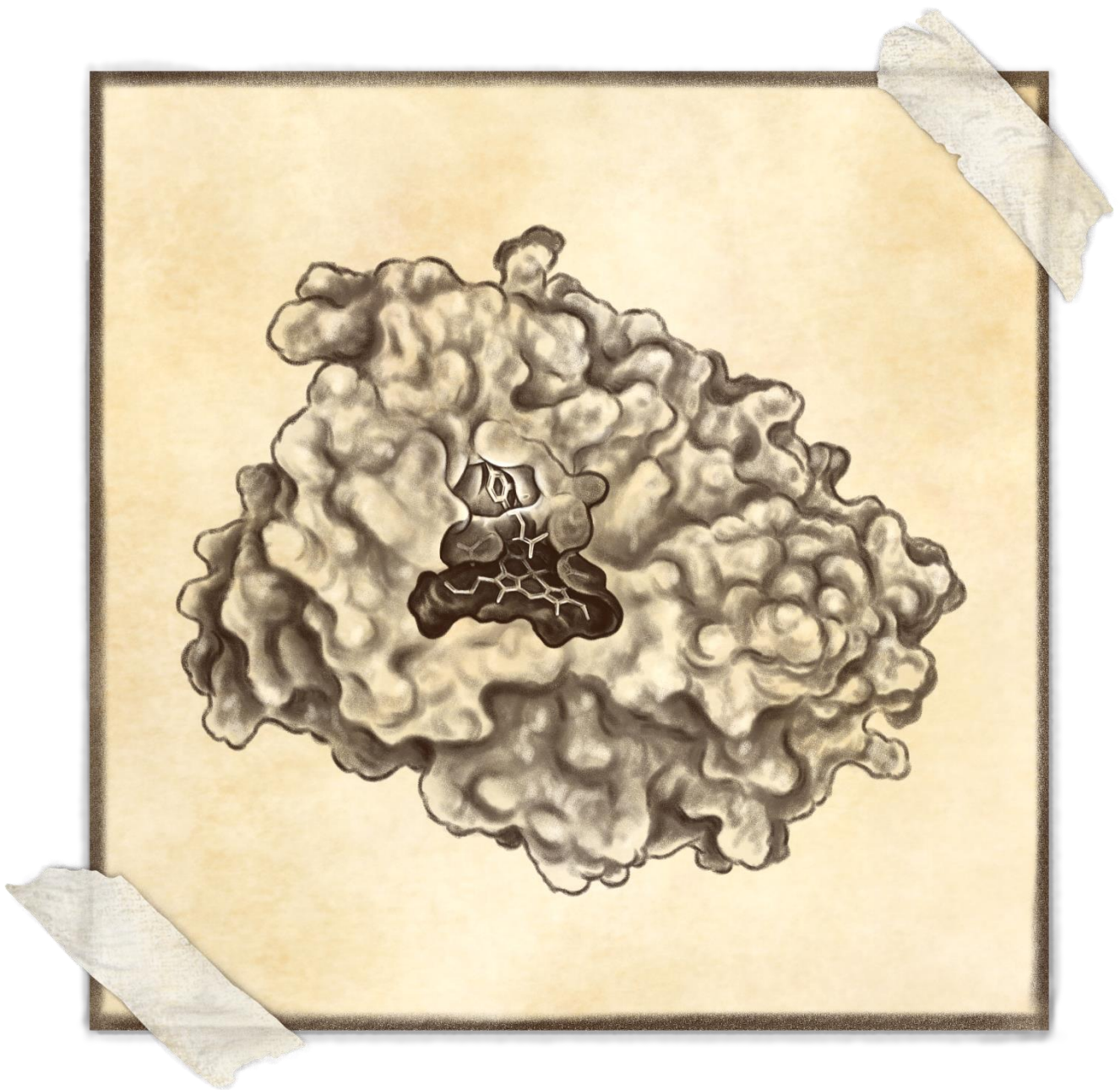
CHAPTER IV.

IN THE ACTIVE SITE OF THE MARMESIN SYNTHASES

~

Homology modelling and identification of amino acids that influence the substrate specificity and efficiency of CYP76F112

A. RISE OF THE MARMESIN SYNTHASES: INTRODUCTION AND STRATEGY

According to my previous results, the marmesin synthase activity has probably emerged in the Moraceae during a bloom of the CYP76F subfamily. In order to better understand how these enzymes evolved such a function, it would be interesting to identify amino acids responsible for the specificity of this reaction. Tracking such amino acids across the phylogeny of the CYP76F subfamily might help to clarify the emergence of the marmesin synthase and the furanocoumarin pathway in the Moraceae.

The strategy I set up to identify amino acid(s) responsible for the specificity of CYP76F112 was based on a comparison between CYP76F112 and CYP76F111, its closest *F. carica* homologue. Both enzymes share 62.2% protein sequence identity but, unlike CYP76F112, CYP76F111 does not accept DMS as a substrate, nor any of the (furano)coumarins tested in **Chapter II**. The sequence variations between these enzymes therefore appeared as a good starting point to highlight some differences that would explain the metabolization – or not – of DMS.

One powerful approach that can be used to identify essential amino acids involved in the interaction between an enzyme and its substrate consists in elucidating the tridimensional (3D) structure of the enzyme with X-ray or NMR-based technologies. However, as explained in **Chapter I** (**II. D**), most "typical" plant P450s are membrane-bound proteins, which makes such structure determination quite difficult to be done (Rupasinghe and Schuler 2006). For instance, the only crystal structures of "typical" plant P450s determined so far are those of CYP90B1 (*A. thaliana*) and CYP76AH1 (*Salvia miltiorrhiza*), which both required engineering to remove their membrane anchor before they were crystalised (Gu et al. 2019; Fujiyama et al. 2019). For these reasons, alternative methods are generally used to assess P450 structures. One of these alternatives consists in combining homology modelling and substrate docking analyses. These techniques provide *in silico*-based knowledge about the structure of an enzyme and its putative interaction with a substrate. Such a strategy is often reliable, but must be validated via site-directed mutagenesis experiments. Many examples are reported in Rupasinghe and Schuler (2006), more recent ones can be found in (Vazquez-Albacete et al. 2017; Li et al. 2019; Durairaj et al. 2019; Kuhlman and Bradley 2019; Sun et al. 2020).

> **In this chapter, my objective was to identify some amino acids that might be essential for the specificity of the marmesin synthase activity.**

For this purpose, I used homology modelling and docking approaches to realise a comparison between CYP76F112 and CYP76F111. This led me to highlight some amino acids which may be critical for the positioning of the DMS in the active site of these CYP76Fs. In a second time, I tested the *in-silico* models and predictions by using a site-directed mutagenesis approach.

B. THE MOLECULAR SHAPE OF YOUR P450: 3D MODELLING AND DOCKING

The first step of the structural analysis of the marmesin synthase consisted in using homology modelling to create 3D models of CYP76F111 and CYP76F112. Then, I used docking techniques to position the heme moiety and the DMS in the active sites of both enzymes.

B.1. Homology modelling of CYP76F111 and CYP76F112

B.1.a. Approach: the homology modelling technique

It has long been shown that two proteins sharing a high amino acid identity tend to adopt a similar structural fold (Chothia and Lesk 1986). Therefore, the 3D structure of a given protein (the template) can be transferred by homology to another protein (the query) that has a similar enough sequence. Of course, modelled structures are less accurate than the structures resulting from X-ray or NMR-based technologies. Yet, enough crystal structures of various P450s are now available to make reasonable predictions on new P450s. For review and detailed explanations about homology modelling, see Kirton *et al.* (2002); Kemp *et al.* (2005); Vazquez-Albacete *et al.* (2017) and Waterhouse *et al.* (2018).

In this study, the 3D models of CYP76F111 and CYP76F112 were generated using the SWISS-MODEL (https://swissmodel.expasy.org), an automated protein homology modelling server (Guex et al. 2009; Bienert et al. 2017; Bertoni et al. 2017; Waterhouse et al. 2018; Studer et al. 2020). To build a homology model of a given protein, the first step consists in identifying the best template available: a "good" template is a protein which 3D structure has already been described, and which amino acid sequence is as similar as possible to that of the query. When using the SWISS-MODEL, the identification of the best template(s) available in databases and the building of the 3D homology model of the query are automated. The expected quality of the model is estimated by using the Global Model Quality Estimate (GMQE) and the Qualitative Model Energy ANalysis (QMEAN) scoring functions. The GMQE reflects the expected accuracy of the model: it varies between 0 (not reliable) and 1 (highly reliable). The QMEAN is a composite estimator that reflects the global and local absolute quality estimates: it varies between -4 (low quality) and 0 (high quality) (Benkert et al. 2011; Biasini et al. 2014; Waterhouse et al. 2018).

B.1.b. Building the 3D models of CYP76F111 and CYP76F112

The protein selected as a template to build the 3D models of CYP76F111 and CYP76F112 was CYP76AH1 (PDB ID: 5YLW), a ferruginol synthase from *Salvia miltiorrhiza* recently crystallised by Gu *et al.* (2019). As the N-terminal transmembrane domain of CYP76AH1 was engineered to allow the

crystallisation, it was not included in the template. Consequently, the coverage between the truncated CYP76AH1 and the two CYP76Fs was only of 93%. Yet, apart from the missing N-terminal extremity, CYP76F111 and CYP76F112 respectively share 43.66% and 45.09% amino acid identity with CYP76AH1, which ensured a good sequence alignment (not shown). In addition, as explained in **Chapter I** (**II. D**), P450s generally share a highly conserved 3D structure, even when they are from different clans. In these conditions, it was reasonable to assume that the 3D structure of CYP76AH1 could be transferred to other CYP76s such as CYP76F111 and CYP76F112 with a high accuracy.

The 3D homology models I obtained are presented in **Figure 46**. The value of the GMQE and QMEAN associated to these models – respectively, 0.74 and -1.36 for CYP76F112, and 0.74 and -1.59 for CYP76F111 – were considered as correct (https://swissmodel.expasy.org), which made the two models reliable. In addition, despite a sequence identity a bit lower than 50%, the 3D models of CYP76F111 and CYP76F112 overlapped almost perfectly the CYP76AH1 model (**Figure 46**). In particular, the α-helices and β-sheets were extremely conserved. Therefore, these homology models can reasonably be considered as very accurate, and confidently used for the following docking experiments.
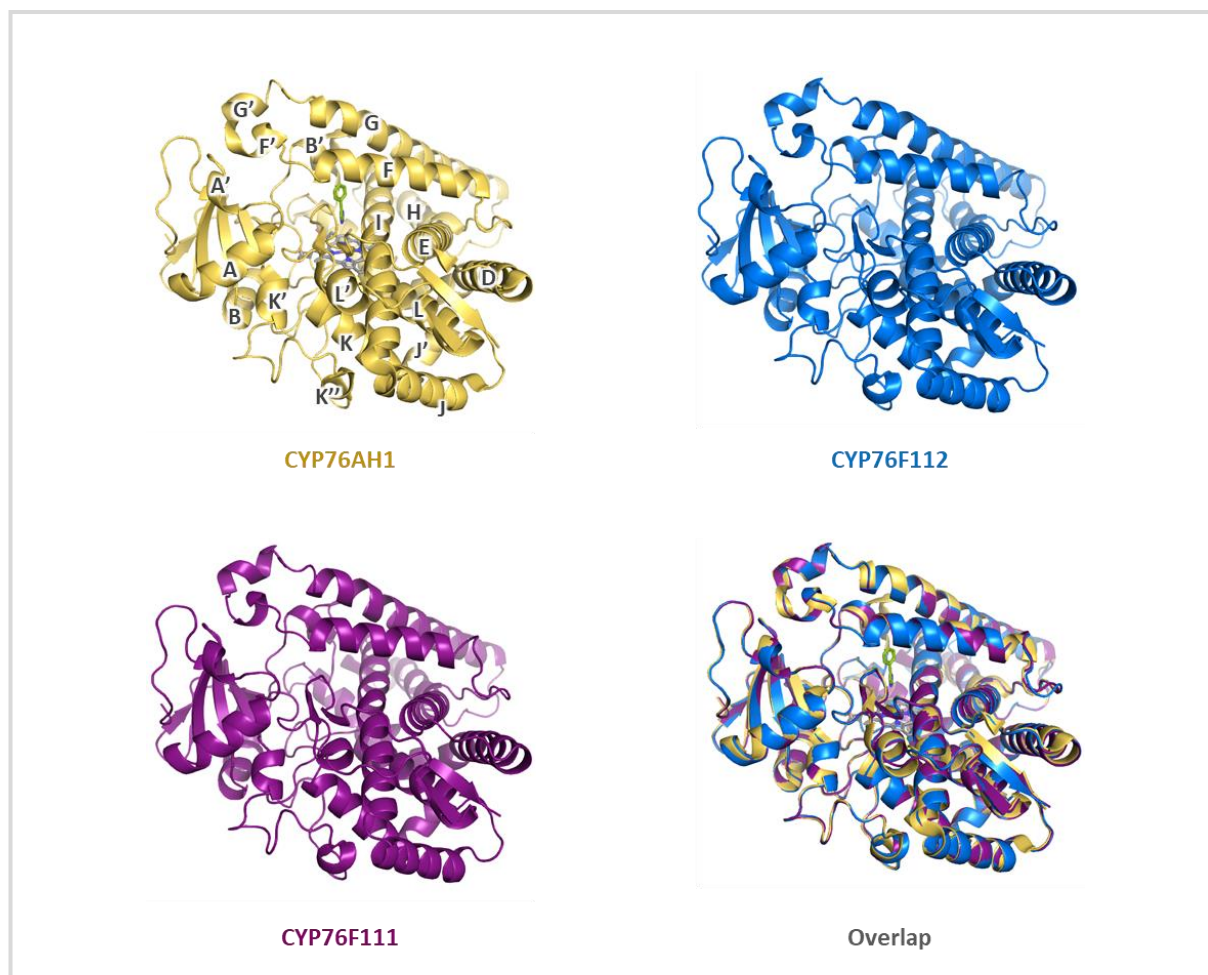
Figure 46 3D models of CYP76AH1, CYP76F112 and CYP76F111. The 3D models of CYP76F112 and CYP76F111 were generated by homology with that of CYP76AH1 (PDB ID: 5YLW). CYP76AH1 is shown in complex with 4-phenylimmidazole (PDB ID: 5YM3) and its α-helices have been annotated as described in Gu *et al.* (2019). The helices hardly visible in the back of the model are not annotated.

B.2. Docking experiments within CYP76F111 and CYP76F112

B.2.a. Approach: the docking technique

Molecular docking is a computational method that allows to predict noncovalent binding between two molecules, by relying on their 3D structure. It is often used to dock a small molecule (the ligand) within the catalytic site of a macromolecule (the receptor), but a similar procedure can also be applied to predict the binding between two macromolecules.

In this study, the substrate docking experiments were performed with the Autodock Vina software (http://vina.scripps.edu/ – Trott and Olson 2009). Prior to perform a docking, some essential parameters have to be defined. First, the area in which the ligand has to be docked can be restricted to a "search box" (the receptor grid) that is smaller than the entire receptor: restricting the receptor grid minimises the computational resources and time that are required. Then, the ligand and receptor can be defined as rigid or (partially) flexible bodies. As flexible residues are computationally demanding, enzymes are usually treated as mainly rigid, with only a few flexible amino acids. Finally, once the receptor grid and the rigid/flexible status of the molecules are defined, the docking is automatically performed by Autodock Vina, which generates a file containing the best models of the docked ligand. These models are ranked according to their energy conformation / binding affinity.

The dockings within CYP76F111 and CYP76F112 were performed in two steps. Indeed, as P450s are heme-containing enzymes, it was necessary to first dock a heme (the ligand) within the 3D homology models of CYP76F111 and CYP76F112 (the receptors). Then, a molecule of DMS (the ligand) was docked within CYP76F111 and CYP76F112 in complex with their heme (the receptors).

B.2.b. Docking of the heme within CYP76F111 and CYP76F112

As described in **Chapter I** (**II.D**), all P450s share the same global architecture and they all contain an absolutely conserved cysteine, which thiolate group interacts with the iron of the heme (Atkins and Sligar 1988; Hasemann et al. 1995; Werck-Reichhart et al. 2002; Rupasinghe et al. 2003; Mestres 2004; Mclean et al. 2011). For these reasons, the positioning of the heme within CYP76F111 and CYP76F112 must be similar to that of other P450s whose structure has been elucidated, such as CYP76AH1. Therefore, by analogy to CYP76AH1, I defined a restricted but realistic receptor grid around the conserved cysteine of both CYP76F111 and CYP76F112 (**Chapter VII**, **B.6.c.2**). This grid was bigger than the heme but small enough to reduce the computation work. During the docking, the heme was treated as flexible and the P450 as rigid – except for the conserved cysteine which was defined as flexible because of its interaction with the heme.

When the heme was docked within CYP76F112, two models were generated. Both were ranked with close affinities of -12.6 and -12.4 kcal/mol, and their positioning in the enzyme was almost identical. The model with the best affinity (-12.6 kcal/mol) was selected as the most probable binding mode (**Figure 47**). I repeated the process to dock a heme within CYP76F111 and obtained similar results. In the best models obtained with CYP76F111 and CYP76F112, the heme was positioned almost exactly like the heme of CYP76AH1 (**Figure 47**), which suggested a successful docking.
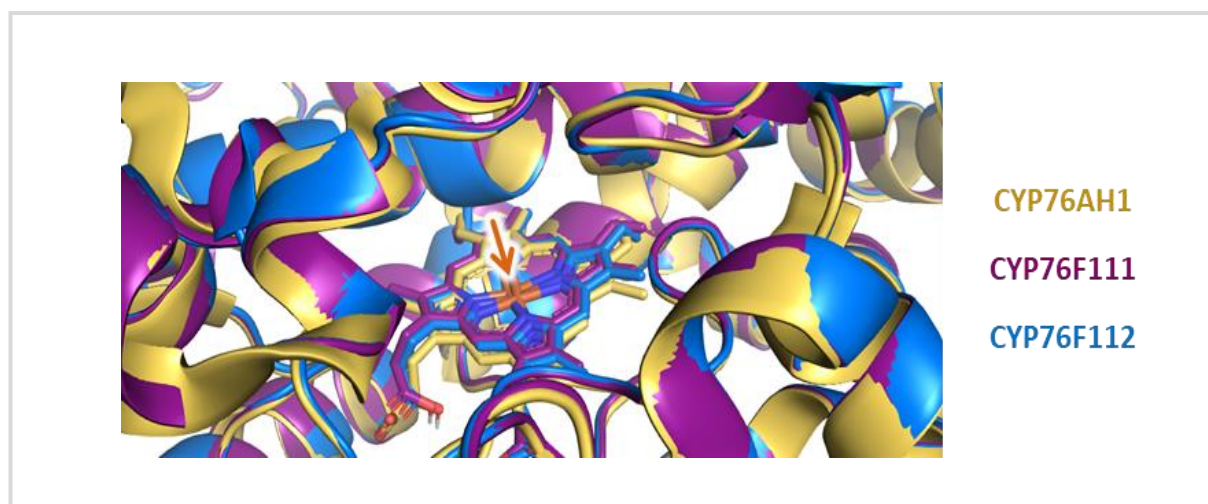
Figure 47 Hemes docked in the 3D models of CYP76F112 and CYP76F111 (best models). Both models can be overlapped with that of CYP76AH1 (PDB ID: 5YLW). The arrow points toward the heme irons.

B.2.c. Docking of the DMS within CYP76F111 and CYP76F112

B.2.c.1. Approach: definition of the grid receptor and the flexible residues

For an easier understanding of the choices I made, it is noteworthy to remind that the substrate-binding pocket of a P450s is usually placed "above" the heme, on the other side of the conserved cysteine (**Chapter I. II.D** – Rupasinghe et al. 2003; Poulos and Johnson 2015). During a typical reaction, a dioxygen molecule binds to the heme iron and one of its oxygen atoms is transferred on the substrate (Werck-Reichhart and Feyereisen 2000; Zangar et al. 2004; Mclean et al. 2011). Consequently, to dock the DMS within CYP76F111 and CYP76F112 complexed with their respective heme, I constructed a receptor grid around the heme iron, which was extended "above" the heme (**Chapter VII**, **B.6.c.2**). During the docking, the DMS was treated as flexible. On the contrary, the receptor (*i.e.* the P450 complexed with its heme) was defined as rigid, except for a set of 15 amino acids that were treated as flexible (**Supp. Table 8**). These 15 amino acids were chosen because they are located in the active site of the enzyme, and were determined to be closer than 6Å from the docked DMS in several preliminary docking experiments (not shown).

B.2.c.2. Docking of the DMS within CYP76F112

When the DMS was docked within CYP76F112 associated with its heme, 9 models were generated. These models were ranked with affinity ranging from -8.9 to -7.0 kcal/mol (**Supp. Table 9**). The model with the lowest energy conformation (-8.9 kcal/mol) was selected as the most probable binding mode (**Figure 48**). In this model, the position of the DMS was perfectly conform to what is expected from a P450 that converts DMS into marmesin. Indeed, the carbons C2' and C3', which are the two carbons of the DMS involved in the reaction (highlighted on **Figure 48** with purple and blue arrows), were positioned just above the heme, at respectively 4.3 and 4.6 Å from the iron. Such a distance is consistent with the reactive range described for other P450s (Larbat et al. 2007; Hritz et al. 2008; Li et al. 2009; Du et al. 2017; Fujiyama et al. 2019; Durairaj et al. 2019; Sun et al. 2020).

On the contrary, in the 8 other models displaying higher energy conformation (**Supp. Table 9**), the DMS was either "badly orientated" (the carbon C2' and C3' were not directly located above the iron) and/or at more than 6 Å from the heme (**Supp. Table 9**).
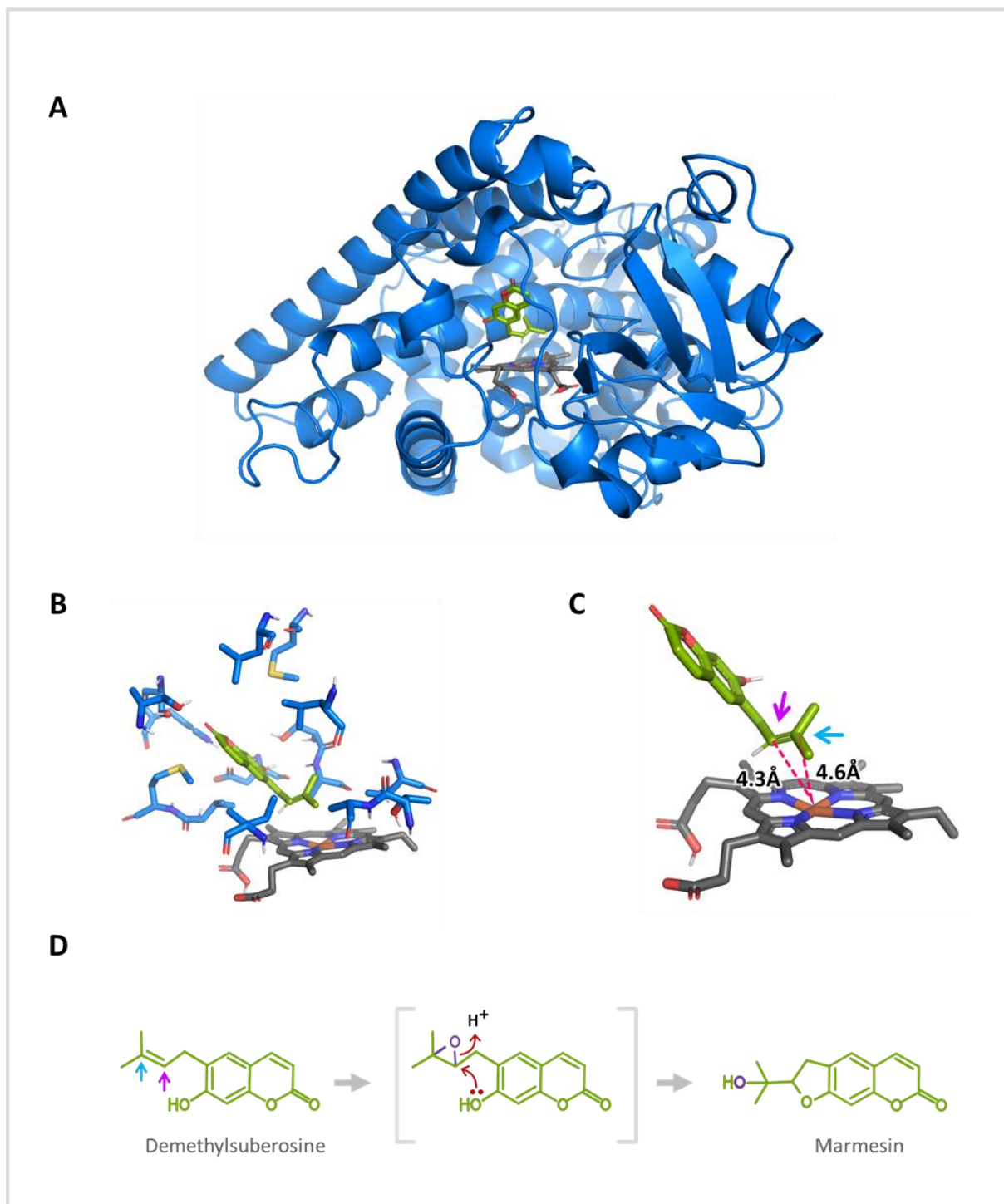
Figure 48 Docking of the DMS within CYP76F112. A 3D model of CYP76F112 (blue) in complex with a heme (grey) and the DMS (green). **B** Zoom on the docking site. **C** Zoom on the DMS. **D** Proposed mechanism of the P450-mediated conversion of the DMS into marmesin, adapted from Mizutani and Sato (2011). Also see **Chapter I**, **Figure 19**. The carbons C2' and C3' of the DMS, respectively highlighted with purple and blue arrows, are the ones involved in the reaction.

Lastly, it should be noted that some of the flexible amino acids moved during the docking (**Figure 49**). The most notable movements are those of the threonine T102 and the serine S105, which side chains get reoriented toward the oxygen from the keton group (C=O) of the DMS. In addition, the methionines and the leucines M117 and M211, L215 and L487 get reoriented a bit, which positions them closer to the DMS. The other amino acids barely or did not move.

Figure 49 Movement of the flexible amino acids during the docking of the DMS within CYP76F112. The blue model represents the docking site and the heme before the docking of the DMS. It is superimposed with the docking site and the heme after the docking of the DMS, in green.

B.2.c.3. Docking of the DMS within CYP76F111

When the DMS was docked within CYP76F111, 9 models were generated. These models were ranked with affinity ranging from -8.2 to -6.7 kcal/mol (**Supp. Table 9**). The model that has the lowest energy conformation (-8.2 kcal/mol) was selected as the most probable binding mode (**Figure 50**). In this model, the DMS was "turned upside down" in the active site of CYP76F111, compared to its position in CYP76F112. As a consequence, the C2' and C3' carbons are positioned far from the heme iron, making the conversion of DMS into marmesin impossible.

In the other models, which had higher energy conformations, the DMS was also positioned in ways that made the reaction very unlikely (**Supp. Table 9**). This means that none of the models obtained with CYP76F111 would allow the conversion of DMS into marmesin.

In summary, the best model obtained with CYP76F112 placed the DMS in a position that strongly suggested it could be converted into marmesin. On the contrary, none of the models obtain with CYP76F111 suggested that the reaction could happen.

The results of this *in-silico* analysis are consistent with the findings of **Chapter II**; which may be seen as another argument in favour of the reliability and accuracy of the models.
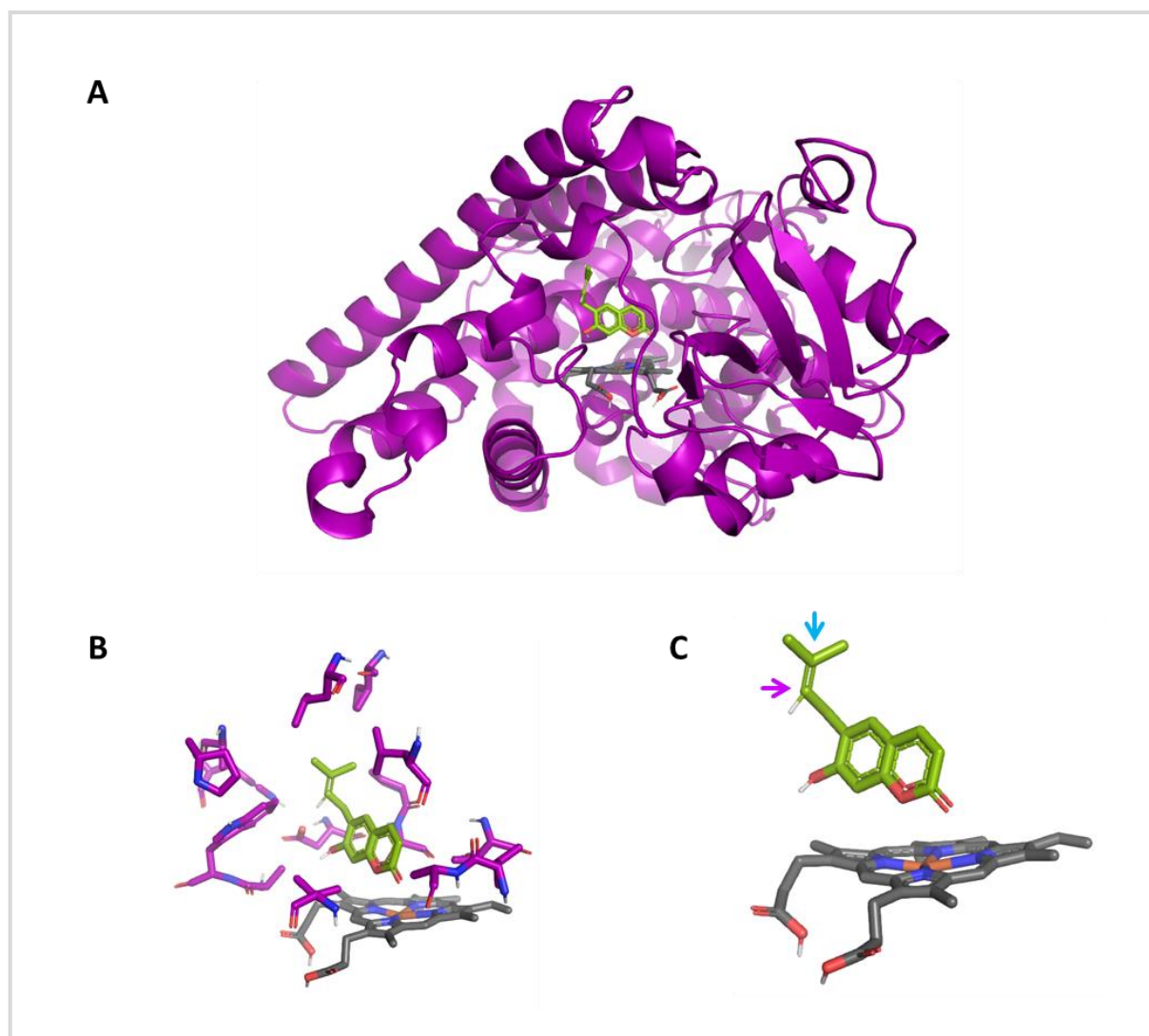
Figure 50 Docking of the DMS within CYP76F111. A 3D model of CYP76F111 (purple) in complex with a heme (grey) and the DMS (green). **B** Zoom on the active site: the visible amino acids are the one treated as flexible during the docking. **C** Zoom on the DMS. The carbons C2' and C3' of the DMS are highlighted with purple and blue arrow, respectively (see **Figure 48**).

C. FINDING KEY AMINO ACIDS INFLUENCING THE DOCKING OF THE DMS

The second step of this analysis consisted in identifying some amino acids that might be responsible for the differential positioning of the DMS within the active sites of CYP76F111 and CYP76F112. To do so, I relied on a comparison of CYP76F112 and CYP76F111, and modified their sequence to visualise the impact of some mutations on the docking of the DMS.

C.1. Comparison of the docking sites of CYP76F111 and CYP76F112

The comparison of CYP76F111 and CYP76F112 was based on the following assumption: the amino acids that have the strongest influence on the positioning of the DMS within CYP76F111 and CYP76F112

should be the ones that are the closest to the docked DMS. According to this hypothesis, these amino acids might all belong to the set of 15 flexible residues defined in the previous part (the "docking site"), for it includes the only 13 residues positioned at less than 5Å from the DMS (**Supp. Table 8**).

A simple comparison between the respective docking sites of CYP76F111 and CYP76F112 showed that, out of the 15 amino acids, 7 were identical in both enzymes (**Table 7**). These amino acids were of no interest to explain the differential positioning of the DMS within the two enzymes. On the contrary, the 8 remaining amino acids were different in CYP76F111 and CYP76F112: this means they harboured different properties and/or sizes that may significantly impact (or not) the positioning of the DMS into the active site (**Table 7**).

Table 7 Comparison of the docking site of CYP76F111 and CYP76F112. The amino acids (AA) described in a single row are equivalent in the two enzymes. The ones highlighted in blue are variable in CYP76F111 and CYP76F112 but are all hydrophobic. The ones highlighted in purple are variable and exhibit different properties. "Hydro." stands for "hydrophobic", "Neg." for "negative". The interesting properties (distance to the DMS, movement during the docking) are in red, the others are in grey. The four amino acids of interest (A, B, C, D) were identified through preliminary docking experiments which are described in the following paragraphs.

SRS	Amino acid in...		Comparison CYP76F112 -CYP76F111	Dist. to the DMS as positioned in CYP76F112		Move during docking?	Interest?
	CYP76F112	CYP76F111		F112	F111		
SRS1	T102 - Polar	P105 - Other	Different AA	<3Å	<4Å	Yes	A
	S105 - Polar	A108 - Hydro.	Different AA	<3Å	<4Å	Yes	B
	H110 - Polar	H113 - Polar	Same AA	<5Å	<5Å	No	
	V116 - Hydro.	A119 - Hydro.	Different AA	<4Å	<5Å	No	
	M117 - Hydro.	W120 - Hydro.	Different AA	<4Å	<3Å	Yes	C
SRS2	M211 - Hydro.	I214 - Hydro.	Different AA	<6Å	<8Å	Yes	
	L215 - Hydro.	L218 - Hydro.	Same AA	<5Å	<5Å	Yes	
SRS4	D302 - Neg.	D307 - Neg.	Same AA	<3Å	<3Å	No	
	S305 - Polar	P310 - Other	Different AA	<4Å	<5Å	No	D
	A306 - Hydro.	A311 - Hydro.	Same AA	<4Å	<4Å	No	
	T310 - Polar	T315 - Polar	Same AA	<4Å	<4Å	No	
SRS5	T370 - Polar	V375 - Hydro.	Different AA	<6Å	<7Å	No	
	A371 - Hydro.	A376 - Hydro.	Same AA	<5Å	<5Å	No	
	I375 - Hydro.	V380 - Hydro.	Different AA	<4Å	<4Å	No	
SRS6	L487 - Hydro.	L490 - Hydro.	Same AA	<5Å	<5Å	Yes	

C.2. Influence of the variable amino acids during the docking of the DMS

To evaluate the impact of the 8 variable amino acids and only select the most relevant ones, I modified them (individually and simultaneously) in the sequences of CYP76F111 and CYP76F112, replacing them by their equivalent from the other enzyme. Then, I generated 3D homology models for each of these virtual mutants, as explained in **B.1**. Such punctual modifications did not modify the overall structure of the mutants, compared to the original enzymes (not shown). Finally, I docked a heme and the DMS in every mutant, as explained in **B.2.b** and **B.2.c.** The results obtained with the most interesting mutants – and only those – are presented in the following paragraphs.

C.2.a. Preliminary results: identification of 4 key amino acids

In a first attempt, I generated homology models by simultaneously modifying the 8 variable amino acids, or only one of them at a time, in CYP76F111 and CYP76F112. This preliminary approach led to identify 4 amino acids of interest that were predicted to impact the positioning of the DMS within the active sites of CYP76F111 and CYP76F112. The modification of the 4 other variable amino acids did not impact the docking. For this reason, I will now focus on the 4 residues of interest that seemed to play a critical role during the docking.

In CYP76F112, the 4 amino acids of interest are T102, S105, M117 and S305. Their respective equivalents in CYP76F111 are P105, A108, W120 and P310. For an easiest understanding, the positions of these amino acids have been respectively named A, B, C and D (**Table 7**). In CYP76F112, the amino acids A and B are polar residues that are very close to the DMS (<3Å) and move during the docking. Their respective equivalent in CYP76F111 are not polar and a bit further from the docked DMS. The amino acid C is hydrophobic in both enzymes, it moves during the docking, but it is bigger and therefore closer to the DMS in CYP76F111 rather than in CYP76F112. Lastly, the amino acid D is polar in CYP76F112 but not in CYP76F111. Contrary to the amino acids A and B, the D is not as close to the DMS (<4Å) and it does not move during the docking (**Table 7**).

The mutated versions of CYP76F111 and CYP76F112 will be designated by the name of the original enzyme, followed by the letters corresponding to the position of the modified amino acids: for instance, "CYP76F112-A" corresponds to the sequence of CYP76F112 in which the threonine in position A has been replaced by a proline – its equivalent from CYP76F111 (**Table 7**). A summary of the sequence modifications applied to every mutant described in the following paragraphs can be found in **Supp. Table 10**.

C.2.b. Simultaneous modification of the 4 amino acids A, B, C and D

The first investigated mutants were CYP76F111-ABCD and CYP76F112-ABCD – which correspond to the simultaneous replacement of the 4 amino acids in both enzymes (**Supp. Table 10**). When the DMS was docked within CYP76F112-ABCD, 9 models were generated (**Supp. Table 11**). The two models with the lowest energy conformation (-8.6 kcal/mol for both of them) were selected as the most probable binding modes (**Figure 51**). In the first model, the DMS was docked in a position that was very similar to its position within CYP76F111: the DMS was "upside down", and the carbons C2' and C3' were far
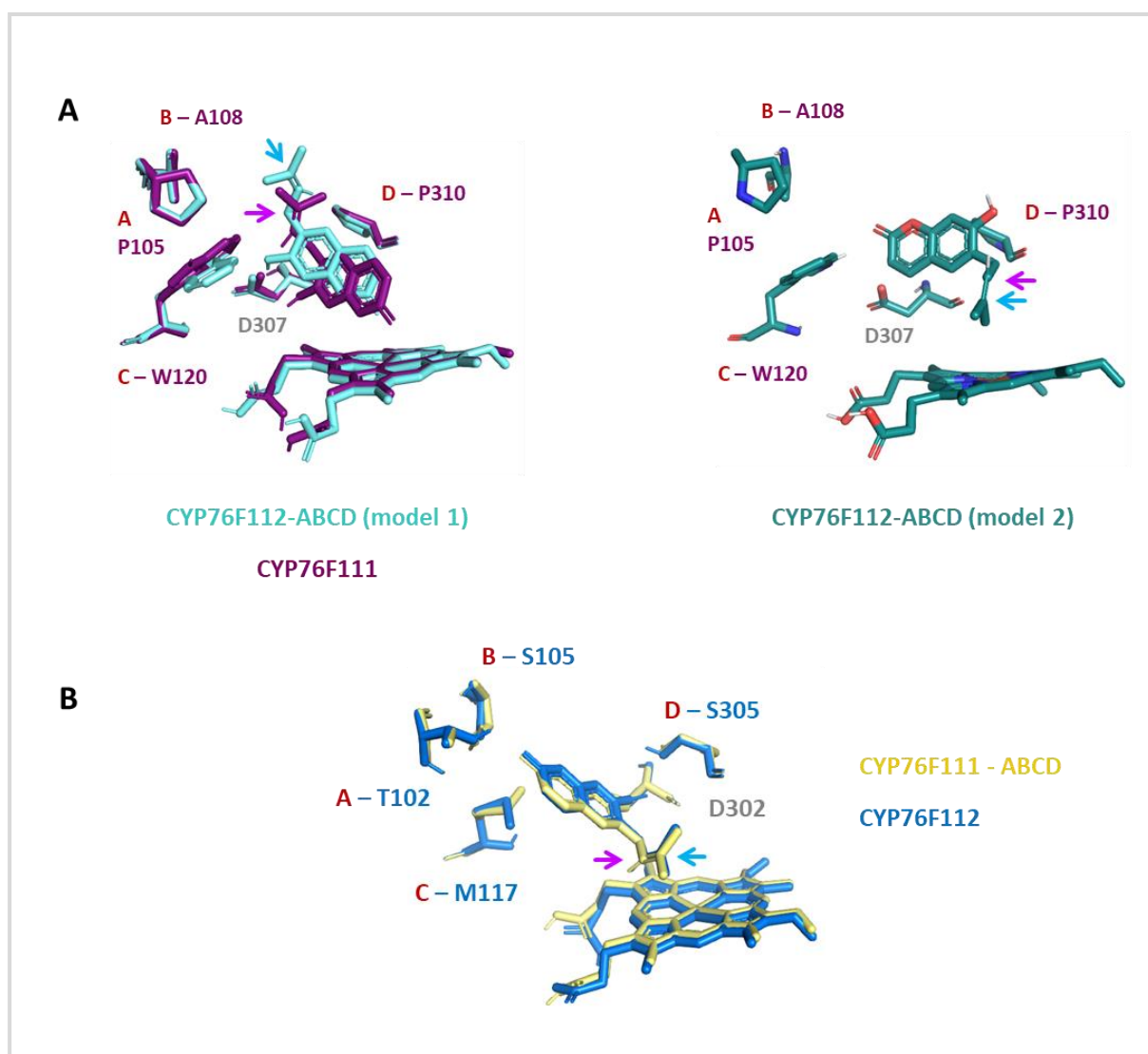
Figure 51 Docking of the DMS within CYP76F111-ABCD and CYP76F112-ABCD. The amino acids shown in this figure are A, B, C and D. The carbons C2' and C3' of the DMS are highlighted with purple and blue arrow, respectively (see **Figure 48**). **A** Two best models corresponding to the docking within CYP76F112-ABCD. The first best model is overlapped with that of CYP76F111. **B** Best model generated during the docking of the DMS within CYP76F111-ABCD, overlapped with that of CYP76F112.

from the heme iron. In the second-best model, the DMS was also docked in a position that was quite different from its position in CYP76F112: the carbons C2' and C3' were not the ones that were the closest to the heme iron, which would make the conversion into marmesin very unlikely. According to these two best models, the position of the DMS within CYP76F112-ABCD would not allow its conversion into marmesin. The other models with higher energy conformations (**Supp. Table 11**) were also checked, but none of them gave a better positioning (not shown). These results suggested that CYP76F112-ABCD would not be able to convert DMS into marmesin.

Likewise, the DMS was docked within CYP76F111-ABCD. The model with the lowest energy conformation (-8.0 kcal/mol – **Supp. Table 11**) was selected as the most probable binding mode (**Figure 51**). This time, the DMS was docked in the exact same position than within CYP76F112. This model therefore suggested that CYP76F111-ABCD may be able to convert DMS into marmesin.

In brief, the docking simulations with these first mutants suggested that the simultaneous replacement of the amino acids in position A, B, C and D might be enough to make CYP76F112 lose its marmesin synthase activity, and to create this activity in CYP76F111.

C.2.c. Individual modifications of A, B, C and D

The role of the amino acids in position A, B, C and D have been further investigated with single modifications: the associated mutants were CYP76F112-A, CYP76F112-B, CYP76F112-C and CYP76F112-D (**Supp. Table 10**).

First, the DMS was docked within CYP76F112-A and CYP76F112-B. In the most probable binding mode of each enzyme (-8.8 kcal/mol and -8.7 kcal/mol, respectively – **Supp. Table 11**), the position of the DMS was the same than within CYP76F112 (**Figure 52**). These results suggested that the individual mutations of the residues in position A or B may have no impact on the docking of the DMS within CYP76F112, and that CYP76F112-A and CYP76F112-B should be able to convert DMS into marmesin.
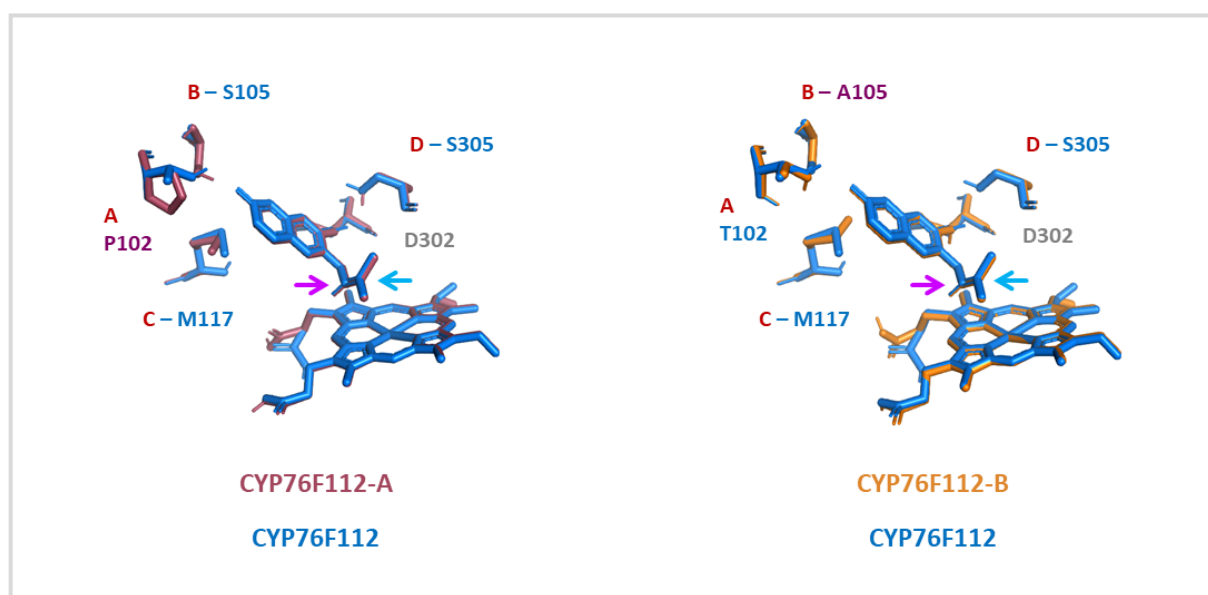
Figure 52 Docking of the DMS within CYP76F112-A and CYP76F112-B. The best models generated during the docking within the 2 mutants are overlapped with that of CYP76F112. The carbons C2' and C3' of the DMS are highlighted with purple and blue arrow, respectively (see **Figure 48**).

Then, the DMS was docked within CYP76F112-D. In the two first models (same affinity of -8.2 kcal/mol – **Supp. Table 11**), the DMS was docked in a position that would not allow its conversion into marmesin (**Figure 54**). Indeed, these two models were identical to the two best models obtained during the docking of the DMS within CYP76F112-ABCD. This suggested that CYP76F112-D would not be able to convert DMS into marmesin. Yet, in the third best model (-8.0 kcal/mol), the DMS was docked in the same position than in CYP76F112. So, CYP76F112-D may in fact exhibit a marmesin synthase activity.
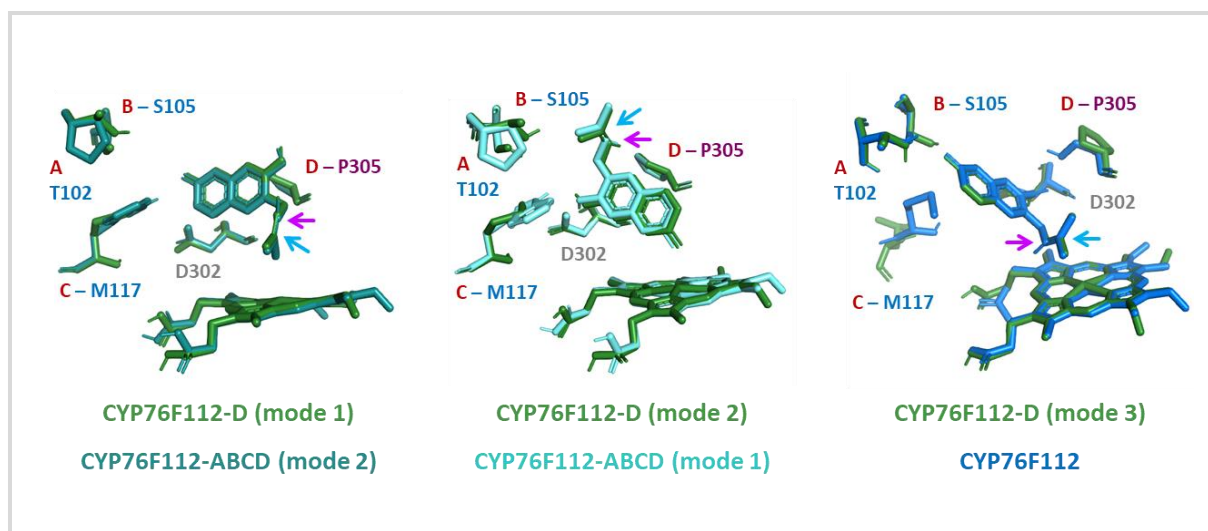
Figure 54 Docking of the DMS within CYP76F112-D. The best modes corresponding to the docking within CYP76F112-D are overlapped with that of CYP76F112-ABCD and CYP76F112. The carbons C2' and C3' of the DMS are highlighted with purple and blue arrow, respectively (see **Figure 48**).

Finally, regarding the mutant C, things were different: when the DMS was docked within CYP76F112-C, the best model (-8.6 kcal/mol – **Supp. Table 11**), was quite similar to that of CYP76F112, but could not be perfectly overlapped with it (**Figure 53**). Indeed, in CYP76F112-C, the DMS was docked a bit farther from the heme rather than in CYP76F112. In particular, the carbons C2' and C3' were positioned at 5.7Å and 4.8Å from the heme iron – compared to 4.3Å and 4.6Å in CYP76F112, respectively. Moreover, in this mutant, the carbons C2' and C3' were not located directly above the heme anymore. The extremity of the prenyl group seemed to interfere or prevent any direct interaction between the oxygen that bound on the heme iron and these carbons. This model suggested that the position of the
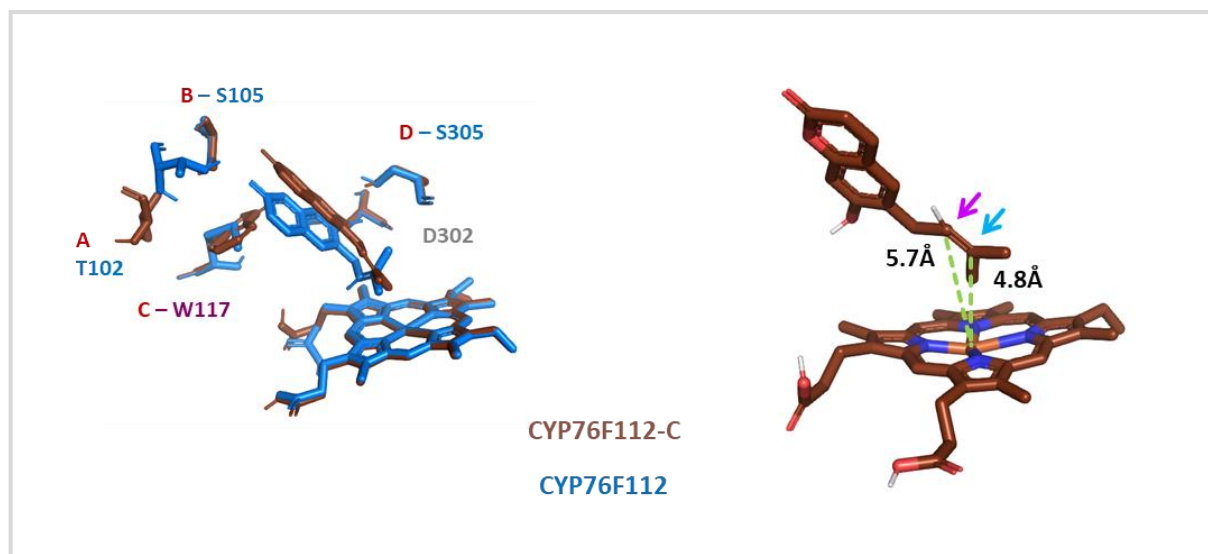
Figure 53 Docking of the DMS within CYP76F112-C. The best model corresponding to the docking within the mutant (in brown) is overlapped with that of CYP76F112 (in blue). The carbons C2' and C3' of the DMS are highlighted with purple and blue arrow, respectively (see **Figure 48**).

DMS within CYP76F112-C would not allow its conversion into marmesin. The other models generated during the docking, with higher energy conformations, were also checked, but none of them seemed to place the DMS in a "better" way (**Supp. Table 11**). So, according to these results, a single mutation on the amino acid in position C may be enough to make CYP76F112 lose its marmesin synthase activity.

C.2.d. Simultaneous modifications of A and B

At this point, the amino acids in position A and B did not seem to have an essential role during the docking of the DMS. But then, I docked the DMS in a last additional mutant: CYP76F112-AB (**Supp. Table 10**). In the best model (-8.1 kcal/mol – **Supp. Table 11**), the DMS was docked "upside down", in a position that was quite similar to that in CYP76F111 (**Figure 55**). This position might prevent the conversion of DMS into marmesin. The other models generated during the docking (higher energy conformations) were also checked, but none of them seemed to place the DMS in a position allowing its metabolization (**Supp. Table 11**). These results suggested that CYP76F112-AB would not be able to convert DMS into marmesin. This would mean that, if the individual mutations of A or B does not affect the docking of the DMS within CYP76F112, their combined mutation probably does.
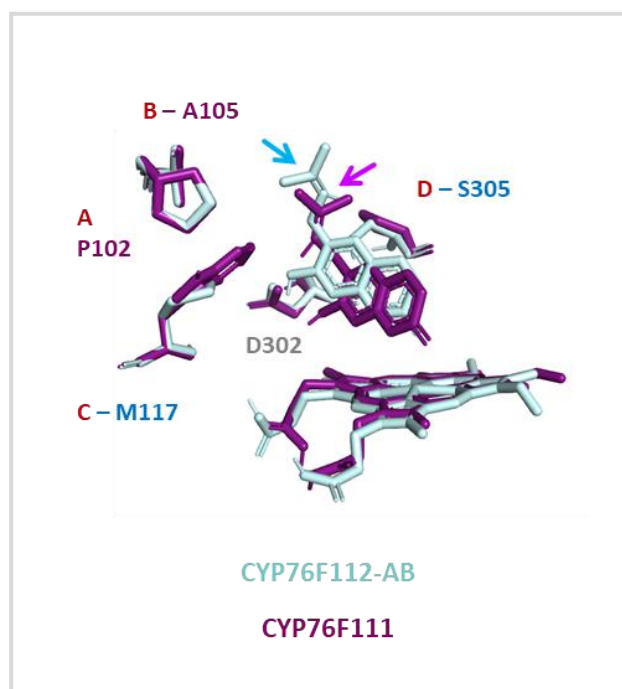
Figure 55 Docking of the DMS within CYP76F112-AB. The best model is overlapped with that of CYP76F111. The C2' and C3' of the DMS are highlighted with purple and blue arrow, respectively (see **Figure 48**).

In summary, the docking experiments permitted to highlight 4 residues in position A, B, C and D that might contribute to explain the differential position of the DMS within the substrate-binding pocket of CYP76F111 and CYP76F112.

The combined mutations of these 4 amino acids might prevent the marmesin synthase activity of CYP76F112, and allow it in CYP76F111.

Individually, a mutation on the amino acids in position C or D might be enough to impact the activity of CYP76F112. On the contrary, it seems that the amino acids in position A and B should both be modified to alter it.

D. SITE-DIRECTED MUTAGENESIS: 4 AMINO ACIDS, AND NOTHING ELSE MATTERS?

In a last step, I investigated the impact of the amino acids A, B, C and D on the marmesin synthase activity, using an experimental approach. This was done by using site-directed mutagenesis, associated to heterologous expression and enzymatic functional characterisation.

D.1. The choice of the mutants

To perform this validation, 7 mutants were of particular interest (**Table 8**). CYP76F111-ABCD and CYP76F112-ABCD were required to evaluate the simultaneous effect of the 4 amino acids. CYP76F112-A, CYP76F112-B, CYP76F112-C and CYP76F112-D were necessary to confirm their individual effect. Lastly, CYP76F112-AB was required to test the combined effect of A and B.

If the results of the docking were correct, the marmesin synthase activity could be impacted (completely lost or reduced) in CYP76F112-ABCD, CYP76F112-AB, CYP76F112-C and CYP76F112-D (**Table 8**). In CYP76F112-A and CYP76F112-B, it would not be impacted, or just slightly. The impact on the activity could be reflected by altered K_m or K_{cat} values. Finally, CYP76F111-ABCD may be able to convert DMS into marmesin.

Table 8 Preliminary hypotheses related to the 7 mutants included in the experimental validation

Mutant	Docking of the DMS	Expected impact on the activity
CYP76F112-A	Correct positioning in the best model	No impact (or reduced activity)
CYP76F112-B	Correct positioning in the best model	No impact (or reduced activity)
CYP76F112-C	No model with a correct positioning	Reduced or no activity
CYP76F112-D	Correct positioning in the third model	Reduced activity
CYP76F112-AB	No model with a correct positioning	Reduced or no activity
CYP76F112-ABCD	No model with a correct positioning	Reduced or no activity
CYP76F111-ABCD	Correct positioning in the best model	Marmesin synthase activity

D.2. Synthesis and expression of the mutants

In order to test the activity of the 7 mutants, their nucleotide sequences have first been synthesised and inserted into the pYeDP60expression vector. These nucleotide sequences are those of *CYP76F111* and *CYP76F112*, in which the codon(s) corresponding to the amino acid(s) to modify have been replaced by the codon(s) from the other enzyme. To allow highlighting the expression of the proteins, a His-tag has also been added before their respective stop codons (see **Chapter II. C.2**). To be comparable, the original genes of *CYP76F111* and *CYP76F112* – initially cloned into the pYeDP60_GW version – were also cloned into the pYeDP60 vector, as explained in the materials and methods section (**Chapter VII**, **B.1.n**). Then, the 9 vectors containing *CYP76F111*, *CYP76F112* and the seven mutants

(**Supp. Figure 5**) have been introduced into the WAT21 yeast strain. The recombinant proteins were heterologously expressed in yeast, and successfully collected by preparing yeast microsomes (**Figure 56** – also see **Chapter II**, **C.2**). It should be noted that the production of the CYP76F111 and CYP76F111-ABCD enzymes was less successful that that of CYP76F112 and its associated mutants.
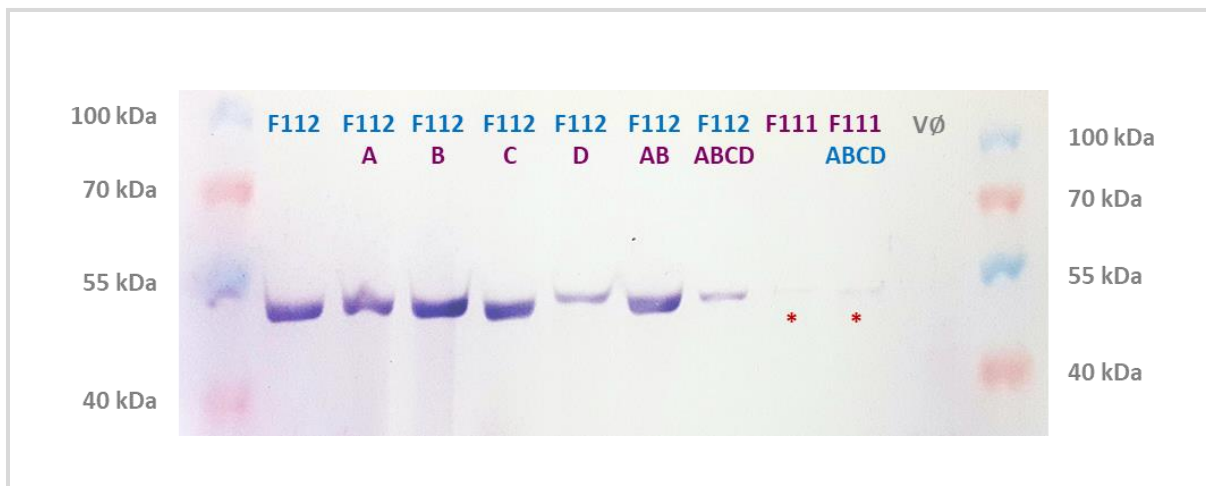
Figure 56 Evaluation of the expression of the CYP76F mutants by immunodetection. Western Blot performed on the microsomes collected after the expression of the CYP76F mutants in the yeast WAT21, directed against the His-tag. VØ referred to the negative control (empty vector). The two red asterisks highlight the two mutants for which the signal was weak but detectable.

D.3. Functional characterisation: influence of the 4 amino acids

The microsomes prepared from the yeasts transformed with each enzyme have been incubated in presence of DMS and NADPH. All the incubations were performed with fresh microsomes produced within the day (**Chapter II. D.1**). After the incubations, the reaction products were analysed by UHPLC-MS. As expected, the results showed that CYP76F112 converted DMS into marmesin. On the contrary, CYP76F111 did not metabolise DMS. These controls confirm that the change of the expression vector did neither impact the production of functional enzymes, nor the activity of these enzymes.

Regarding the mutants, only 4 were able to metabolise DMS: CYP76F112-A, CYP76F112-B, CYP76F112-AB and CYP76F112-D. For each of these 4 mutants, DMS was converted into a single reaction product, which was identified as marmesin by direct comparison of its retention time and mass spectrometry spectrum with those of the standard molecule (not shown). The 3 others mutants – CYP76F112-C, CYP76F112-ABCD and CYP76F111-ABCD – did not metabolise DMS (summarised in **Table 9**). This means that, as hypothesised from the results of the docking experiments (**Table 8**), the individual (and combined) mutation of the amino acid C was enough to prevent the metabolization of DMS by CYP76F112. On the contrary, the combined mutation of the amino acids A, B, C and D in CYP76F111 was not enough to allow this enzyme to metabolise DMS.

The 4 mutants which were able to convert DMS into marmesin have been further characterised by determining their kinetic parameters – as it was done for CYP76F112 in **Chapter II**, **D.2.a.3**. Briefly, I

first dosed the P450s present in the microsomal solutions by using the differential CO spectrum method (**Supp. Figure 6**). Then, I incubated known amounts of P450s in presence of variable concentrations of DMS: the modalities of the incubations, different for every mutant, are described in **Supp. Table 12** and **Chapter VII**, **B.2.f.3**. The reaction products were then concentrated and the quantity of formed marmesin was quantified by UHPLC-MS analyses (**Supp. Table 12**). Finally, Michaelis-Menten saturation curves were plotted (**Supp. Figure 7**). The kinetic parameters associated to the conversion of DMS into marmesin by the mutants are reported in **Table 9**.

Table 9 Summary of the site-directed mutagenesis experiments. Activity and kinetic parameters of the 7 mutants. For comparison, the kinetic parameters of CYP76F112 have also been reported.

Enzyme	Marmesin synthase activity	K_m (nm)	K_{cat} (min^{-1})	K_{cat}/K_m ($\mu M^{-1}.min^{-1}$)
CYP76F112	**Yes**	**32.2 ± 3.9**	**22.1 ± 0.9**	**687**
CYP76F112-ABCD	No	---	---	---
CYP76F111-ABCD	No	---	---	---
CYP76F112-A	Yes	90.3 ± 13.8	3.9 ± 0.3	44
CYP76F112-B	Yes	34.4 ± 3.8	27.8 ± 1.1	809
CYP76F112-C	No	---	---	---
CYP76F112-D	Yes	114.3 ± 7.2	275.7 ± 5.9	2412
CYP76F112-AB	Yes	117.8 ± 14.4	28.8 ± 1.7	244

Compared to CYP76F112, the mutant CYP76F112-A had a reduced affinity (apparent K_m = 90.3 ± 13.8 nm) for the DMS and a lower catalytic constant (apparent K_{cat} = 3.9 ± 0.3 min^{-1}). This means CYP76F112-A needed a higher concentration of substrate to be as active as CYP76F112, and metabolised a smaller number of molecules of DMS in a given time. Consequently, the catalytic efficiency (K_{cat}/K_m) of this mutant (44 $\mu M^{-1}.min^{-1}$) was about 15 times lower than that of CYP76F112 (**Table 9**). This implies that, contrary to what had been expected from the docking experiments (**Table 8**), the individual mutation of the amino acid A did impact CYP76F112 activity.

The K_m associated to the mutant CYP76F112-B was not significantly different than that of CYP76F112 and its K_{cat} was just slightly higher (**Table 9**). This means the individual modification of the amino acid B barely impacted CYP76F112 activity. This is consistent with the hypotheses resulting from the docking experiments (**Table 8**).

No metabolization of DMS was expected for CYP76F112-AB (**Table 8**), but this mutant did convert DMS into marmesin. The apparent K_m value of the mutant CYP76F112-AB (117.8 ± 14.4 nM) was about 4 times higher than for CYP76F112 and CYP76F112-B. It might be a bit higher than that of CYP76F112-A, but the difference was not significant. The apparent K_{cat} associated to CYP76F112-AB was slightly

higher than that of CYP76F112 (**Table 9**). Therefore, the combined mutations of the amino acids A and B significantly impacted CYP76F112 activity: in particular, the impact of this double mutation on the K_m is similar to that of the single mutation of A, while the very slight increase of the K_{cat} reminds that of the individual mutation of B.

Finally, with an apparent K_m value of 114.3 ± 7.2 nM, the affinity of the DMS for the mutant CYP76F112-D was about 4 times lower than for CYP76F112. On the contrary, and unexpectedly, the apparent K_{cat} associated to CYP76F112-C was more than 10 times higher than CYP76F112 apparent K_{cat}. This means that, in a given time, the mutant can metabolise more molecules of DMS than CYP76F112. Because of its high K_{cat} value, this mutant also had a higher catalytic efficiency (2412 $\mu M^{-1}.min^{-1}$, **Table 9**). Thus, the individual mutation of the amino acid D significantly impacted CYP76F112, which is consistent with the hypotheses resulting from the docking experiments (**Table 8**). In particular, this single mutation led to a decreased affinity which was comparable to that of the double AB mutation, but to an improved catalytic constant.

In brief, the results obtained with the site-directed mutagenesis experiments confirmed most of the hypotheses resulting from the docking experiments, but not all of them. **This shows a relative robustness of the modelling and the docking approach – since we were able to predict the influence of some single amino acid mutations – but also some limits that will be further discussed latter.**

One the one hand, by modifying the amino acid C in CYP76F112, I was able to completely prevent the metabolization of DMS. As expected, the single modification of the amino acid B did not significantly impact CYP76F112 activity. Yet, the respective mutations of the amino acids A, AB, and D did impact the affinity and catalytic constant of CYP76F112.

On the other hand, by simultaneously modifying the amino acids A, B, C and D in CYP76F111, I could not allow this enzyme to metabolise DMS. This implies that other amino acids which were not studied here also play an important role in CYP76F112 marmesin synthase activity.

E. GOTTA DOCK THEM ALL: ADDITIONAL DOCKINGS WITH THE F112-LIKE

Finally, as the docking approach seemed quite robust, I performed additional *in silico* analyses with some of the F112-like highlighted in **Chapter III**. As a reminder, the "F112-like clade" is a clade composed of 17 *Ficus* CYP76Fs that includes CYP76F112, the two putative marmesin synthases from *F. erecta* and *F. religiosa*, and CYP76F111. It can be subdivided into 4 subclades, as shown in **Figure 57** (also see **Chapter III.D.1**). The F112-clade might include several marmesin synthases, distributed within the 4 subclades, or concentrated in the subclade that include CYP76F112. Therefore, I modelized some enzymes from the F112-like clade and performed additional docking experiments with the DMS.

Four enzymes were selected for these additional dockings (highlighted with black dots on **Figure 57**). First, I selected G02_76F_F.erecta and G02_76F_F.religiosa, because they belong to the same subclade than CYP76F112, but are not its closest homologues. They might thus help us to guess if this whole subclade is composed of marmesin synthases – or not. Then, I selected G04_76F_F.erecta and G11_76F_F.erecta, which respectively belongs to two different subclades. These enzymes might provide information about the repartition of the marmesin synthase activity between the subclades. As CYP76F111 did not metabolise DMS, I did not choose any additional enzyme from its subclade.
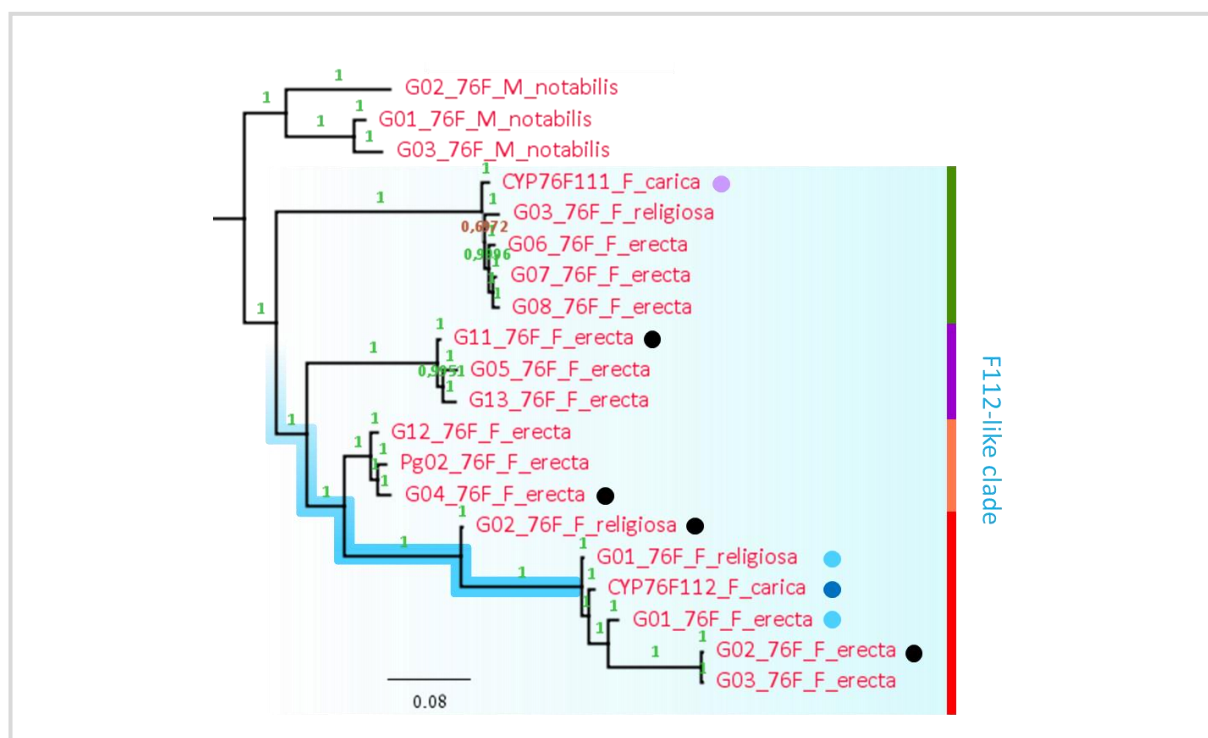
Figure 57 Enzymes from the F112-like clade tested in the additional docking experiments. The phylogenetic tree is a detail from **Figure 38** (**Chapter III**). The F112-clade is in blue; it can be subdivided into 4 subclades that are coloured in green, purple, orange and red. The *F. carica* marmesin synthase is highlighted with a deep blue circle, the putative marmesin synthases from *F. religiosa* and *F. erecta* are highlighted with light blue circles. CYP76F111, that does not metabolise DMS, is highlighted with a purple circle. The enzymes included in the additional docking experiments are highlighted with black circles. The putative emergence of the marmesin synthase activity defined in **Figure 44** (**Chapter III**) is highlighted on the branches with a blue path.

The 3D homology models of these 4 additional F112-like have been generated as explained previously (**B.1**). Then, a heme and the DMS have been docked in every model (**B.2**).

In the best models obtained with the enzymes G02 from *F.erecta* and *F.religiosa* (**Supp. Table 13**), the DMS was docked in the same position than in CYP76F112 (**Figure 58**). This suggested that both enzymes might display a marmesin synthase activity. By extension, it can also be assumed that the entire subclade that contains CYP76F112, G01-03 from *F.erecta* and G01-02 from *F.* religiosa (highlighted in red on **Figure 57**) might be able to convert DMS into marmesin.
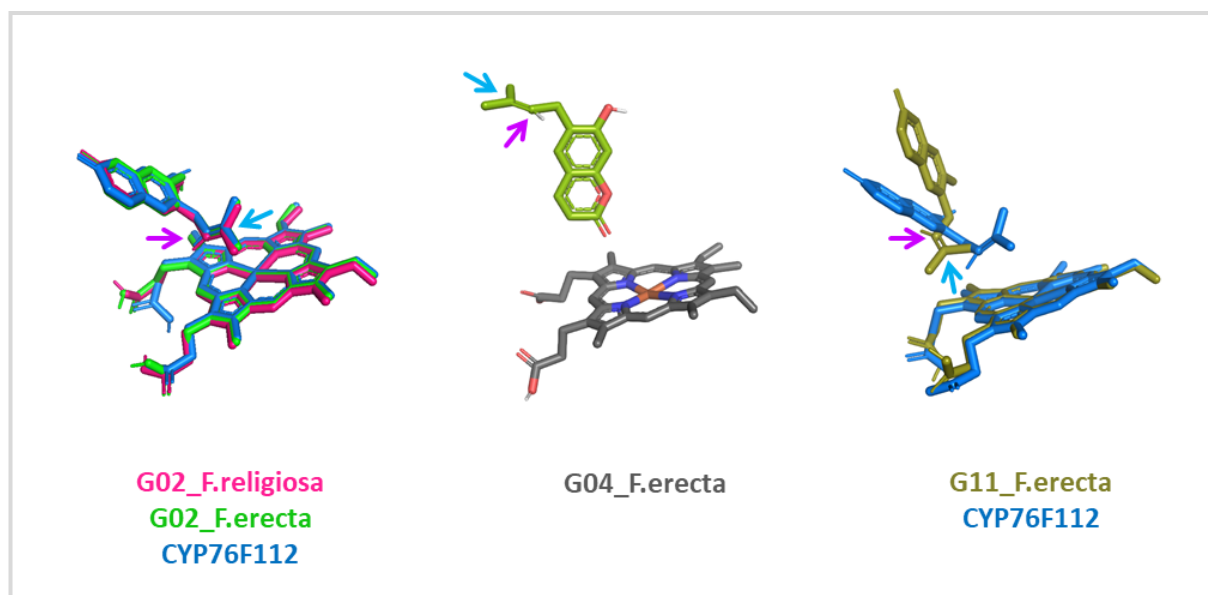
Figure 58 Docking of the DMS within some of the F112-like. The carbon C2' and C3' of the DMS are highlighted with purple and blue arrows, respectively (see **Figure 48**).

On the contrary, in the best models obtained with G04_76F_F.erecta and G11_76F_F.erecta (**Supp. Table 13**), the positioning of the DMS above the heme would not allow its conversion into marmesin (**Figure 58**). The other models were also checked, but none of them gave a better positioning of the DMS (**Supp. Table 13**). These enzymes probably do not convert DMS into marmesin.

In brief, these additional docking experiments allowed me to highlight a subclade of CYP76Fs that contains CYP76F112 and 5 other enzymes that should all be able to convert DMS into marmesin. On the contrary, the F112-like from the other subclades might not be able to metabolise DMS.

Consequently, it can be hypothesised that the subclade that contains CYP76F112 is exclusively composed of marmesin synthases, and that the marmesin synthase activity is specific and restricted to this subclade. But of course, it should be reminded that these *in silico* analyses are only predictions which would require further experimental confirmation.

F. DISCUSSION: A SINGLE AMINO ACID IS MISSING, AND ALL BEGINS ANEW

In this chapter, my objective was to point out amino acids responsible for the specificity of CYP76F112, in order to clarify the emergence of the marmesin synthase activity. By generating homology models and performing docking analyses on CYP76F111, CYP76F112, and modified versions of these enzymes, I identified four residues that seemed to influence the positioning of the DMS within the active site of the CYP76Fs. Subsequent site-directed mutagenesis experiments specified the impact of these four residues on the marmesin synthase activity.

F.1. CYP76F111 and CYP76F112: reliable modelling and accurate docking

F.1.a. Overall architecture of CYP76F111 and CYP76F112

The 3D models of CYP76F111 and CYP76F112 have been built by homology with that of CYP76AH1: the use of a CYP76 as a template, with enzymes that usually share such a high structure conservation, ensured a very precise and reliable modelling. The overall structure of CYP76AH1 consists in 12 main α-helices (α-A to α-L) and 3 β-sheets (β-1 to β-3) (Gu et al. 2019). As these structural elements are also present in the 3D models of CYP76F111 and CYP76F11, a structurally annotated illustration of CYP76F112 can be proposed, by analogy to that of CYP76AH1 (**Figure 59**).

In CYP76AH1, two access channels were described: a wide channel located between the F/G loop and the β-domain, and a small one located close to the first one (Gu et al. 2019). Similar channels can be found in the 3D models of CYP76F111 and CYP76F112, and the big channel looks wide enough to let the DMS enter into the substrate-binding pocket (**Figure 59**).

F.1.b. Confronting the in-silico dockings with the in-vitro experiments

When the DMS was docked within the 3D model of CYP76F112, the best model positioned it above the heme, with its carbons C2' and C3' in close vicinity to the iron (**Figure 48**). This model was the only one in which the position of the DMS seemed to allow its conversion into marmesin. So, as CYP76F112 has indeed a marmesin synthase activity (**Chapter II**), we can suggest that the best docking model may illustrate the real positioning of the DMS within CYP76F112 when it is converted into marmesin.

On the contrary, the DMS docked into CYP76F111 was positioned "upside down", in a way that would not allow the reaction to happen. In **Chapter II**, CYP76F111 did not metabolise DMS: this may have been caused by inadequate reaction conditions or non-functional enzymes. Yet, the models produced by these docking experiments tend to rule out this possibility, in favour of the following one: the DMS was not metabolised by CYP76F111 because it cannot be correctly positioned in its active site. Or, more simply, the DMS is really not a substrate for CYP76F111.

F.2. The impact of the residues A, B, C and D on the marmesin synthase activity

F.2.a. Key residues from the SRSs

Numerous docking analyses and site-directed mutagenesis experiments performed on various P450s have demonstrated the role of various amino acids in the determination of P450 catalytic properties – such as substrate specificity, selectivity, or efficiency. The great majority of these amino acids belong to the 6 substrate recognition sites (SRSs – see **Chapter I**, **II.D.1.d**), but it seems that there is not a unique rule to predict which amino acid will be responsible for the properties of a given P450. Indeed, as P450s have highly variable primary sequences and metabolise a large diversity of substrates, every P450 might interact differently with its substrate(s). As a result, the amino acids reported to influence a P450 activity are almost never the same, have various properties, can be located at different positions in the active site, and involve diverse type of interaction with their respective substrate
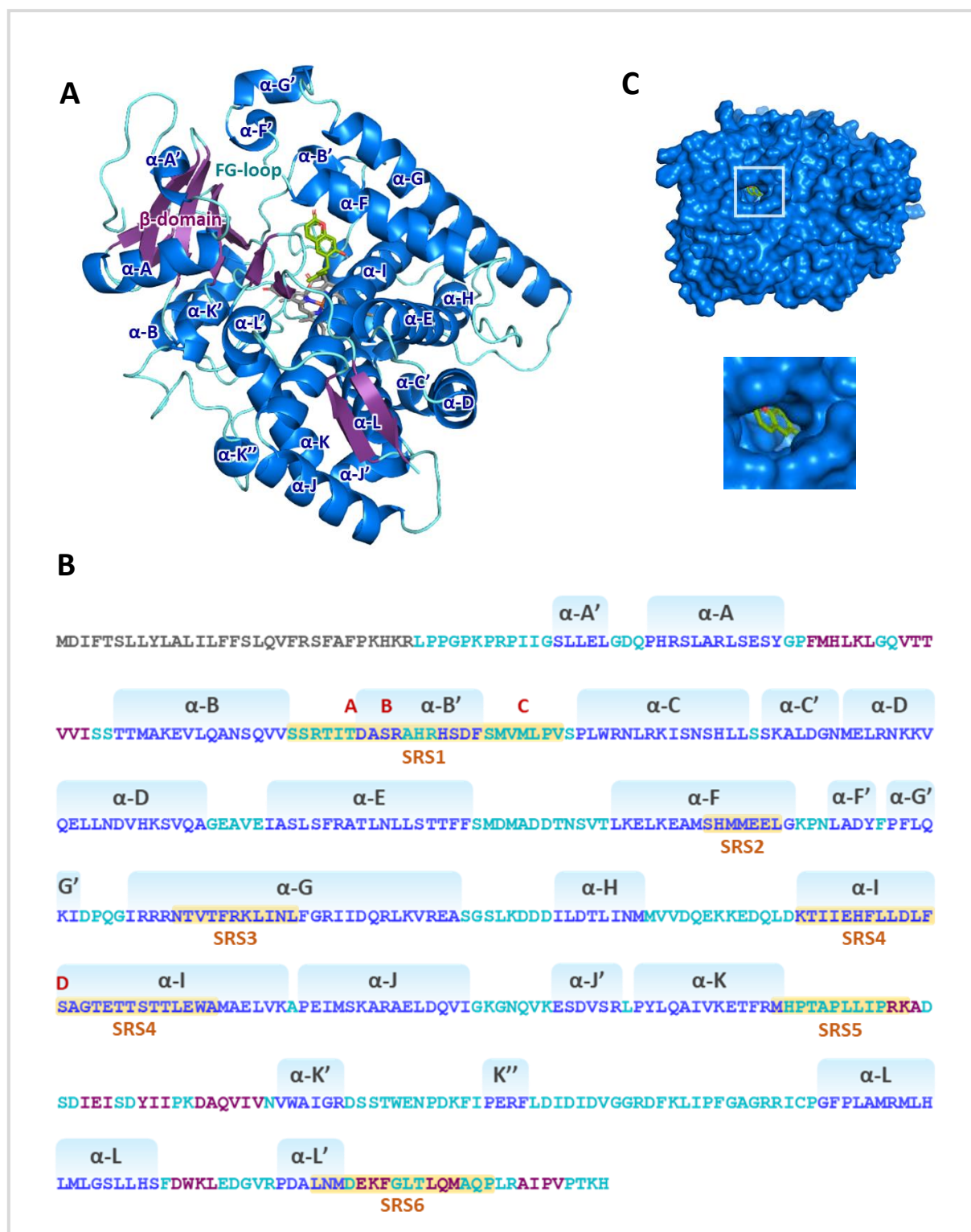
Figure 59 Overall structure of CYP76F112, complexed with a heme and a molecule of DMS. The α-helices are in deep blue, the β-sheets in purple, the loops in turquoise, the heme in grey, and the DMS in green **A** Annotated 3D model of CYP76F112. The α-helices α-B and α-C which are hardly visible, on the back of the model. **B** Amino acid sequence of CYP76F112. The part of the sequence that were not modelled are in grey. The SRSs and α-helices are detailed. The residues A, B, C and D are in red. **C** Zoom on the DMS docked into the substrate-binding pocket, view from the access channel.

(Schalk et al. 1999; Schalk and Croteau 2000; Kahn et al. 2001; Sawada et al. 2002; Schoch et al. 2003; Takahashi et al. 2005; Rupasinghe and Schuler 2006; Li et al. 2008a; Lee et al. 2008; Seifert et al. 2009; Sirim et al. 2010; Roberts et al. 2010; Vazquez-Albacete et al. 2017; Sun et al. 2020). However, some systematic analyses conducted on many P450s highlighted a few hotspot for activity and selectivity (Sirim et al. 2010).

In this study, the docking analyses permitted to highlighting 4 key residues from CYP76F111 and CYP76F112 that seemed to influence the positioning of the DMS within the active site: the amino acids in position A (P105/T102), B (A108/S105), C (W120/M117) and D (P310/S305) (**Table 10**). The residues A, B and C belong to the SRS1 while the residue D is located into the SRS4 (**Figure 59**).

Table 10 Main characteristics of the amino acids in position A, B, C and D. Also see **Table 7**.

AA	SRS	Amino acid		Side chain		Dist. to DMS		Move during docking?	Impact of the individual mutation on CYP76F112?
		F112	**F111**	**F112**	**F111**	**F112**	**F111**		
A	SRS1	T102	P105	Polar	Other	<3Å	<4Å	Yes	Reduced K_m and K_{cat}
B	SRS1	S105	A108	Polar	Hydro.	<3Å	<4Å	Yes	Almost no impact
C	SRS1	M117	W120	Hydro.	Hydro.	<4Å	<3Å	Yes	No activity
D	SRS4	S305	P310	Polar	Other	<4Å	<5Å	No	Reduced K_m Increased K_{cat}

As described in **Chapter I** (**II.D.1**), the SRS1 corresponds to the highly variable BC-loop (*i.e.* the loop located between the helices α-B and α-C) while the SRS4 is located in the middle of the α-I helix. The SRS1 and SRS4 are both positioned over the heme, flanking the substrate-binding cavity. Some of their residues have been predicted to be in contact with the substrate and to interact with it (Rupasinghe et al. 2003; Rupasinghe and Schuler 2006; Seifert et al. 2009; Sirim et al. 2010; Schuler and Rupasinghe 2011). Moreover, several studied demonstrated that the SRS1 (Seifert et al. 2009; Sirim et al. 2010; Roberts et al. 2010; Vazquez-Albacete et al. 2017) and the SRS4 (Schalk et al. 1999; Sawada et al. 2002; Schoch et al. 2003) contribute to determine the activity and specificity of some P450s.

In the present study, the site-directed mutagenesis experiments confirmed that the individual and combined modification of the residues A, B, C and D impact CYP76F112 activity by modifying its substrate specificity (amino acid C) and kinetic parameters (amino acids A, B and D). These results therefore constitute another evidence of the influence of the SRS1 and SRS4 on P450s catalytic properties. In addition, based on the results of the dockings and the site-directed mutagenesis experiments (**Table 10**), it is now possible to propose some hypotheses regarding to the respective role of these residues toward the positioning of the DMS within the active sites of the CYP76Fs and their impact on the marmesin synthase activity.

F.2.b. The residues A and B might stabilise the DMS in the active site

As a reminder, the single mutation of the amino acid A reduced the affinity of the DMS for CYP76F112 (K_m increased about 3 times – **Table 9**). This means that the possession of a threonine in position A (as in CYP76F112) instead of a proline (as in CYP76F111) allows the enzyme to remains active with lower concentrations of DMS. The presence of the threonine in CYP76F112 is therefore consistent with the concentration of DMS in the fig tree, which is probably very low (discussed in **Chapter II**, **E.2.a**). Similarly, the replacement of this threonine by a proline reduce the catalytic constant of CYP76F112 (**Table 9**), which means that the threonine may help converting more molecules of DMS in a given time. This would also be consistent with the low amount of DMS in the fig tree, since the presence of a threonine in position A would help metabolising DMS quickly, preventing its accumulation.

The individual mutation of the amino acid B barely impacted CYP76F112 activity, but the combined mutation of A and B had a different effect on CYP76F112 activity than the single mutation of A (**Table 9**). On the one hand, the K_m of CYP76F112-AB (117.8 ± 14.4 nm) was not significantly different than that of CYP76F112-A (90.3 ± 13.8 nm – **Table 9**). The reduced CYP76F112-AB affinity might thus be entirely due to the mutation of the amino acid A. But on the other hand, if the K_{cat} value – which was reduced by the individual mutation of A – was barely affected by the double AB mutation. This suggests that the amino acids A and B might somehow act in synergy, making difficult to predict their combined effect on the enzyme efficiency. The evolutionary implications of these mutations will be discussed in **F.3**; for now, let us try to understand how the amino acids in position A and B impact the affinity of the marmesin synthase for the DMS.

In the 3D models of CYP76F111 and CYP76F112, the residues A and B are localised side by side (**Figure 60**). In CYP76F112, these amino acids – the threonine T102 and the serine S105 – are both polar (**Table 10**). During the docking, they were both reoriented toward the DMS (**Figure 49**), which was positioned in their close vicinity (<3Å). In particular, in the best docking model, the hydroxy groups at the extremity of the side chains of T102 and S105 pointed toward the ketone group of the DMS (**Figure 60**). On the contrary, in CYP76F111, the residues A (P105) and B (A108) are not polar.
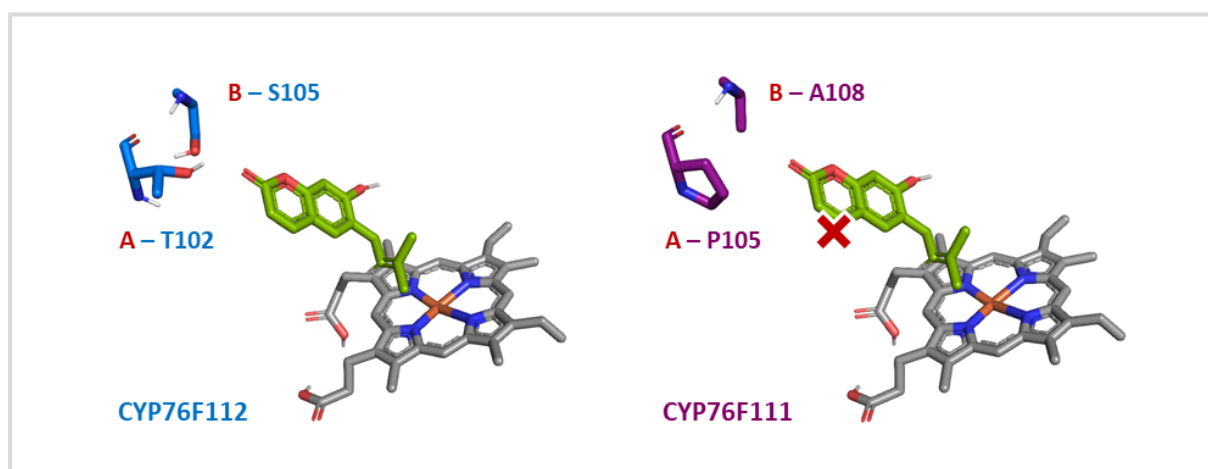
Figure 60 Influence of the residues A and B. Residues A and B from the 3D models of CYP76F112 (blue) and CYP76F111 (purple). For an easier understanding, the DMS (green) and the heme (grey) are shown in the two models as docked within CYP76F112 ("correct" position).

Taken together, these results suggest that the hydroxy groups of the polar T102 and, to a lesser extent, S105, might establish hydrogen bonds with the ketone group from the DMS. In CYP76F112, this interaction might serve to attract the ketone group and stabilise it, which means these polar residues probably help positioning and stabilising the DMS above the heme iron. On the contrary, in CYP76F111, the non-polar P105 and A108 would not favour a correct positioning of the DMS, even though they do not prevent it either. Therefore, we can assume that the polar T102 (and S105) might also contribute to CYP76F112 high affinity, by helping positioning and stabilising the DMS it in the active site, without being indispensable for the reaction to happen. T102 (and S105) might thus be seen as amino acid(s) that are not directly responsible for substrate specificity, but that improve the enzyme efficiency.

F.2.c. The residue D might stabilise the DMS in the active site

The case of the amino acid D is quite similar to that of the amino acids A and B. As a reminder, the mutant CYP76F112-D metabolised DMS, but its apparent K_m was of 114.3 ± 7.2 nM, compared to the 32.2 ± 3.9 nM of CYP76F112 (**Table 9**). The presence of a serine (as in CYP76F112) instead of a proline (as in CYP76F111) in position D therefore contributes to the high affinity of CYP76F112, which is again consistent with the putatively low concentration of DMS in the fig tree. Yet, unexpectedly, the mutation of D lead to an increase of the catalytic constant. Indeed, the apparent K_{cat} of CYP76F112 was of 22.1 ± 0.9 min^{-1}, but the apparent K_{cat} of CYP76F112-D was about 10 times higher with a value of 275.7 ± 5.9 min^{-1} (**Table 9**). This means the mutant is able to convert 10 times more molecules of DMS that CYP76F112 in a given time. But here again, this might not be completely inconsistent with the low concentration of DMS in the fig tree: with its apparent K_m of 32.2 ± 3.9 nM, CYP76F112 might be seen as an enzyme that is "improved" to deal with very low concentrations of substrate. With such low concentrations, being able to metabolise more molecules in a given time might be irrelevant – or at least less interesting than increasing the affinity. For these reasons, the value of CYP76F112 catalytic activity may not have had a strong impact on the evolution of its sequence. Consequently, a replacement of the polar D by a non-polar residue would probably not be retained by evolution because of the lower affinity it would cause.

In CYP76F112, the residue D (S305) is polar (**Table 10**). In the best docking model, it is positioned close to the DMS (<4Å) and the hydroxy group at the extremity of its side chain points toward the hydroxy group of the DMS (**Figure 62**). On the contrary, in CYP76F111, the residue D (P310) is not polar. This suggests that, in CYP76F112, the hydroxy group from the polar S305 might establish hydrogen bonds with the hydroxy group from the DMS, which would help attracting and stabilising the DMS in its correct orientation. In CYP76F111, P310 – which cannot establish hydrogen bonds with the DMS – would not be able to fulfil this function, but it would not prevent the reaction either. So, as for the residues A (and B), we can assume that the polar S305 might interact with the DMS, contributing to position and stabilise it, thus increasing CYP76F112 high affinity. But once again, it is not indispensable for the reaction to happen.

F.2.d. The residue C contribute to shape the substrate-docking site

The case of the residue C is quite different because the mutant CYP76F112-C could not metabolise DMS at all (**Table 9**). This implies that the replacement of the methionine in position C (as in CYP76F112) by a tryptophan (as in CYP76F111) completely prevents the marmesin synthase activity.
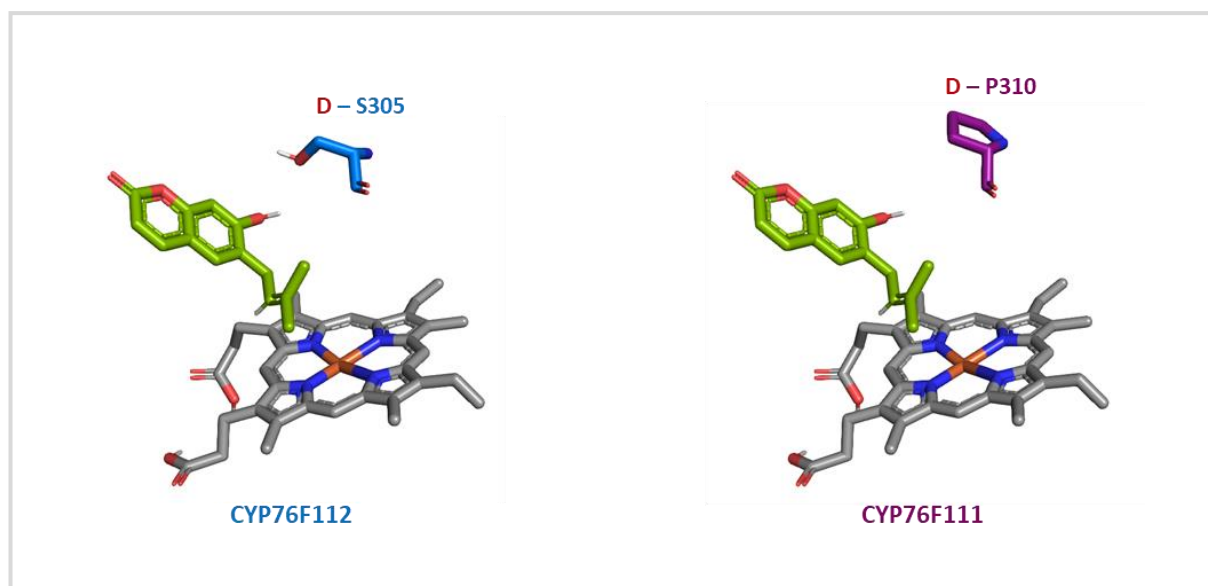
Figure 62 Influence of the residue D. Residue D from the 3D models of CYP76F112 (blue) and CYP76F111 (purple). For an easier understanding, the DMS (green) and the heme (grey) are shown in the two models as docked within CYP76F112 ("correct" position).

This time, the influence of the residue C is not linked to its polarity, since the amino acids C from both CYP76F112 and CYP76F111 are hydrophobic (**Table 10**). In CYP76F112, this residue C is a small methionine (M117). In the best docking mode, it is localised "bellow" the distal end of the DMS. On the contrary, in CYP76F111, this residue is a large tryptophan that slightly extends above the heme. Therefore, as shown in **Figure 61**, the replacement of M117 by a W seems to be enough to physically push the DMS away, preventing the conversion of DMS into marmesin. This suggests that the residue C contributes to shape the substrate-binding pocket: compared to the small M117 that allows a correct positioning of the DMS, W120 seems to reduce the substrate-binding pocket, creating a physical hindrance that totally prevents the reaction.
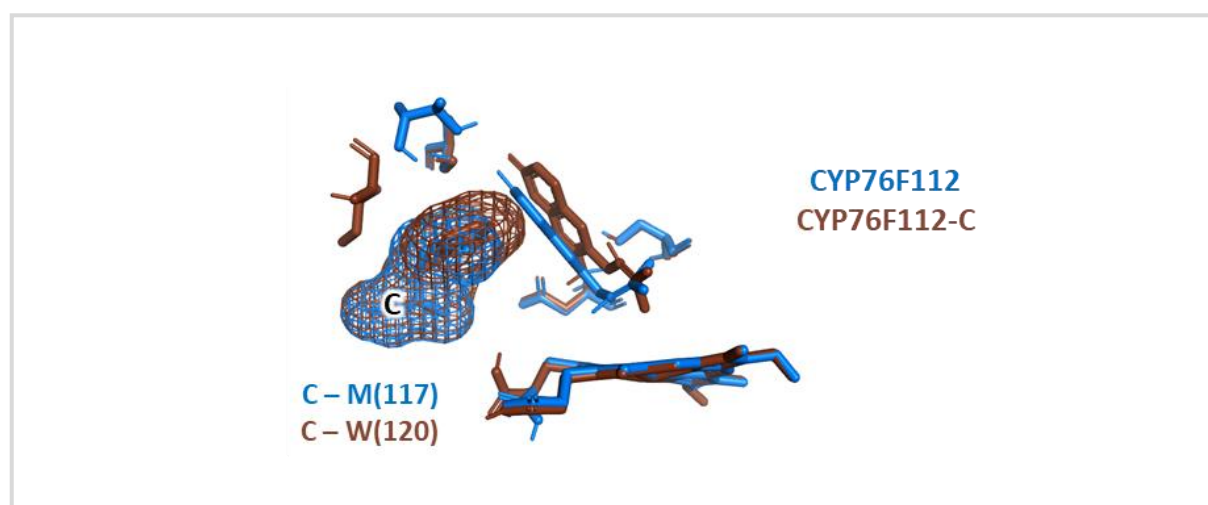
Figure 61 Influence of the residue C. The residues C from the 3D models of CYP76F112 and CYP76F112-C are shown as « mesh » to better visualise their respective size.

F.2.e. Summary of the influence of the residues A, B, C and D

This study draws the light on the residues A, B and D, which polarity and position in the active site might allow them to establish hydrogen bonds with the DMS. In CYP76F112, the polar T102 (A) and maybe the polar S105 (B) would both stabilise the ketone group on one extremity of the DMS, while S305 (D) would provide stability to the hydroxy group on the other side of the substrate (**Figure 63**). These three amino acids may therefore, contribute to attract and "lock" the DMS in a correct orientation, preventing its movement in the substrate-binding pocket. By doing so, these polar amino acids increase the affinity of the DMS for CYP76F112. Therefore, even though they are not absolutely indispensable for the marmesin synthase activity, these three amino acids allow CYP76F112 to remain relatively active with only a few nanomolar of substrate, which may correspond to the physiological concentration range of DMS in the fig tree.

On the contrary, the influence of the residue C is certainly linked to its size and shape. In CYP76F112, the small M117 may not particularly interact with the DMS, but its replacement by a large W is enough to physically disturb the positioning of the DMS – which completely prevent the reaction (**Figure 63**). Therefore, we can assume that any relative of CYP76F112 which would have a W in position C would not be able to convert DMS into a marmesin.
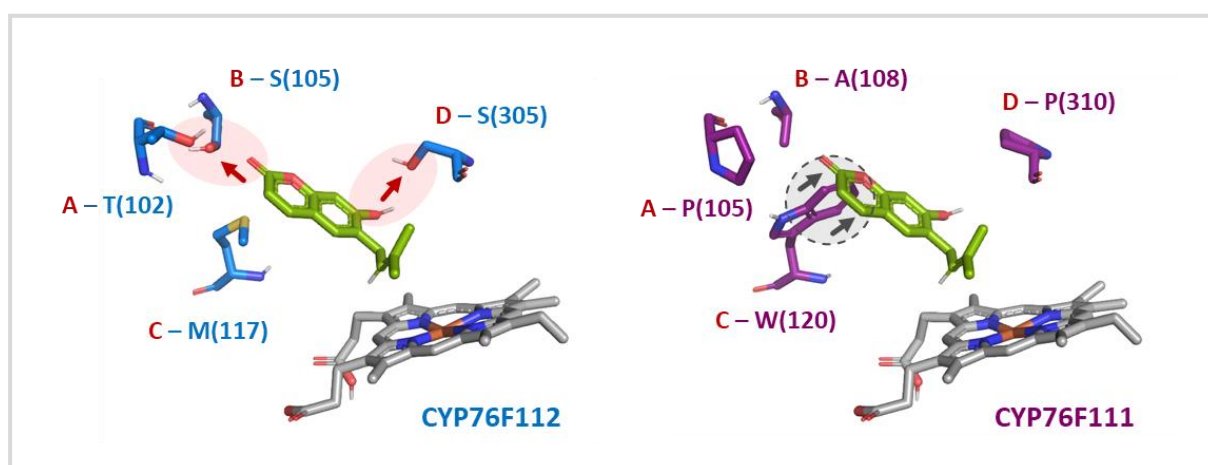
Figure 63 Proposed influence of the amino acids A, B, C and D during the docking of the DMS. The heme and the DMS are shown as docked in CYP76F112. In CYP76F112, the hydroxy group of the polar A/B and D might respectively establish hydrogen bonds with the ketone and hydroxy groups of the DMS. In CYP76F111, the tryptophan in position C physically repulses the DMS away from the heme.

More generally, the substrate-binding pocket of CYP76F112 is mostly defined by hydrophobic amino acids, excepts for two polar regions that contain the residues T102 (A), S105 (B), H110, S305 (D) and T310 (**Table 7**). The residues H110 and T310, respectively located near A/B and D, are conserved in CYP76F111. CYP76F112 substrate-binding pocket therefore consists of a hydrophobic cavity with two polar regions that probably interact with the ketone and hydroxy groups of the DMS, orientating and stabilising it during the reaction. The substrate-binding pocket is quite narrow and would probably not allow the binding of substrates substantially larger than the DMS (**Supp. Figure 8**).

F.2.f. The residue C: a hotspot position?

As mentioned above, because of the variability of sequence, substrate and reactions performed by the P450s, the amino acids reported to influence a P450 activity are almost never the same. For the residues A, B and D, I could not find any study that described the influence of equivalent residues located at the same position.

However, the residue C might correspond to a hotspot position that had already been described in other P450s from bacteria, fungi and animals. Indeed, one residue in the SRS1 is assumed to be a key residue for substrate specificity and selectivity. This residue corresponds to the amino acid F87 in CYP102A1 (*Bacillus megaterium*). It is located close to the heme and contributes to form the substrate binding cavity (Seifert et al. 2009; Sirim et al. 2010). Because of the BC-loop variability, the equivalents of this amino acid can be difficult to identify in new P450s via sequence alignment. Yet, in the 3D models, this residue has been described as being almost always located at the same position, pointing toward the heme, and staying rigid during substrate binding (Sirim et al. 2010). So, to identify this residue in CYP76F111 and CYP76F112, I superimposed their 3D models with that of CYP102A1 (PDB ID: 1BU7). Even though the bacterial CYP102A1 did not perfectly overlap with the CYP76Fs, this superposition allowed me to identify two successive residues in CYP76F111 and CYP76F112 that showed somehow similar position and orientation than F87 from CYP102A1 (**Supp. Figure 9**). These amino acids were V116 and M117 in CYP76F112, and A119 and W120 in CYP76F111. In other words, they are the residue C (M117/W120) and the one that precedes it (V116/A119). Among these two residues, only V116/A119 remained rigid during the docking. Yet, the influence of the residue V116/A119 was tested during preliminary docking experiment, but it seemed to have no impact on the positioning of the DMS and was set aside (not shown). The residue C might therefore be the one that plays a role similar to that of F87 from CYP102A1. And as described in other P450s (Seifert et al. 2009; Sirim et al. 2010), it seems to contribute shaping the substrate-binding pocket, which influences the specificity of CYP76F112.

F.2.g. The limits of the models and the importance of the access channel

Finally, even if the modelling and docking experiments are quite accurate, it should not be forgotten that they are only predictions that do not exactly reflect the reality of the enzymes. For instance, with the model of the mutant CYP76F112-AB, the DMS was never docked in a way that seemed to allow its conversion into marmesin. Yet, the corresponding mutant enzyme did it. As another example, the DMS was docked in a correct position within the mutant CYP76F111-ABCD – even though the associated enzyme never metabolised DMS. This might be due to inadequate experimental conditions that resulted in a non-functional enzyme, or most likely, to some docking limits.

Indeed, during the docking experiments, many parameters were not taken into account. For instance, to minimise the computation, the enzyme was treated as mainly rigid, with only a few flexible amino acids: this limited flexibility would not allow to predict important conformational changes, that could make the heme iron more accessible for the DMS. In addition, the oxygen atom that binds to the heme iron during the reaction was not included in the model, nor the water molecules from the buffer, nor the influence of the pH, etc. Such parameters could be included in more complex models.

Moreover, during the docking experiments, the role of the access channel was not taken into account: the DMS was docked within the active site (the grid receptor) regardless of its possibility to access it. This could explain why the DMS was correctly docked within the 3D model of CYP76F111-ABCD, while the associated enzyme never metabolises DMS. Indeed, the SRS2 and SRS3, which are assumed to form the substrate access channel (Larbat et al. 2007), are quite different in CYP76F111 and CYP76F112 (**Figure 39**, **Chapter III)**. And even though the global structure of CYP76F111 and CYP76F112 is exactly the same, (α-helixes, β-sheets, loops), the precise shape of their respective access channel and substrate-binding site looks quite different (**Supp. Figure 8**). It would therefore be possible that the DMS cannot enter at all in the active site of CYP76F111, nor in that of the mutant CYP76F111-ABCD. Similarly, in the additional docking experiments, the psoralen was docked in a promising position in several CYP76Fs – including in CYP76F112 which cannot metabolise it (**Chapter II**). So, here again, the divergence between the docking and *in vitro* experiments might be due to the access channel: the psoralen might be too big to enter in CYP76F112 active site, even if the positioning it would adopt within this active site would allow its hydroxylation into bergaptol.

Finally, in this analysis, I only focused on the amino acids that were the closest to the substrate. Yet, amino acids from the distal active pocket can also influence the catalytic properties of some P450s (Sun et al. 2020). Therefore, even though I identified 4 residues that do influence the catalytic properties of CYP76F112, it should not be forgotten that these 4 residues alone are not sufficient to entirely create the marmesin-synthase activity in another CYP76F. Other amino acids from the close or distal active pocket, from the access channel, or even from the membrane anchor region (Larbat 2006; Larbat et al. 2007) must also contribute to this activity.

F.3. Recent amino acids for a recent marmesin synthase activity

In this chapter, I identified four residues in position A, (B), C and D that influence the marmesin synthase activity of CYP76F112. In order to better understand the emergence of this activity, I reported the amino acids corresponding to these 4 positions throughout the CYP76F phylogeny (**Figure 64**). Across this phylogeny, and except for a restricted subclade, the properties of the amino acids in position A, B, C and D are not very variable: the residues in position A are almost always prolines, the ones in position B are always hydrophobic, the ones in position C are mostly "big" hydrophobic tryptophan (or "big" phenylalanine), while the ones in position D are almost never polar (**Figure 64**). On the contrary, the residues in position A, B and D are polar in a restricted subclade of 6 CYP76Fs that contains CYP76F112, its two orthologues from *F. erecta* and *F. religiosa* (G01_76F_F.erecta and G01_76F_F.religiosa), G02_76F_F.religiosa, G02_76F_F.erecta and G03_76F_F.erecta. This subclade will be called the marmesin-synthase subclade. Similarly, in this marmesin-synthase subclade, the hydrophobic residues in position C are only small methionine or leucine.

This suggests that the ancestral amino acid in position A should have been a proline, that only evolved into a polar threonine in the marmesin-synthase subclade. Similarly, the ancestral amino acids in position B, C and D should have been different than their actual equivalent from the marmesin-synthase subclade. In other words, the amino acids in position A, B, C and D have evolved recently in the marmesin-synthase subclade. As a CYP76F-like with a big tryptophan in position C would not be able to convert DMS into marmesin, we can reasonably propose the marmesin synthase activity appeared specifically in the marmesin-synthase subclade. This would be consistent with the results of

the *in vitro* and docking experiments. Indeed, as CYP76F112 converts DMS into marmesin, and as the DMS was docked in a correct position in G02_76F_F.religiosa and G02_76F_F.erecta, it can be assumed that the 6 enzymes from the marmesin-synthase subclade might convert DMS into marmesin (highlighted in blue on **Figure 64**). On the contrary, CYP76F110 and CYP76F111 do not metabolise DMS, and the DMS was dock in an incorrect position in G11_76F_F.erecta and G04_76F_F.erecta. **Thus, no CYP76F outside the marmesin-synthase subclade would be able to convert DMS into marmesin**.
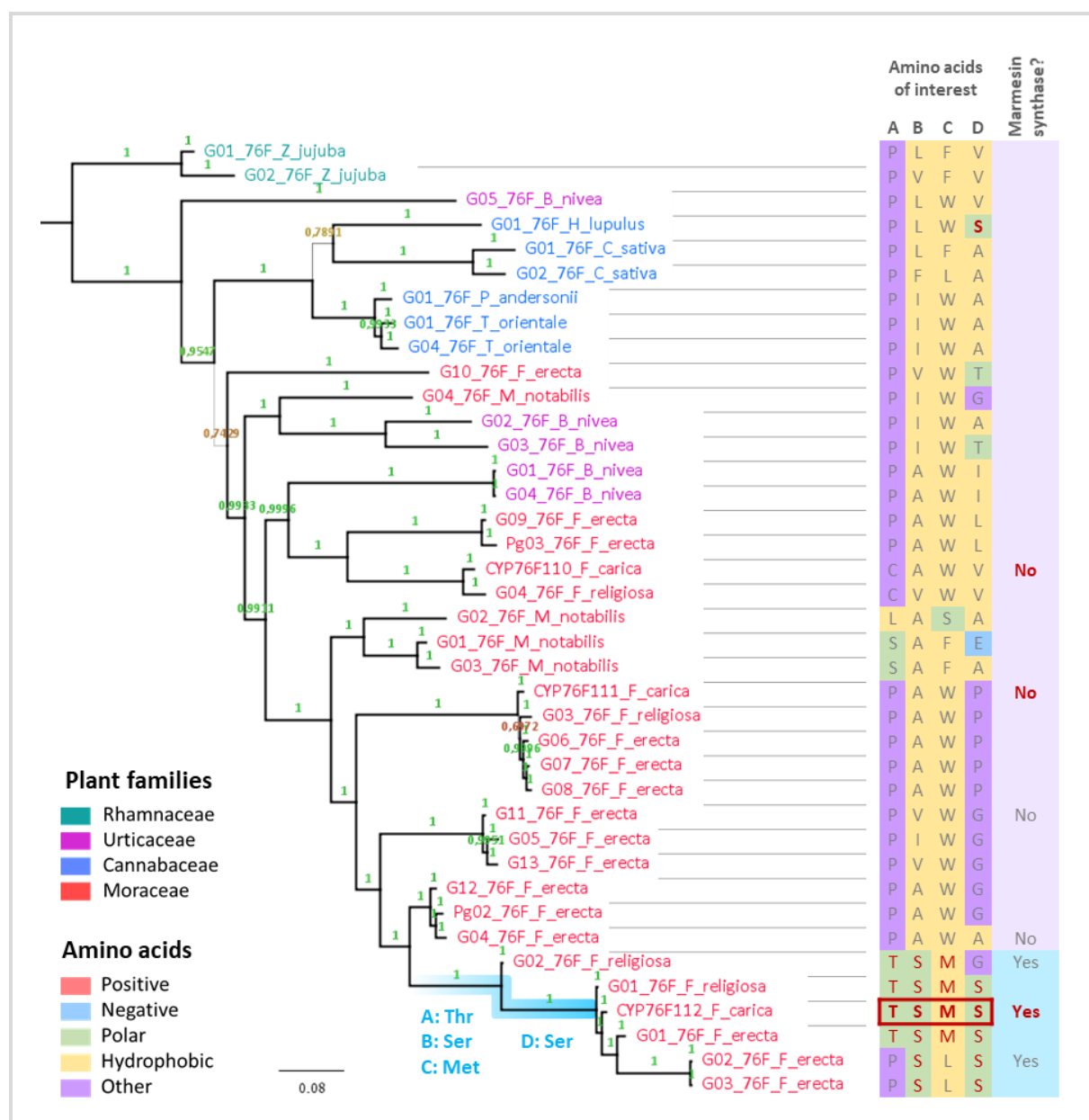
Figure 64 Evolution of the 4 residues of interest and emergence of the marmesin synthase activity. The phylogenetic tree is the one presented in **Figure 38** (**Chapter III**). The marmesin synthase activity is detailed: the (putative) marmesin synthases are in blue, the enzymes that might not have this activity are in purple. The activities confirmed by in vitro experiments (**Chapter II**) are highlighted in red, the ones suggested by the dockings are in grey. The putative emergence of the marmesin synthase activity is highlighted on the branches with a blue path. The mutation of the 4 residues in position A, B, C and D are reported below this blue path.

Taken together, this information now allows us to confidently propose the following hypotheses:

The marmesin-synthase activity emerged recently in the Moraceae family, through a bloom of the CYP76Fs.

In particular, the acquisition of this activity and the remarkable affinity of DMS for CYP76F112 was made possible by the evolution of the residues in position A, (B), C and D.

The residue in position C should have initially been a tryptophan that evolved into a methionine: this modified the shape of the substrate-binding pocket, opening a space without which it was physically impossible for the DMS to be correctly positioned above the heme iron.

In addition, the amino acids A, B and D evolved and became polar, which seems to have improved the enzyme affinity. As the physiological concentration of DMS in the fig tree might be very low, this increased affinity might have conferred a significant evolutive advantage that was favoured over an increased catalytic constant, and therefore retained by evolution. CYP76F112 might thus be seen as an enzyme that has become particularly adapted to low the substrate concentration it might encounter *in planta*.

Finally, despite the essential role of these 4 residues, it should not be forgotten that the emergence of the marmesin synthase activity would not have been possible without other mutations. For instance, it is very likely that mutations affecting the substrate access channel, or maybe the transmembrane domain, had also played a critical role for the acquisition of the marmesin synthase activity.

Chapter V
The cost of furanocoumarins

CHAPTER V.

THE COST OF FURANOCOUMARINS

~

Metabolic engineering: insertion of the furanocoumarin pathway in the tomato genome, in order to assess the metabolic cost of psoralen production

A. INTRODUCTION: DO TOMATOES DREAM OF TOXIC FURANOCOUMARINS?

As described in **Chapter I** (**I.B.3**), the production of specialised metabolites such as furanocoumarins has a cost that might affect plant physiology and negatively impact growth and reproduction. This cost has sometimes been measured: for instance, in parsnip, the induced synthesis of furanocoumarins in damaged leaflets represents an energetic cost of 12.6 µg glucose.cm^{-2} (Zangerl et al. 1997).

To better assess this metabolic cost, we chose to insert the beginning of the furanocoumarin pathway in the genome of a plant that does not naturally produce these molecules. Preliminary experiments were initiated in the LAE by G. Galati, who aimed to insert the two first genes of the furanocoumarin pathway in the genome of tobacco (*Nicotiana tabacum*), leading to the synthesis of DMS from the plant wide *p*-coumaroyl CoA. Therefore, G. Galati generated a plasmid containing the *PsDiox* and *PsPT1* genes, and used it to initiate tobacco transformations (Galati 2019). With the identification of CYP76F112, it became possible to continue this work. Indeed, the addition of a marmesin synthase and a psoralen synthase such as *CYP71AJ3* (Larbat et al. 2007) makes it possible to heterologously produce psoralen, which is actually the first toxic furanocoumarin of the pathway (**Figure 65**).
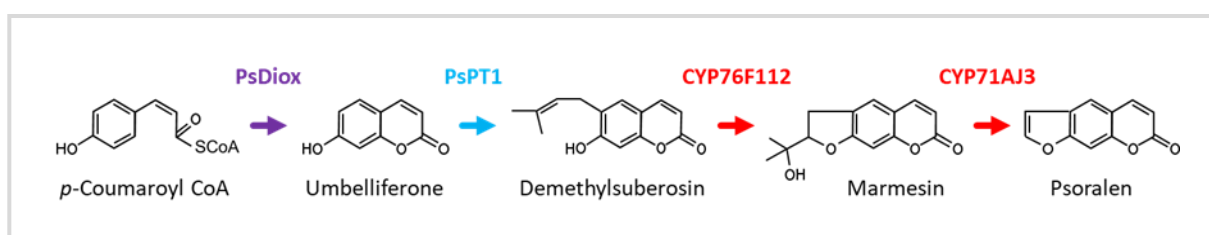
Figure 65 Genes associated to the production of psoralen, to be included in the tomato genome.

Accordingly, my first goal was to complete the multigene plasmid constructed by G. Galati – which contains the *PsDiox* and *PsPT1* ORFs – with the *CYP76F112* and *CYP71AJ3* ORFs. Together, these 4 genes encode for enzymes that allow the successive conversion of *p*-coumaroyl CoA into umbelliferone, DMS, marmesin and then, into psoralen (**Figure 65**). It should be noted that, except for the conversion of DMS into marmesin, several genes have been associated to these successive reactional steps (**Chapter II.E**). *PsDiox*, *PsPT1*, and *CYP71AJ3* were selected for this work because they all belong to the same species, *Pastinaca sativa* (Galati 2019).

My second goal consisted in inserting these 4 genes in the genome of a plant that is not a natural furanocoumarin-producer. The general idea was to set up a tool allowing the comparison between the growth and defence of plants producing – or not – of furanocoumarins. For this purpose, we chose tomato (*Solanum lycopersicum*), a species that does not naturally produce furanocoumarins, that is a model plant, and has been studied for a long time at the LAE.

My last objective was to insert the first 4 genes of the furanocoumarin biosynthesis pathway in the genome of tomato, in order to assess the metabolic cost of psoralen production.

For this purpose, I used a multi-gene cloning technology called GoldenBraid to construct a plasmid harboring the firsts 4 genes of the furanocoumarin pathway. Then, I used transgenesis methods along with *in vitro* culture to introduce these genes in the tomato genome. If the transformations were successful, the transgenic psoralen-producing tomatoes could later be used to study the trade-offs between growth and defence associated to psoralen production.

B. BRICK BY BRICK: THE GOLDENBRAID MULTI-GENIC CONSTRUCTIONS

B.1. Approach: which plasmid to construct, and how?

B.1.a. Overall strategy

To insert several genes in the genome of a plant, it is generally easier to perform a single transformation with a multigene plasmid rather than to adopt a multipartite transformation strategy which would require multiple and lengthy steps of plant regeneration and seed production. Therefore, the first step of this analysis consisted in constructing a unique plasmid containing the *PsDiox*, *PsPT1*, *CYP76F112* and *CYP71AJ3* genes.

Several prerequisites were to keep in mind while constructing such a plasmid. Firstly, as the genes were to be transferred into the tomato genome, they had to be inserted in a binary vector usable for plant transformation. Secondly, to allow the identification of the transformed plants, a selection marker such as an antibiotic resistance gene had to be included in the plasmid, and transferred into the tomatoes. The gene selected for this purpose was a kanamycin resistance (*KanR*) gene. Lastly, as the *PsDiox*, *PsPT1*, *CYP76F112*, *CYP71AJ3* and *KanR* genes were to be expressed in the tomato, they needed to be included into transcriptional units (TUs), under the control of promoters and terminators. To ensure a strong expression, we chose to place every gene under the control of a constitutive 35S promoter, and a tNOS terminator. By doing so, we expected a psoralen production that would have been important enough to be detectable and to significantly impact the growth of the tomato.

To construct such a plasmid – illustrated in **Figure 66** – we chose to use the GoldenBraid (GB) technology, which is extensively described in Sarrion-Perdigones *et al.* (2011, 2013, 2014). GB is a standardised multi-cloning system that allows to easily construct and assemble multiple TUs into
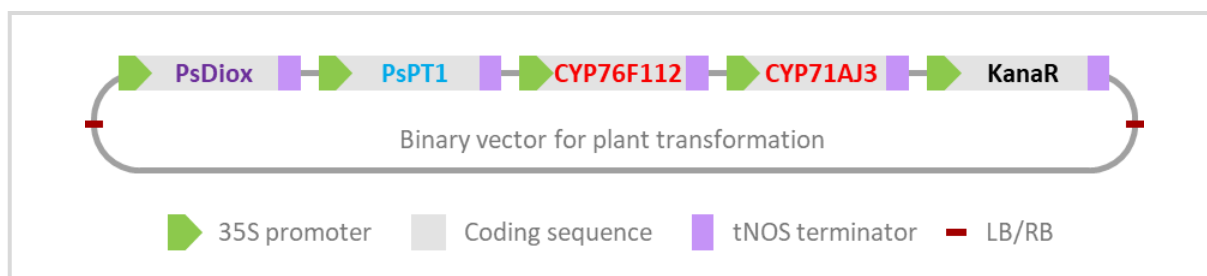
Figure 66 Schematic representation of the plasmid to be constructed and used for tomato transformation. LB and RB respectively stand for the left and right borders that flank the TDNA.

binary vectors usable for plant transformations. It can thus be used to assemble an entire biochemical pathway into a single plasmid, and to insert it in the genome of a host plant with a single transformation step. In the following paragraphs, I will provide a simplified overview of this technology, and will only detail the mechanisms associated to the steps and material used in this study.

B.1.b. Presentation of the GoldenBraid cloning system

The GB method consists in repeatedly assembling DNA sequences from entry vectors into a single destination vector (**Figure 67**). To construct a biosynthesis pathway, the first step consists in assembling basic DNA parts – also called "GB parts" – into TUs: for every TU, a promoter, a coding sequence and a terminator from 3 entry vectors (pUPD plasmids) are ordinarily assembled into a single destination vector (PDGB1α plasmid). Then, the vectors containing different TUs become entry vectors and are binarily combined in new destination vectors (PDGB1Ω plasmids) to form higher order modules (**Figure 67**) which can end up constituting entire biochemical pathways. This process is based on a succession of restriction-ligation reactions, using type IIS restrictions enzymes.
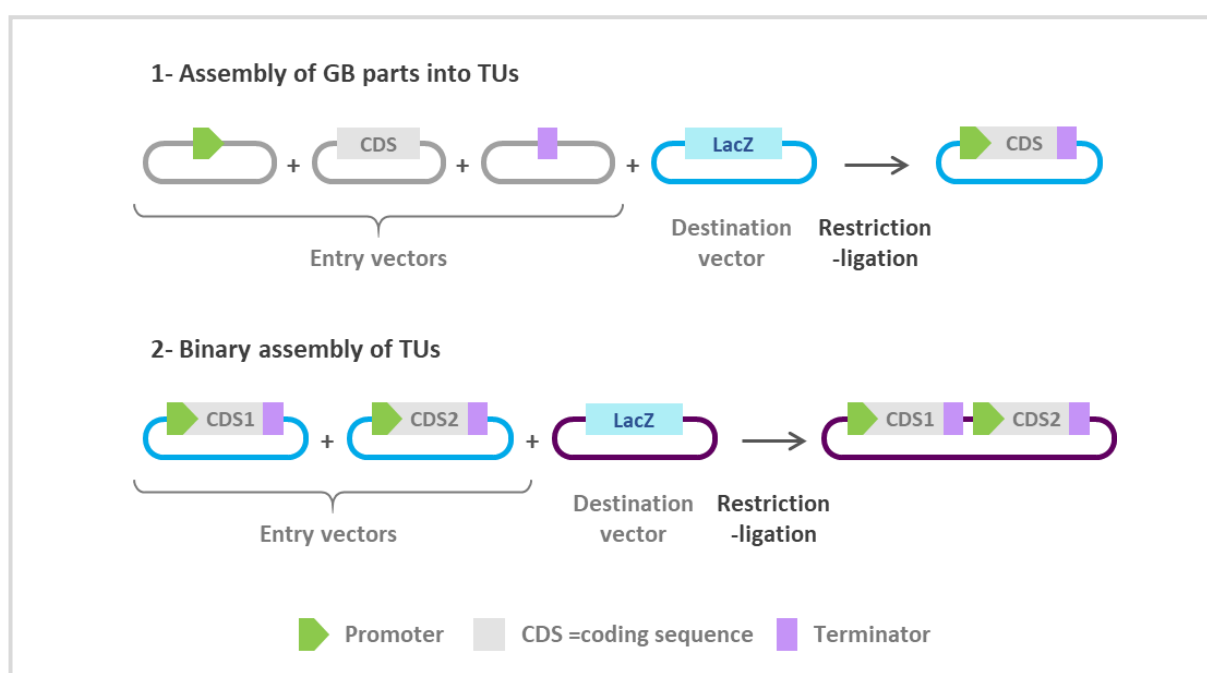
Figure 67 Schematic and simplified illustration of the GoldenBraid cloning approach.

Type IIS restriction enzymes are a particular type of restriction enzymes that have distinct recognition and cleavage sites: they cleave outside their recognition site, at a defined distance away from it (a few nucleotides), and regardless of the sequence of the actual cleavage site. Many type IIS restriction enzymes leave a short overhang, such as *Bsa*I and *Bsm*BI that both generate a 4-nucleotide overhang (**Figure 68**). These features allow to generate sticky ends which sequences can be chosen by the experimenter. In addition, by positioning the recognition sites in opposite directions in entry and destination vectors, and by using the right combination of complementary ends, the recognition sites disappear from the final vector – which means it is not cleavable anymore by the restriction enzyme. Therefore, by using a unique enzyme, it is possible to easily assemble multiple DNA parts in a predefined order, all in a single restriction-ligation step (single reaction mixture – **Figure 68**). These features constitute the very basis of the GB technique.
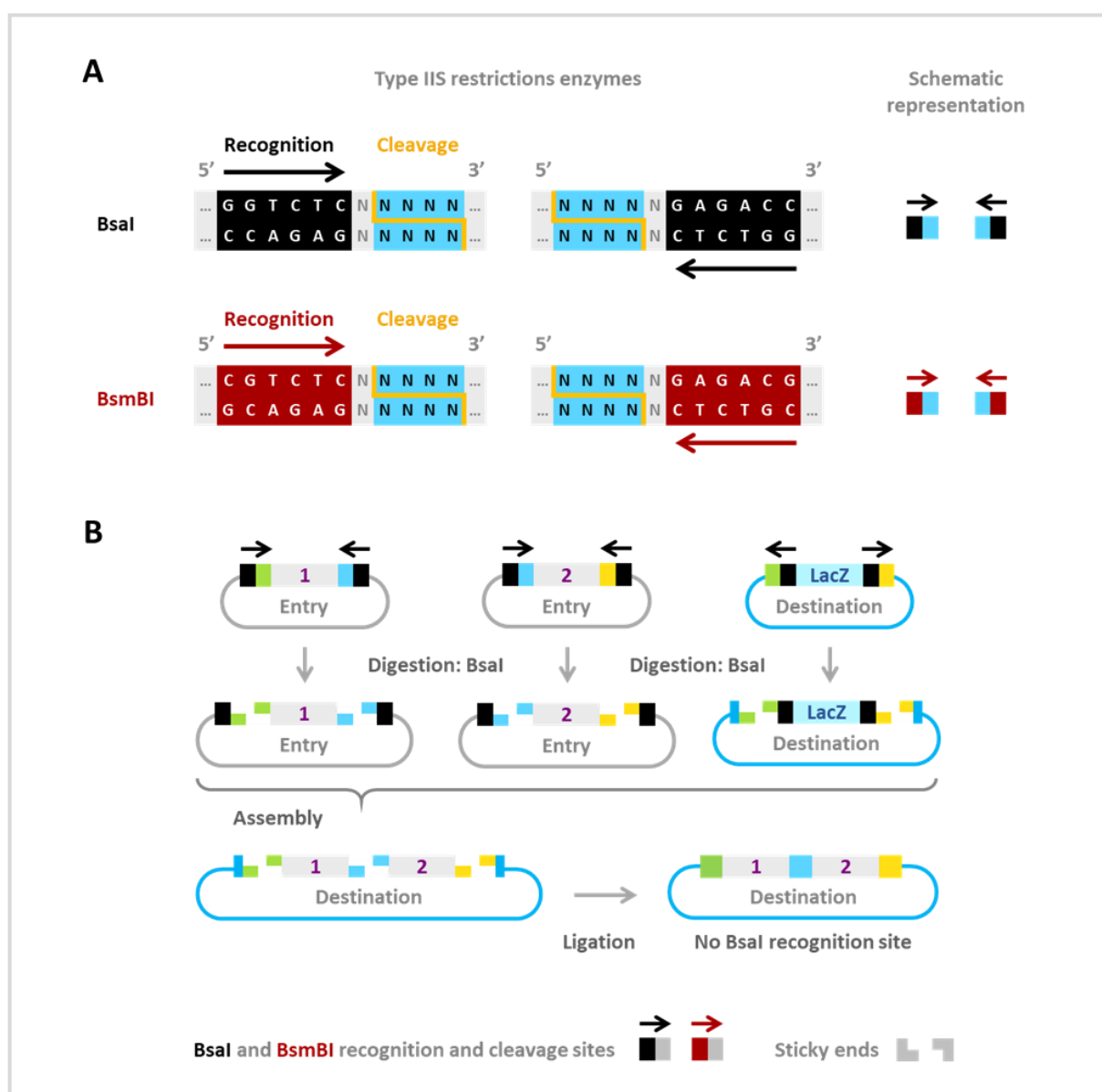
Figure 68 Type IIS restriction enzymes. A Recognition and cleavage site of the enzymes *Bsa*I and *Bsm*BI. **B** Example of a single step restriction-ligation reaction using *Bsa*I. The complementary sticky ends share the same colors.
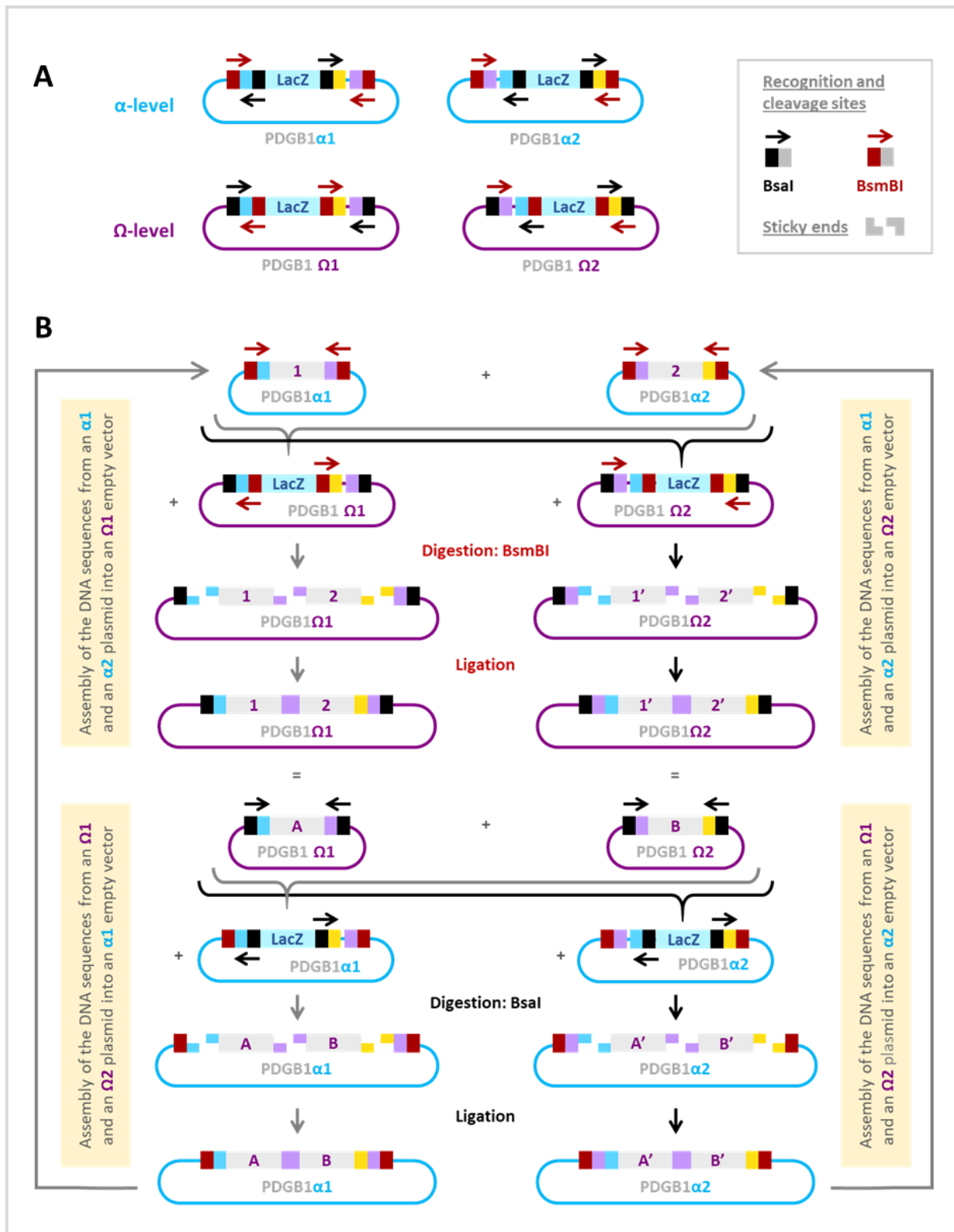
Figure 69 The GoldenBraid loop. A Destination vector used in the GB loop**. B** Repeated restriction-ligation reactions. The complementary sticky ends share the same colors. Based on Sarrion-Perdigones *et al*. (2011, 2013).

The angular stone of the GB cloning system lies in the use of two types of destination plasmids that contain recognition sites for two different type IIS enzymes: the α- and Ω-level plasmids (PDGB1α and PDGB1Ω), which both possess two *Bsa*I and *Bsm*BI recognition sites (**Figure 69**), and exists in two versions – α1 and α2, Ω1 and Ω2. These plasmids were designed so that each plasmid level can become the entry or destination vector of the other type of plasmid, depending on the restriction enzyme that is used. Upon *Bsm*BI digestion, the DNA sequences (GB parts or TUs) inserted between the *Bsm*BI recognition sites of an α1 and an α2 plasmid can be orderly assembled and inserted into an Ω-level plasmid (Ω1 or Ω2). During this assembly, the *Bsm*BI recognition sites disappear: the resulting plasmid thus contains the newly assembled DNA sequences from the two α-level plasmids, flanked by two *Bsa*I recognition sites from the Ω-level plasmid (**Figure 69**). Such an Ω-level plasmid can be reused as an entry vector for a subsequent assembly: upon *Bsa*I digestion, the DNA sequences inserted between the *Bsa*I recognitions site of an Ω1 and an Ω2 plasmid (entry vectors) can be assembled and inserted into an α-level plasmid (α1 or α2). And once again, the resulting α-level plasmid can be reused as an entry vector, and assembled with its complementary α-level plasmid in a new Ω-level vector (**Figure 69**). The α- and Ω-level plasmids can therefore alternate indefinitely, in an endless iteration of binary assemblies that allows to incorporate more and more TUs into the next-level plasmid.

B.1.c Using the GoldenBraid technology, or how to construct the desired plasmid

To use the GB cloning system, the first thing to do consists in adapting the initial GB parts (promoters, coding sequences, terminators) to the GB rules. This is called "domestication", and it consists in two essential points: removing any internal *Bsa*I or *Bsm*BI recognition site, and adding specific "GB extensions" on each side of every GB part. These flanking extensions are composed of *Bsm*BI recognition sites and of the nucleotides corresponding to the overhangs that will determine the relative position of every GB parts/TU(s) during a binary or multipartite assembly.

Once the GB parts are domesticated, they are cloned into an initial entry vector (usually the pUPD plasmid) through a restriction-ligation using *Bsm*BI. Then, the second part of the GB strategy can begin: first, a promoter, a coding sequence and a terminator from 3 entry plasmids are digested by *Bsa*I and assembled into an α-level vector, constituting a first TU. Then, 2 TUs from α1 and α2 plasmids are combined into a Ω-level plasmid, etc.

According to this general scheme, the plasmid harbouring the psoralen biosynthesis pathway had to be constructed as follow. In a first time, the coding sequences of *PsDiox*, *PsPT1*, *CYP76F112*, *CYP71AJ3*, *KanaR*, but also the 35S promoter and the tNOS terminator had to be domesticated, and individually inserted into pUPD plasmids. As a second step, each coding sequence had to be combined with a 35S promoter and a tNOS terminator into an α1- or an α2-level plasmid, as shown on **Figure 70**. The resulting TUs will be called [PsDiox], [CYP76F112], [PsPT1], [CYP71AJ3] and [KanaR]. Lastly, these TUs were to be repeatedly combined in α- and Ω-level plasmids to give the final 5-TUs plasmid containing [PsDiox+PsPT1+CYP76F112+CYP71AJ3+KanaR] (**Figure 70**).

Some of these steps were already realised by G. Galati, who constructed the Ω1 plasmid containing the TUs [PsDiox+PsPT1] and domesticated the *CYP71AJ3* coding sequence (Galati 2019). In addition, the 35S and the tNOS used in this study were already domesticated and inserted into pUPD plasmids.
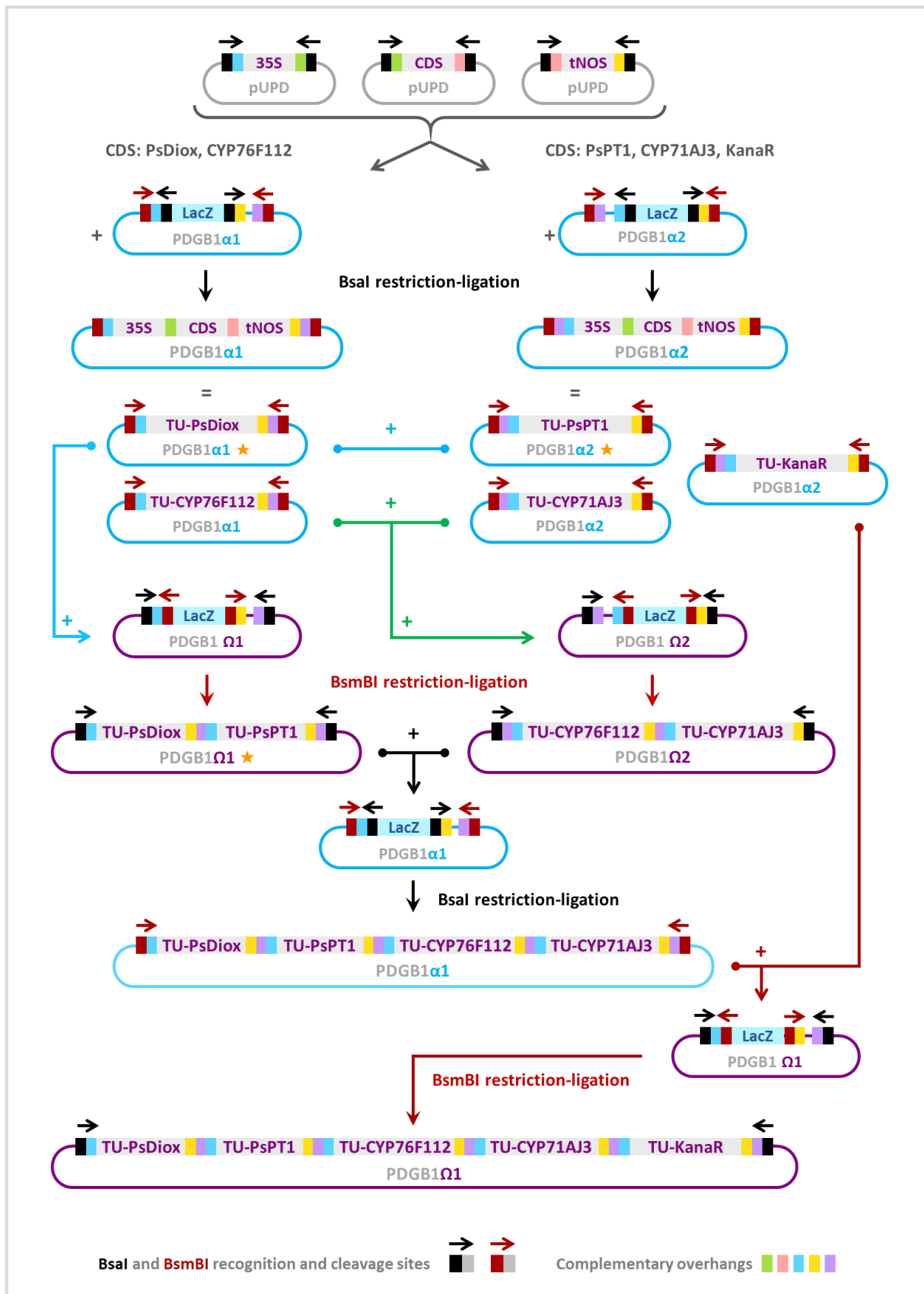
Figure 70 Assembly of the 5-TUs plasmid harboring the psoralen biosynthesis pathway. The initial domestication of the GB parts and their insertion into pUPD vectors are not shown. The plasmids highlighted with an orange star were already constructed by G. Galati (2019).

Therefore, my actual work consisted in (1) domesticating the *CYP76F112* and *KanaR* genes, (2) inserting *CYP76F112*, *CYP71AJ3* and *KanaR* into pUPD vectors, and combining them with 35S promoters and tNOS terminators, and (3), combining these newly-assembled TUs, together and with the plasmid of G. Galati, to generate the final 5-TUs plasmid harboring the psoralen biosynthesis pathway (**Figure 70**).

B.2. Construction of the plasmid

The multiple steps corresponding to the domestication of the genes (**Supp. Figure 10**) and the repeated assemblies that led to the 5-TUs plasmid (**Figure 70**) have been performed as explained in the materials and methods section (**Chapter VII**, **B.4**).

The generated 5-TUs plasmid is illustrated on **Figure 71**. Its annotated sequence is reported in **Supp. Figure 11**. Not only the presence of the 5 TUs was confirmed by a digestion and multiple PCRs (**Figure 71**), but the 5 coding sequences were also sequenced, which confirmed their integrity. In other words, the correct construction of the final 5-TUs plasmid has been fully checked and validated.

In summary, I successfully used the GoldenBraid technology to construct a PDGB1Ω1 plasmid harbouring the psoralen biosynthesis pathway.

This plasmid contains 5 TUs composed of a 35S promoter, a coding sequence and a tNOS terminator (**Figure 71**). The 5 coding sequences are *PsDiox*, *PsPT1*, *CYP76F112* and *CYP71AJ3* for psoralen biosynthesis, and *KanaR* for subsequent plant selection.

Moreover, as the PDGB vectors have been design to be directly usable for *Agrobacterium*-mediated plant transformations, the 5 TUs are flanked by the right and left borders (RB/LB) that define the T-DNA which can be inserted in the plant.

C. THE TOMATOES OF EVIL: TOMATO TRANSFORMATION AND REGENERATION

C.1. Approach and choice of the negative control

In order to study the trade-offs between growth and defence, my second goal consisted in creating a line of transgenic psoralen-producing tomatoes which growth and defence could be compared to those of tomatoes that do not produce furanocoumarins. For this purpose, I aimed to insert the psoralen biosynthesis pathway in the genome of the tomato by performing *Agrobacterium*-mediated stable transformations with the 5-TUs plasmid. The tomato line selected for this work was WVa106 (*Solanum lycopersicum* var. *cerasiformae* 'West Virginia 106'), a variety that has an undetermined growth and which is commonly used for transformation.
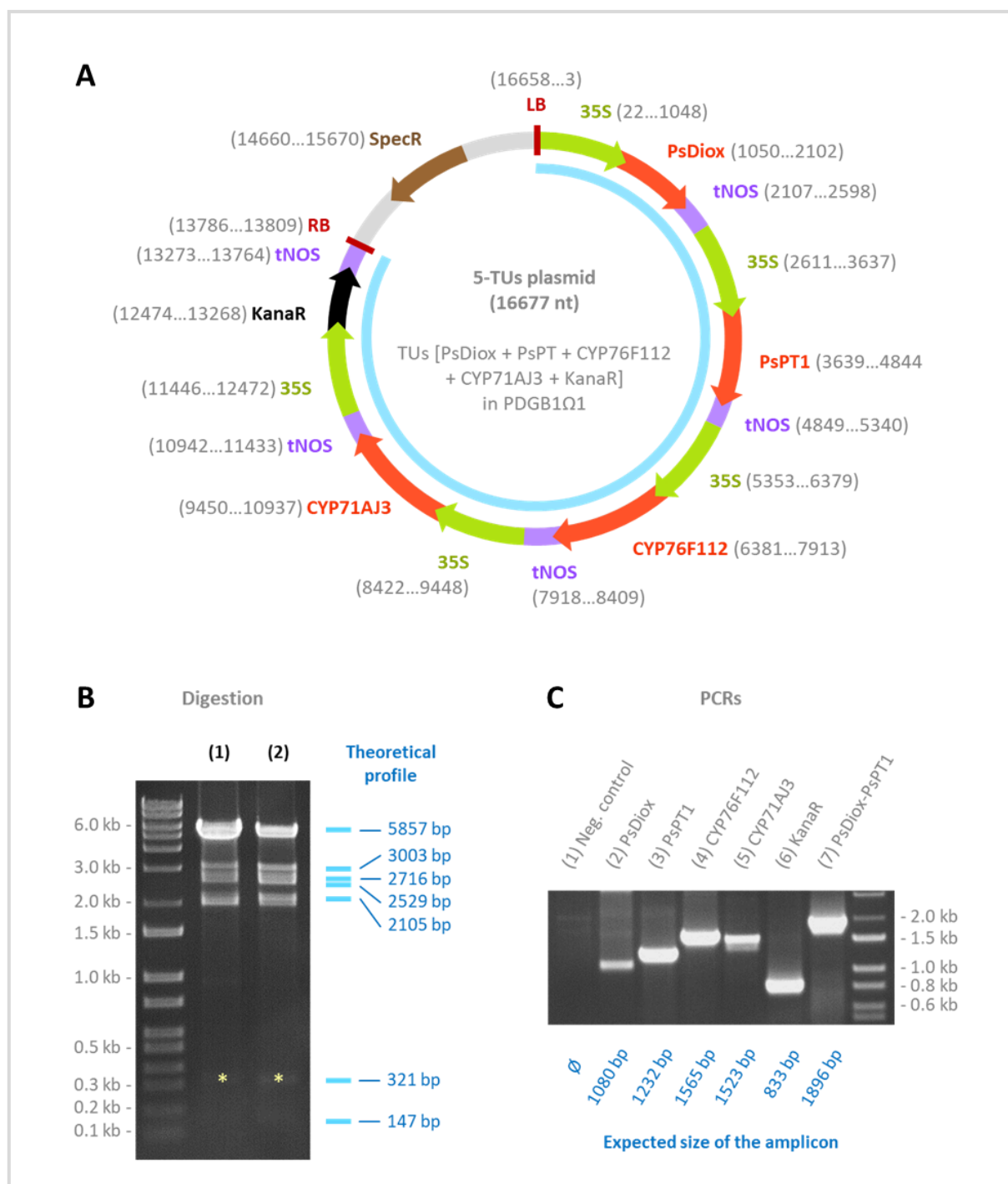
Figure 71 5-TUs plasmid harbouring the psoralen biosynthesis pathway. **A** Plasmid map of the final 5-TUs plasmid. SpecR refers to the gene conferring spectinomycin resistance to the bacteria. RB and LB refers to the right and left borders delimiting the T-DNA to be inserted in the plant, highlighted in blue. **B** Validation of the 5-TUs plasmid: digestion by *Pvu*II and *Sac*I. (1) and (2) correspond to two 5-TUs plasmid solutions, extracted from two individual bacterial colonies transformed with the 5-TUs product of the final GB restriction-ligation step. As often, the longer fragments are a lot more visible than the smaller ones, which are barely to not visible. The digestion profile is therefore considered as correct. **C** Validation of the 5-TUs plasmid by PCR. (1) Negative control using the MYB8 pair of primers, which do not correspond to any part of the 5-TUs plasmid. (2-6) Individual amplification of the 5 coding sequences corresponding to the 5 TUs. (7) Amplification of a fragment that overlaps the PsDiox and PsPT1 genes. The primer sequences are available in **Supp. Table 14**.

It should be reminded that the 5-TUs plasmid includes a kanamycin resistance gene. As this selection marker may affect the tomato physiology, the psoralen-producing tomatoes that were to be generated could not be simply compared to wild-type tomatoes. On the contrary, to avoid any bias, transgenic tomatoes that would not produce furanocoumarins but would simply resist to kanamycin also had to be generated and used as negative controls. To construct the 5-TUs plasmid, I initially assembled each TU in a PDGB1α vector (**B.2.b**), which means that a binary vector containing the TU [KanaR] was already available. This KanaR plasmid could thus be used in *Agrobacterium*-mediated transformation to insert the TU [KanaR] – and only that – in the tomato genome. Its complete and annotated sequence is reported in **Supp. Figure 12**; the map of the plasmid is presented in **Supp. Figure 13**.

> Therefore, I aimed to create two lines of transgenic WVa106 tomatoes: one line that simply resists to kanamycin (transformed with the KanaR plasmid), and one line that produces psoralen and resists to kanamycin (transformed with the 5-TUs plasmid).

C.2. Generation of the transgenic tomatoes

C.2.a. Tomato transformation and regeneration

The 5-TUs and the KanaR plasmids were respectively electroporated into the *Agrobacterium tumefaciens* EHA105 strain. Then, these *Agrobacterium* were used to transform WVa106 tomatoes, which were subsequently regenerated using *in vitro* culture techniques, as described in the materials and methods section (**Chapter VII**, **B.5**).
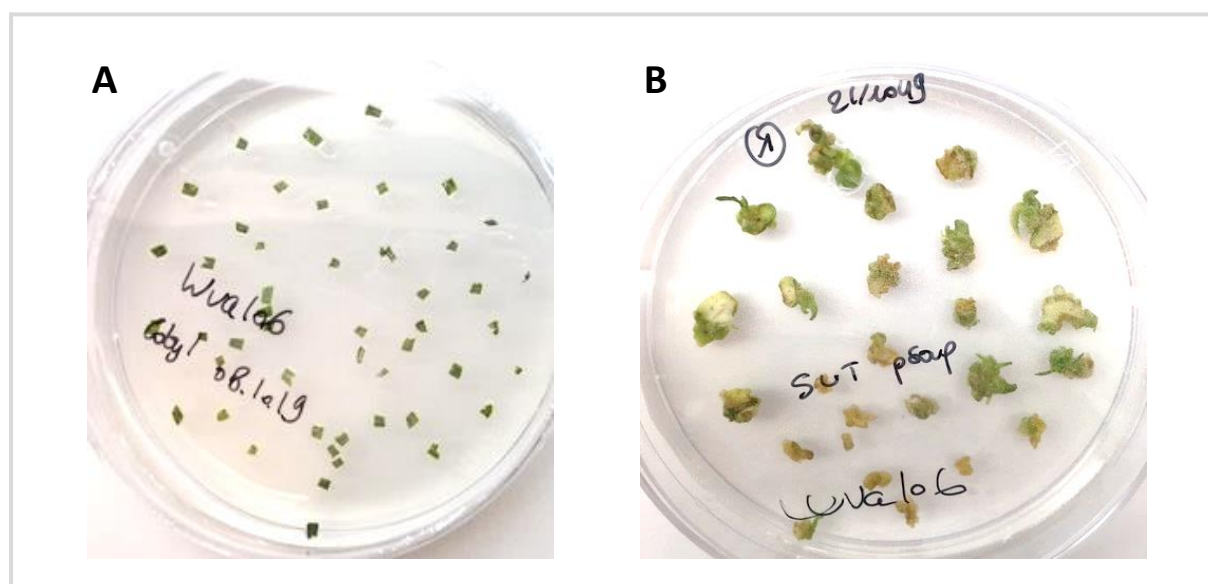
Figure 72 Tomato transformation and regeneration. A Fragments of cotyledons on preculture medium, ready for the *Agrobacterium*-mediated transformation. **B** Fragments of cotyledons on regeneration medium, 18 days after the transformation. Some cotyledons have formed calli and started to produce green shoots. The others, probably not transformed, turned yellowish and died.

In brief, seeds of the Wva106 tomato line were sterilised and cultured until plantlets with well-developed cotyledons were obtained. The cotyledons were excised from the plantlets, cut in 3 parts and cultured for 24h in the dark in order to prepare them for the transformation (**Figure 72**). The fragments of cotyledons were then incubated in a suspension containing the *Agrobacterium* transformed with the 5-TUs or with the KanaR plasmid. They were subsequently placed on a co-cultivation medium and cultured for 48h, which corresponds to the actual transfection step during which the T-DNA from the binary plasmid is randomly inserted into the plant genome. After that, the fragments of cotyledons were transferred on a regeneration medium, which purpose was to induce the formation of callus and shoots. This medium also contained kanamycin to ensure the selection of the transformed cells. The fragments of cotyledons were cultured until they formed calli and started to produce green shoots (**Figure 72**). At this point, no phenotypic difference was observable between the tomatoes transformed with the 5-TUs plasmid and the ones transformed with the KanaR plasmid.

When a green shoot was developed enough to form a plantlet – composed of a stem, at least two well-developed leaves and an apical meristem – it was excised from its callus and placed in a rooting medium. Plantlets were cultured until they were well-rooted. Three different plantlets phenotypes were observed for the tomatoes transformed with the 5-TUs plasmid, as well as for the ones transformed with the KanaR plasmid (**Figure 73**). These phenotypes were therefore independent of the plasmid used for the transformation. Most of the plantlets had a "normal" phenotype: they were entirely green, grew roots and developed normally. Others plantlets presented necroses on their leaves, and/or developed calli and/or roots on their aerial parts – which was probably caused by the *in vitro* culture conditions. Despite their atypical features, these plantlets could form roots and they developed normally. Moreover, in the next cultivation steps, these atypical features progressively reduced and disappeared. Lastly, a few plantlets started to turn yellow and, after some time, became completely white (albino phenotype). Most of these plantlets could not develop roots, and they all died quickly. The occurrence of the albino phenotype is quite common when performing *in vitro* plant transformation or anther culture, and can be caused by many genetic and environmental factors (Cho et al. 1998; Zhang et al. 2000; Kumari et al. 2009; Song et al. 2012; Yan et al. 2019). Therefore, dying albinos were discarded and will not be further discussed.
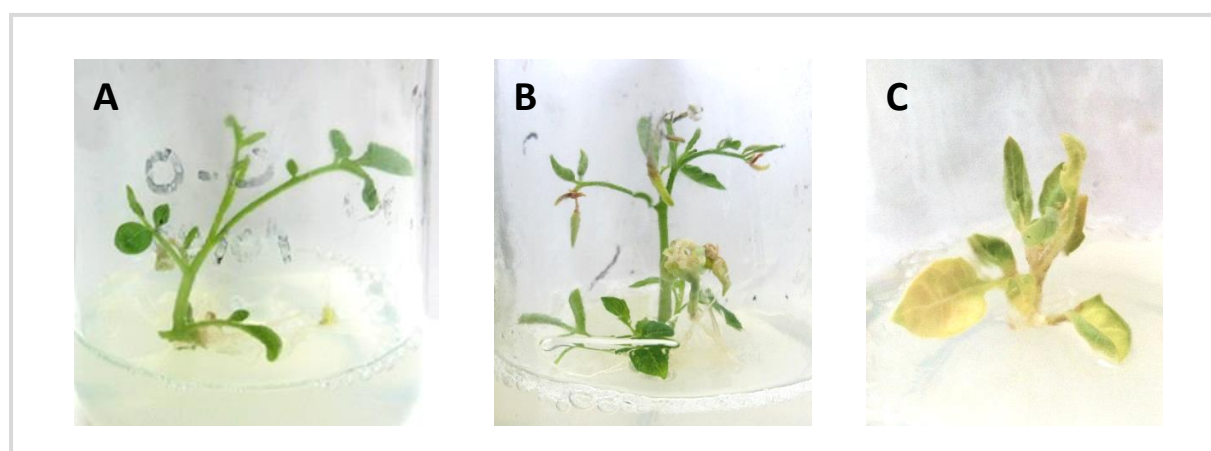
Figure 73 Plantlets in the rooting medium, about 3-4 months after the transformation. A "Normal" phenotype: green plantlet growing and forming roots. **B** Plantlet growing callus on its aerial parts, with necrosis on its leaves. **C** "Albino" phenotype: plantlet starting to turn yellowish and white.

Finally, once well-rooted, the plantlets were removed from the gel, planted into soil and maintained in a phytotron. After a short adaptation to non-sterile conditions and months of growth, the plantlets turned into fully developed tomato plants (T0 generation), flowered, and produced fruits (**Figure 74**). The fruits were harvested, their seeds were carefully collected, and dried for a long storage. They constitute the T1 generation. Once again, no phenotypic difference was noticed between the tomatoes transformed with the 5-TUs and the KanaR plasmids.
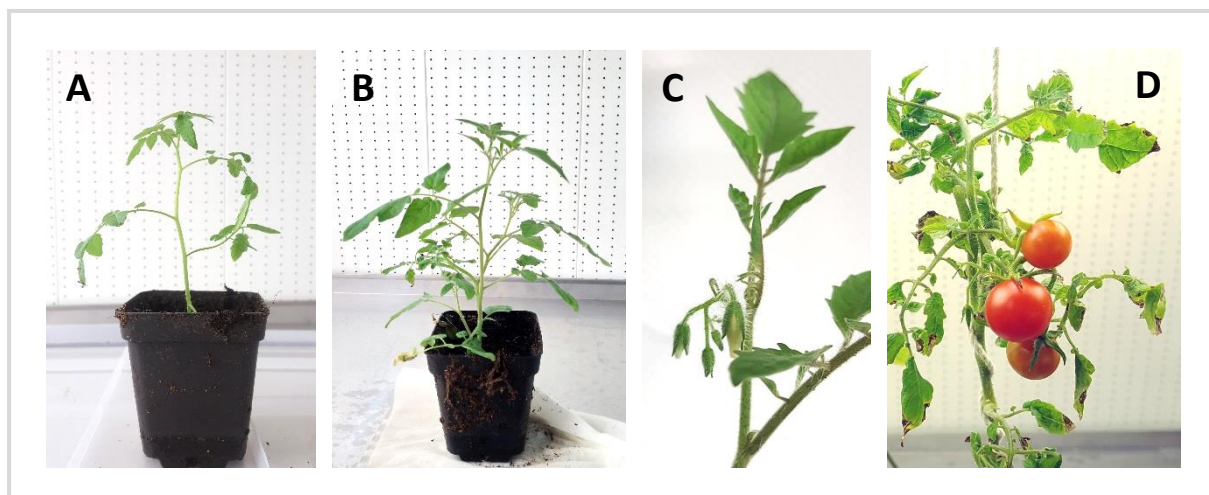
Figure 74 Development of the transformed tomatoes, after their removal from the gel. A Tomato plant after 1 week in the soil. **B** Tomato plant after 2 weeks in the soil. **C** First flower buds, after 1 month in the soil. **D** Ripen tomatoes, 2 months after the apparition of the buds.

C.2.b. Confirmation of the transformations

In order to verify the correct insertion of the T-DNA into the tomato genome, I sampled a leaf from every tomato plant from the T0 generation, extracted its DNA, and performed PCRs on it. To avoid causing damages to the young tomatoes, the sampling was performed on the *in vitro* plantlets that had at least 6 leaves, or on plants that were already into the soil. It should be reminded that transformed plants from the T0 generation can be chimeric. Therefore, even if a given tomato plant is tested positive, its offspring might not possess the transgene(s) and will require subsequent testing.

The tomatoes transformed with the KanaR plasmid were tested by PCR with a pair of primers designed to amplify the *KanaR* gene. A second pair of primer designed to amplify the *Solanum lycopersicum* transcription factor *MYB13* (accession number XM_004242311.4) served as a positive control (**Supp. Table 14**). The *KanaR* gene was amplified in all the tomato plants transformed with the KanaR plasmid, but not in the wild-type WVa106 (**Figure 75**) – which means that the *KanaR* gene was successfully inserted into the genome of all the regenerated plants. In addition, to further validate these PCR results, I randomly selected 4 of these tomatoes (2b, 5b, 7b and 14) and sequenced the fragments amplified with the KanaR primers. The results of the sequencing confirmed the presence and the integrity of the *KanaR* gene. I therefore obtained 28 transgenic tomato plants from 10 different calli (*i.e* independent transformation events), which contain the *KanaR* gene in their genome. For now, 8 of them from 5 calli have successfully flowered and given fruits and seeds. A higher number of fruits from the other tomato plants should be harvested in the coming weeks / months.
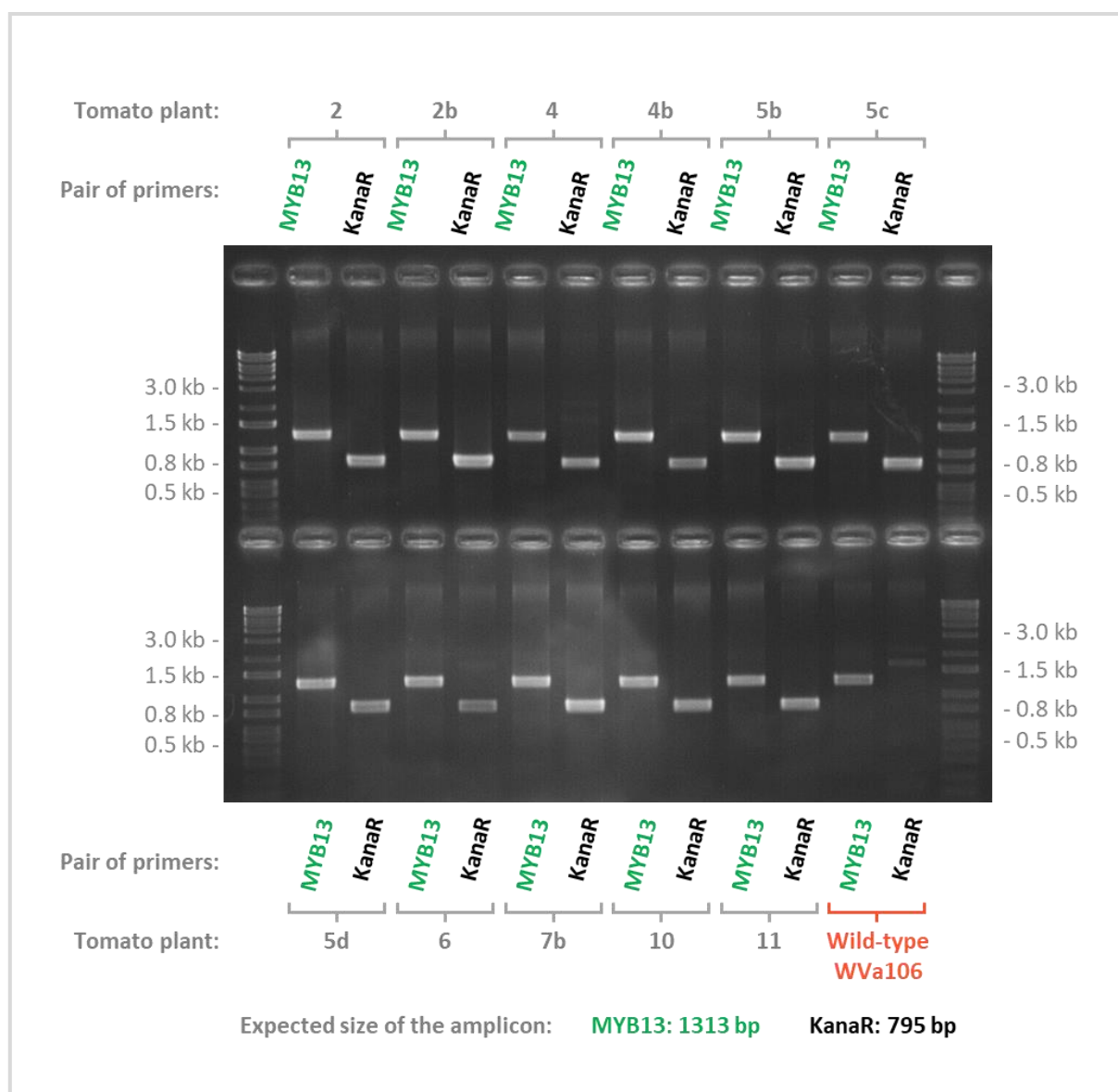
Figure 75 Tomatoes transformed with the plasmid KanaR. PCRs on the tomato DNA, using the pairs of primers KanaR and MYB13 (positive control) described in **Supp. Table 14**. The tomato plants with the same number but a different letter (for instance, 4 and 4b) originate from the same callus. A wild-type WVa106 tomato has been used as a negative control for the KanaR primers.

The tomatoes transformed with the 5-TUs plasmid were tested by PCR with 5 different pairs of primers, respectively designed to amplify the *PsDiox*, *PsPT1*, *CYP76F112*, *CYP71AJ3* and *KanaR* genes – and the additional MYB13 pair of primers as a positive control (**Supp. Table 14**). The *KanaR* gene and the *MYB13* sequence were successfully amplified in most of the plants, but none of the 4 genes from the psoralen biosynthesis pathway was found in any of the tomatoes (not shown). A sequencing performed with 2 plants confirmed that the fragment amplified with the KanaR primers was the *KanaR* gene. I also extracted the phenolic compounds from some of these tomato plants and analysed these extracts by UHPLC-MS, but no furanocoumarin was detected (not shown). This means that the tomatoes regenerated after the transformation with the 5-TUs plasmid were indeed transgenic, since they contained the *KanaR* gene (which allowed them to survive on the selective medium) but they did not contain any of the genes from the psoralen biosynthesis pathway.

In summary, I successfully used the KanaR plasmid to insert the TU [KanaR] into the genome of WVa106 tomatoes, creating transgenic plants resistant to kanamycin. On the contrary, the transformation with the 5-TUs plasmid was not successful since I did not manage to insert the psoralen biosynthesis pathway into the tomato genome. Because of the time required to transform and growth tomatoes, I could not try this experiment a second time.

D. DISCUSSION: GET A BETTER PLASMID, DON'T GIVE UP THE TRANSFORMATIONS

In this last chapter, my goal was to assess the metabolic cost of psoralen production. This objective has not been fulfilled, but I have successfully carried out the preliminary steps and obtained first results that will help to pursue this project.

D.1. Summary of the preliminary results

To assess the metabolic cost of psoralen production, I first used the GoldenBraid technology to construct a plasmid containing the first 4 genes of the furanocoumarin pathway and a selection marker. For this purpose, I successfully domesticated the *CYP76F112* and *KanaR* genes, constructed transcriptional units associated to the *CYP76F112*, *CYP71AJ3* and *KanaR* genes, and assembled them together and with a previously generated plasmid (Galati 2019). By doing so, I obtained a binary-vector containing the *PsDiox*, *PsPT1*, *CYP76F112*, *CYP71AJ3* and *KanaR* genes, under the control of individual 35S promoters and tNOS terminators (**Figure 71**). In the process, I also generated a plasmid containing the TU [KanaR], which represents the necessary negative control for the transformed plants. This work also constitutes a new demonstration of the simplicity and efficiency of the GB cloning system to repeatedly construct and assemble multiple TUs to form a biosynthesis pathway.

The second step of this work consisted in using the 5-TUs and the KanaR plasmids in *Agrobacterium*-mediated tomato transformations. On the one hand, I successfully inserted the TU KanaR into the genome of WVa106 tomatoes, obtaining a T0 generation of tomatoes resistant to kanamycin. These transgenic tomatoes produced seeds, which constitute the T1 generation. This means that, even though the plants from this T1 generation will need to be tested to only select the ones that does contain the transgene, the negative control tomatoes are ready to be used. On the other hand, the *Agrobacterium*-mediated transformation performed with the 5-TUs plasmid only led to a partial insertion of the T-DNA into the tomato genome, since only the *KanaR* gene was found in the regenerated plants. Numerous hypotheses can be proposed to explain this partial insertion.

D.2. Limits, hypotheses and recommendations for a future continuation

D.2.a. A matter of size?

The most intuitive hypothesis that could explain why the T-DNA from the 5-TUs plasmid was not entirely inserted in the tomato genome would be its size. Indeed, the T-DNA of the 5-TUs plasmid –

that contains 5 promoters, 5 coding sequences and 5 terminators – has a length of 13,782 bp. Such a big T-DNA might be more difficult to insert in a genome than the 2,357 bp T-DNA from the KanaR plasmid. However, such a big insertion, using GB plasmids, has already been reported. For instance, Fresquet-Corrales *et al.* (2017) used the GB technology to construct a plasmid that was quite similar to our 5-TUs plasmid, since it also contained 5 TUs respectively composed of a promoter, a coding sequence (2 genes coding for enzymes of interest, 2 transcription factors, and a gene conferring resistance to hygromycin), and a terminator. Their plasmid has been used for *Agrobacterium*-mediated tobacco transformation, and the full set of transgenes has been successfully inserted in the tobacco genome – which permitted to activate the anthocyanin and proanthocyanidin pathways in the transgenic plants (Fresquet-Corrales et al. 2017). As an even more drastic example, Sonawane *et al.* (2017) used the GB cloning system to generate a plasmid harbouring a kanamycin resistance gene and the 11 genes that constitute the entire cholesterol pathway – every coding sequence being individually flanked by a promoter and a terminator. Their plasmid therefore contained 12 TUs, which have been successfully inserted in the genome of *Arabidopsis*, leading to the creation of two independent lines of transgenic plants that accumulate a higher amount of cholesterol than the wild-type *Arabidopsis* (Sonawane et al. 2017).

These examples show that the use of the GB technology is a very effective approach to engineer plants and introduce multiple TUs / entire biosynthesis pathway in their genome. Therefore, the T-DNA size of our 5-TUs plasmid should not be a major barrier to the successful insertion of the psoralen biosynthesis pathway in the tomato genome. Yet, because of this important size, partial insertions of the T-DNA are more likely to happen than complete insertions.

D.2.b. Reordering the transcriptional units

During an *Agrobacterium*-mediated transformation, the insertion of a T-DNA inside a plant genome results from a polar transfer: the T-DNA is inserted from the right (RB) to the left border (LB), and deletions of the T-DNA near the 3' end (LB) are common (Wang et al. 1984; Lee and Gelvin 2008). For this reason, to ensure the complete insertion of the gene(s) of interest, the selection marker is usually placed near the LB – which means that partial insertions result in the removal of the selection marker (Lee and Gelvin 2008). In our 5-TUs plasmid, the kanamycin resistance gene was placed near the RB (**Figure 71**). The TU [KanaR] is therefore the first one to be inserted in the tomato genome, which partially explains why it is the only one that have been identified in the plants transformed with the 5-TUs plasmid: the plants that were regenerated only got a partial insertion of the T-DNA, that stopped "before" the second TU associated to CYP71AJ3.

Positioning the TU [KanaR] near the RB border was therefore a mistake, but thanks to the reusability and exchangeability of the GB constructs, this could be easily corrected by switching two destination vectors. Indeed, the TU [KanaR] would have to be assembled in an α1 vector instead of an α2, while the TUs [PsDiox+PsPT1+CYP76F112+CYP71AJ3] would have to be assembled in an α2 vector instead of an α1 (**B.2**). The two resulting plasmids could then be assembled into an Ω1 vector, which would give a plasmid similar to the actual 5-TUs plasmid, except that the TU [KanaR] would be positioned near the LB. The relative order of the TUs would therefore be [CYP71AJ3+CYP76F112+PsPT1+PsDiox+**KanaR**] – from RB to LB – instead of the actual [**KanaR**+CYP71AJ3+CYP76F112+PsPT1+PsDiox].

Such a modification would ensure that all the plants regenerated on the kanamycin-containing medium possess the entire psoralen biosynthesis pathway. For comparison, Fresquet-Corrales *et al.* (2017), who constructed the plasmid containing 4 genes and transcription factors of the anthocyanin and proanthocyanidin pathways, positioned their antibiotic resistance gene near the LB. Sonawane *et al.* (2017), who constructed the plasmid containing the 11 genes of the cholesterol pathway did the same. In addition, in both studies, the antibiotic resistance gene was positioned in the opposite orientation than the other TUs, which positioned its terminator closer to the LB, ensuring the integrity of the promoter and the resistance gene if short deletion occur near the LB. This could also be done for our 5-TUs plasmid by assembling the TU [KanR] in the opposite orientation, via the use of an α1R destination vector (pDGB1_ α1R – http://www.gbcloning.org) instead of an α1 vector.

D.2.c. Avoiding the repetitive use of identical promoters and terminators

Another point that might be critical to ensure the transfer of the entire T-DNA would be to avoid the repetition of the same 35S promoters and tNOS terminator in the 5 TUs. Indeed, such repetitive use of the same GB parts might favour DNA recombinations that would result in the loss of one or several transgenes. It could therefore be helpful to re-assemble the 5 TUs with different promoters and terminators. To avoid the domestication steps, these promoters and terminators might be chosen in the pre-existing GBCollection (http://www.gbcloning.org – Sarrion-Perdigones et al. 2014; Fresquet-Corrales et al. 2017). Yet, this point might not be critical, because some GB plasmids with repetitive promoters and terminators have already been successfully used for plant transformation. For instance, to come back to the previous examples, the 11 genes of the cholesterol pathway present on the plasmid generated by Sonawane *et al.* (2017) are all under the control of the same promoters and terminators, which means these GB parts are both repeated 11 times.

D.2.d. Using inducible instead of 35S promoters

However, I would highly recommend to avoid using the 35S promoter for the 4 TUs associated to the psoralen biosynthesis pathway, for two reasons.

Firstly, even though the 35S is a strong and constitutive promoter commonly used for plant transformation (Franck et al. 1980), it can also be responsible for interferences that may lead to misinterpretation of the results obtained with the transgenics plants. For instance, it can alter the transgene expression or upregulate some genes adjacent to the T-DNA borders, which can change the phenotype of the transgenic plants, and lead to the obtention of results that differ from those obtained with different promoters (Weigel et al. 2000; Yoo et al. 2005). In addition, 35S promoter siRNAs have sometimes been detected in T-DNA insertions mutant plants, causing a transcriptional silencing of the transgenes expressed with the 35S promoter (Mlotshwa et al. 2010). The transcriptional silencing of these transgenes might also vary with the promoter methylation heterogeneity (Matsunaga et al. 2019). This silencing might contribute to explain why the KanR gene was the only one found in the genome of the tomatoes transformed with the 5-TUs plasmid. Indeed, as the risk of silencing increase with the number of 35S copy, the insertion of several TUs in the genome of a given tomato might have resulted in the transcriptional silencing of the transgenes, including the KanR gene. Silencing this selection marker would have caused an early death of the plants. Therefore, the only tomatoes that survived are probably the ones with the TU [KanR] but the lowest number of 35S copies, favouring

the partial insertions of the T-DNA. Once again, this point might be important but not critical, since the 11 transgenes present on the cholesterol pathway plasmid generated by Sonawane *et al.* (2017) are all under the control of a 35S promoter, and they were all expressed in the transgenic plants, resulting in an increase of the cholesterol amount. Similarly, the GB plasmid constructed by Fresquet-Corrales *et al.* (2017) contained 4 TUs under the control of the 35S promoter. Nonetheless, in order to avoid bias and/or silencing of the genes of the psoralen biosynthesis pathway, it would still be better to avoid too many repetitions of the 35S promoter.

Secondly, it should be reminded that furanocoumarins, and particularly psoralen, are toxic compounds. For this reason, a constitutive expression of the 4 genes of the psoralen pathway driven by the strong 35S promoter might result in autotoxicity and kill the tomatoes. It is therefore possible that the entire T-DNA of the 5-TUs plasmid was inserted in the genome of a few tomatoes, which started to produce psoralen and died in the very early stages of the regeneration. Plant transformation and regeneration is a sensible thing, and the transgenic plants already have to grow on a kanamycin-containing medium. Therefore, the production of additional toxic compounds by the plants themselves, during the critical first steps of the regeneration, is certainly to avoid. Consequently, I would highly recommend to place the 4 genes of the psoralen pathway under the control of an inducible promoter. As having 4 repetitions of the same promoter should not be a major problem and would ensure a uniform expression of the 4 genes, a single promoter might be chosen for these 4 TUs. On its side, the *KanaR* gene can be kept under the control of the 35S promoter, because it needs to be constitutively expressed since the very beginning of the regeneration steps. In addition, I successfully generated kanamycin-resistant tomatoes with the KanaR plasmid – which suggests that a single copy of the 35S promoter does not cause major problems in tomato plants.

D.2.e. Additional TUs to prevent autotoxicity?

Also linked to the toxicity of furanocoumarins, we could also imagine to construct an even more complex plasmid which would not only allow the tomatoes to produce furanocoumarins, but would also provide the solutions to prevent autotoxicity. As explained in **Chapter I**, **I.D.3**, the plants that naturally produce furanocoumarins have evolved strategies to avoid autotoxicity. These strategies include a spatio-temporal variation of the furanocoumarin production – which could be more or less mimicked by the use of inducible promoters – but also strategies to inactivate or store furanocoumarins away from the sensitive tissues and/or cell compartments. Following this example, it would be interesting to generate transgenic tomatoes that would be able to inactivate or safely store furanocoumarins, because these strategies also represent a significant cost linked to the establishment of defence.

To do this, our 5-TUs plasmid could be completed with additional TUs associated to genes that would code, for instance, for enzymes that can glycosylate furanocoumarins (inactivation), or for furanocoumarin transporters. Unfortunately, to the best of my knowledge, no enzyme that can specifically glycosylate furanocoumarins have ever been described. Similarly, and despite the work that have been done on transporters that can secrete coumarins (Ziegler et al. 2017; Lefèvre et al. 2018), no furanocoumarin specific transporter have been reported so far.

In conclusion, I was not able to generate psoralen producing tomatoes, but my negative results permitted to highlight some weakness in our initial strategy. Accordingly, I proposed hypotheses and recommendations to reconstruct the 5-TUs plasmid. The main modifications would consist in putting the 4 genes of the psoralen pathway under the control of inducible promoters, and in positioning the TU [KanaR] in the opposite orientation, near the LB. Such a plasmid is presented on **Figure 76**. Using this modified plasmid in *Agrobacterium*-mediated transformation would give higher chances to introduce the entire psoralen biosynthesis pathway in the tomato genome.

If these future transformations were successful, the psoralen-producing tomatoes could finally be used to study the metabolic cost of psoralen production. This could be done by comparing the physiological parameters related to the growth and defence of the putative psoralen-producing tomatoes with those of the kanamycin resistant tomatoes I generated as a negative control. To avoid bias linked to chimeric plants and to the *in vitro* culture itself, this comparison should not be performed on the T0 but on the T1 generation.
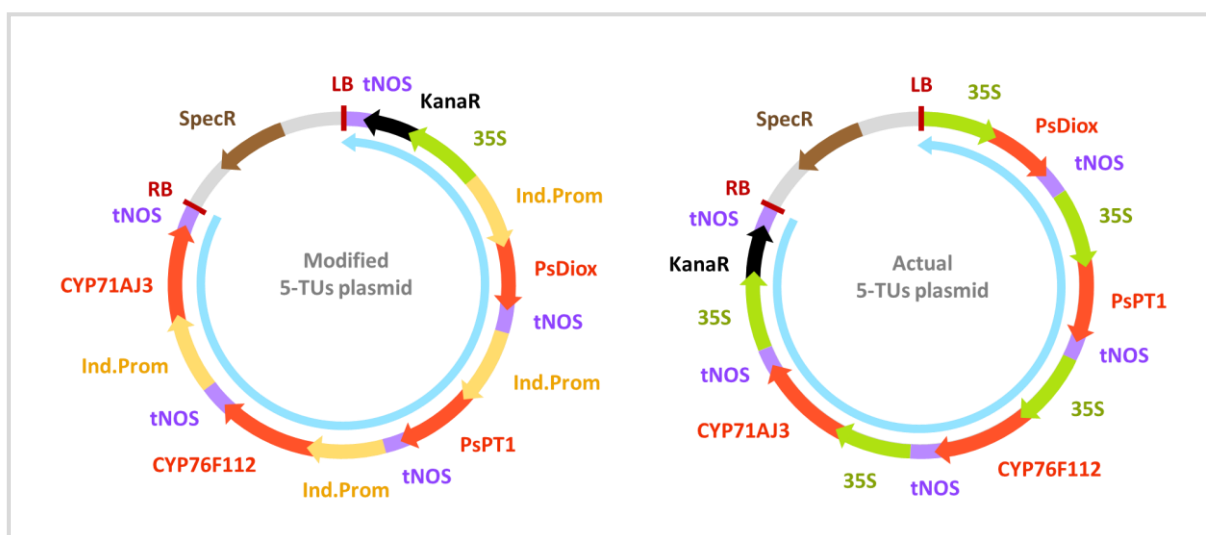
Figure 76 Plasmid map of the modified 5-TUs plasmid that should be constructed, in comparison to the actual one. SpecR refers to the gene conferring spectinomycin resistance to the bacteria. RB and LB refers to the right and left borders delimiting the T-DNA to be inserted in the plant, highlighted in blue. "Ind.Prom" stands for inducible promoter.

Chapter VI
General conclusion
and perspectives

CHAPTER VI.
GENERAL CONCLUSION AND PERSPECTIVES

Through evolution, plants have evolved many physical and chemical defence mechanisms that allow them to cope with pests and pathogens. For instance, plant chemical defences mainly rely on the production of a large array of specialised metabolites which are often species-specific. Amongst them, furanocoumarins constitute a family of toxic compounds that can be found in distant plant taxa such as the Apiaceae, Rutaceae, Fabaceae and Moraceae families. The emergence of furanocoumarins in higher plants has been made possible by the establishment of a biosynthesis pathway that involves at least four enzyme families: the dioxygenases, the prenyltransferases, the cytochromes P450s and the methyltransferases. While this pathway has been biochemically characterised for decades, some enzymatic steps are still to be resolved. Moreover, many questions remain about the emergence of the furanocoumarin pathway and the consequences of its establishment for the plants themselves. My PhD aimed to improve our knowledge of plant defence through the example of the furanocoumarin biosynthesis. It was articulated around two distinct axes: pursuing the elucidation of the furanocoumarin pathway, and assessing the metabolic cost of furanocoumarin production.

A. INTO THE UNKNOWN STEPS OF THE FURANOCOUMARIN PATHWAY

To pursue the elucidation of the furanocoumarin pathway, my first objective was to identify new *Ficus carica* genes involved in furanocoumarin production. By relying on a transcriptomic approach, I identified 10 putative P450s that were preferentially expressed in the latex of *F. carica* petioles. Then, I successfully cloned 8 of these candidates, heterologously expressed them in yeast, and assessed their activity by performing a metabolic screening with many coumarin and furanocoumarin substrates. This led me to the functional characterisation of three P450s with original activities: CYP76F112, CYP82J18 and CYP81BN4, which respectively metabolise demethylsuberosin (DMS), auraptene, and cnidilin. These three enzymes constitute the first known P450s from *F. carica* – and even the Moraceae family – that can metabolise coumarins and furanocoumarins.

A.1. CYP76F112: a recent marmesin synthase that opens many prospects

A.1.a. The marmesin synthase activity

CYP76F112 is a P450 with a high substrate and reaction specificity. It was found to convert a single substrate, DMS, into a single product, marmesin. With an apparent K_m value of 32.2 ± 3.9 nM, DMS possesses a particularly high affinity for CYP76F112. Therefore, DMS is most certainly the physiological substrate of CYP76F112, which makes this enzyme the first known marmesin synthase – and a very efficient one. To get the evidence required to state this with certainty, it would be possible to study the consequences of CYP76F112 silencing on the accumulation of furanocoumarins in *F. carica*.

Similarly, as furanocoumarins are defensive compounds, it would be interesting to inquire the variations of CYP76F112 expression patterns under biotic stress conditions. Regardless, the identification of a marmesin synthase clarifies the last unresolved enzymatic step from the beginning of the pathway, which opens new perspectives for the use and study of furanocoumarins. Moreover, as CYP76F112 is the first known CYP76 involved in the furanocoumarin pathway, it also provides new insights into the emergence of furanocoumarins in the fig tree, and in higher plants in general.

To inquire the emergence and evolution of CYP76F112, I inferred the gene-family phylogeny of the CYP76F subfamily in the Nitrogen Fixing Clade. This analysis revealed that CYP76F112 results from a recent bloom of the CYP76Fs that happened in the Moraceae family. This taxa-specific expansion happened via multiple tandem duplication events, followed by functional divergence. In *F. erecta*, and probably in all other *Ficus* species, most of the duplicates are clustered on a single chromosome. Then, I performed docking and site-directed mutagenesis experiments, which led me to highlight 4 key amino acids – A, B, C and D – from the SRS1 and SRS4 that influence CYP76F112 activity. The amino acid in position C (M117) was shown to be absolutely essential for the marmesin synthase activity, because its replacement by a bigger residue completely prevents the conversion of DMS into marmesin. On the contrary, the amino acids in position A, (B) and D (T102, (S105) and S305) contribute to CYP76F112 high affinity, but they are not critical for its activity. Their influence seems to be linked to their polar side chain, which allows them to establish hydrogen bonds with the DMS: by doing so, they might help in attracting, positioning and stabilising the DMS above the heme – which corresponds to the catalytic site of the enzyme. Finally, I tracked the evolution of these 4 amino acids throughout the CYP76F gene-family phylogeny. The evolutionary patterns strongly suggest that the 4 amino acids observed in CYP76F112 do not correspond to ancestral states, but to recent synapomorphies which are shared by a small *Ficus*-specific subclade composed of CYP76F112 and 5 close relatives from *F. erecta* and *F. religiosa*. All enzymes from this subclade might be marmesin synthases, and the marmesin synthase activity is probably restricted to this subclade.

In summary, we can now confidently assume that the marmesin synthase activity emerged recently in the Moraceae family, after the divergence of the clades that now contains the *Morus* and *Ficus* genera. In particular, the acquisition of the marmesin synthase activity, but also the high affinity of CYP76F112 for DMS, have been made possible by the evolution of the residues in positions A, (B), C and D – and most certainly by other mutations, for instance in the substrate access channel or maybe the transmembrane domain.

A.1.b. The furanocoumarin pathway, a case of convergent evolution

The evolutionary history of CYP76F112 also enlightens us about the emergence of the furanocoumarin pathway in higher plants. Until now, two hypotheses were proposed for the emergence of this pathway: on the one hand, it was suggested that the pathway may have appeared only once in the plant kingdom, but would have been lost in many taxa. This hypothesis implies that the genes involved in furanocoumarin production should be homologs in phylogenetically distant plant taxa such as the Moraceae and the Apiaceae family. For comparison, such a "single gain followed by massive losses" scenario has been proposed for the nitrogen-fixing root nodules trait in the NFC (van Velzen et al. 2019). On the other hand, it was proposed that the furanocoumarin biosynthesis pathway may have emerged multiple times, by convergent evolution. Under this hypothesis, distant plant taxa may have

recruited enzymes belonging to very different families to catalyse the same reactional steps of the furanocoumarin biosynthesis pathway. Such a repeated yet independent emergence of enzymes with the same function seems to be quite common in plants (Pichersky and Gang 2000; Weng 2014).

As the marmesin synthase activity is a recent acquisition of the Moraceae family, CYP76F112 has necessarily emerged independently from the marmesin synthases of Fabaceous, Apiaceous and Rutaceous species – which are not yet known. This is in perfect accordance with the independent emergence of moraceous and apiaceous UDTs proposed by Munakata *et al.* (2020), and it strongly suggests that the moraceous furanocoumarin pathway evolved independently from that of distant plant families. In other words, we can now propose with strong confidence that the marmesin synthase activity – and probably the whole furanocoumarin pathway – emerged multiple time in higher plants, through convergent evolution. As the apiaceous marmesin synthase is assumed to be a P450 (Hamerski and Matern 1988b), this would even be a case of parallel evolution.

Of course, as this study focused on the Moraceae family, the question remains open concerning the unique or multiple emergence of the furanocoumarin pathway in the Apiaceae, Rutaceae and Psoraleae families. Indeed, it is theoretically possible for these three plant families to share a common ancestor that produced furanocoumarins; an ancestor from which they inherited a common furanocoumarin pathway. Yet, like the Moraceae, the Psoraleae family belongs to the Nitrogen Fixing Clade. The Psoraleae are therefore phylogenetically closer to the Moraceae than to the Apiaceae and Rutaceae (**Figure 77** – Stevens 2001; The Angiosperm Phylogeny Group 2016). For this reason, the hypothetical furanocoumarin-producing ancestor shared by the Apiaceae, Rutaceae and Psoraleae families would also be an ancestor of the Moraceae family. In that case, the recent and independent emergence of CYP76F112 and FcPT1 in the Moraceae family could be explained by a partial replacement of some ancestral enzymes by more efficient ones, or by a complete loss of the putative ancestral furanocoumarin pathway, followed by the establishment of a new one. Even though it would
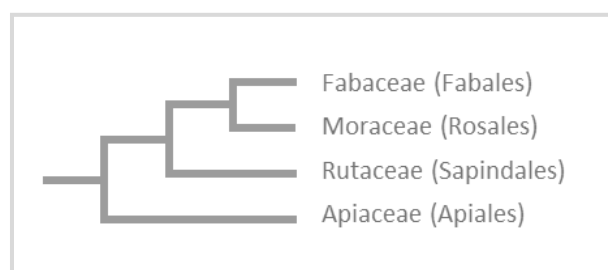
Figure 77 Simplified phylogenetic relationship between the Moraceae, Psoraleae, Rutaceae and Apiaceae families. Based on The Angiosperm Phylogeny Group (2016) and Stevens (2001).

not be completely impossible, this hypothesis is clearly not the most parsimonious. Likewise, the Rutaceae family shares a more recent common ancestor with the Fabaceae and Moraceae families rather than with the Apiaceae family (**Figure 77** – Stevens 2001; The Angiosperm Phylogeny Group 2016). Consequently, a common ancestral pathway shared by the Apiaceae and the Rutaceae but not the Moraceae and Fabaceae families would also be quite unlikely. In addition, one should be reminded that the hydroxylation of xanthotoxin is catalysed by a P450s from the CYP71 family in Apiaceae, but by a P450 from the CYP82 family in Rutaceae (Dugrand-Judek 2015; Krieger et al. 2018; Limones-Mendez et al. 2020). So, even though no precise phylogenetic analyses were done with these enzymes, this information also suggests an independent emergence of these enzymes in the Apiaceae and the Rutaceae families. Last but not least, as most of the plant taxon that are phylogenetically close to the furanocoumarin producing families do not produce furanocoumarins, a putative unique ancestral pathway would have been

associated to a very high number of independent losses. For all of these reasons, it can be hypothesised that the furanocoumarin pathway has emerged independently in each of the four main furanocoumarin producing plant families – Apiaceae, Rutaceae, Psoraleae and Moraceae. But for now, this is just a hypothesis that would need to be confirmed by additional phylogenetic studies.

A.2. CYP81BN4 and CYP82J18: promiscuous and non-species-specific enzymes?

In this study, I also reported the respective metabolization of cnidilin (a prenylated derivative of psoralen) and auraptene (a simple coumarin) by CYP81BN4 and CYP82J18. Compared to the marmesin synthase, these two enzymes were of lower interest for the study of core furanocoumarin biosynthesis. Indeed, the reactions they catalyse do not correspond to the main steps of the furanocoumarin pathway, and the activities described here might only be opportunistic reactions. Still, I inferred the gene-family phylogenies of both the CYP81BNs CYP82Js in the Nitrogen Fixing Clade. These analyses revealed that the CYP81BN subfamily did not particularly diversify in the Moraceae, and that the closest CYP81BN4 homologues tend to remain low-copy. Consequently, CYP81BN4 may have the same physiological function than its closest homologues from *Morus*, which does not produce furanocoumarins. As for CYP82J18, it seems to belong to an old and well-conserved subfamily that remained single or low-copies in most species of the NFC. Therefore, it probably originates from an ancient sequence that predates the NFC, which ancestral function might have been conserved by purifying selection. Consequently, neither CYP81BN4 nor CYP82J18 seem to originate from a recent taxa-specific diversification that would have been in favour of the acquisition of a new activity. Yet, in any case, while their physiological functions are unsure, it is prudent not to rely on the evolutive pattern of these enzymes to discuss the emergence of the furanocoumarin pathway.

Several approaches can be envisaged to complete the study of CYP81BN4 and CYP82J18. First, if both enzymes converted their respective substrates into multiple products, these products have not been chemically characterised. Their identification could be achieved with analyses such as NMR-based experiments. Then, one of the arguments against the physiological metabolization of cnidilin and auraptene by CYP81BN4 and CYP82J18 lies in the fact that neither cnidilin nor auraptene has ever been detected in the fig tree. Therefore, to determine if these compounds (and their precursors and potential derivatives) exist in the fig tree or are indeed absent from it, a deep *Ficus* phytochemical analysis could also be carried out. Lastly, it would be interesting to perform a wider metabolic screening on both enzymes with compounds that would not be related to furanocoumarins, but are naturally present in the fig tree – for instance, by performing enzymatic assays with crude *Ficus* extracts. Such a screening might lead to the identification of more likely physiological substrates for both CYP81BN4 and CYP82J18.

A.3. New prospects to pursue the elucidation of the furanocoumarin pathway

A.3.a. A complete genome for *Ficus carica*

Very recently, Usai *et al.* (2020) published the complete genome of *F. carica*, assembled at a chromosome level. This genome contains about 333 Mbp, distributed on 13 chromosomes (Usai et al. 2020). This publication will therefore allow us to complete the phylogenetic analysis I performed when

only partial *F. carica* genomic data were available. In particular, it will be very interesting to screen the complete *F. carica* genome, to identify all the CYP76Fs present on it, to include the new sequences in the CYP76F dataset, and to use it to build new phylogenetic gene trees. Preliminary BLAST searches performed on this genome have already shown that most of the CYP76Fs found in *F. erecta* have a *F. carica* equivalent. And as in the *F. erecta* genome, most of these CYP76Fs are clustered on a single *F. carica* chromosome – which tends to confirm the results and interpretations of the present study. Yet, it seems that the CYP76Fs from *F. carica* do not perfectly mirror those of *F. erecta*. For instance, the genes G06, G07 and G08 from *F. erecta* – which are very similar and certainly duplicates of each other's (**Chapter III**, **Figure 38**) – only seem to possess a single common relative in *F. carica*. This would suggest that the three copies found in *F. erecta* result from duplications that happened after the divergence of the *F. carica* and *F. erecta* species. Similarly, *F. carica* might possess some CYP76Fs copies that do not exist in *F. erecta* – implying that duplications also happened in *F. carica* after the divergence of the two species. Of course, these are only preliminary results based on a quite superficial screening of the *F. carica* new genome, but they would already suggest that the diversification of *Ficus* CYP76Fs was still going on after the *Ficus* genus had diverged from its closest neighbouring genera. For all these reasons, a deep screening and analysis of the newly published *F. carica* genome will certainly provide complementary information to the present study.

A.3.b. Finding the ancestral substrate of the *Ficus* CYP76Fs

If we can reasonably state that the marmesin synthase activity of CYP76F112 results from a recent bloom of CYP76Fs, we have no idea of the function associated to the ancestral gene which underlay these duplications. As the marmesin synthase catalyses the first P450-mediated step of the furanocoumarin pathway, it is most likely that its ancestral substrate was not a furanocoumarin – which would correspond to a downstream step of the pathway. Moreover, to the best of my knowledge, the only members of the CYP76F subfamily described so far are involved in terpenoid metabolism (**Chapter II**, **Figure 32**). Therefore, the ancestral substrate of *Ficus* CYP76Fs may also be a terpenoid. This would mean that the acquisition of the marmesin synthase activity was associated to a significant modification of the CYP76Fs substrate nature – from terpenoids to phenolic compounds. Under this hypothesis, CYP76F112 could be seen as an example of plant P450s ability to evolve and get recruited in new specialised pathways, allowing plants to adapt to fluctuating environmental constraints such as pest and pathogen attacks. In order to confirm (or not) this hypothesis, it would thus be very interesting to find out the physiological function of CYP76F112 ancestor. This would also allow us to better understand how the mutations that happened in the CYP76F sequences permitted the transition from the ancestral substrate to DMS. In other words, identifying such an ancestral substrate would provide a deeper understanding of the recruitment of the CYP76Fs in the furanocoumarin biosynthesis pathway. This would thus provide new data linked to the evolution of specialised biosynthesis pathway and the recruitment of their associated genes from other pathways.

One strategy that could be used to infer this ancestral activity would be to use phylogenetic tools to reconstruct the ancestral sequence of the moraceous CYP76Fs, to synthesise this putative ancestral gene, to heterologously express it, and to perform a metabolic screening on the resulting enzyme. Another strategy would be to characterise CYP76F110, CYP76F111 and/or other *Ficus* and non-*Ficus* CYP76Fs. Indeed, unless the function of the marmesin synthase ancestor is not useful anymore, or fulfilled by different enzyme(s), it should have been retained by at least one of the other *Ficus* CYP76Fs.

This ancestral function may also be found in CYP76Fs from the neighbouring taxa such as the *Morus* genus. In this study, I reported that CYP76F111 did not metabolise any furanocoumarin, and branched very early in the CYP76Fs *Ficus*-specific clade. For these reasons, it might have retained an ancestral activity that predated the *Ficus-Morus* divergence. It would thus be interesting to perform a wider metabolic screening on CYP76F111, with a large range of metabolites that are not related to furanocoumarins but are naturally present in the *Ficus* and *Morus* genera. If such screening were successful, the positive substrate(s) could then be tested on *Morus* CYP76Fs.

A.3.c. Other *Ficus* CYP76Fs potentially involved in the furanocoumarin pathway

However, even though CYP76F110 and CYP76F111 seem to be unable to metabolise furanocoumarins, CYP76F112 may not be the only enzyme resulting from the Moraceae-specific CYP76F bloom that is involved in the biosynthesis of furanocoumarins. This means that other CYP76Fs may catalyse other reactional steps of the pathway. Therefore, the numerous CYP76Fs found in the genome of *F. erecta* – as well as the other CYP76Fs that are to be found in the new *F. carica* genome – represent interesting candidates for a further elucidation of the furanocoumarin biosynthesis pathway. In particular, the genes G02 and G03 from *F. erecta* look promising since they are close CYP76F112 relatives but possess a leucine instead of a methionine in position C. As a leucine is smaller than a methionine, this mutation may not prevent the marmesin synthase activity. Yet, given the importance of the amino acid C and its impact on substrate specificity, these G02 and G03 genes may encode enzymes that metabolise other furanocoumarin substrates.

The various *Ficus* CYP76Fs would thus have to be cloned and characterised. A preliminary alternative to this functional characterisation would consist in using modelling and docking approaches to predict which CYP76F(s) might be able to metabolise furanocoumarin substrate(s). For this purpose, and as the docking analysis of CYP76F112 has given quite reliable results, I started a preliminary work on some of the CYP76Fs identified in *F. erecta*. Briefly, I performed additional docking experiments with the 4 F112-like modelized in **Chapter IV, E**, and with CYP76F112 for comparison. As substrates, I chose osthenol, marmesin and psoralen – which are respectively the angular equivalent of DMS, and the two intermediates of the furanocoumarin pathway that follow DMS (**Figure 78**). These 3 substrates might thus be the most likely substrates for enzymes that are phylogenetically close to CYP76F112. The docking experiments were performed as explained in **Chapter IV**, **B**.

The dockings performed with osthenol and marmesin did not gave any satisfying result (not shown), but the docking performed with psoralen might be promising. Indeed, in the best models (**Supp. Table 15**) generated during the docking of psoralen within G02_F.erecta, G02_F.religiosa, and G11_F.erecta, the psoralen was positioned in a way that might allow its conversion into bergaptol (**Figure 78**). These results suggest these three enzymes might be bergaptol synthases. Yet, as the psoralen was docked in the same position within CYP76F112 – which does not metabolise it – these results should be considered very carefully. So, as explained previously, docking experiments are only predictions, which have limits and do not always reflect the reality of the enzymes. A functional characterisation is therefore required. But despite their limits, these preliminary dockings make the F112-likes promising candidates for further elucidation of the furanocoumarin pathway, and psoralen will certainly be the most interesting substrate to include in the metabolic screening.
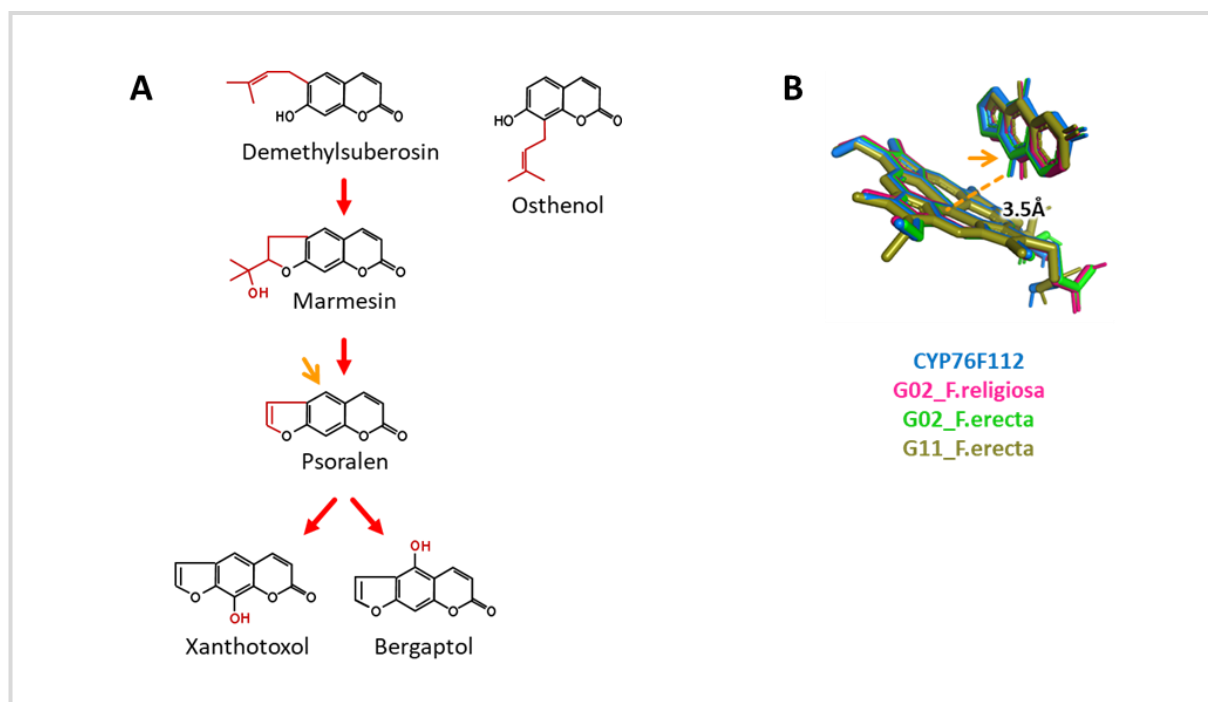
Figure 78 Potential substrates and additional docking experiments. A The major linear furanocoumarins that can be hydroxylated by P450s are the DMS, marmesin and psoralen. Osthenol is the angular equivalent of DMS. **B** Docking of the DMS within some of the F112-like. The carbon C5 of the psoralen, highlighted with an orange arrow, is the carbon that must be hydroxylated to form bergaptol. It is positioned just above the heme, at 3.5 Å from the iron.

A.3.d. Other P450 families potentially involved in the furanocoumarin pathway

If the CYP76Fs that result from the same diversification than CYP76F112 constitute the most obvious candidates to pursue the elucidation of the furanocoumarin pathway in the fig tree, it is also possible that other P450s families got recruited in this pathway. For instance, even though it might not be its physiological function, CYP81BN4 can metabolise cnidilin, which does create an evolutionary potential for furanocoumarin biosynthesis. Therefore, to identify other *Ficus* P450s from the furanocoumarin pathway, we should not limit ourselves to the CYP76 family. On the contrary, it would be interesting to investigate other P450 families from the CYP71 clan.

To identify interesting candidates, several approaches can be considered. The first one would be the transcriptomic approach, which has already proven itself useful. Yet, I already screened the *F. carica* RNA-seq library described in Kitajima *et al.* (2018) and, as far as I know, it is the only *Ficus* transcriptomic database currently available. Another possibility would be to rely on a potential clustering of the genes involved in the furanocoumarin pathway, and to search for other P450s located on the same chromosomal region than CYP76F112. Such an approach has already been successfully used in *Pastinaca sativa* (Roselli et al. 2017), but it might not be relevant in the fig tree. Indeed, considering the recent origin of the marmesin synthase, and the fact that CYP76F112 and FcPT1 are localised on different *F. carica* chromosomes – respectively, the chromosomes 11 and 7 described by Usai *et al.* (2020) – it is likely that the *Ficus* genes involved in the furanocoumarin pathway do not (yet) form a single metabolic cluster. The third approach that could be considered would be to rely on a

comparison between *Ficus* and *Morus* CYPomes. Indeed, CYP76F112 results from a taxa-specific expansion and diversification that happened after the *Ficus-Morus* divergence. And given that the *Morus* genus do not produce furanocoumarin, it is likely that the other *Ficus* genes involved in furanocoumarin biosynthesis would also result from a taxa-specific bloom that happened after the *Ficus-Morus* divergence. In their study, Ma *et al.* (2014) screened the genomic resources of *Morus notabilis* to predict its CYPome. By doing so, they identifying 174 putative P450 genes distributed in 47 families – which includes 98 P450s belonging to the CYP71 clan (Ma et al. 2014). Since the recent publication of *F. carica* and *F. erecta* genome (Shirasawa et al. 2019; Usai et al. 2020), it would be possible to use a similar approach to predict the *F. carica* (and/or *F. erecta*) CYPome(s). Then, a comparative analysis of *Morus* and *Ficus* CYPomes – that could be done by building a phylogenetic gene tree with all P450s from *M. notabilis* and *F. carica* (and/or *F. erecta*) – could lead to the identification of P450 families that would be more extended in *Ficus* than in *Morus*. This would suggest a taxa-specific bloom, and would make these P450 families promising candidates for a further elucidation of the furanocoumarin pathway.

A.3.e. The marmesin synthases in other plant families

To go even further in the discussion, it would be interesting to identify the marmesin synthases from taxonomically distant plant taxa such as the Apiaceae, Rutaceae or Psoraleae families. A BLAST search performed on the partial genetic data available for these species did not revealed the presence of CYP76s that would be close to the marmesin synthase – which was to be expected given the recent emergence of CYP76F112. Therefore, other marmesin synthases must have been recruited from other P450 families. Until now, in the Apiaceae family, all the known P450s involved in the furanocoumarin pathway belong to the CYP71 family. It is therefore possible that they all evolved from a single ancestral CYP71 gene that underwent multiple duplications, and that also resulted in the emergence of a marmesin synthase. However, the apiaceous marmesin synthase might also belong to a P450 family that would be completely different from all the currently known P450 families involved in furanocoumarin biosynthesis (CYP71, CYP82 and CYP76). The same is also valid for the Rutaceae and Psoraleae families. So, to pursue the elucidation of the furanocoumarin pathway in higher plants, the approach that consists in searching new genes by similarity has to be completed with approaches based on transcriptomic, phylogenetic, modelling, and/or interspecies comparisons – as it was done or suggested in the present study.

A.3.f. Other enzymes families: *Ficus* methyltransferases and dioxygenases

Finally, if this study mainly focused on cytochromes P450s, they are not the only enzymes involved in the furanocoumarin biosynthesis pathway. Until now, three other enzymes families were identified: the prenyltransferases, methyltransferases and dioxygenases (**Figure 12**). Several prenyltransferases involved in the pathway have already been characterised, including FcPT1 from *F. carica* (Munakata et al. 2020). Yet, less information is available concerning the involvement of methyltransferases and dioxygenases in furanocoumarin production, and none is related to the Moraceae family (Hehmann et al. 2004; Matsumoto et al. 2012; Vialart et al. 2012; Roselli et al. 2017).

Dioxygenases mediated reactions might be quite similar to P450 mediated reactions, and some enzymatic steps might either be catalysed by a dioxygenase or a P450. For instance, the conversion of

scopoletin into fraxetin is mediated by a P450 (CYP71AZ4) in *Pastinaca sativa* (Krieger et al. 2018) and by a dioxygenase (At3g12900) in *Arabidopsis thaliana* (Siwinska et al. 2018). For this reason, it would be possible for the marmesin synthase activity to be associated to a dioxygenase in non-*Ficus* species; and enzymatic steps such as the hydroxylation of marmesin or psoralen may even be catalysed by dioxygenases in the fig tree. Therefore, inquiring these other enzymes families is also essential since every plant family might have evolved its own furanocoumarin pathway, by recruiting very different enzymes depending on the evolutionary opportunities they were offered.

By relying on a transcriptomic approach similar to that used for the P450s, I identified 7 genes coding for putative dioxygenases, and 7 for putative methyltransferases (**Supp. Table 2**). Some preliminary work has been performed at the LAE: the candidates have been synthesised, heterologously expressed in bacteria, and a metabolomic screening allowed to assess their activity. The recent preliminary results revealed that at least two methyltransferases can metabolise furanocoumarins, and at least one dioxygenase can metabolise coumarins. These enzymes will thus provide precious data and insight into furanocoumarin production.

A.3.g. Application and study of plant biosynthesis pathways

CYP76F112 nicely completes the set of enzymes that allows the production of toxic psoralen from the plant wide *p*-coumaroyl CoA. Therefore, it opens many prospects for the use and study of the furanocoumarin biosynthesis pathway. For instance, it can be used to study the costs and impacts linked to the establishment of psoralen biosynthesis in a given plant – which constituted the second axis of my PhD. Another direct application made possible by the identification of CYP76F112 would be to engineer microorganisms such as yeasts to produce high amounts psoralen. This could be of medicinal interest, since the biological properties of furanocoumarins could be used to fight cancer or diseases such as vitiligo. Yet, as their toxicity also apply to humans, furanocoumarins still need to be studied for a safe use (Chauthe et al. 2015; Zabolinejad et al. 2020). On the contrary, the knowledge of the furanocoumarin biosynthesis pathway could also be used to extinguish furanocoumarin production in natural furanocoumarin-producing plants. As an example, the genes of the furanocoumarin pathway could be targeted in breeding and selection schemes or with genome editing approaches to reduce the quantity of furanocoumarins found inside agronomical plants. This would help reducing the health issues associated to the contact and/or ingestion of these toxic compounds.

Finally, and more broadly, the identification of CYP76F112 also provide new perspectives about the study of enzymes, biosynthesis pathways and the production of molecules of interest in general. For instance, contrary to most P450s described so far in the furanocoumarin pathway, CYP76F112 is an enzyme that was very easy to produce and to store, since it was extremely resistant to freezing and defrosting. Performing deeper analyses of CYP76F112 structure might thus help us to understand what makes a P450 stable – such as its folding or its membrane anchor. This could help to engineer P450s used to produce molecules of interest, in order to make them more sable, improve and ease the production. In addition, the reaction catalysed by CYP76F112 is not a common hydroxylation but an oxidation followed by a cyclisation; which is quite atypical and could also be a subject for deeper enzymatic studies. As a last example, identifying new P450s from the furanocoumarin pathway also contribute to increase our knowledge of enzymes that perform rather similar reactions. And when several enzymes performing the same reaction step (such as the conversion of DMS into marmesin)

will be characterised, it will be possible to compare some of their characteristics such as their relative efficiency in different cellular environments and plants. This could provide general knowledge about the functionality or the structure-function relationship of plant enzymes.

B. TOO MUCH FURANOCOUMARINS WILL COST YOU

The second objective of my PhD was to assess the metabolic cost of furanocoumarin production by inserting the beginning of the furanocoumarin pathway in the genome of tomato (*Solanum lycopersicum*), a plant that does not naturally produce these molecules. To do so, I used a multi-gene cloning technology, transgenesis, and *in vitro* culture methods to introduce 4 genes of the furanocoumarin pathway into the genome of tomato. The idea was to generate psoralen-producing tomatoes which growth and defence might be compared to negative control tomatoes that do not produce furanocoumarins.

By using the GoldenBraid technology, I constructed two binary plasmids usable for plant transformation. One of these plasmids contains a single transcription unit associated to a kanamycin resistance gene. The second plasmid contains 5 TUs associated to a kanamycin resistance gene and the first 4 genes of the furanocoumarin pathway. Then, I used these plasmids in *Agrobacterium*-mediated transformations to introduce these genes in the tomato genome. On the one hand, I successfully generated kanamycin-resistant tomatoes, which constitute the necessary negative controls for the subsequent growth and defence analysis. On the other hand, I did not succeed to insert the psoralen biosynthesis pathway in the tomato genome but my work constitutes a pioneer work that will be continued. In particular, it led me to the identification of limits in my initial strategy, and the establishment of recommendations to reconstruct the 5-TUs plasmid prior any new attempt at tomato transformations. If psoralen-producing tomatoes could be generated, they could be compared with the kanamycin-resistant tomatoes I obtained, to study the metabolic cost of psoralen production.

~

As a general conclusion, this study contributed to decipher the furanocoumarin biosynthesis pathway. In particular, the characterisation and analysis of the first known marmesin synthase enlightens us on the multiple emergences of the furanocoumarin pathway in higher plants and opens many prospects for future uses and study of furanocoumarins. More broadly, it also opens new tracks that could lead to a better understanding of the P450s and their plasticity, of the recruitment of enzymes in specialised biosynthesis pathways, and of the acquisition and evolution of new enzymatic activities.

Yet, as interesting as they could be, it should not be forgotten that furanocoumarins only represent one of the many mechanisms that constitute the fascinating world of plant defence. This means that many other stories are still to be written if we intend to understand what happens in these little green things that lives all around us.

Chapter VII
Materials and methods

Chapter VII.

Materials and methods

A. Materials

A.1. Plant material

A.1.a. *Ficus carica*

The *Ficus carica* plants used for RNA extraction come from seeds provided by Plant Advanced Technologies (Vandoeuvre-lès-Nancy, France).

A.1.b. *Solanum lycopersicum*

The seeds of the *Solanum lycopersicum* var. *cerasiformae* 'West Virginia 106' (WVa106) used for plant transformation were provided by the Fruit Biology and Pathology's laboratory (Laboratoire Biologie du fruit et pathologie, UMR 1332, INRAE – Bordeaux, France).

A.2. Bacterial strain

A.2.a. *Escherichia coli* MC1022

The *E. coli* strain MC1022 (genotype: *ara*D139, Δ(*ara*, *leu*)7697, Δ(*lacz*)M15, *galU*, *galK*, *strA*) is used for cloning and plasmid amplification. It is derived from the strain M182 (genotype: Δ(*lacI*POZY) X74, *galK*, *galU*, *strA*) and allows for quick and efficient transformation and amplification of large plasmids.

A.2.b. *Escherichia coli* ccdB Survival™

The *E. coli* strain ccdB Survival™ (genotype F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lacX*74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G *ton*A::P$_{trc}$-*ccd*A) was purchased at Invitrogen™. It is used for cloning and plasmid amplification: as it possesses a resistance to the toxic ccdB gene product, it allows the amplification of vectors such as pYeDP60_GW® (**A.4.b**).

A.2.c. *Agrobacterium tumefaciens* EHA105

The *Agrobacterium tumefaciens* strain EHA105 (Hood et al. 1993) is used for plant stable transformation. This hypervirulent agrobacterium strain carries in its genome the *Rif$_R$* gene that confers rifampicin resistance.

A.3. Yeast strain: *Saccharomyces cerevisiae* WAT21

The *Saccharomyces cerevisiae* strain WAT21 is used to produce all *Ficus* P450s. It derives from the strain W303 and had been engineered to allow the heterologous expression of plant P450s. The gene encoding the endogenous NADPH P450-reductase had been replaced by *AthR2*, one of the gene from *A. thaliana* encoding for a P450-reductase. This gene is placed under the transcriptional control of the *GAL10-CYC1* promoter, which is inducible by galactose but repressed by glucose. Therefore, when this strain is transformed with a pYeDP60 plasmid containing a P450, and grown on galactose, it overexpresses both the P450 and the *A. thaliana* P450-reductase, allowing the study of the P450 activity in the microsomal fraction. Like the W303 strain, the WAT21 strain expresses an endogen cytochrome b_5 which can contribute to the P450 activity, and possesses a mutation on the *ADE2* gene that prevents it to grow on medium lacking adenine (Pompon et al. 1996; Urban et al. 1997).

A.4. Vectors

A.4.a. pCR®8/GW/TOPO™

The pCR®8/GW/TOPO™ TA Cloning Kit is provided by Invitrogen™. It contains the ready-to-use linearised pCR®8/GW/TOPO™ vector (**Figure 70**), which possesses 3´-T overhangs (deoxythymidine residue) that prevent circularisation and allow the direct ligation of PCR products amplified with the *Taq* polymerase – which adds a single A (deoxyadenosine residue) to the 3' ends of PCR products. This is referred to as TA Cloning. In addition, a topoisomerase I is covalently bound to the pCR®8/GW/TOPO™ vector, releasing the energy required to complete the ligation. This vector therefore allows for quick and efficient cloning of PCR products. It also contains a pUC origin allowing high-copy replication in *E. coli*, and a spectinomycin resistance gene for an efficient selection. The PCR products ligated in the plasmid can be sequenced by using the universal M13 forward primer combined with the reverse T7 promoter/priming site. Finally, since it possesses *att*L1 and *att*L2 sites, this plasmid also delivers access to the Gateway® technology, allowing the transfer by recombination of the PCR product into other Gateway® destination vectors such as pYeDP60_GW® (**A.4.b**). Consequently, the pCR®8/GW/TOPO™ vector was used to clone, amplify, and sequence various PCR products, but also for subsequent recombination.
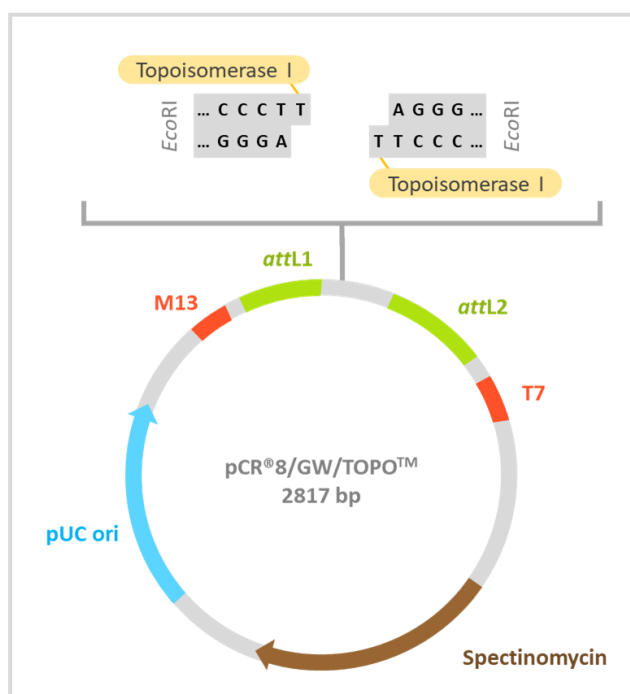
Figure 79 Simplified representation of the linearised pCR®8/GW/TOPO™ vector.

A.4.b. pYeDP60 and pYeDP60_GW®

The pYeDP60 vector (**Figure 80**) was provided by Dr D. Pompon from the Membrane Protein Engineering laboratory (Laboratoire d'Ingénierie des Protéines Membranaires, CNRS - Gif-sur-Yvette, France). This plasmid was specifically designed for the expression of P450s in various *Saccharomyces cerevisiae* strains such as WAT21 (Urban et al. 1990, 1997). It contains a 2μ origin allowing replication in yeast, and the URA3 and ADE2 selection markers to complement the auxotrophy of their yeast host. It also contains an expression cassette constituted by a polylinker *Bam*HI/*Sma*I/*Kpn*I/*Sac*I/*Eco*RI framed by the *GAL10-CYC1* promoter upstream, and by the PGK terminator downstream. The *GAL10-CYC1* promoter is inducible by galactose and repressed by glucose, which allows to control the expression of a gene of interest cloned within the polylinker. The presence of a ColE1 replication origin and an ampicillin resistance gene also allow the subcloning, amplification and selection of the pYeDP60 vector in *E. coli*.

The pYeDP60_GW® vector (**Figure 80**) is directly derived from pYeDP60. It has been constructed by Dr C. Krieger during her PhD at the LAE (Krieger 2014) and results from the addition of the Gateway® technology to the pre-existing pYeDP60 vector. For this purpose, the RfA recombination cassette harbouring the *att*R1 and *att*R2 sites has been inserted in the polylinker, using the *Bam*HI restriction site. This cassette also contains a chloramphenicol resistance gene and the *ccdB* gene. The amplification of the empty pYeDP60_GW® plasmid must therefore be done in bacteria that resist to the ccdB toxin, such as the *E. coli* strain ccdB Survival™. The addition of the Gateway® technology allows for quick and efficient subcloning of genes of interest from various destination vectors.

The pYeDP60_GW® vector is used to express the various P450s cloned from *F. carica*; the pYeDP60 vector is used to express CYP76F111, CYP76F112 and their associated mutants.
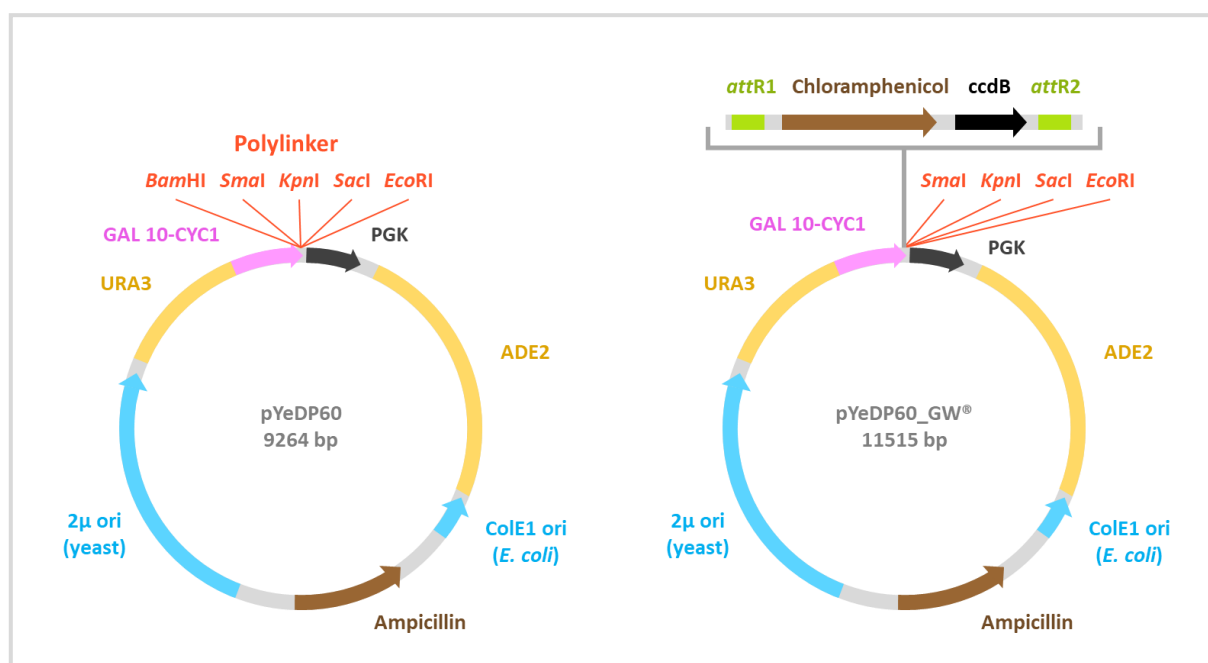
Figure 80 Simplified representation of the pYedp60 and pYeDP60_GW® vectors.

A.4.c. GoldenBraid commercial vectors

The following vectors are commercial and standardised GoldenBraid plasmids described on https://gbcloning.upv.es/ (Sarrion-Perdigones et al. 2011, 2013). They are all replicated in *E. coli* and contain a resistance gene allowing the selection of the transformed bacteria. They were provided by Addgene (Diego Orzaez Lab Plasmids – Cambridge, MA, USA) and are used for GB construction and tomato stable transformation.

A.4.c.1. The pUPD vectors: pUPD, pUPD-35S and pUPD-tNOS

The pUPD vector (**Figure 81**) is the universal GB part domesticator plasmid. It allows the cloning of every basic GB parts (promotor, coding sequence, terminator) and their subsequent assembly into transcriptional units. Its selection marker is an ampicillin resistance gene. The pUPD vector contains the *LacZ* gene, flanked upstream and downstream by *Bsm*BI and *Bsa*I restriction sites. The *Bsm*BI recognition sites allow the cloning of a domesticated GB part within the plasmid, instead of the *LacZ* gene (destination vector). This recognition site is destroyed during the cloning. The *Bsa*I recognition sites allow the assembly of the GB parts from several pUPD plasmids (entry vectors) into a pDBG1_α plasmid – which is the next level vector. The complete sequence of the pUPD vector can be found in https://gbcloning.upv.es/feature/pUPD/.

The pUPD-35S vector (GB0030 – **Figure 81**) is derived from the empty pUPD vector. It contains the CaMV 35S promoter, inserted via *Bsm*BI restriction ligation. Therefore, it does not contain the *LacZ* gene and the *Bsm*BI recognition sites anymore. It is used to construct transcriptional units. The complete sequence of the pUPD-35S vector can be found in https://gbcloning.upv.es/feature/GB0030/ (Addgene plasmid # 68163 ; http://www.addgene.org/68163/ ; RRID:Addgene_68163).

The pUPD-tNOS vector (GB0037 – **Figure 81**) is derived from the empty pUPD vector. It contains the Nos terminator, inserted via *Bsm*BI restriction ligation. Therefore, it does not contain the *LacZ* gene and the *Bsm*BI recognition sites anymore. It is used to construct transcriptional units. The complete sequence of the pUPD-tNOS vector can be found in https://gbcloning.upv.es/feature/GB0037/ (Addgene plasmid # 68188 ; http://www.addgene.org/68188/ ; RRID:Addgene_68188).
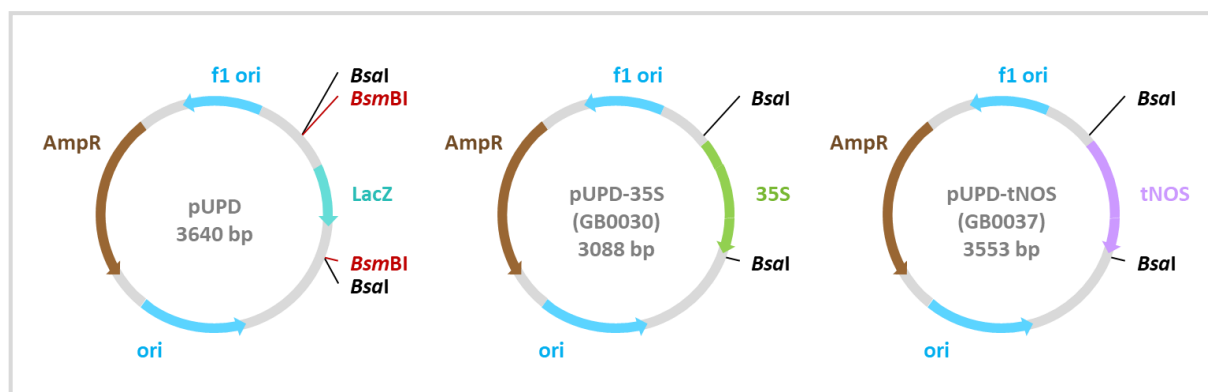
Figure 81 Simplified representation of the pUPD, pUPD-35S and pUPD-tNOS vectors.

A.4.c.2. The α-level vectors: pDGB1_α1 and pDGB1_α2

The pDGB1_α1 and pDGB1_α2 plasmids (**Figure 82**) are GB binary α-level vectors used for GB assembly, which can directly be used for *Agrobacterium*-mediated plant transformation. Their selection marker is a kanamycin resistance gene. They both contain the *LacZ* gene, flanked upstream and downstream by *Bsa*I and *Bsm*BI restriction sites; and they differ by the overhangs left upon *Bsm*BI restriction. The *Bsa*BI recognition sites allow the assembly of several GB parts / transcriptional units into higher order modules within the plasmid, instead of the *LacZ* gene. This recognition site is destroyed during the cloning. The *Bsm*BI recognition sites allow the ordered assembly of a module from the pDGB1_α1 plasmid and a module from the pDGB1_α2 plasmid into a pDBG1_Ω vector. These binary vectors contain left and right borders (LB and RB) that define the T-DNA which can be inserted in a plant genome by *Agrobacterium tumefaciens* strain such as EHA105. However, their replication in *Agrobacterium* require the presence of a helper plasmid such as the pSoup plasmid. The complete sequence of the pDGB1_α1 vector can be found in https://gbcloning.upv.es/feature/pDGB1_alpha1/ (Addgene plasmid # 68224 ; http://www.addgene.org/68224/ ; RRID:Addgene_68224). The complete sequence of the pDGB1_α2 vector can be found in https://gbcloning.upv.es/feature/pDGB1_alpha2/ (Addgene plasmid # 68225 ; http://www.addgene.org/68225/ ; RRID:Addgene_68225).
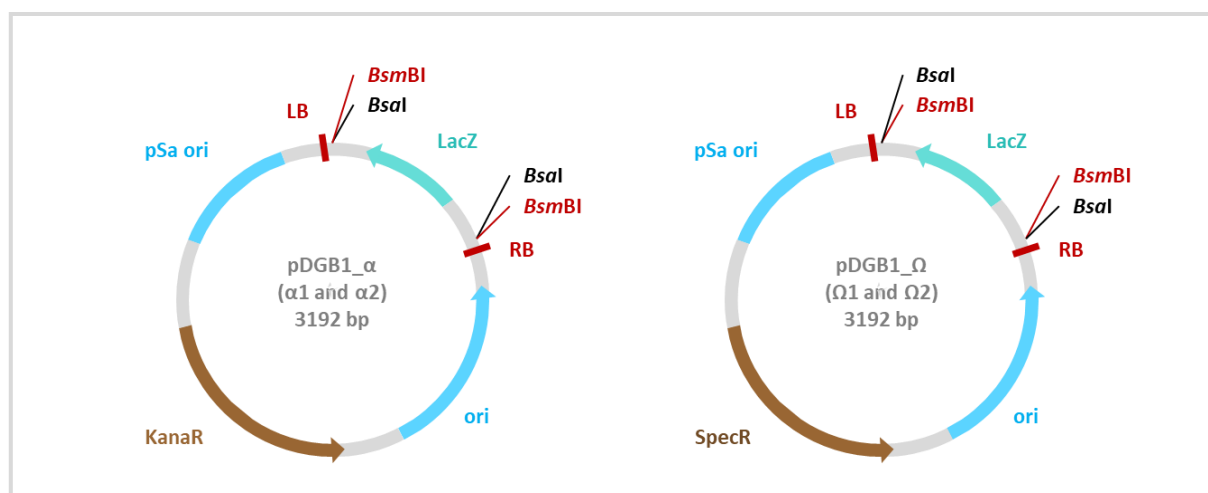
Figure 82 Simplified representation of the pDGB1_α and pDGB1_Ω vectors.

A.4.c.3. The Ω-level vectors: pDGB1_Ω1 and pDGB1_Ω2

The pDGB1_Ω1 and pDGB1_Ω2 plasmids (**Figure 82**) are GB binary Ω-level vectors used for GB assembly and that can directly be used for *Agrobacterium*-mediated plant transformation. Their selection marker is a spectinomycin resistance gene. They both contain the *LacZ* gene, flanked upstream and downstream by *Bsm*BI and *Bsa*I restriction sites; and differ by the overhangs left upon *Bsa*I restriction. The *Bsm*BI recognition sites allow the assembly of several transcriptional units into higher order modules within the plasmid, instead of the *LacZ* gene. This recognition site is destroyed during the cloning. The *Bsa*I recognition sites allow the ordered assembly of a module from the pDGB1_Ω1 plasmid and a module from the pDGB1_Ω2 plasmid into a pDBG1_ α vector. These binary vectors contain left and right borders (LB and RB) that define the T-DNA which can be inserted in a

plant genome by *Agrobacterium tumefaciens* strain such as EHA105. However, their replication in *Agrobacterium* require the presence of a helper plasmid such as the pSoup plasmid. The complete sequence of the pDGB1_Ω1 vector can be found in https://gbcloning.upv.es/feature/pDGB1_omega1/ (Addgene plasmid # 68234 ; http://www.addgene.org/68234/ ; RRID:Addgene_68234). The complete sequence of the pDGB1_Ω2 vector can be found in https://gbcloning.upv.es/feature/pDGB1_omega2/ (Addgene plasmid # 68225 ; http://www.addgene.org/68235/ ; RRID:Addgene_68225).

A.4.d. Recombinant GoldenBraid vectors

The following plasmids have been constructed by Dr G. Galati during his PhD at the LAE (Galati 2019). They result from the insertion by restriction-ligation of domesticated GB parts into commercial pUPD and PDGB1-Ω1vectors. They are used for GB assembly.

A.4.d.1. pDGB1_Ω1 [PsDiox+PsPT1]

The plasmid pDGB1_Ω1 [PsDiox+PsPT1] (**Figure 83**) is referred to as [35S-PsDiox-tNOS : 35S-PsPT1-tNOS] in Galati (2019). It results from the insertion of two transcriptional units into the empty pDGB1_Ω1 vector, through *Bsm*BI restriction-ligation. Therefore, the pDGB1_Ω1 backbone does not contain the *LacZ* gene and the *Bsm*BI recognition sites anymore. The two transcriptional units are respectively composed of the domesticated coding sequences of PsDios and PsPT1, flanked by a 35S promotor and a tNOS terminator.

A.4.d.2. pUPD-CYP71AJ3

The plasmid pUPD-CYP71AJ3 (**Figure 83**) results from the insertion of the domesticated sequence of CYP71AJ3 into the pUPD plasmid, through *Bsm*BI restriction-ligation. Therefore, the pUPD backbone does not contain the *LacZ* gene and the *Bsm*BI recognition sites anymore.
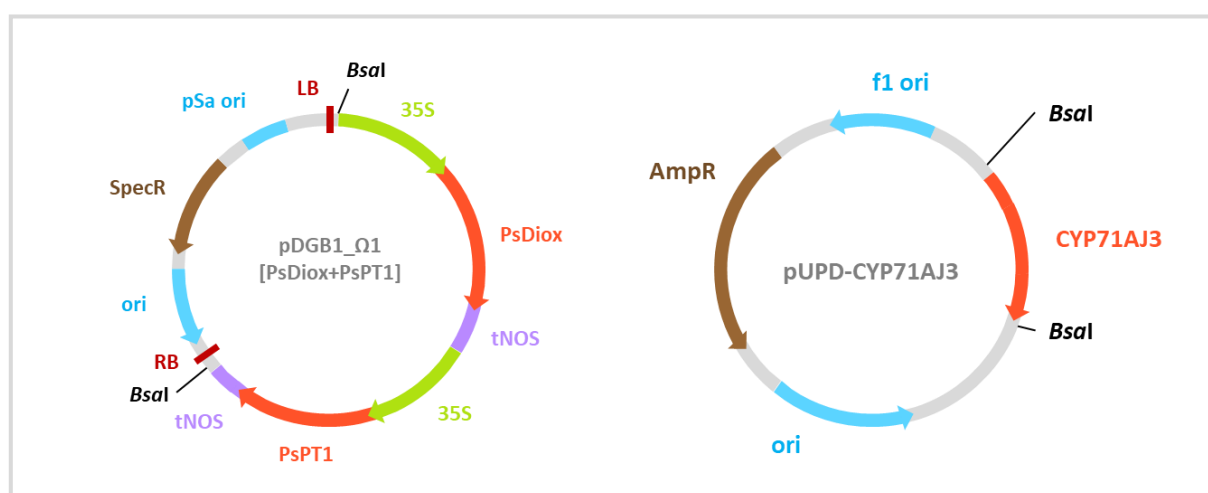
Figure 83 Simplified representation of pDGB1_Ω1 [PsDiox+PsPT1] and pUPD-CYP71AJ3.

A.4.e. pSoup

The pSoup vector (**Figure 84**) was provided by Mark Smedley (John Innes Center – Norwich, United Kingdom). It is derived from the pRK2 vector and is used for *Agrobacterium* co-transformation with pGreen family plasmids such as the PDGB vectors. It carries the *pSa* replicase gene that interacts in *trans* with the *pSa* replication origin from PDGB1-α and PDGB1-Ω vectors to allow their replication in *Agrobacterium* strains. It also contains a tetracycline resistance gene as a selection marker.
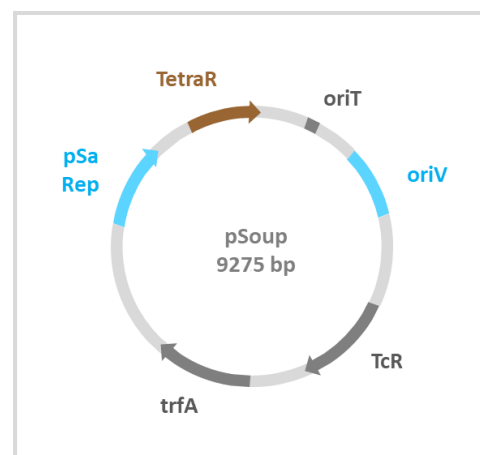
Figure 84 Simplified representation of the pSoup vector.

A.4.f. plCSL11024 vector

The plCSL11024 vector (pICH47732::NOSp-NPTII-OCST – **Figure 85** – Addgene plasmid # 51144 ; http://www.addgene.org/51144/ ; RRID: Addgene_51144 – Level 1 KAN and BASTA, unpublished) contains a gene that confers kanamycin resistance to plant. This gene is cloned and used in the GoldenBraid assembly. The vector can be amplified in *E. coli* and carries an ampicillin resistance gene as a selection marker.

A.5. Culture media

Unless otherwise specified, the media are sterilised with an autoclaving cycle of 121°C for 20 min. The liquid media are stored at room temperature; the solid media are stored at 4°C.
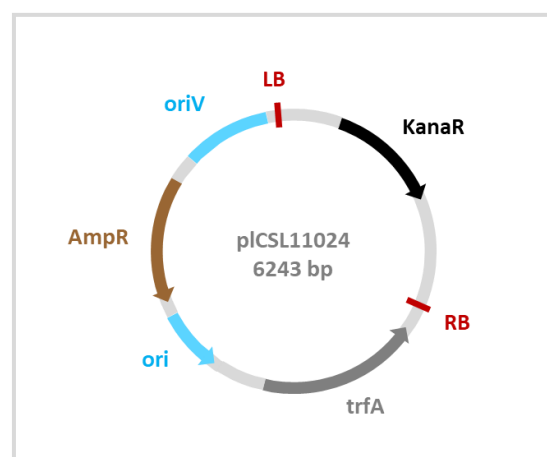
Figure 85 Simplified representation of the plCSL11024 vector.

A.5.a. Bacteria culture medium: LB medium and associated antibiotics

The LB medium (Lysogeny Broth) is used for the culture of *E. coli* and *A. tumefaciens*. It is prepared by dissolving 25 g of commercial LB powder (Sigma-Aldrich – Saint-Louis, MO, USA) into 1 L of distilled water. It is composed of tryptone (10 g/L), yeast extract (5 g/L) and NaCl (10 g/L). To prepare solid medium, 10 g/L of agar (Sigma-Aldrich) is added.

The LB medium is supplemented with one or several antibiotics to allow the selection of recombinant bacteria. The antibiotics are added to the medium after autoclaving. In the case of solid medium, they are added when the medium is at a temperature of about 50°C, just before it is poured into sterile petri dishes. The antibiotics are listed in **Table 2**; they are used at the indicated final concentration, after dilution of the stock solution in the LB medium.

Table 11 Antibiotics used for bacteria culture.

Antibiotic	Supplier	Reference	Solvent	Concentration of the stock solution (g/L)	Final concentration (mg/L)
Ampicillin	Sigma-Aldrich	A9518-25G	Distilled water	100	**100**
Kanamycin	Sigma-Aldrich	K1377	Distilled water	50	**50**
Rifampicin	Duchefa	R0146	DMSO	20	**20**
Spectinomycin	Sigma-Aldrich	S4014	Distilled water	100	**100**
Tetracycline	Duchefa	T0150	Ethanol	50	**5**

A.5.b. Yeast culture media

Three culture media are used for the culture of the *S. cerevisiae* strain WAT21. Their composition is described in **Table 12**. To prepare solid medium, 7 g/L of agar is added. The YPGA medium is a complete medium, complemented with adenine, that is used to grow the WAT21 strain prior transformation. The SGI medium is a selective medium that does not contain adenine. It is used to select the yeasts transformed with the recombinant pYeDP60 and pYeDP60_GW® vectors, which both contain the *ADE2* gene conferring adenine autotrophy. The YPGE medium is a complete medium used to perform P450 heterologous expression in yeast. It contains glucose that is used by the yeasts as the preferential carbon source, and ethanol that is used as a secondary source when glucose is entirely consumed.

Table 12 Composition of the yeast culture media. The compounds in white constitute the main solution which is sterilised by autoclaving. The ones highlighted in yellow are sterilised separately (filtration 0.2µm for the tryptophan and adenine, autoclaving for the glucose solution) and subsequently added to the main solution. The galactose solution (blue) is sterilised by filtration 0.2µm and added during the experiment, when the glucose is entirely depleted.

	Supplier	Reference	YPGA	SGI	YPGE
Yeast extract (g/L)	Conda pronadisa	1702	10		10
Bactopeptone (g/L)	Becton Dickinson	211677	10		10
Bactocasaminoacids (gL/)	Becton Dickinson	223050		1	
Yeast nitrogen base (g/L)	Roth	HP26		7	
Tryptophan (mg/L)	Alfa Aesar	A10230		20	
Adenine (mg/L)	Sigma-Aldrich	A2786	30		
Glucose (g/L)	Fisher	10335850	20	20	5
Ethanol (v/v)	Carlo Erba	3086612			3%
Galactose (g/L)	Roth	4987			20

A.5.c. Tomato *in vitro* culture media

Six culture media are used for the *Agrobacterium*-mediated transformation of tomatoes. The references of the chemicals and the composition of the stock solutions required to prepare these media are described in **Table 13**. These stock solutions are sterilised by filtration (0.2 μm) and stored at -20°C.

Table 13 Reference of the chemicals and composition of the stock solutions required to prepare the tomato *in vitro* culture media.

Chemicals	Supplier	Reference	Stock solution	
			Solvent	Concentration
MS salts (Murashig & skoog medium Basal salt mixture)	Duchefa	M0221	---	---
KH$_2$PO$_4$	VWR Chemicals	16B020015	---	---
MES	Sigma-Aldrich	M3671-50	---	---
Agar HP696	Kalys-Biotech	A7921	---	---
Sucrose	Duchefa	S0809	---	---
Thiamine HCl	Sigma-Aldrich	T-3902	Distilled water	**10 mg/mL**
IAA (Indole acetic acid)	Sigma-Aldrich	I2886	EtOH 70%	**1 mg/mL**
BAP (6-benzylaminopurine)	Sigma-Aldrich	B9395	Distilled water + 2% NaOH	**10 mg/mL**
Acetophenone (3',5'-Dimethoxy-4'-hydroxyacetophenone)	Aldrich Chemistry	D134406-5G	EtOH 70%	**0.098 mol/L**
Kanamycin	Duchefa	K0126	Distilled water	**75 mg/mL**
Timentin (Ticarcilin disodium w/Clavula)	Toku-E	TO-F020-10G	Distilled water	**150 mg/mL**
MS vitamins (Murashig & skoog vitamins mixture)	Duchefa	M0409.0250	Distilled water	**10 g/L**

The media used for the transformation and regeneration are described in **Table 14**. The suspension medium is a liquid medium used to prepare the *Agrobacterium* suspension that is required for the transformation itself. The five other media are solid media optimised for the germination, transformation and regeneration of *Solanum lycopersicum* var. *cerasiformae* 'West Virginia 106'. For every solid medium, a main solution is sterilised by autoclaving, and the sterile stock solutions are added before pouring, when the medium is at a temperature of about 50°C.

Table 14 Composition of the tomato transformation *in vitro* culture media.

Chemicals	Germination medium	Preculture medium	Suspension medium	Co-culture medium	Regeneration medium	Rooting medium	
MS salts	1.1 g/L	4.41 g/L	4.41 g/L	4.41 g/L	4.41 g/L	2.2 g/L	Added to the main solution, sterilised by autoclaving
KH$_2$PO$_4$	---	200 mg/L	200 mg/L	200 mg/L	---	---	
MES	---	---	---	---	700 mg/L	700 mg/L	
Agar HP696	8 g/L	8 g/L	---	8 g/L	8 g/L	8 g/L	
Sucrose	15 g/L	20 g/L	20 g/L	20 g/L	30 g/L	10 g/L	
Adjust pH of the solutions to 5.8 (KOH) prior autoclaving							
Thiamine HCl	---	0.9 mg/L	0.9 mg/L	0.9 mg/L	---	---	Sterile stock solution added to the medium after it is autoclaved
IAA	---	1 mg/L	---	1 mg/L	1 mg/L	---	
BAP	---	2 mg/L	---	2 mg/L	2 mg/L	---	
Acetophenone	---	$200*10^{-6}$ mol/L	$200*10^{-6}$ mol/L	$200*10^{-6}$ mol/L	---	---	
Kanamycin	---	---	---	---	150 mg/L	---	
Timentin	---	---	---	---	300 mg/L	75 mg/L	
MS vitamins	---	---	---	---	100 mg/L	50 mg/L	

A.6. Bioinformatic tools

A.6.a. Databases

A.6.a.1. Ficus carica *RNA-seq library*

A *Ficus carica* differential RNA-seq library is screened to identify P450 candidates. It was provided by Dr R. Munakata and Dr S. Kitajima (Kyoto University – Japan). It was constructed from the latexes of fruits, petioles, and lignified trunks of *F. carica*. It contains RNA sequences (contigs) associated to their expression level and relative expression level in the three latexes (Kitajima et al. 2018).

A.6.a.2. Public online databases

Genetic resources from various plants species are accessed, screened and downloaded on 1KP (https://db.cngb.org/onekp/ – One Thousand Plant Transcriptomes Initiative 2019; Carpenter et al. 2019) and GenBank (https://www.ncbi.nlm.nih.gov/ – Benson et al. 2012). These databases provide a free access to transcriptomic, proteomic and genomic information and to the associated BLAST tools.

The Cytochrome P450 Homepage (https://drnelson.uthsc.edu/CytochromeP450.html – Nelson 2009) provides sequence information on cytochrome P450s. It is used to identify published P450s, in order to constitute a P450 reference dataset.

The Protein Data Bank (PDB – http://www.rcsb.org/ – Berman et al. 2000; Burley et al. 2019) is accessed to retrieve the 3D structures of P450s with published references. They are identified throughout the document with their PDB ID.

The 3D models of the heme (protoporphyrin IX containing Fe – CID: 4971), DMS (CID: 5316525), marmesin (CID: 334704), psoralen (CID: 6199) and osthenol (CID: 5320318) used for the docking are downloaded on PubChem (https://pubchem.ncbi.nlm.nih.gov/).

A.6.b. Software

A.6.b.1. Software used for molecular biology and basic sequence analyses

The software BioEdit (BioEdit Sequence Alignment Editor, version 7.0.5.3 – Hall 1999) is used to perform basic manipulations of nucleotide and protein sequences such as pairwise alignments, identity and similarity calculation, local BLASTs, translations and six-frame translations.

The software Serial Cloner (version 2.6.1 – SerialBasics – http://serialbasics.free.fr/Home/Home.html) is used to annotate plasmids and perform basic plasmid manipulations such as virtual digestions.

The software SigmaPlot (version 12 – Systat – https://systatsoftware.com/products/sigmaplot/) is used to determine enzyme kinetic parameters.

The software LabSolution (Shimadzu – https://shimadzu.com.au/labsolutions) and Xcalibur™ (version 2.1.SP1.Build1160 – Thermofisher) are used to acquire and process the data from UHPLC-MS analyses and LTQ Orbitrap analyses, respectively.

A.6.b.2. Software used for phylogenetic analyses

The software Geneious (version 8.1.9 and Geneious prime 2019 – https://www.geneious.com/) is used to align multiple nucleotide sequences using the translation alignment tool.

The software Mesquite (version 3.6 – https://www.mesquiteproject.org/ – Maddison and Maddison 2019) is used to check sequence alignment and to partition, parameter and export aligned sequences in nexus file format usable for Bayesian inference tree building.

The software MrBayes (version 3.2.7 – Ronquist and Huelsenbeck 2003; Ronquist et al. 2012) is used to check the correct setting of the Bayesian Markov chain Monte Carlo (MCMC) analysis prior running it on CIPRES.

The software Tracer (version 1.7.1 – http://beast.community/tracer – Rambaut et al. 2018) is a graphical tool used for analysing the trace files generated by Bayesian MCMC runs.

The software FigTree (version 1.4.4 – http://beast.community/figtree) is a graphical tool used to visualise and annotate the various phylogenetic trees built in this study.

A.6.b.3. Software used for modelling and docking

The software PyMol (The PyMOL Molecular Graphics System, version 2.3.3, Schrödinger, LLC – https://pymol.org/2/) is a graphic tool used to visualise 3D molecule structures.

The software AutoDock Vina (version 1.1.2 – http://vina.scripps.edu/ – Trott and Olson 2009) is used to perform the molecular docking of the heme and substrates into the various modelled P450s. To run AutoDock Vina, the MGLTools (version 1.5.7 – https://ccsb.scripps.edu/mgltools/) are required.

A.6.c. Online tools

The Oligo Calculator (http://mcb.berkeley.edu/labs/krantz/tools/oligocalc.html) allows the calculation of the T_m temperature and the GC proportion of given nucleotide sequences. It is used to design the primers used for PCRs.

The NetGene2 Server (http://www.cbs.dtu.dk/services/NetGene2/ – Hebsgaard 1996) is a service producing neural network predictions of splice sites in *A. thaliana*. It is used to predict the introns in genomic sequences.

The CIPRES Science Gateway (version 3.3 – https://www.phylo.org/) is a public platform that allows the inference of large phylogenetic trees by providing access to NSF XSEDE's computational resources. It is used to build the Bayesian inference trees.

The SWISS-MODEL (https://swissmodel.expasy.org/ – Guex et al. 2009; Benkert et al. 2011; Bienert et al. 2017; Bertoni et al. 2017; Waterhouse et al. 2018; Studer et al. 2020) is an automated protein structure homology-modelling server. It is used for the modelling of the CYP76Fs.

B. Methods

B.1. Common molecular biology and microbiology methods

B.1.a. Plant tissue grinding

Prior starting the experiment, the work surface and the material are cleaned with a bleach solution (2%) and then with ethanol (70%). Fresh plant tissues (leaves) are instantly frozen by dipping in liquid nitrogen, and ground using clean mortar and pestle. 25 to 100 mg of frozen ground plant tissue are weighted and transferred to a 1.5 mL microcentrifuge tube. This powder can be immediately used for RNA, DNA or phenolic compounds extraction, or stored at -80°C for later extractions.

B.1.b. Extraction and purification of plant RNA

The total RNAs from frozen ground plant tissues are extracted using the E.Z.N.A.® Plant RNA Kit (Omega Bio-tek), according to the standard protocol provided by the supplier. The elution is performed twice in 40 µL of DEPC water provided in the kit. To check if the extraction went well and the integrity of the RNAs, a sample of the extracted total RNAs is separated by performing an agarose gel electrophoresis. The residual genomic DNA that might remain in the RNA solution is removed by performing a digestion with the Amplification Grade DNAse I (Sigma-Aldrich) according to the recommendation provided by the supplier. Finally, the concentration of total RNA in the solution is determined by spectrometry (**B.1.l**). The resulting RNA solution can be used directly or stored at -20°C.

B.1.c. Synthesis of complementary DNA

Complementary DNA (cDNA) are synthesised from 100 ng of total RNA (free from genomic DNA) by using the High Capacity RNA-to-cDNA™ Kit (Applied Biosystems™), according to the recommendation provided by the supplier. The reaction mix – composed of 100 ng of RNA, the RT buffer (1x) and the retrotranscriptase (RT enzyme mix 1x) – is incubated for 1h at 37°C and then for 5min at 95°C. The resulting cDNA is diluted 5 times in ultrapure water; it can be used directly or stored at -20°C.

B.1.d. Extraction of plant genomic DNA

The genomic DNA from frozen ground plant tissues are extracted using the E.Z.N.A ® SP Plant DNA kit (Omega Bio-tek), according to the standard protocol provided by the supplier. The elution is performed twice in 50 µL of elution buffer provided in the kit. The concentration of genomic DNA in the resulting solution is determined by spectrometry (**B.1.l**). The gDNA can be used directly or stored at -20°C.

B.1.e. Amplification of DNA fragments by PCR

DNA fragments were amplified by PCR, in a C1000 Touch™ Thermal Cycler (Bio-rad), using the polymerases described below. The many pairs of primers used for PCR are described in the annexes (**Supp. Table 3**, **Supp. Table 14**, **Supp. Figure 10**), their associated annealing temperature (T_m) are calculated using the Oligo Calculator online tool (**A.6.c**). For every PCR reaction, the reaction mixture is composed of a premix at a final 1x concentration, the matrix DNA (20-200ng), forward and reverse primers (0.2 µM), and ultrapure water for a final volume of 15 to 50 µL.

B.1.e.1. PrimeSTAR® Max DNA Polymerase

The PrimeSTAR® Max DNA Polymerase (Takara Bio inc.) is a premix that includes a high-performance DNA polymerase, a dNTP mixture and an optimised buffer. The polymerase contained is this premix possesses a high accuracy, sensitivity, specificity and fidelity, with proofreading activity. It is used to amplify fragments that needs to be cloned or sequenced. The DNA fragments it amplifies have blunt ends. The PCR conditions used with the PrimeSTAR® Max DNA Polymerase premix are the following: an initial denaturation at 98°C for 1 min, 25 (to 40 cycles) composed of a denaturation step (98°C for 10 s), an annealing step (T_m for 15 s) and an elongation step (72°C for 10 s/kb), followed by a final elongation step at 72°C for 5 min.

B.1.e.2. SapphireAmp® Fast PCR Master Mix 2X

The SapphireAmp® Fast PCR Master Mix 2X (Takara Bio inc.) is a premix that includes a hot start DNA polymerase, a dNTP mixture, an optimised buffer and a blue gel loading dye. The polymerase contained in the SapphireAmp® Fast PCR Master Mix 2X is optimised for fast PCRs but, contrary to the previous enzyme, it has a lower fidelity. Therefore, it is used for colony PCR screening and for amplifying fragment that are not meant to be cloned or sequenced. The PCR conditions used with this premix are the following: an initial denaturation at 94°C for 1 min, 25 (to 40 cycles) composed of a denaturation step (98°C for 10 s), an annealing step (T_m for 10 s) and an elongation step (72°C for 10 s/kb), followed by a final elongation step at 72°C for 2 min.

B.1.f. DNA extraction from agarose gel

Agarose gel bands containing DNA fragments revealed by UV light can be cut with a clean scalpel and transferred into a 2 mL microcentrifuge tube. DNA is then extracted from the gel using the kit Nucleospin® Gel and PCR Clean Up (Macherey-Nagel), according to the standard protocol provided by the supplier. The elution of the extracted DNA is performed twice in 30 µL of elution buffer provided in the kit. DNA can be used directly or stored at -20°C.

B.1.g. DNA digestion using restriction enzymes

Simple digestions of DNA samples (usually, purified plasmids) are performed with various restriction enzymes listed in **Table 15**. Restriction reactions are generally performed in 20 µL of reaction mixture composed of 1X reaction buffer, 1U of restriction enzyme and 100-500 ng of DNA diluted in distilled water. All restriction enzymes used in this study for simple digestions are compatible with the FastDigest buffer. Therefore, reactions can be performed with several restriction enzymes at the same time in the same buffer. They are incubated for 20-45 min at 37°C – depending on the quantity of DNA to digest and the number of restriction sites.

Table 15 Restriction enzymes used for simple digestions of DNA samples.

Enzyme	Supplier	Reference	Restriction site
*Bam*HI	Thermo Scientific™	FD0054	G↓GATC↑C
*Eco*RI	Thermo Scientific™	FD0274	G↓AATT↑C
*Hind*III	Thermo Scientific™	FD0504	A↓AGCT↑T
*Pvu*II	Thermo Scientific™	FD0634	CAG↑↓CTG
*Sac*I	Thermo Scientific™	FD1133	G↑AGCT↓C

B.1.h. Cloning techniques

B.1.h.1. Cloning of a PCR-amplified fragment in pCR™8/GW/TOPO™

The pCR™8/GW/TOPO™ vector allows TA Cloning® (**A.4.a**), but the DNA fragments amplified with the high fidelity PrimeSTAR® Max DNA Polymerase (**B.1.e.a**) have a blunt end. Therefore, prior cloning them in the pCR™8/GW/TOPO™ vector, such fragments are amplified using the *Taq* DNA polymerase contained in the PCR Master Mix 2X (ThermoFisher Scientific – K0171), which adds a 3'-A overhang to PCR products. For this purpose, a reaction mixture composed of 1X of MasterMix (which includes the *Taq* DNA Polymerase and dNTPs) and the amplified fragment (100-600 ng/µL) is incubated for 40 min at 72°C. Then, the resulting adenylated fragment is cloned into the pCR™8/GW/TOPO™ vector by using the pCR™8/GW/TOPO® TA Cloning Kit (Invitrogen™). To do this, a reaction mixture is prepared in a final volume of 3 µL: it is composed of 100-400 ng/µL of the adenylated fragment, 0.5 µL of salt solution provided in the cloning kit but diluted 4 times, and 2 ng/µL of the pCR™8/GW/TOPO™ vector. This reaction mix is incubated at room temperature (20-25°C) for about 2-3 hours. It is better used fresh for bacteria transformation, but can also be stored at -20°C for later uses.

B.1.h.2. Recombination into the pYeDP60_GW® vector

Genes of interest cloned into the pCR™8/GW/TOPO™ vector can be transferred to the pYeDP60_GW® vector (**A.4.b**) using the Gateway® technology. The recombination is performed using the Gateway LR Clonase™ II Enzyme Mix (Thermofischer Scientific), according to recommendations provided by the supplier: 50-150 ng of entry plasmid (recombinant pCR™8/GW/TOPO™) are mixed with 150 ng of destination vector (empty pYeDP60_GW®) in TE buffer (pH=8.0) for a final volume of 8 µL. Then, 2 µL of LR Clonase™ II enzyme are added to the mixture, briefly homogenised and centrifugated (2x2 s each), and incubated at 25°C for 1h. Finally, 1 µL of the Proteinase K solution provided in the kit is added and the reaction mixture is incubated at 37°C for 10 min. The resulting solution is better used fresh for bacteria transformation, but can also be stored at -20°C for later transformation.

B.1.i. Preparation of electrocompetent bacteria

The following preparations of competent bacteria are performed in sterile condition.

B.1.i.1. Preparation of electrocompetent Escherichia coli

When *E. coli* are not in isotonic conditions, they are very fragile and can be easily lysed. Therefore, except for the (pre)cultures, the following steps are performed on ice, with cold materials, and the cells have to be gently manipulated (no mechanical knocks or violent centrifugation / resuspension). Some *E. coli* (**A.2**) from a glycerol stock (stored at -80°C) are plated on solid LB medium (**A.5.a**) and incubated at 37°C until the appearance of isolated colonies. A single colony is picked and used to inoculate 30 mL of liquid LB medium, which are incubated for 16 h at 37°C with an orbital shaking of 180 rpm. A volume of 10 mL of this preculture is used to inoculate 200 mL of liquid LB medium, which are incubated at 37°C and 180 rpm until the OD_{600nm} reaches 0.6 to 0.8. The bacterial culture is then placed in ice for 30 min and immediately centrifugated for 15 min at 4°C and 2200g. The supernatant is discarded and the cell pellet is gently resuspended into 20 mL of sterile cold water. The centrifugation

and resuspension steps are repeated once in sterile cold water, and twice in sterile cold glycerol (10%). The bacterial solution is placed in ice for 30 min and aliquoted in sterile tubes (45 µL in each tube) that are directly used for transformation, or frozen in liquid nitrogen and stored at -80°C.

B.1.i.2. Preparation of electrocompetent Agrobacterium tumefaciens *EHA105*

Some *A. tumefaciens* strain EHA105 (**A.2.c**) from a glycerol stock (stored at -80°C) are plated on solid LB medium supplemented with rifampicin (**A.5.a**) and incubated at 28°C until the appearance of isolated colonies. A single colony is picked and used to inoculate 5 mL of liquid LB medium supplemented with rifampicin (**A.5.a**), which are incubated for 16h at 28°C with an orbital shaking of 200 rpm. 2 mL of this preculture are used to inoculate 50 mL of liquid LB medium supplemented with rifampicin, which are incubated at 28°C and 200 rpm until the OD_{600nm} reaches 0.6 to 0.8. The bacterial culture is then placed in ice for 30 min and centrifugated for 10 min at 4°C and 3000 g. The supernatant is discarded and the cell pellet is gently resuspended into 45 mL of sterile cold water. The centrifugation and resuspension steps are repeated once in 45 mL of sterile cold water, once in 40 mL of sterile cold glycerol (10%), and once in 500 µL of sterile cold glycerol (10%). The bacterial solution is finally aliquoted in sterile tubes (40 µL in each tube) that are directly used for transformation, or frozen in liquid nitrogen and stored at -80°C.

B.1.j. Transformation of competent bacteria

Electroporation is a transformation method that consists in introducing a plasmid in bacteria whose membrane structure has been temporally destabilised and made more permeable by the application of a brief but high voltage electrical field. For this purpose, 1 to 10 ng of cold plasmid are mixed in a cold aliquot of electrocompetent bacteria (*E. coli* or *A. tumefaciens*), and transferred into a cold electroporation cuvette which electrodes are spaced 2 mm apart. The cuvette is placed in a MicroPulser Electroporator (Bio-rad) and submitted to an electrical shock of 2.5 kV. Immediately after the shock, 300 µL of liquid LB medium are added into the cuvette, gently mixed with the bacteria, and transferred into a 2 mL microcentrifuge tube. The resulting bacterial solution is incubated for 45 min at 37°C and 180 rpm (*E. coli*), or for 3-4h at 28°C and 180 rpm (*A. tumefaciens*), before it is plated on a solid LB medium supplemented with the adequate antibiotic(s). The plate is incubated at 37 or 28°C (*E. coli* or *A. tumefaciens*, respectively), until the appearance of isolated colonies. In order to identify the transformed bacteria, colony PCRs (**B.1.e.2**) can be performed on isolated colonies.

B.1.k. Isolation of plasmid DNA from bacteria

A colony of *E. coli* containing a plasmid of interest is used to inoculate 4 mL of liquid LB medium supplemented with the appropriate antibiotics. The culture is incubated overnight at 37°C with an orbital shaking of 180 rpm, until a saturated bacterial solution is obtained. The plasmid DNA from the bacteria are then extracted using the kit NucleoSpin® Plasmid (Macherey-Nagel), according to the standard protocol provided by the supplier. The plasmid DNA is eluted in 2x25 µL of elution buffer (Buffer AE provided in the kit); it can be directly used or stored at -20°C for a later use.

In the case of plasmid DNA to extract from *A. tumefaciens*, a spatula tip of lysozyme (Thermo Scientific™ – 89833) is added to the resuspension buffer (Buffer A1 provided in the kit), for a final

concentration of about 10 mg/mL. This buffer is incubated for 5-10 min at room temperature to improve the lysis of the agrobacteria, before pursuing the standard NucleoSpin® Plasmid protocol.

B.1.l. Spectrophotometry quantification

The concentration of nucleic acid in a given solution is quantified by measuring the $OD260_{nm}$ with a spectrophotometer BioPhotometer (Eppendorf). First, a blank is done with 70 µL of distilled water (or another adequate solution). Then, 5 µL of sample are diluted into 65 µL of distilled water (or another adequate solution), and the $OD260_{nm}$ of the resulting solution is measured.

The OD_{600nm} of bacterial and yeast cultures is determined using an Ultrospec® 10 Cell Density Meter (Biochrom). First, a blank is done with 1 mL of adequate sterile culture media. Then, the OD_{600nm} is measured in 1mL of (diluted) microbial culture.

B.1.m. Sequencing

DNA sequencing is realised by sending DNA sample to Macrogen Europe B.V. (Amsterdam, the Netherlands). Sequencing of the genes of interested cloned into the pCR™8/GW/TOPO™ vector is performed by using the universal M13F and/or T7 Prom primers. Sequencing of the genes of interested cloned into the pYeDP60 and pYeDP60_GW® vectors is performed using the primers YeEco (CAACACCTGGCAATTCCTTACC) and/or YeBam (TCTTTCCTTATACATTAGGTCC) which are to be sent to the sequencing company. Additional sequencings have been performed using the primers described in **Supp. Table 14**.

B.1.n. Synthesis of the CYP76F mutants

The nucleotide sequences of the His-tagged CYP76F111, CYP76F112 and their associated mutants (**Supp. Figure 5**) are synthesized and inserted into the expression vector pYeDP60 (**A.4.b**) by the company GeneCust (Boynes, France – https://www.genecust.com/fr/). The cloning of these coding sequences into the pYeDP60 shuttle vector is performed using the *Bam*HI and *Eco*RI restriction enzymes. Therefore, prior synthesis, a *Bam*HI and an *Eco*RI restriction site are respectively added at the 5' and 3' end of the coding sequence; and two internal restriction sites are removed from CYP76F111 and replaced by equivalent codons (**Supp. Figure 1**, **Supp. Figure 5**).

B.2. Methods linked to P450 heterologous expression, assay and characterisation

B.2.a. Yeast transformation

B.2.a.1. Preparation of competent S. cerevisiae *WAT21*

The following steps are performed in sterile conditions. Some *S. cerevisiae* strain WAT21 (**A.3**) from a glycerol stock (stored at -80°C) are plated on solid YPGA medium (**A.5.b**) and incubated at 30°C for 48-72 h, until the appearance of isolated colonies. A single colony is picked and used to inoculate 10 mL of liquid YPGA medium, which are incubated for 18 h at 30°C with an orbital shaking of 180 rpm. The

yeast preculture is then diluted into 50 mL of liquid YPGA, with an OD_{700nm} of 0.2. This culture is incubated at 30°C and 180 rpm for about 4-5 h, until an OD_{700nm} of 0.8 is reached. The yeast solution is then centrifugated for 5 min at 3000 g (room temperature). The supernatant is discarded; the cell pellet is gently resuspended into 1 mL of sterile distilled water and transferred into a 2 mL microcentrifuge tube. It is centrifugated for 1 min at 3000 g, the supernatant is discarded and the cell pellet is gently resuspended into 1 mL of lithium acetate 100 mM / TE 1X (Tris 10 mM, EDTA 1 mM, pH=7.5). It is centrifugated again for 1 min, the supernatant is discarded and the cell pellet is finally resuspended into 200 μL of lithium acetate 100 mM / TE 1X. Competent yeast are ready and need to be immediately transformed; they cannot be stored for later transformation.

B.2.a.2. Transformation of S. cerevisiae *WAT21*

In a sterile 1.5 mL microcentrifuge tube, 50 μL of freshly prepared competent yeast are mixed with 50-200 μg of plasmid of interest, 10 μL of salmon sperm 10 mg/mL (preliminarily denatured by being incubated 100°C for 30min, and immediately placed on ice) and lithium acetate 100 mM / TE 1X diluted in PEG4000 (50%), for a final volume of 500 μL. This reaction mixture is incubated for 30 min at 30°C and 180 rpm, then for exactly 15 min at 42°C, and immediately placed in ice for 2 min. It is centrifugated for 2 min at 3000 g, the supernatant is carefully discarded (using a micropipette) and the cell pellet is gently resuspended into 1 mL of ultrapure sterile water. The yeast solution is centrifugated again for 1 min at 3000 g, the supernatant is carefully discarded and the cell pellet is resuspended into 300 μL of YPGA. It is incubated for 2 h at 30°C (no agitation), which allows the transformed yeast to express the selection marker gene. After the incubation, the yeast solution is centrifugated, the supernatant is discarded and the cell pellet is resuspended into 300 μL of SGI. This washing step is repeated, and the yeast are plated on selective SGI medium. Plates are incubated for 3-4 days at 30°C until the appearance of isolated colonies of about 3 mm diameter.

B.2.b. Isolation of plasmid DNA from yeast

Colony PCRs performed on yeasts generally give mitigated results. Therefore, to check the presence of the plasmid of interest within the yeast colonies resulting from the transformation, it is better to first extract the plasmid DNA from the yeast, and to perform PCR on the obtained plasmid solution. For this purpose, a colony of *S. cerevisiae* transformed with the plasmid of interest is used to inoculate 1 mL of liquid SGI medium, and incubated overnight at 30°C and 180 rpm, until a saturated yeast solution is obtained. The solution is centrifugated for 1 min at 3000g, the supernatant is discarded and the cell pellet is resuspended into 30 μL NaOH 0.05 N by vortexing. The solution is incubated for 10 min at 95°C and immediately placed on ice for 1 min. Then, 30 μL HCl 0.05 N are added to the solution and mixed by pipetting (still on ice). The solution is finally centrifugated for 1 min at 13000 g and the supernatant is used for PCR with the adequate primers.

B.2.c. Heterologous expression of P450s in *S. cerevisiae*

The methods for heterologously expressing P450s in yeasts and for preparing microsomes are adapted from Pompon *et al.* (1996). They are used to express P450 genes cloned into the pYeDP60 and pYeDP60_GW® vectors, subsequently transformed in *S. cerevisiae* strain WAT21.

B.2.c.1. Yeast culture and P450 expression

The following steps are performed in sterile conditions. A colony of yeast containing the plasmid of interest is picked and used to inoculate 10 mL of SGI medium (**A.5.b**), which are incubated for 24 h at 30°C with an orbital shaking of 180 rpm. A volume of 2 mL of this yeast preculture is used to inoculate 50 mL of YPGE medium, and incubated for 24 h at 30°C and 180 rpm. The YPGE medium contains a low concentration of glucose (5 g/L), which is used as the preferential carbon source, allowing the production of biomass and repressing the *GAL10-CYC1* promoter present into the pYeDP60 and pYeDP60_GW® vectors (**A.4.b**) and in the yeast genome (**A.3**). When the glucose is entirely consumed by the multiplying yeasts, they use the ethanol present in the YPGE medium as a secondary carbon source. After the 24 h of culture, it is assumed that the glucose is entirely consumed, which means that the *GAL10-CYC1* promoter is not repressed anymore. Then, 5 mL of sterile galactose 200 g/L are then added into the 50 mL of culture, which induces the *GAL10-CYC1* promoter, and therefore the expression of the P450 and the *AthR2* P450-reductase. The culture is incubated at 30°C and 180 rpm for 15-24 h. Finally, the produced P450 and its associated reductase can be collected by preparing yeast microsomes.

B.2.c.2. Preparation of yeast microsomes

To prevent the degradation of the proteins, the following steps are performed at 4°C. Sterile conditions are not necessary anymore. First, the yeast culture is transferred into a 50 mL Falcon tube and centrifugated at 7000 g for 10 min. The supernatant is discarded and the cell pellet is resuspended into 10 mL of cold TEK (**Table 16**). It is centrifugated again at 7000 g for 10 min, the supernatant is discarded, and the cell pellet is resuspended into 2 mL of cold TES (**Table 16**). Cold glass microbeads (Glass beads 400-625 µm – Sigma) are added into the Falcon until they surface the yeast solution. Then, the Falcon is manually yet vigorously shaken to produce microsomes by breaking the yeasts against the microbeads: shaking is performed for 5 x 1 min, and the Falcons are placed in ice between every shaking step to avoid unwanted heating. Next, 10 mL of cold TES are added into the solution containing the crushed yeasts and the microbeads. The solution is gently homogenised and, once the microbeads are decanted, the supernatant containing the fragmented yeasts is transferred into a new 50 mL Falcon tube. This is repeated 3 times, which results in a 40 mL TES solution containing fragmented yeasts, free from microbeads. The solution is centrifugated at 10000 g for 15 min to remove the bigger cell debris. The supernatant – which contains the microsomes – is filtered on paper towel to remove some last big debris; the pellet is discarded. Microsomes are precipitated by the drop-by-drop addition of 2.5 mL of cold $MgCl_2$ 1 M, and the final volume of the solution is adjusted to 50 mL by the addition of cold TES. The yeast extract is then incubated in ice for 30 min and centrifugated at 15000 g for 20 min. The supernatant is discarded. The microsome-containing pellet is resuspended into 2 mL of cold TEG (**Table 16**), transferred into a Potter-Elvehjem glass tube and gently homogenised with a PTFE pestle. The microsome solution containing the P450 of interest and the P450-reductase is ready to use; it is better used fresh for enzyme assay, but can also be aliquoted and stored at -20°C for later uses.

Table 16 Solutions used for the preparation of microsomes. The compounds in yellow have to be added to the solution just before it is used. The others can be prepared in advance and stored at 4°C.

Compound	TEK	TES	TEG
Tris HCl pH = 7.5	50 mM	50 mM	50 mM
EDTA	1 mM	1 mM	1 mM
KCl	100 mM	---	---
Sorbitol	---	0.6 M	---
Glycerol	---	---	30% (v/v)
BSA (bovine serum albumin)	---	1% (p/V)	---
β-mercaptoethanol	---	20 mM	---

B.2.d. Western-Blotting: confirmation of the presence of the P450s of interest

Once the yeast microsomes are prepared, the presence of the proteins of interest in the solution can be checked by performing an SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). In the case of His-tagged proteins, immunodetection approaches such as Western-Blot are preferred.

B.2.d.1. Polyacrylamide gel electrophoresis in denaturing conditions

SDS-PAGE is a technique that allows the separation of proteins from a complex mix according to their molecular weight. Sodium dodecyl sulfate (SDS) is a strong detergent that totally denatures proteins and masks their intrinsic charge. Therefore, by adding SDS in a polyacrylamide gel, proteins can be separated by the application of an electric field according to their molecular weight only. First, the protein samples are mixed with a loading buffer 1X (Tris HCl 125 mM pH=6.8, glycerol 20%, SDS 5%, β-mercaptoethanol 5%, dithiothreitol 400 mM, bromphenol blue 0.05%) and incubated at 98°C for 5 min. Then, they are loaded into the polyacrylamide gel composed of a concentration gel (to concentrate the protein samples) on top of a resolution gel (to separate them). The concentration gel is made of Tris HCl pH=6.8 125 mM, 5% acrylamide, 1% SDS (Sodium dodecyl sulfate), 0.1% APS (Ammonium persulfate) and 0.1% TEMED (Tetramethylethylenediamine) in a final volume of 5 mL. The resolution gel is made of Tris HCl pH=8.8 375 mM, 10% acrylamide, 1% SDS, 1% APS and 0.08% TEMED in a final volume of 10 mL. In order to determine the molecular weight of the proteins after the migration, a molecular weight marker is added in one well of the gel (ProSieveTM QuadColorTM protein marker, 4.6 kDa - 300 kDa, Lonza Rockland Inc). Finally, the gel is placed in an electrophoresis chamber, and an electrical field of 120 V is applied for about 1 h.

B.2.d.2. Transfer of the proteins to a polyvinylidene difluoride membrane

To make them accessible to antibody detection, the proteins separated in the polyacrylamide gel need to be transferred to a polyvinylidene difluoride (PVDF) membrane (AmershamTM HybondTM P 0.45 PVDF – 10600023 – GE Healthcare) that allows their immobilisation. For this purpose, a PVDF

membraned is placed in top of the gel containing the separated proteins. To maximise the contact between the membrane and the gel, they are sandwiched between two layers of filter paper, which are themselves sandwiched between pads and inserted into a support grid. The whole thing is then inserted into an electrophoresis chamber containing transfer buffer (Tris/Glycine 1X, ethanol 20%), and submitted to an electrical field of 100 V for 45 min. During this transfer, the proteins contained in the gel move onto the membrane, but maintain the exact organisation they had within the gel.

B.2.d.3. Immunodetection

The PVDF membrane on which the proteins have been transferred is washed for 5 min in a PBS/Tween20 1X buffer prepared from a stock solution of PBS 10X (87.6 g/L NaCl, 2 g/L KH_2PO_4, 14.2 g /L NaH_2PO_4, and 2 g/L KCl) diluted in distilled water and supplemented with 1% Tween20, pH=7.4. This washing buffer is then discarded, and the membrane is incubated for 1 h into a new buffer containing PBS /Tween20 1X supplemented with 5% of milk powder that will saturate all non-specific sites. After the incubation, primary antibodies directed against the His-tag (6xHistidine Epitope Tag antibody – Acris – Origene) are added to the buffer, for a final concentration of 1/2000. The membrane is incubated in the presence of the antibodies for at least 2 h, and washed three times for 10 min in the PBS /Tween20 1X buffer, which allows the elimination of all antibodies that are not fixed on the proteins immobilised onto the membrane. The membrane is subsequently placed in a new buffer containing PBS /Tween20 1X supplemented with 5% of milk powder and secondary antibodies directed against the primary antibodies (Anti-Rabbit IC (whole molecule)-Alkaline Phosphatase produced in goat – Sigma-Aldrich). To allow the allow detection of the His-tagged proteins, the secondary antibody is linked to the alkaline phosphatase, which is a reporter enzyme. The membrane is incubated in this solution for 90 min. It is then incubated two times in a buffer containing Tris 0.1 M and NaCl 0.1 M (pH=9.6). Finally, it is incubated for 15 min into a solution containing NBT (nitroblue tetrazolium chloride – 100 mg/mL) and BCIP (5-bromo-4-chloro-3-indole-phosphate – 100 mg/mL). During this incubation, the BCIP is phosphorylated by the alkaline phosphatase linked to the secondary antibodies. To increase its stability, it forms dimers of 5'-5'-dibromo-4-4'-dichloro-indigo that reduces the NBT, turning it into blue formazan. The presence of blue formazan on the membrane, visible with the naked eye, reveals the presence of His-tagged proteins on the membrane.

B.2.e. Quantification of functional P450s with the differential CO spectrum method

The differential CO spectrum method (Omura and Sato 1964a, b) relies on the ability of carbon monoxide to interact with the heme iron from P450s. This interaction results in a shift of the iron state, which modifies the P450 absorption spectrum. This can be visualised by the appearance of a characteristic peak around 450nm.

In order to determine the concentration of functional P450s, the microsomal solutions are diluted in a final volume of 2 mL of TEG: the microsomal solutions collected from the yeast producing CYP76F112, CYP76F112A, CYP76F112B and CYP76F112AB have been diluted 40 times. The solution associated to CYP76F112D have been diluted with a factor 4.28. The microsomal solutions are reduced by the addition of a spatula tip of sodium dithionite. The sample is gently homogenised by pipetting, divided into two spectrophotometer cuvettes, and a baseline is measured between 400 and 500 nm. Then, in order to saturate the P450s, compressed CO is bubbled for about 30 s into the solution contained in

one of the two cuvettes. A differential CO spectrum is finally measured between the control cuvette and the one saturated in CO, between 400 and 500 nm. The appearance of the characteristic peak at 450 nm indicates the presence of functional P450s in the solution. Even though yeasts possess 3 endogenous P450s, they are only expressed when the yeast is stressed. Therefore, it can be assumed that the peak at 450 nm only corresponds to the exogenous plant P450s. The molar concentration of functional P450s in the diluted microsomal solution (C) is determined by measuring the height of the peak between 450 and 480 nm (ΔDO) and by using the molar extinction coefficient $\varepsilon = 91$ mM^{-1} cm^{-1}. The associated formula is the following: $\Delta DO = \varepsilon$ (mM^{-1} cm^{-1}) x C (mM) x L (cm) – with L=1 cm.

B.2.f. Enzymatic assay and functional characterisation

B.2.f.1. Functional screening

A functional screening is performed with various substrates (**Chapter II**, **Table 4**) on every P450 candidate, by using fresh microsomal solutions produced within the day, prior to any freezing attempts. For this purpose, every microsomal solution is individually incubated with each substrate: in a 2 mL microcentrifuge tube, 200 µM NADPH (Sigma-Aldrich) and 100 µM substrate are mixed into NaPi buffer (Na$_2$HPO$_4$, NaH$_2$PO$_4$) 0.1 M, pH=7, for a final volume of 100 µL. For the controls, these reactions are performed without the NADPH cofactor. Then, 20 µL of microsomal solution are added to the reaction mixture and incubated for 30 min in a ThermoMixer® (Eppendorf) at 27°C and with an agitation of 600 rpm. After the incubation, 50 µL of acetonitrile supplemented by 1% HCl are added to the mix, which stops the reaction. The reaction mixture is centrifugated for 10 min at maximum speed, the supernatant is collected, filtered at 0.2 µm and transferred into a chromatography vial for subsequent UHPLC-MS analyses (**B.3.b**).

B.2.f.2. Determination of the optimal conditions: temperature and pH

The optimal temperature of CYP76F112 and CYP82J18 are determined by incubating the microsomes at temperatures ranging from 20 to 45°C. For every incubation, 200 µM NADPH and 50 µM substrate (DMS for CYP76F112, auraptene for CYP82J18) are mixed into NaPi buffer 0.1M, pH=7, for a final volume of 90 µL. Then, 10 µL of microsomal solution (diluted 5 times for CYP76F112) are added to the reaction mixture and incubated for exactly 5 min (CYP76F112) or 30 min (CYP82J18) at 600 rpm, with a given temperature comprised between 20 and 45°C. The reactions are stopped by the addition of acetonitrile HCl 1%. The reaction mixtures are centrifugated for 10 min at maximum speed, the supernatants are collected, filtered at 0.2 µm and transferred into a chromatography vial for subsequent UHPLC-MS analyses (**B.3.b**). Incubations are performed in triplicates.

The optimal pH is determined by incubating the microsomes in a pH range of 4 to 10. For every incubation, 200 µM NADPH and 50 µM substrate (DMS for CYP76F112, auraptene for CYP82J18) are mixed into one of the buffers presented in **Table 17**, for a final volume of 90 µL. Then, 10 µL of microsomal solution (diluted 5 times for CYP76F112) are added to the reaction mixture and incubated for exactly 5 min (CYP76F112) or 30 min (CYP82J18) at 27°C and 600rpm. The reactions are stopped by the addition of acetonitrile HCl 1%. The reaction mixtures are centrifugated for 10 min at maximum speed, the supernatants are collected, filtered at 0.2 µm and transferred into a chromatography vial for subsequent UHPLC-MS analyses (**B.3.b**). Incubations are performed in triplicates.

Table 17 Buffers used to determine the optimal pH.

Nature of the buffer	pH of the buffer				
Sodium acetate 0.1 M	4	4.6	5	5.6	---
NaPi 0.1 M	5.9	6.5	7	7.6	8.1
Tris HCl 0.1 M	8	8.5	9	---	---
Glycine NaOH 0.1 M	9	9.5	10	---	---

B.2.f.3. Determination of the kinetic parameters

CYP76F112 and its associated mutants (A, B, AB and D) have an apparent K_m value lower than 0.1 µM, which is below the detection limit of the UHPLC-MS used in the LAE. To determine their kinetic parameters, a specific protocol allowing to test very low DMS concentrations has been developed.

The incubations are performed in 2mL microcentrifuge tubes, in a final volume of 1 mL. For every incubation, 990 µM of reaction mixture containing DMS and an excess of NADPH diluted into NaPi buffer 0.1 M, pH=7 is prepared and pre-heated at 27°C. Then, 10 µM of diluted microsomal solution, pre-heated at 27°C, are added to the mixture and incubated at 27°C and 600 rpm for a short but very precise time. The incubation duration and the quantity of functional P450s in the solution are adapted to avoid the metabolization of more than 50% of the initial DMS with the lowest tested DMS concentration. The modalities of the incubations for every P450s are presented in **Table 18**. At the end of the incubation time, 300 µL acetonitrile HCl are added to the mix to stop the reactions. For every modality, 5 to 15 incubations of 1 mL are performed. This is repeated in duplicates – making a total of 10 to 30 incubations of 1 mL for every P450 and modality (**Supp. Table 4**, **Supp. Table 12**).

Table 18 Modalities of the 1 mL incubations performed to determine the P450 kinetic parameters.

Enzyme	Concentration in the final 1 mL reaction mixture		Dilution of the microsomal solution	Quantity of P450s added in the final 1mL reaction mixture	Incubation time
	NADPH	DMS			
CYP76F112	20 µM	10 to 300 nM	20	0.54 pmol	1 min
CYP76F112-A	20 µM	10 to 300 nM	5	2.08 pmol	2 min
CYP76F112-B	20 µM	10 to 300 nM	10	0.322 pmol	45 s
CYP76F112-AB	20 µM	10 to 300 nM	10	0.266 pmol	2 min
CYP76F112-D	20 µM	10 to 1000 nM	5	0.035 pmol	3 min

Then, the 5 to 15 incubations performed for each modality and duplicate are pooled together, and they are concentrated 50 to 150 times. For this purpose, 0.2 volume of ethyl acetate is added to the pulled solution and vigorously mixed by vortexing (maximum speed) for 45s – during which the phenolic compounds are extracted from the NaPi buffer and transferred into ethyl acetate. The

solution is centrifugated for 1 min at 5000 g. As a result, an organic phase made of ethyl acetate and containing the furanocoumarins is distinctly visible on top of the NaPi buffer. This ethyl acetate phase is collected and transferred into new (but multiple) 2 mL microcentrifuge tubes (1.5 mL in each tube). A second extraction is performed with the addition of 0.2 volumes of ethyl acetate into the initial reaction mix. It is vortexed and centrifugated again, before the ethyl acetate phase is collected and transferred into new 2 mL microcentrifuge tubes. The tubes resulting from the first and second extractions are entirely evaporated (Concentrator plus – Eppendorf – 30°C, V-AL, for about 2-3 h), and the furanocoumarin pellets are resuspended into 200 µL methanol (100%). The tubes are vortexed for 5min (maximum speed) to ensure the correct resuspension of the furanocoumarins; and the methanol solutions corresponding to one modality (the 5 to 15 initial incubations) are all pooled into a single 2 mL microcentrifuge tube. This tube is entirely evaporated again (30°C, V-AL, for about 2-3h), and the furanocoumarin pellet is finally resuspended into 100 µL methanol (100%). This final 100 µL methanol solution therefore contains all furanocoumarins that were present into the initial 5 to 15 incubations of 1 mL performed with the same conditions – which correspond to a concentration 50 to 150 times. It is filtered at 0.2 µm and transferred into a chromatography vial for subsequent UHPLC-MS analysis.

The marmesin present in the concentrated solution is quantified by using the peak area obtained with the MS data. To convert this peak area into a quantity of marmesin (mol), a calibration line obtained by measuring the peak area of known concentration of standard marmesin is used ($y = 1.41E^{14} x$). The quantity of marmesin measured in the concentrated solution is then divided by the concentration factor (50 to 150 times) to obtain the quantity of marmesin formed in 1 initial reaction mixture of 1 mL. The specific activity of the P450 is finally calculated, and expressed in mol of marmesin formed per minute per mol of functional P450.

To determine the kinetic parameters of the enzymes, the specific activity of the P450s are entered into the SigmaPlot software (**A.6.b.1**). This software allows to plot a Michaelis-Menten model curve (specific activity as a function of the initial substrate concentration), to determine the Michaelis-Menten equation fitting the experimental data and the associated apparent K_m and K_{cat} values.

B.3. Metabolic analyses

B.3.a. Extraction of phenolic compounds from plant grinded sample

Up to 100 mg of plant grinded tissues (**B.1.a**) are transferred into a 2 mL microcentrifuge tube and mixed in 800 µL of methanol 80%. The tube is vortexed for 1 min (maximum speed) and placed into an ultrasonic bath (Elmasonic S70 – Elma) at 37 kHz for 10 min. Ultrasounds improve the plant cell lysis, increasing the efficiency of the extraction. The solution is centrifugated for 30 min at 15000 g. The supernatant is collected and transferred into a new 2 mL microcentrifuge tube. A second extraction is performed on the cell pellet with the addition of another 800 µL of methanol 80%, followed by vortexing, ultrasonic bath and centrifugation. The supernatant is collected and pooled with the first one – for a volume of 1.6 mL of solution. This solution is evaporated overnight (Concentrator plus – Eppendorf), and the pellet containing the phenolic compounds is resuspended in a final volume of 100 µL of methanol 80%. The phenolic solution is filtered at 0.2 µm and transferred into a chromatography vial for subsequent UHPLC-MS analyses.

B.3.b. UHPLC-MS analyses

The metabolomic analyses of plant samples and microsomal incubations are performed using the Ultra Performance Liquid Chromatography (UHPLC) Nexera2 (Shimadzu, Kyoto, Japan), equipped with a UV detector SPDM20A (Shimadzu) coupled to a simple quadrupole mass spectrometer LCMS2020 (Shimadzu). Phenolic compounds are separated into a reverse-phase chromatography column C18 (Kinetex XB-C18 150 × 2.10 mm length; 2.6 µm particles – Phenomenex) surmounted by a pre-column (1290 Infinity, Agilent Technologies). These two parts are placed into a chamber at 40°C. The elution solvents are a solution of methanol 99.9% / formic acid 0.1% or acetonitrile 99.9% / formic acid 0.1%, and a solution of ultrapure water 99.9% / formic acid 0.1%. The mobile phase gradient is shown on **Figure 86**. The mobile phase flow is of 0.2 mL/min; the injection volume of 10 µL. The detection of the compounds is performed by the UV detector for wavelengths ranging from 200 to 600 nm. Once separated, the compounds are analysed by mass spectrometry, using the SIM mode (Selected Ion Monitoring), which allows to monitor specific compounds and increase the accuracy of the method. The mass spectrometer is connected to the UHPLC through a dual ion source (DUIS) set on positive mode, which combines electrospray ionisation (ESI) and atmospheric pressure chemical ionisation (APCI). The electrospray ionisation receives a voltage of 4.5 kV. The temperature of the heating block, the entry and the desolvation line of the mass spectrometer are respectively set on 400°C, 350°C and 250°C. Data are recorded and analysed on the LabSolution software (**A.6.b.1**). The phenolic compounds present in the samples are identified by direct comparison of their retention time and m/z ratio with those of standard molecules.
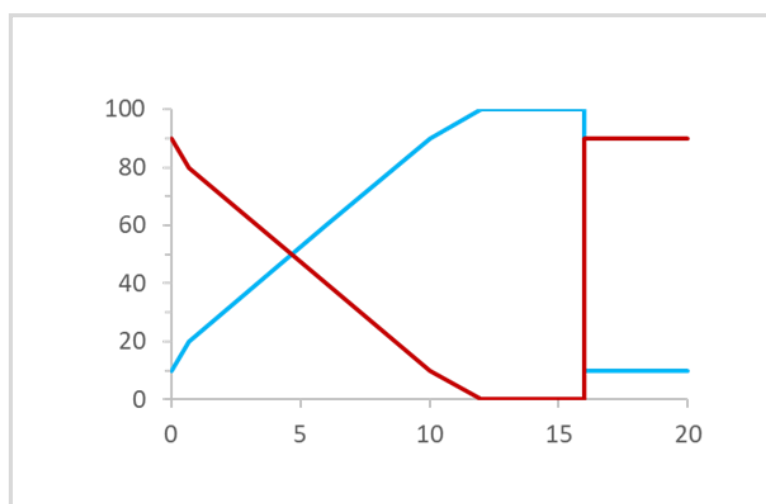
Figure 86 Mobile phase gradient for UHPLC-MS analyses. The gradient of the solution of methanol or acetonitrile is in blue. The gradient of the water solution is in red.

B.3.c. Orbitrap-IDX analyses

The use of a liquid chromatograph apparatus (Thermo Vanquish, Thermo Scientific) coupled to a tandem mass spectrometer (Orbitrap-IDX, Thermo Scientific) allows for the determination of the spectroscopic fragmentation patterns of the reaction product of CYP76F112 and a standard of marmesin. Compounds are separated into a reverse-phase chromatography column C18 (Kinetex XB-C18 150 × 2.1 mm length; 2.6 µm particles – Phenomenex). The elution solvents are a solution of methanol 99.9% / formic acid 0.1% and a solution of ultrapure water 99.9% / formic acid 0.1%. The mobile phase gradient is shown on **Figure 87**. The injection volume is of 10 µL. The tandem mass spectrometer includes a linear ion trap that allows the discrimination of compounds according to their

m/z ratio. It is linked to the liquid chromatography system through an ion source ESI set on positive mode, which receives a voltage of 4.5 kV and is heated to 300°C. The flowrates of the outer-coaxial gas, auxiliary gas and barrier gas are respectively set on 40, 10 and 10 arbitrary units/min. In the ion source entry, the tension applied to the transfer capillary, the split and the front lens are respectively set on 36, -44 and -3.5 V. The analysis covers m/z ratios ranging from 120 to 1000, but can be performed on specific m/z ratios. Data are recorded and analysed on the Xcalibur™ (software (**A.6.b.1**).
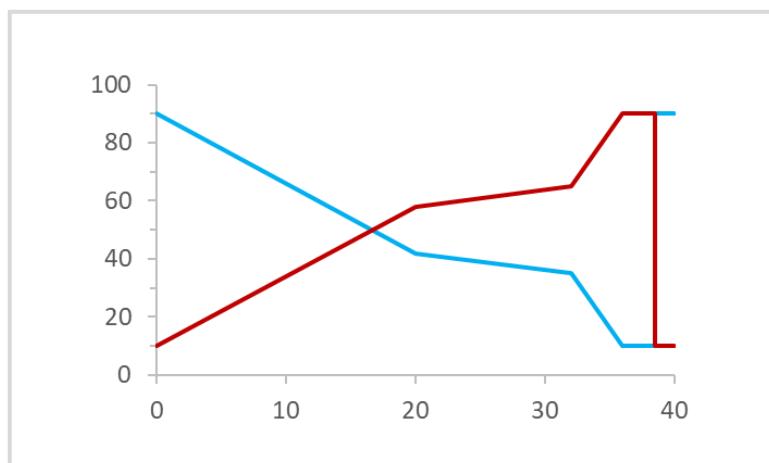
Figure 87 Mobile phase gradient for Orbitrap-IDX analyses.

B.4. The GoldenBraid cloning technique

The GB technology is a standardised multi-cloning system that consists in repeatedly assembling DNA sequences from entry vectors into a single destination vector (**Chapter V**, **Figure 67**). It relies on a succession of restriction-ligation reactions using type IIS restrictions enzyme (Sarrion-Perdigones et al. 2011, 2013, 2014). The enzymes used in this study are *Bsa*I and *Bsm*BI (**Chapter V**, **Figure 68**); the pUPD and pDGB vectors, which serves as entry and destination vectors, are described in (**A.4.c**).

B.4.a. Domestication of the genes of interest

The first step of the GB methodology consists in domesticating the GB parts (promoters, coding sequences, terminators) to the GB rules. All GB parts used in this study were already domesticated, except for the *CYP76F112* and *KanaR* genes. The domestication of the *CYP76F112* and *KanaR* genes consists in removing their internal *Bsa*I and *Bsm*BI recognition site, and flanking them with "GB extensions" of 16 and 19 nucleotides – respectively, GCGC**CGTCTC**GCTCGA upstream the start codon and GCTTCGAGC**GAGACG**GCGC downstream the stop codon. These extensions contain a *Bsm*BI recognition site (in bold), and the sequences corresponding to the 4-nucleotide overhangs required for the subsequent assemblies.

B.4.a.1. Domestication of CYP76F112

The *CYP76F112* gene contains one *Bsm*BI recognition site (CGTCTC, nucleotides 284 to 290), and no *Bsa*I recognition site (**Supp. Figure 1**). This single *Bsm*BI internal recognition site is removed by replacing one of its codons (GTC, coding for a valine) by a synonymous codon (GTA) which has a similar usage frequency in the tomato (Sablok et al. 2013) – in which CYP76F112 was to be expressed. The removal of the internal restriction site and the addition of the flanking GB extensions is done through an overlapping PCR which methodology is illustrated in **Figure 88**. For this purpose, two pairs of primers

are used: the first pair consists in two external primers that includes the GB extensions to be added on either side of the gene. The second pair of primers is internal, and incorporates a nucleotide mismatch to mutate the GTC from the *BsmB*I recognition site into a GTA (**Supp. Figure 10**).

First, 2 PCRs are individually performed. Their respective reaction mixtures are composed of the PrimeSTAR® Max premix (**B.1.e.1**) at a final 1x concentration, 100 ng of matrix DNA (purified pCR™8/GW/TOPO™ vector containing the *CYP76F112* gene), 0.2 µM of primers, and ultrapure water for a final volume of 15 µL. One PCR is performed with the primers 1 and 2 from **Supp. Figure 10**; it allows the amplification of the beginning of the *CYP76F112* gene with the addition of a GB extension upstream the start codon, and the removal of the *BsmB*I internal recognition. The second PCR is performed with the primers 3 and 4 from **Supp. Figure 10**; it allows the amplification of the end of the *CYP76F112* gene with the addition of a GB extension downstream the stop codon, and the removal of the *BsmB*I internal recognition (**Figure 88**). For
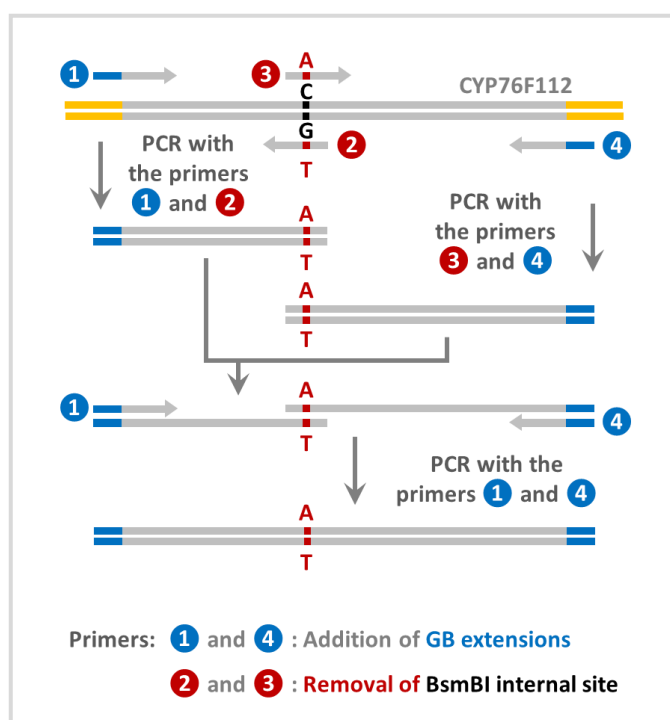
Figure 88 Domestication of *CYP76F112*. The CYP76F112 sequence is in grey, initially inserted into the pCR™8/GW/TOPO™ vector (yellow). The nucleotide to be modified is in black. The GB extensions are in blue.

both PCRs, the conditions are the following: an initial denaturation at 98°C for 1 min is followed by 5 cycles composed of a denaturation step (98°C for 10 s), an annealing step (50°C for 15 s – allowing the aspecific hybridisation of the primers despite the nucleotide mismatches) and an elongation step (72°C for 15 s). This is followed by 25 cycles composed of a denaturation step (98°C for 10 s), an annealing step (60°C for 15 s – specific hybridisation) and an elongation step (72°C for 15 s); and a final elongation step at 72°C for 3min. The PCR-amplified fragments, respectively corresponding to the two halves of the domesticated *CYP76F112* gene, are isolated on an agarose gel and extracted (**B.1.f**).

Then, the two halves of the domesticated *CYP76F112* gene are used to perform an overlapping PCR: the reaction mixture is composed of the PrimeSTAR® Max premix at a final 1x concentration, 50 ng of each of the two halves of the domesticated *CYP76F112*, 0.2 µM of primers 1 and 4 from **Supp. Figure 10**, and ultrapure water for a final volume of 15 µL. The conditions of the PCR are the following: an initial denaturation at 98°C for 1 min, 25 cycles composed of a denaturation step (98°C for 10 s), an annealing step (55°C for 15 s) and an elongation step (72°C for 20 s), followed by a final elongation step at 72°C for 3 min. The resulting PCR-amplified fragment corresponds to the domesticated *CYP76F112* gene. It is then cloned into the pCR™8/GW/TOPO™ empty vector (**B.1.h.1**), inserted and amplified into *E. Coli* (**B.1.j** and **B.1.k**). Finally, a sequencing (**B.1.m**) confirms the addition of the GB extensions, the removal of the internal recognition site, and the absence of unwanted mutations.

B.4.a.2. Domestication of the KanaR gene

The plant *KanaR* gene used in this study is amplified from the plasmid pICSL11024 (**A.4.f**). As this gene does not contain neither *BsmB*I nor *Bsa*I recognition site, its domestication only consists in adding the GB extensions. This is done by amplifying it with the PrimeSTAR® Max DNA Polymerase (**B.1.e.1** – standard protocol), by using specific primers (**Supp. Figure 10**). The PCR-amplified fragment is then cloned into the pCR™8/GW/TOPO™ empty vector (**B.1.h.1**), inserted and amplified into *E. Coli* (**B.1.j** and **B.1.k**). A sequencing (**B.1.m**) confirms its correct domestication.

B.4.b. Cloning into the pUPD vector

The second step consists in assembling the newly domesticated *CYP76F112* and *KanaR* genes – but also the *CYP71AJ3* gene which was previously domesticated by G. Galati (2019) – into TUs. To start the assembly, these three genes first need to be cloned into the empty pUPD plasmid (**A.4.c.1**), which is done by performing a digestion with the *BsmB*I enzyme coupled with a ligation with the T4 DNA ligase. This restriction-ligation corresponds to a one-step reaction performed in a single tube. The reaction mixture is composed of 1µL T4 DNA ligase 5U/µL (Thermoscientific), 1 µL T4 DNA ligase buffer 10x (Thermoscientific), 1 µL *BsmB*I 10 U/µL (Thermoscientific), 100 ng of pCR™8/GW/TOPO™ vector containing the domesticated gene of interest, and 100 ng of empty pUPD plasmid diluted in a final volume of 10 µL. The reaction mixture is incubated for 25 cycles composed of a digestion step (37°C for 2 min) and a ligation step (16°C for 5 min). During the digestion steps, *BsmB*I cleaves the entry and destination plasmids as illustrated on **Figure 89**. By doing so, it generates the overhangs that allow the
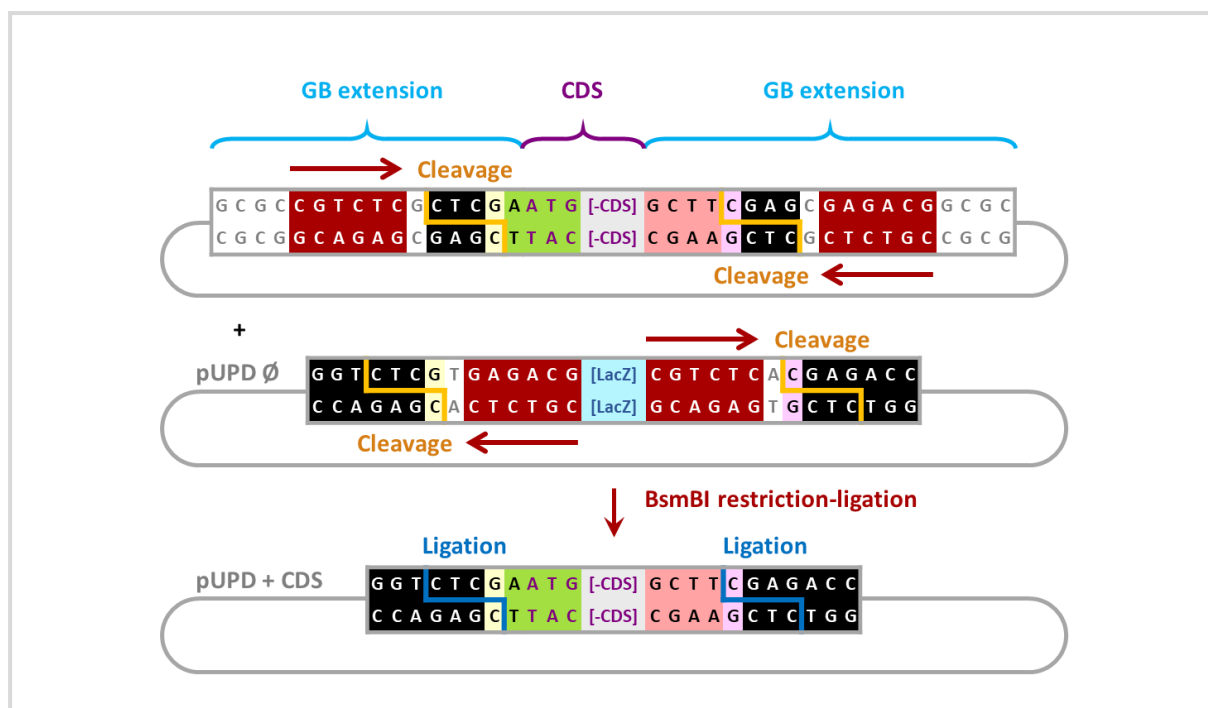
Figure 89 Cloning of a domesticated coding sequence (CDS) into the pUPD vector, via *BsmB*I restriction-ligation. The GB extensions and the start codon of the CDS are detailed. The *BsmB*I and *Bsa*I (partial) recognition sites are respectively in red and black. The other colors refer to the complementary overhangs, as shown in **Chapter V**, **Figure 70**.

insertion of the gene of interest into the pUPD plasmid, which can then be ligated by the T4 DNA ligase during the ligation steps. During the reaction, the *BsmB*I recognition sites are destroyed; the resulting recombinant plasmids include the genes of interest flanked by *Bsa*I recognition sites. They are introduced into *E. coli* for further amplification (**B.1.j** and **B.1.k**). The presence of the genes of interest into the pUPD vectors is confirmed by PCR (**B.1.e**) and digestions using *Bsa*I and/or other restriction enzymes presented in **B.1.g**.

B.4.c. Assembly of simple transcriptional units

Once inserted into pUPD plasmids, the coding sequence of every gene of interest (CYP76F112, CYP71AJ3 and KanaR) is combined with a 35S promoter and a tNOS terminator (**A.4.c.1**) into a single PDGB1-α plasmid (**A.4.c.2**) to form a simple transcriptional unit. The TU associated to CYP76F112 is assembled into a PDGB1-α1 plasmid while the TUs associated to CYP71AJ3 and KanaR (**Supp. Figure 13**) are assembled into PDGB1-α2 plasmids. Similarly to the previous step, this is done by performing a digestion with the *Bsa*I enzyme coupled with a ligation with the T4 DNA ligase. The reaction mixture is composed of 1μL T4 DNA ligase 5U/μL, 1 μL T4 DNA ligase buffer 10x, 1 μL *Bsa*I 10 U/μL (Thermoscientific), 100 ng of the pUPD plasmid containing the gene of interest, 100 ng of the pUPD plasmid containing the 35S, 100 ng of the pUPD plasmid containing the tNOS, and 100 ng of empty PDGB1-α plasmid (PDGB1-α1 or -α2) diluted in a final volume of 10 μL. The reaction mixture is incubated for 25 cycles composed of a digestion step (37°C for 2 min) and a ligation step (16°C for 5 min). During the digestion steps, *Bsa*I cleaves the entry and destination plasmids as illustrated on **Figure 90**, allowing the ordered assembly of the different GB parts into a functional TU: the coding sequence is assembled right after the 35S, with a single nucleotide between the 35S and the start codon of the coding sequence. The tNOS is assembled after the coding sequence, with only 4 nucleotides between them. During this multipartite assembly, the *Bsa*I recognition sites disappear, but the TU remains flanked by the *BsmB*I recognition sites from the α-level plasmid in which it is inserted (**Figure 90**). Once again, the resulting recombinant plasmids are introduced into *E. coli*, amplified, and the presence of the TUs into the α-level plasmids is confirmed by PCR and *BsmB*I digestions.

B.4.d. Repeated assembly of multiple transcriptional units

The next step consists in repeatedly combining the different TUs to create the 5-TUs plasmid harbouring the psoralen biosynthesis pathway.

B.4.d.1. Assembly of two transcriptional units: CYP76F112 and CYP71AJ3

First, the PDGB1-α1 plasmid containing the TU [CYP76F112] is assembled with the PDGB1-α2 plasmid harbouring the TU [CYP71AJ3] into an empty PDGB1-Ω2 plasmid, using the *BsmB*I restriction enzyme and the T4 DNA ligase. The reaction mixture is composed of 1μL T4 DNA ligase 5U/μL, 1 μL T4 DNA ligase buffer 10x, 1 μL *Bsm*BI 10 U/μL, 100 ng of the PDGB1-α1 plasmid containing the TU [CYP76F112], 100 ng of the PDGB1-α2 plasmid containing the TU [CYP71AJ3], and 100 ng of empty PDGB1-Ω2 plasmid diluted in a final volume of 10 μL. The reaction mixture is incubated for 25 cycles composed of a digestion step (37°C for 2 min) and a ligation step (16°C for 5 min). During the digestion steps, *Bsm*BI cleaves the plasmids as illustrated on **Figure 91**, allowing the ordered assembly of two TUs. In the resulting PDGB1-Ω2 plasmid, the TUs [CYP76F112] and [CYP71J3] are 12 nucleotides apart, and

flanked with *Bsa*I recognition site. Once again, the resulting recombinant plasmid is introduced into *E. coli*, amplified, and the presence of the 2 TUs is confirmed by PCR and *Bsa*I digestions.
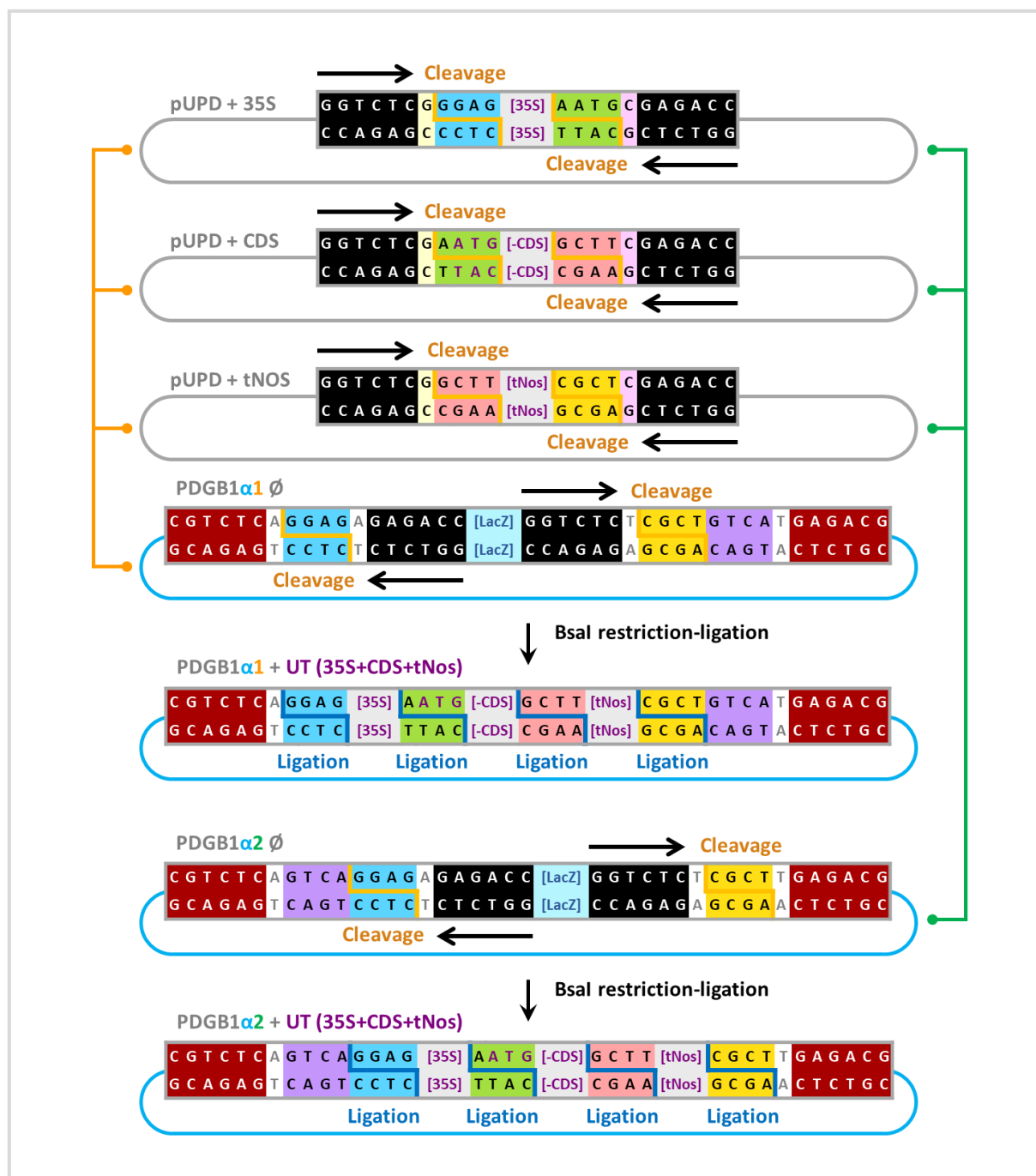
Figure 90 Assembly of a 35S, a coding sequence (CDS) and a tNOS into an α1 or α2 plasmid. The start codon of the CDS is detailed. The *BsmB*I and *Bsa*I recognition sites are respectively in red and black. The other colors refer to the complementary overhangs, as shown in **Chapter V**, **Figure 70**.
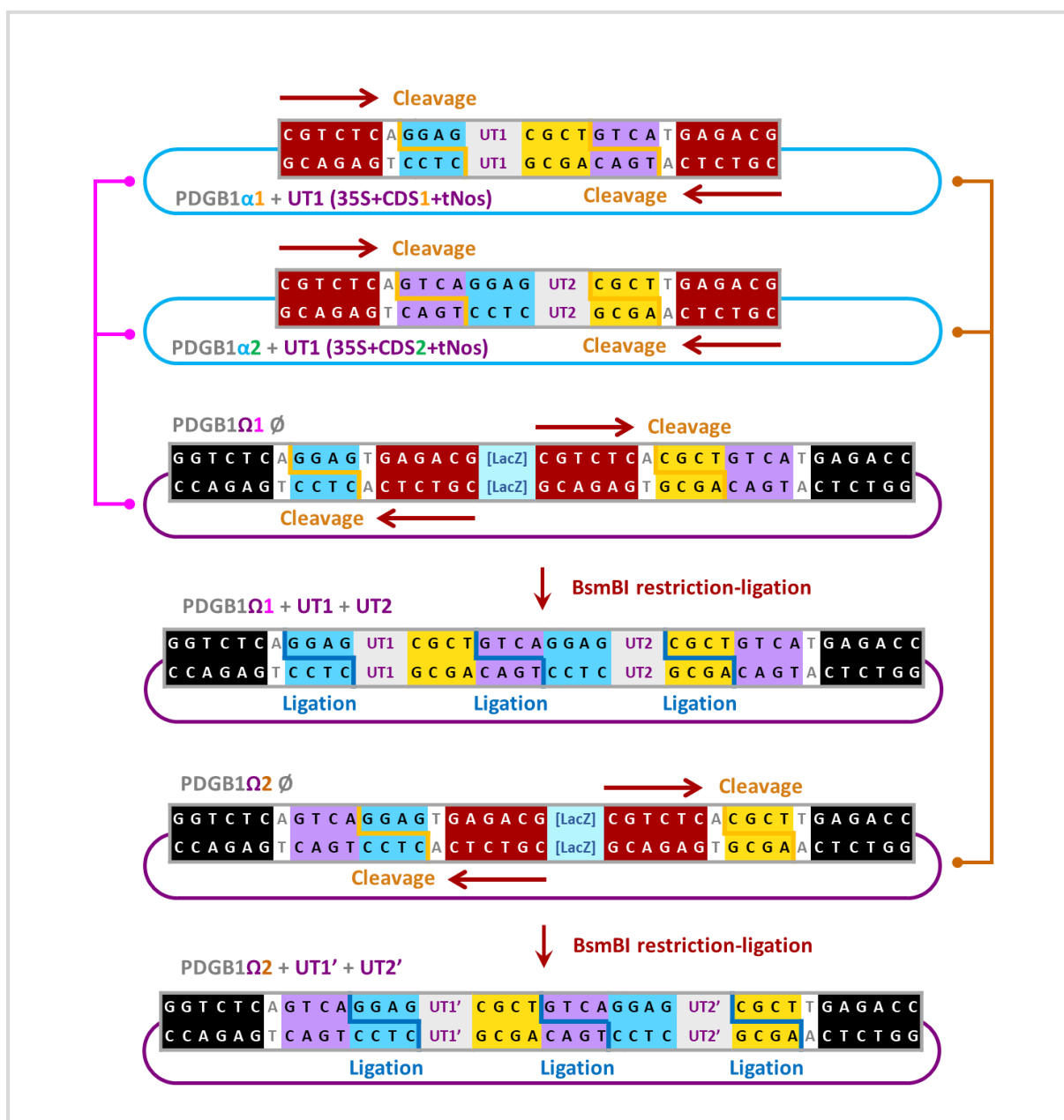
Figure 91 Assembly of 2 UTs into an Ω-level plasmid. The *BsmB*I and *Bsa*I recognition sites are respectively in red and black. The other colors refer to the complementary overhangs, as shown in **Chapter V**, **Figure 70**.

B.4.d.2. Assembly of four transcriptional units: PsDiox, PsPT1, CYP76F112 and CYP71AJ3

During his previous work, G. Galati used a methodology similar to that shown in **Figure 90** to generate TUs associated to *PsDiox* and *PsPT1* into PDGB1-α1 and PDGB1-α2 plasmids, respectively. Then, he combined these two TUs into a PDGB1-Ω1 plasmid, as shown on **Figure 91**. Therefore, his PDGB1-Ω1 plasmid containing the TUs [PsDiox+PsPT1] can be combined with the newly generated PDGB1-Ω2 plasmid harbouring the TUs [CYP76F112+CYP71AJ3]. The reaction mixture is composed of 1μL T4 DNA ligase 5U/μL, 1 μL T4 DNA ligase buffer 10x, 1 μL *Bsa*I 10 U/μL, 100 ng of the PDGB1-Ω1 plasmid containing the TUs [PsDiox+PsPT1], 100 ng of the PDGB1-Ω2 plasmid harbouring the TUs

[CYP76F112+CYP71AJ3], and 100 ng of empty PDGB1-α1 plasmid diluted in a final volume of 10 µL. The reaction mixture is incubated for 25 cycles composed of a digestion step (37°C for 2 min) and a ligation step (16°C for 5 min). During the digestion steps, *Bsa*I cleaves the entry and destination plasmids as illustrated on **Figure 92**, allowing the ordered assembly of four TUs [PsDiox+PsPT1+CYP76F112+ CYP71AJ3], with only 12 nucleotides between every TU. This composite part of 4 TUs is flanked with *Bsm*BI recognition site from the PDGB1-α1 plasmid. Once again, the resulting recombinant plasmid is introduced into *E. coli*, amplified, and the presence of the 4 TUs is confirmed by PCR and *Bsm*BI digestions.
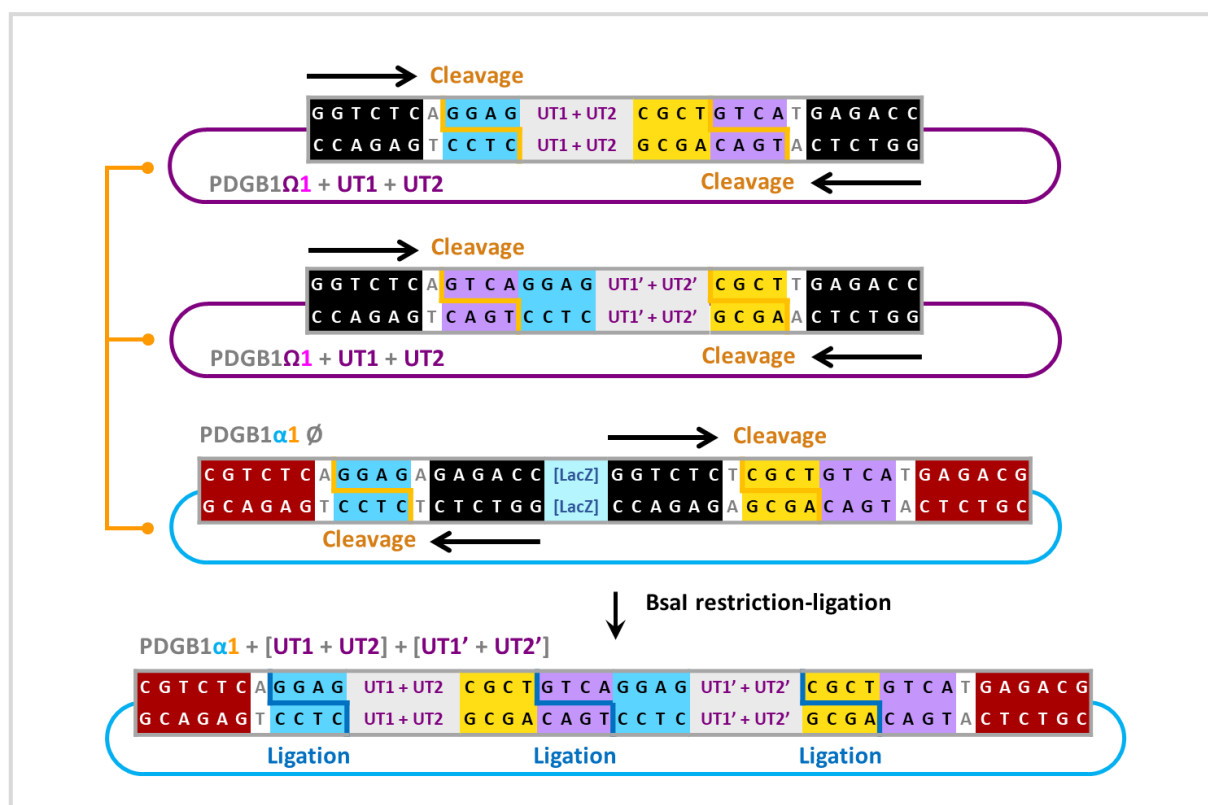
Figure 92 Assembly of 4 UTs into an α-level plasmid. The *Bsm*BI and *Bsa*I recognition sites are respectively in red and black. The other colors refer to the complementary overhangs (**Chapter V**, **Figure 70**).

B.4.d.3. Assembly of 5 transcriptional units to construct the final 5-TUs plasmid

Lastly, the PDGB1-α1 plasmid containing the 4 TUs is combined with the PDGB1-α2 plasmid containing the TU [KanaR] into a PDGB1-Ω1 plasmid. The methodology is the same than for the previous assembly of the α-level plasmids into Ω-level plasmid described in **B.4.d.1** and **Figure 91** – except than the reaction mixture is composed of 1µL T4 DNA ligase 5U/µL, 1 µL T4 DNA ligase buffer 10x, 1 µL *Bsm*BI 10 U/µL, 100 ng of the PDGB1-α1 plasmid containing the 4 TUs, 100 ng of the PDGB1-α2 plasmid containing the TU [KanaR], and 100 ng of empty PDGB1-Ω1 plasmid diluted in a final volume of 10 µL. The resulting plasmid contains the 5 TUs associated to the *PsDiox*, *PsPT1*, *CYP76F112*, *CYP71AJ3* and *KanaR* genes. It is introduced into *E. coli*, amplified, and the presence of the 5 TUs is confirmed by multiple PCRs, by digestion, and by sequencing (**Chapter V**, **Figure 71**).

B.5. Tomato stable transformation and regeneration

This protocol has been developed to perform stable transformation of tomatoes from the WVa106 line (**A.1.b**). It was provided by Pr Michel Hernould (Laboratoire Biologie du fruit et pathologie, UMR 1332, INRAE – Bordeaux, France). The following steps are performed in sterile conditions.

B.5.a. Preparation of the tomatoes to be transformed

B.5.a.1. Sterilisation of the tomato seeds

WVa106 seeds are sterilised by being immerged and gently agitated for 15 min in 200 mL sterile water supplemented with 2.5% bleach. They are washed 3 times with sterile water. Then, the sterile seeds can either be dried and stored for months, or immediately used and left to germinate.

B.5.a.2. Germination of the sterile tomato seeds

Sterile WVa106 seeds are delicately transferred into sterile pots containing germination medium (**A.5.c**). The seeds are place on top of the medium, they do not need to be buried. Up to 10-15 seeds can be transferred into every sterile pot. The pots are placed into a box that is half-closed and maintained for about 7 days in a growth chamber (Fitoclima 1200 – Aralab) at 24°C under a 16h light/8h dark cycle (150 µmol m^{-2} s^{-1}).

B.5.a.3. Preparation of the cotyledons

The WVa106 seeds are ready to be used when they have germinated and produced seedlings whose cotyledons are several millimetres long and fully expanded. When it is the case, the cotyledons are excised from the seedling by sectioning the base of the hypocotyls with a scalpel. The manipulation of the cotyledons has to be done very carefully and with great delicacy to avoid damaging the tissues (precise and sharp cuts, no tearing); but it also has to be done quickly to avoid the tissues to dry-out.

The cotyledons are transferred on a sterile hard surface (petri dish) and sectioned into 3 parts as shown on **Figure 93**. The basis and the tips of the cotyledons are cut off to generate a clean injury that will ease the adsorption of the bacterial suspension. The 3 fragments from the middle part of the cotyledons are gently transferred onto petri dishes (15 cotyledons per petri dish) containing preculture medium (**A.5.c**), with their adaxial surface on contact with the medium. The fragments are cultured for 24h in the dark, at 24°C before being used for transformations.
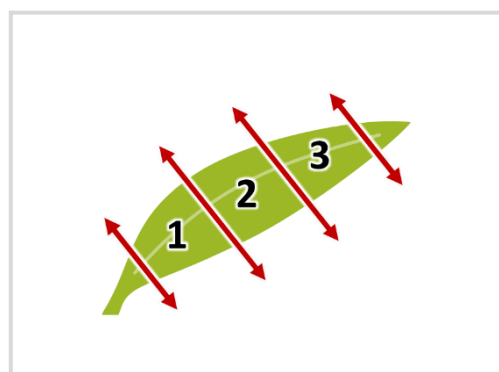
Figure 93 Sectioning of the cotyledons.

B.5.b. Preparation of the *Agrobacterium* suspension to transform the tomatoes

The plant transformation is mediated by *Agrobacterium tumefaciens.* Therefore, *A. tumefaciens* have to be preliminary transformed with the appropriate plasmid(s). They also have to be cultured and prepared in parallel to the cotyledons, so that the fragments of cotyledons are ready to be transformed when the agrobacteria are ready to transform them.

B.5.b.1. Co-transformation of the Agrobacterium

Two lines of transgenic tomatoes had to be created: one with the 5-TU plasmid and the second with the KanaR plasmid, which both have a PDGB backbone. The PDGB vectors have been designed for *Agrobacterium*-mediated plant transformations, but as they are based on the pGreenII vector, they have to be transformed in an *Agrobacterium* strain that also carry the helper plasmid pSoup (Hellens et al. 2000; Sarrion-Perdigones et al. 2014). Therefore, two individual co-electroporation of the *Agrobacterium tumefaciens* EHA105 strain are performed (**A.2.c** and **B.1.j**): one with the pSoup and the 5-TUs plasmids, the second with the pSOUP and the KanaR plasmids.

B.5.b.2. Preparation of the Agrobacterium *suspension*

Three days before tomato transformation, an isolated colony of the desired *A. tumefaciens* line is plated on solid LB medium supplemented with the appropriate antibiotics (**A.5.a**), and incubated for 48 h at 28°C. Then, one day before tomato transformation, a pipette tip of *A. tumefaciens* is picked and used to inoculate 10 mL of liquid LB medium, supplemented with the appropriate antibiotics. This preculture is incubated overnight at 28°C with an orbital shaking of 200 rpm.

The day of the tomato transformation, 5 mL of the *A. tumefaciens* preculture are used to inoculate 50 mL of liquid LB medium supplemented with the appropriate antibiotics. The culture is incubated at 28°C and 200 rpm until the OD_{600nm} reaches 0.4 to 0.6. It is then centrifugated for 10 min at 7000 g. The supernatant is discarded and the cell pellet is gently resuspended into 40 mL of sterile suspension medium (**A.5.c**). The OD_{600nm} of the *Agrobacterium* suspension is measured, and the solution is diluted again into sterile suspension medium for a final volume of 30 mL with an OD_{600nm} of 0.08. The *Agrobacterium* suspension is ready and have to be immediately used for plant transformation.

B.5.c. Transfection of the cotyledon fragments

The fragments of cotyledons that have been cultured in the dark for 24h are gently dipped into 30 mL of *Agrobacterium* suspension contained in a 50 mL Falcon. The Falcon is wrapped with aluminium foil and incubated for 30 min at room temperature, with a soft agitation every 5 min. After the incubation, the fragments of cotyledons are carefully transferred and blotted dry onto sterilised paper towel. They are then transferred onto petri dishes containing co-culture medium (**A.5.c**), with the adaxial surface of the cotyledon on contact with the medium. They are subsequently cultured for 48 h in the dark, at 24°C, which corresponds to the actual transfection step during which the T-DNA from the PDGB plasmids is randomly inserted into the tomato genome.

B.5.d. Regeneration and selection of the transgenic cotyledons

After the 48 h of co-culture, the fragments of cotyledons are delicately transferred onto petri dishes containing regeneration medium (**A.5.c** – about 15 fragment per petri dish), with the adaxial surface of the cotyledon on contact with the medium. The regeneration medium purpose is to induce the formation of callus and shoots. It also contains kanamycin to ensure a selection of the transformed plant cells, and timentin to inhibit the growth of the *Agrobacterium*. The fragments of cotyledons are maintained in a growth chamber at 24°C under a 16h light/8h dark cycle (150 µmol m^{-2} s^{-1}). Every 7 days, they are subcultured onto fresh regeneration medium and maintained under the same light/temperature conditions. After a few weeks, some cotyledon fragments should develop into calli; the fragments that are not transformed and therefore do not contain the kanamycin resistance gene turn yellowish and die. Then, the calli start to produce green shoot primordia that soon turn into plantlets.

B.5.e. Rooting of the transgenic plantlets

When green shoots are developed enough to form a plantlet – composed of a stem, at least two well-developed leaves and an apical meristem – it is excised from its callus. For this purpose, the stem of the plantlet is cut-off with a scalpel, bellow the first leaves, but at least 5 mm above the callus (**Figure 94**). The excised plantlets are transferred into sterile pots containing rooting medium (**A.5.c** – one plantlet per pot). The basis of their stem needs to be buried into the medium to ensure a good rooting. The plantlets are maintained in a growth chamber at 24°C under a 16h light/8h dark cycle (150 µmol m^{-2} s^{-1}). Every 2 weeks, they are subcultured onto fresh rooting medium and maintained under the same light/temperature conditions. The plantlets continue to growth in the medium and, after a few weeks, they should start to develop roots.
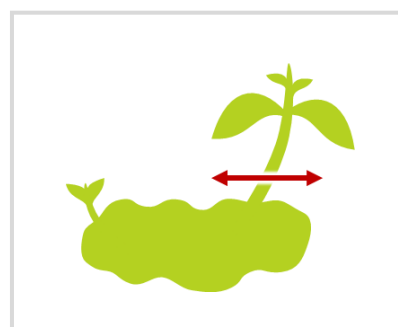
Figure 94 Plantlet excision.

B.5.f. Transfer into the soil and growth of fully developed tomato plants

Once plantlets are well developed and well rooted (at least 8 cm high, numerous and long roots), they can be transferred into soil. Started from this moment, sterile conditions are not necessary anymore.

Rooted plantlets are carefully removed from the gel, their roots are softly cleared of as much gel as possible, and they are planted into soil in flower pots. They are grown in a phytotronic room (Froids et Mesures) at 24°C under a 16h light/8h dark cycle (400 µmol m^{-2} s^{-1}) and a relative humidity of 70%. They are watered twice a week, and supplemented with NPK once a week. After a short adaptation to non-sterile conditions and months of growth, the plantlets should turn into fully developed tomato plants, flower, and produce fruits. Bunches of flower buds are enclosed into pollination bags; the bags are removed when the flowers are faded and the tomato fruits become visible. When the fruits are ripe, they are harvested, opened, and their seeds are carefully collected, washed and dried. They can be stored for months.

B.6. Bioinformatic analyses

B.6.a. Identification of P450 candidates from the *F. carica* RNA-seq library

To identify putative P450 into the *F. carica* RNA-seq library (**A.6.a.1**), it is screened by performing local tBLASTn against the *CYP71AJ3* sequence, using the BioEdit software (**A.6.b.1**) and an e-value of 1000. The resulting hits are kept if they are at least 80 amino acid long, with at least 20% identity. Using such low selection criteria ensures the identification of every contig that may correspond to a P450. By doing so, 50 contigs are selected. They are compared together by using the "CAP contig assembly program" from BioEdit, which allows to easily visualise the overlapping contigs. Small contigs that entirely overlap with bigger ones (at least 99% sequence identity) are excluded. Contigs with overlapping extremities are merged into a single sequence.

In order to identify "complete P450 genes", the (merged) contigs are translated using the "Six-Frame translation" tool from Bioedit (default parameters). A sequence is defined as a complete P450 gene if it corresponds to a coding sequence (starting by an ATG and ending with a stop codon) of about 1500 bp, which associated amino acid sequence shows most of P450 conserved features (such as the absolutely conserved cysteine or the PERF motif – **Chapter I**, **D.1.c**). As contigs comes from an RNA-seq library, they do not contain any introns. The sequences that seems to correspond to complete P450 genes are extracted from their associated contigs and saved separately. The contigs that are not long enough to contain entire genes are used as query to perform BLASTn searches (Mega BLAST, default parameters) against the *Ficus religiosa* cDNA library available on 1KP (**A.6.a.2** – taxid:66387). Some lengthy sequences from *F. religiosa* sharing at least 90% identity with one or several partial contigs from *F. carica* are used to assemble and complete the *F. carica* sequences: the *F. religiosa* sequences overlapping with several *F. carica* contigs that do not or hardly overlap with each other are used as a frame to assemble and merge the contigs. When necessary, these *F. religiosa* sequences are used to "fill the gap" between two *F. carica* contigs. In some cases, the *F. religiosa* sequences are also used to complete the beginning and/or the end of the *F. carica* contigs, allowing the artificial reconstruction of a complete P450 gene. Finally, the new sequences that seems to correspond to complete genes are extracted and saved along with the first ones – for a final number of 13 complete genes. The contigs from which no complete gene can be reconstructed are eliminated.

Numerous P450 proteins sequences belonging to various P450 families from the CYP71 clan are downloaded from the Cytochrome P450 Homepage and the NCBI Protein database (**A.6.a.2**). They are all saved into a single FASTA file to constitute a P450 reference dataset. This dataset is used to perform a local BLASTx on BioEdit (e-value of 1) against every *F. carica* sequence identified as a complete P450 gene. Each *F. carica* sequence is assumed to belong to the same P450 family than the resulting top hit – with which it should share at least 40% protein identity. This preliminary information is used to eliminate candidates from ancient and/or very conserved families involved in core pathways (*i.e.* of low interest for the biosynthesis of furanocoumarins): the CYP73 and CYP75 families.

Finally, every P450 candidate is associated to the expression patterns (expression and relative expression in the latexes of leaves, trunks and fruits) of the main contig from the *F. carica* RNA-seq library from which it derives.

B.6.b. Phylogenetic analyses

B.6.b.1. Constitution of the CYP76F, CYP81BN and CYP82J initial datasets

The CYP76F, CYP81BN and CYP82J datasets are constituted by respectively using the sequences from *CYP76F112*, *CYP81BN4* and *CYP82J18* to perform exhaustive similarity searches in the genetic databases from 20 plant species (**Chapter III**, **Table 6**) publicly available on 1KP and GenBank. The genomic or transcriptomic resources from these species are either downloaded to perform local BLASTn analyses on BioEdit (e-value of 1), or screened online using the BLASTn tools provided by 1KP and GenBank (default parameters). In both cases, the resulting hits are kept if they are at least 100 nucleotide long, with at least 55% identity. By doing so, we expect to be as exhaustive as possible in the respective search of the 3 P450 subfamilies.

The next step consists in identifying complete P450 genes from the selected hits. As described in **B.6.a**, a sequence is defined as a complete P450 gene if it corresponds to a coding sequence of about 1500 bp, which associated amino acid sequence shows most of P450 conserved features. This time, the sequences coming from genomic databases also have to be cleared of hypothetical intron(s), with correct intron-exon borders. In the best cases, the hits come from well-annotated genetic databases, which means that the complete sequence of its associated gene can be obtained (almost) directly. However, when the genetic resources are poorly or non-annotated, the complete gene sequences have to be manually identified and isolated. For this purpose, several bioinformatic tools are used: first, an alignment with CYP76F112, CYP81BN4 or CYP82J18 on BioEdit allows a global identification of the region that should correspond to the complete the gene (approximative beginning and end of the gene, number of introns that are probably present in the genomic sequence). The NetGene2 Server set on *A. thaliana* is subsequently used to make a precise prediction of the introns. Finally, the "Six-Frame translation" tool from Bioedit allows to precisely identify the start and stop codons, and it confirms that the predicted introns do not cause any frameshift. The resulting coding sequence – freed from every nucleotide upstream the start codon / downstream the stop codon, and from its intron(s) – should align well with CYP76F112, CYP81BN4 or CYP82J18, and its associated protein sequence should correspond to a complete P450. Such sequences are considered as putative complete genes. The sequences showing a correct gene structure, but with a single point mutation creating a premature stop codon or a frameshift are considered as putative pseudogenes.

The putative P450 genes and pseudogenes are used to perform BLAST searches against the P450 reference dataset described in **B.6.a**. The top hit should be a CYP76, CYP81 or a CYP82 with at least 40% protein identity – meaning that the putative (pseudo)gene also belongs to this P450 family. For most of the putative (pseudo)genes, the top hit should even be a CYP76F, CYP81BN or a CYP82J with at least 55% protein identity – meaning it should belong to this subfamily.

The CYP76F, CYP81BN and CYP82J datasets are composed of all the putative P450 genes that are identified in the 20 plant species, and that probably belongs to the CYP76F, CYP81BN and CYP82J subfamilies, respectively. Some putative genes belonging to the same family but to a different subfamily are sometimes kept for comparison and/or to increase the size of small initial dataset – especially in the case of the CYP82J dataset. The putative pseudogenes which contain premature stop codons are included in the datasets; the putative frameshift-containing pseudogenes are eliminated

since they cannot be well aligned with other sequences to generate the following partitioned files. Lastly, a few additional sequences are added to the dataset: they include all the *Ficus* P450s cloned in this study, and some sequences of already published P450s downloaded from the Cytochrome P450 Homepage and the NCBI Protein database (**A.6.a.2**). These additional sequences will constitute the outgroup of the phylogenetic tree and allow for a better distinction of the different P450 subfamilies.

B.6.b.2. Sequence alignment

For each of the datasets, the sequence alignment is performed using Geneious (**A.6.b.2**), relying on Translation Align default settings (Genetic code = standard / Translation frame =1 / Treat first codon as start of coding region / Protein alignment option = geneious alignment / Alignment type = global alignment with free end gaps / Cost matrix=blosum62 / Gap open penalty=12 / Gap extension penalty=3 / Refinement iterations=2). Then, Mesquite (**A.6.b.2**) is used to check and partition (CodonPo1 CodonPos2, CodonPos3) the alignments, and to export them as nexus files.

B.6.b.3. Phylogenetic analyses

Bayesian inference trees are constructed using MrBayes (**A.6.b.2**) at CIPRES (**A.6.c**). First, the Bayesian Markov Chain Monte Carlo (MCMC) analysis is set and tested locally, after which the actual Markov Chain ran on CIPRES using the following parameters: nst=mixed, rates=gamma, shape=(all), 30 million generations, 2 independent runs, 4 chains each, temperature heating 0.05, samples taken every 20000 generations, burn-in time set at 7.5 million samples, partitionning: CodonPos1 CodonPos2, CodonPos3. Model convergence and model likelihood between runs are checked in Tracer (**A.6.b.2**). Trees are drawn and coloured with FigTree (**A.6.b.2**), and labelled in PowerPoint (version 2016 – Microsoft Office). Tree branches supported by posterior probabilities above 0.9 are considered as strong, below 0.7 as weak.

B.6.c. Modelling and docking analyses

B.6.c.1. Homology modelling of the CYP76Fs

Homology modelling of the CYP76Fs is conducted on the SWISS-MODEL protein homology modelling server (**A.6.c**), by using their automated standardised procedure (Guex et al. 2009; Benkert et al. 2011; Biasini et al. 2014; Bienert et al. 2017; Bertoni et al. 2017; Waterhouse et al. 2018). The template search is performed against the SWISS-MODEL template library (SMTL, last update: 2019-10-24, last included PDB release: 2019-10-18). The selected template is chosen according to its sequence identity with the CYP76Fs to be modelled. So far, CYP76AH1 (PDB ID: 5YLW) is the only CYP76 that has been crystallised (Gu et al. 2019). It is therefore chosen as the template used to model all CYP76Fs. The N-terminal transmembrane domain of CYP76AH1, which was engineered to allow its crystallisation, is not included in the template. The models are created by the SWISS-MODEL server based on the target-template alignment using ProMod3 (Remmert et al. 2012). The quality of the model is assessed using the Global Model Quality Estimate (GMQE) and the Qualitative Model Energy ANalysis (QMEAN) scoring functions. GQME value close to 0 are considered as not reliable, close to 1 as highly reliable. QMEAN value close to -4 are considered as weak, 0 as strong. The Ramachandran Plots and the local quality

estimate are also checked. 3D homology models are visualised and the figures are prepared using PyMOL (**A.6.b.3**).

B.6.c.2. Docking experiments

Docking experiments are carried out using AutoDock Vina (**A.6.b.3** – Trott and Olson 2009). Unless otherwise specified, the standard settings are used.

A heme group is docked into the active site of every modelled CYP76F with the following parameters. The receptor file is defined as rigid, except for the absolutely conserved cysteine which is defined as flexible. The receptor grid is defined as a 16x16x16 Å cube around the conserved cysteine (center x,y,z = 55.985, -8.032, 12.013). The resulting models are ranked according to their affinity. The model with the lowest affinity is considered as the most probable binding mode. PyMOL (**A.6.b.3**) is used to visualise the different model, and to merge the files corresponding to the CYP76F and the heme (best position) – creating a single file usable for the next step.

Then, the substrate is docked into the active site of the CYP76F complexed with its heme, with the following parameters. The receptor file is defined as rigid, except for a set of 15 amino acids that are treated as flexible (**Supp. Table 8**). The receptor grid is defined as a 45x35x35 Å block around the heme (center x,y,z = 55.249, -3.558, 7.512). The resulting models are ranked according to their affinity. Finally, PyMOL is used to visualise all resulting models, to compare them, to measure the distance between the substrate and the heme, to identify the amino acids that are the closest to the substrate, and to prepare the figures.

Résumé détaillé
en français

RÉSUMÉ DÉTAILLÉ EN FRANÇAIS

~

A. ETAT DE L'ART ET OBJECTIF DE LA THESE

A.1. Défense des plantes et furocoumarines

Au cœur du royaume végétal, les plantes sont assimilables à de petites forteresses qui luttent depuis des millions d'années contre des ennemis aussi nombreux que variés, allant des plus petits pathogènes aux plus gros herbivores. Pour survivre à cette guerre de tout instant, les plantes ont développé de nombreuses stratégies défensives, qui incluent notamment des mécanismes physiques et chimiques. Les défenses chimiques des plantes reposent sur la production d'un arsenal de protéines et de métabolites spécialisés variés, dont les effets peuvent être répulsifs, incapacitants, ou même létaux. Il existerait plusieurs centaines de milliers de métabolites spécialisés – aussi appelés métabolites secondaires – et chaque plante en possède une combinaison qui lui est propre.

La famille des furocoumarines constitue un exemple de métabolite spécialisé, qui dérive des coumarines. Ces composés ne sont pas présents chez toutes les plantes mais sont restreints à quelques taxons éloignés d'angiospermes ; les principales espèces productrices de furocoumarines appartenant aux familles des Apiacées, Fabacées, Moracées et Rutacées (Pathak et al. 1962; Bourgaud et al. 1989, 2006). Les furocoumarines sont des molécules de défense aux effets toxiques et phototoxiques, efficaces contre une large gamme d'organismes incluant des herbivores, des pathogènes, et d'autres plantes. Cette (photo)toxicité est liée à la capacité des furocoumarines à devenir réactives lorsqu'elles sont excitées par les UV, et à inhiber des enzymes tels que les cytochromes P450 (Musajo et al. 1965; He et al. 1998; Gravot et al. 2004). Les propriétés (photo)toxiques des furocoumarines sont à l'origine de brûlures et autres problèmes de santé chez l'Homme (Pathak et al. 1962; Bonamonte et al. 2010; Dugrand-Judek et al. 2015), mais ouvrent également des perspectives médicales, par exemple dans la lutte contre le cancer ou le vitiligo (Chauthe et al. 2015; Zabolinejad et al. 2020).

La voie de biosynthèse des furocoumarines est étudiée depuis plusieurs décennies (**Figure 12**). Entre les années 1960 et 1990, des expériences de radiomarquage ont permis de caractériser les intermédiaires successifs de cette voie (Floss and Mothes 1966; Brown and Steck 1973; Hamerski and Matern 1988a). Toutefois, la caractérisation moléculaire des gènes qui y sont associés est plus récente, puisque ce n'est qu'à partir de 2004 (Hehmann et al. 2004) que des gènes impliqués dans la voie des furocoumarines ont pu être identifiés, clonés et caractérisés d'un point de vue fonctionnel. Ces gènes codent pour des cytochromes P450, des dioxygénases, des prényltransférases et des méthyltransférases. Cependant, malgré les travaux réalisés, l'élucidation moléculaire de la voie des furocoumarines reste encore incomplète, et de nombreux gènes sont toujours inconnus. Parmi eux, le gène codant pour une enzyme permettant la conversion de déméthylsubérosine (DMS) en marmésine (**Figure 19**) possède un intérêt particulier : en effet, non seulement l'identification d'un tel gène permettrait de disposer du set complet d'enzymes menant à la synthèse de psoralène à partir de *p*-coumaroyl CoA, mais la réaction qui serait catalysée par cette enzyme est également assez atypique.

A.2. Les cytochromes P450s

Les cytochromes P450 (ou P450s) constituent une superfamille enzymatique présente dans presque tous les organismes, et pouvant catalyser des réactions variées. Chez les plantes, les P450s les plus communs catalysent une réaction de mono-oxygénation qui peut s'écrire de la façon suivante : $RH + O_2 + 2e^- + 2H^+ \rightarrow ROH + H_2O$. Afin de comparer les P450s appartenant à des espèces différentes, une classification et une nomenclature standardisée ont été établies : le nom d'un P450 commence par « CYP » (pour <u>c</u>ytochrome <u>P</u>450), est suivi d'un nombre indiquant sa famille (>40% d'identité protéique), d'une lettre représentant sa sous famille (>55%), et d'un nombre indiquant l'ordre de découverte du gène (ex : CYP73AJ1). Le clan, un niveau de classification supplémentaire n'apparaissant pas dans la nomenclature, a été ajouté pour regrouper les familles les plus proches (Nelson et al. 1996, 2004; Nelson 1998, 2018; Nelson and Werck-Reichhart 2011).

Malgré une séquence primaire pouvant être extrêmement variable, les structures secondaires et tertiaires des P450s sont extrêmement bien conservées (**Figure 14**). Le centre catalytique d'un P450 – qui est hémoprotéine – est toujours composé d'un hème et d'une cystéine (Werck-Reichhart and Feyereisen 2000; Poulos and Johnson 2015). En général, les caractéristiques liées à l'interaction avec l'hème et aux propriétés catalytiques sont très conservées d'un P450 à l'autre, tandis que les régions qui contrôlent la spécificité de l'enzyme pour son substrat sont bien plus variables. Les régions responsables de cette spécificité sont au nombre de six et sont nommées Sites de Reconnaissance du Substrat (SRS). Ces SRSs constituent des cibles privilégiées pour des approches de mutagenèse dirigée (Gotoh 1992; Schoch et al. 2003; Schuler and Rupasinghe 2011).

Finalement, les P450s représentent environ 1% du génome des angiospermes, ce qui fait d'eux l'une des principales familles enzymatiques chez les plantes. Avec 127 familles réparties en 11 clans, les P450s végétaux sont également extrêmement diversifiés et prennent part à la production de métabolites primaires et spécialisés variés, essentiels pour la croissance et le développement de la plante aussi bien que pour son adaptation à l'environnement. Les P450s ont par exemple joué un rôle important lors de la colonisation du milieu terrestre, notamment en contribuant à la production de molécules indispensables pour survivre hors de l'eau, telles que la lignine. Enfin, des diversifications plus récentes de P450s, qui ont eu lieu dans des taxons restreints, sont souvent associées à la synthèse de métabolites spécialisés qui ne sont produits que dans un nombre d'espèces limitées (Nelson et al. 2008; Mizutani and Ohta 2010; Nelson and Werck-Reichhart 2011; Mizutani 2012; Nelson 2018).

A.3. Objectifs de la thèse

Mon travail de thèse avait pour but de renforcer notre connaissance des défenses des plantes à travers l'exemple de la biosynthèse de furocoumarines. Cette étude s'est articulée en deux axes. Mon premier objectif était de poursuivre l'élucidation de la voie de biosynthèse des furocoumarines, en identifiant et caractérisant de nouveaux gènes impliqués dans cette voie. En particulier, j'ai concentré mes efforts sur la recherche de gènes de figuier (*Ficus carica*) qui coderaient pour une marmésine synthase – c'est-à-dire une enzyme capable de convertir la DMS en marmésine. Mon second objectif consistait à évaluer le coût métabolique lié à la production de furocoumarines. Pour cela, j'ai cherché à insérer le début de la voie de biosynthèse des furocoumarines dans le génome de la tomate (*Solanum lycopersicum*), une plante qui, naturellement, ne produit pas ces molécules.

B. IDENTIFICATION DE GENES IMPLIQUES DANS LA VOIE DE BIOSYNTHESE DES FUROCOUMARINES

B.1. Approche, choix de la plante modèle et des familles enzymatiques d'intérêt

D'après une étude menée en 1988, la marmésine synthase d'*Ammi majus* (Apiacées) serait un P450 (Hamerski and Matern 1988a). J'ai donc concentré mes recherches sur cette famille enzymatique. Au vu du grand nombre de P450s présents chez les plantes, j'ai restreint mes recherches au clan des CYP71s. Ce choix est basé sur le fait que les P450s de la voie des furocoumarines ayant déjà été caractérisés appartiennent tous au clan 71. Le choix de la plante étudiée s'est porté sur le figuier (*Ficus carica*), une Moracée qui produit des furocoumarines linéaires. A l'heure actuelle, un seul gène de la voie des furocoumarines a été décrit chez le figuier – et même chez les Moracées. Il s'agit de FcPT1, une prényltransférase qui converti l'umbelliférone en DMS (Munakata et al. 2020). Grâce à une collaboration établie avec S. Kitajima (Université de Kyoto, Japon), nous avions accès à une banque différentielle d'ARN-seq construite à partir des latex de fruit, de pétiole, et de tronc de *F. carica*. Ces trois latex contiennent des concentrations différentes en furocoumarines ; le latex des feuilles contenant plus de furocoumarines que les deux autres (Kitajima et al. 2018). A partir de cette information, nous avons émis l'hypothèse que les gènes impliqués dans la biosynthèse de furocoumarines seraient plus exprimés dans le latex des feuilles que dans celui des fruits et des troncs.

Ma stratégie a donc consisté à analyser la banque d'ARN-seq de figuier afin d'identifier des P450s du clan CYP71 qui soient préférentiellement exprimés dans le latex des pétioles.

B.2. Identification, clonage et expression hétérologue des P450 candidats

L'analyse de la banque d'ARN-seq de figuier a été réalisée en utilisant une approche par similarité de séquences. Dans un premier temps, la séquence de CYP71AJ3, la psoralène synthase de panais (Larbat et al. 2009), a été utilisée pour réaliser une recherche par BLAST contre la banque d'ARN-seq. Cela a permis l'identification de nombreuses séquences de tailles variées qui ont ensuite été triées (famille de P450, niveaux d'expression, redondance), assemblées, et au besoin complétées, afin de reconstruire des séquences codantes complètes de P450s (putatives). A l'issu de ce travail, 10 P450s particulièrement intéressants ont été identifiés et définis comme candidats (**Table 2**).

La seconde étape a consisté à vérifier l'existence biologique de ces 10 P450s candidats. Pour cela, j'ai cherché à amplifier par PCR, cloner et séquencer chacun de ces P450s. Ce travail a été réalisé à partir d'ARNm extrait d'une feuille de *F. carica*. Sur les 10 candidats, seuls 9 ont pu être clonés avec succès. L'un des candidats n'a jamais pu être amplifié et a donc été abandonné. Les 9 candidats restants ont été séquencés : 8 possédaient une séquence 99% à 100% similaire avec la séquence attendue, mais le 9ème candidat était perclus de codons stop et a donc été éliminé. Les séquences des 8 candidats retenus ont été identifiés selon la nomenclature standardisée des P450s (**Table 3** – David Nelson). Ces 8 P450 candidats ont ensuite été clonés dans un vecteur d'expression, introduits puis exprimés dans des levures. Enfin, les protéines produites ont été collectées en préparant des microsomes. Par cette approche, les 8 P450 candidats ont été produits avec succès (**Figure 24**).

B.3. Criblage fonctionnel et caractérisation enzymatique

Afin de déterminer la fonction des P450 candidats, un criblage fonctionnel a été réalisé avec 40 substrats différents, incluant des coumarines et furocoumarines. Pour ce faire, les microsomes correspondant à chaque candidat ont été incubés en présence de NADPH (cofacteur de la réaction) et d'un substrat à la fois, et les produits de réaction ont été analysés par UHPLC-MS. Pour 5 des candidats, aucun des substrats testés n'a été converti en un nouveau métabolite. Cependant, des activités enzymatiques originales ont été associées à CYP76F112, CYP82J18 et CYP81BN4.

L'incubation de DMS en présence de CYP76F112 a mené à la formation d'un produit réactionnel unique, qui a été formellement identifié comme étant de la marmésine (**Figure 25**). La caractérisation fonctionnelle de CYP76F112 a permis de montrer que cette enzyme possédait une forte spécificité de substrat et de réaction. Son activité est optimale pour un pH de 7 et une température de 27°C (**Figure 29**). Le K_m apparent associé à la conversion de DMS en marmésine par CYP76F112 est de 32.2 ± 3.9 nM, ce qui témoigne d'une affinité extrêmement élevée. Le K_{cat} apparent associé à cette conversion est de 22.1 ± 0.9 min^{-1}, pour une efficacité catalytique K_{cat}/K_m de 687 $\mu M^{-1}.min^{-1}$ (**Figure 31**).

J'ai également montré que CYP82J18 et CYP81BN4 métabolisaient respectivement l'auraptène (une coumarine – **Figure 26**) et la cnidiline (une furocoumarine linéaire en aval de la voie principale – **Figure 27**) en 4 et 3 produits différents, qui n'ont pas pu être identifiés chimiquement. L'auraptène et la cnidiline ne sont pas des molécules centrales dans la voie de biosynthèse des furocoumarines. Ainsi, en comparaison à CYP76F112 et son activité marmésine synthase, CYP82J18 et CYP81BN4 possédaient un intérêt limité dans le cadre de cette étude. Les analyses réalisées sur ces deux enzymes ont donc été moins poussées. Pour CYP82J18, le pH et la température optimale d'activité ont été déterminés à 7-9 et 22°C ; les paramètres cinétiques de CYP82J18 et CYP81BN4 n'ont pas été déterminés.

B.4. Discussion

CYP76F112, CYP82J18 et CYP81BN4 constituent les premiers P450s connus de *F. carica*, et même des Moracées, qui soient capables de métaboliser des coumarines et furocoumarines. Les éléments collectés à propos de CYP82J18 et CYP81BN4 sont insuffisants pour déterminer si les réactions observées sont physiologiques ou opportunistes ; ce qui limite l'intérêt déjà relatif de ces deux enzymes dans le cadre de mon projet. A l'inverse, la forte sélectivité, spécificité, affinité et efficacité associées à la conversion de la DMS en marmésine par CYP76F112 sont autant d'indices qui laissent à penser que cette réaction correspond à l'activité physiologique de l'enzyme. Cela ferait donc de CYP76F112 la marmésine synthase recherchée, faisant ainsi la lumière sur l'une des dernières étapes inconnues de la voie de biosynthèse des furocoumarines linéaires. De plus, dans cette voie, CYP76F112 est de loin l'enzyme possédant la plus forte affinité pour son substrat. Sa capacité à rester active malgré de faibles concentrations de substrat semble tout à fait cohérente avec la concentration supposément très faible en DMS dans le figuier (intermédiaire fugace). CYP76F112 est également le premier CYP76 connu qui soit impliqué dans la synthèse de furocoumarines, et la DMS semble être un substrat atypique pour les CYP76Fs, qui étaient jusqu'alors décrits dans le métabolisme des terpénoïdes. Cela ouvre de nombreuses perspectives évolutives quant à l'émergence de la voie des furocoumarines dans le règne végétal, et a donc motivé la réalisation d'une étude phylogénétique de gènes poussée.

C. ANALYSE PHYLOGENETIQUE DE CYP76F112, CYP82J18 ET CYP81BN4

C.1. Approche, constitution des jeux de données et construction des arbres

Cette étude phylogénétique de gènes avait pour but de reconstituer l'histoire évolutive de CYP76F112 afin de répondre aux questions suivantes : quand le gène de *CYP76F112* est-t-il apparu ? Dans quelles espèces ? Comment l'évolution de sa séquence a-t-elle permis l'acquisition de l'activité marmésine synthase ? Une analyse phylogénétique de gène similaire a été réalisée sur *CYP81BN4* et *CYP82J18*, dans le but d'obtenir plus d'éléments permettant de discuter de leurs rôles physiologiques.

Ma stratégie consistait à passer au crible des bases de données génétiques appartenant à des espèces proches de *Ficus carica*, afin d'y identifier les séquences nucléotidiques les plus similaires à celles de *CYP76F112*, *CYP81BN4* et *CYP82J18* (recherche par similarité de séquence). Les séquences identifiées devaient ensuite être groupées en trois jeux de données – associés aux trois P450s d'intérêt – utilisables pour des analyses phylogénétiques de familles de gènes. Les espèces incluses dans la recherche de séquences homologues (**Table 6**) appartenaient toutes au clade des espèces fixatrices d'azote (Nitrogen Fixing Clade, NFC), un clade qui inclut notamment l'ordre des Rosales, contenant la famille des Moracées. Les séquences de *CYP76F112*, *CYP82J18* et *CYP81BN4* ont été successivement utilisées pour réaliser des comparaisons de séquences nucléotidiques avec différentes bases de données, en utilisant un algorithme de BLASTn. Pour *CYP76F112*, 123 séquences nucléotidiques correspondant à des séquences putatives codantes complètes de CYP76(F) ont ainsi été identifiés. Ce nombre monte à 93 séquences pour *CYP81BN4* et à 54 pour *CYP82J18*. Les jeux de donnés ont été complétés par l'ajout de séquences de P450s connus, afin notamment de former la racine des arbres phylogénétiques de gènes. Les séquences contenues dans ces trois jeux de données ont finalement été alignées et utilisées pour générer des arbres phylogénétiques à inférence Bayésienne.

C.2. CYP76F112 : analyse phylogénétique et discussion

Les analyses phylogénétiques de gènes réalisées à partir du jeu de données associé à CYP76F112 ont montré que la sous famille des CYP76Fs s'est diversifiée de façon indépendante au sein de nombreux taxons du NFC (**Figure 37**, **Figure 38**). En particulier, une expansion majeure semble avoir eu lieu au sein de la famille des Moracée, et peut être même spécifiquement au sein du genre *Ficus*. En effet, les CYP76Fs sont particulièrement nombreux chez *Ficus*, et *CYP76F112* appartient à un clade qui est exclusivement composé de 17 séquences de *Ficus* – le clade des « F112-like ». Ce clade contient *CYP76F112*, *CYP76F111* (un P450 candidat de *F. carica* pour lequel aucune activité n'avait été détectée), ainsi que 3 séquences de *F. religiosa* et 12 séquences de *F. erecta*. Au moment où l'étude a été réalisée, *F. erecta* était le seul *Ficus* dont le génome était entièrement séquencé. Tous les CYP76Fs identifiés chez *F. erecta* forment un large cluster sur le chromosome 10, et un petit cluster de deux séquences sur le chromosome 9. La localisation génomique de ces séquences présente des motifs caractéristiques qui suggèrent fortement que les nombreux CYP76Fs ont émergé via des duplications en tandem répétées suivies d'une diversification. Au sein du clade des F112-like, une séquence de *F. erecta* et une séquence de *F. religiosa* partagent plus de 98% d'identité avec CYP76F112, leur plus proche relatif. Ces deux séquences sont donc de potentielles marmésine synthases, homologues à celle de *F. carica*, ce qui semble être confirmé par une analyse des SRS de ces enzymes (**Figure 39**).

En résumé tous les résultats de cette analyse tendent vers une unique conclusion : CYP76F112 et les autres marmésine synthases putatives de *Ficus* ont très probablement émergé via une expansion et une diversification importante des CYP76Fs, qui a eu lieu via des duplications répétées au sein d'un taxon restreint de la famille des Moracées. Cette expansion serait particulièrement récente puisqu'elle semble s'être produite après la divergence entre les genres *Ficus* et *Morus* (**Figure 44**, **Figure 45**) – *Morus* ne produisant pas de furocoumarines. Il est donc raisonnable d'en conclure que les CYP76Fs de Moracées n'ont été recrutés que récemment dans la voie de biosynthèse des furocoumarines.

De précédentes études ont montré que les CYP76s constituent une famille ancienne et très diversifiée, qui précèderait la divergence des monocotylédones et dicotylédones, et qui a donné naissance à au moins 34 sous familles différentes. De plus, la famille des CYP76s et ses diverses sous familles semblent s'être multipliées et diversifiées à de nombreuses reprises par duplications de gènes, dans des taxons spécifiques. Les CYP76s ont ainsi été recrutés dans la production de métabolites spécialisés variés (Nelson and Werck-Reichhart 2011; Höfer et al. 2014; Ilc et al. 2018; Bathe and Tissier 2019). Les résultats de la présente étude sont donc en parfait accord avec les études précédentes.

Enfin, en conséquence de son émergence récente, la marmésine synthase de figuier est forcément apparue indépendamment des marmésine synthases (pour l'instant inconnues) d'espèces appartenant aux familles éloignées des Psoralées, Rutacées et Apiacées. Dans une étude récente, Munakata *et al.* (2020) suggéraient que les prényltransférases de Moracées et d'Apiacées impliquées dans la métabolisation d'umbelliférone avaient probablement évolué à partir de séquences ancestrales différentes. Comme ces prényltransférase et la marmésine synthase s'insèrent au début de la voie des furocoumarines, ces enzymes font nécessairement parti des premières enzymes de la voie à avoir été recrutées. Jusqu'à présent, deux hypothèses étaient envisagées pour expliquer la présence des furocoumarines dans des taxons éloignés d'angiospermes : une émergence unique suivie de disparitions répétées, ou des émergences multiples par convergence évolutive. L'analyse phylogénétique de CYP76F112 permet donc pour la première fois de trancher avec certitude en faveur d'une apparition indépendante par évolution convergente – et peut être même parallèle – de la voie des furocoumarines chez les Moracées et les autres familles productrices.

C.3. Les cas de CYP81BN4 et CYP82J18

Pour ce qui est de *CYP81BN4*, les analyses phylogénétiques de gènes ne montrent pas d'extension spécifique de CYP81BN chez les Moracées qui soit comparable à celle des CYP76Fs. Une telle extension aurait constitué un argument en faveur d'une activité « cnidiline hydroxylase » physiologique (**Figure 40**, **Figure 41**). Dans le cas de *CYP82J18*, les analyses phylogénétiques de gène ont montré que la famille des CYP82Js était extrêmement bien conservée dans tout le NFC. En général, seul un à deux gènes de CYP82J existaient dans chacun des génomes étudiés (**Figure 42**, **Figure 43**). Il est donc très probable que l'ancêtre du NFC possédait déjà un CYP82J ancestral, dont la fonction pourrait être très conservée parmi la majorité des espèces composant le NFC. Sous cette hypothèse, CYP82J18 posséderait la même activité physiologique que les autres CYP82Js du NFC. Du fait de la faible chance pour que les réactions décrites chez CYP81BN4 et CYP82J18 correspondent à leurs activités physiologiques, et du faible intérêt de ces réactions dans cette étude, ces deux enzymes ne seront pas analysés plus en détails.

D. EMERGENCE DE L'ACTIVITE MARMESINE SYNTHASE

D.1. Approche

Si l'analyse phylogénétique de *CYP76F112* a permis de montrer que ce gène avait émergé récemment, cette analyse seule est insuffisante pour déterminer quand et comment est apparue l'activité marmésine synthase elle-même. Pour faire la lumière sur ces questions, j'ai donc mis en place de nouvelles expérimentations basées sur des approches de modélisation tridimensionnelle par homologie, ainsi que sur de la mutagenèse dirigée. Mon objectif était d'identifier des acides aminés qui joueraient un rôle clé dans la spécificité ou l'efficacité de CYP76F112 – des acides aminés dont l'évolution pourrait être traquée le long de la phylogénie précédemment établie. Pour identifier de tels acides aminés, je me suis basée sur une comparaison entre CYP76F111 et CYP76F112. CYP76F111 est l'un des candidats P450s cloné et exprimé au début de cette étude, qui partage 62% d'identité protéique avec CYP76F112, mais pour lequel aucune activité métabolique n'a pu être observée. La variation de séquence entre CYP76F111 et CYP76F112 constituait donc un bon point de départ pour identifier des différences pouvant expliquer la possession ou non de l'activité marmésine synthase.

D.2. Modélisation et expérience de docking moléculaire

Dans un premier temps, j'ai utilisé des outils de bioinformatique pour modéliser les structures 3D de CYP76F111 et CYP76F112, par homologie avec la structure de CYP76AH1 (Gu et al. 2019). Un hème et une molécule de DMS ont alors été successivement dockés dans les modèles d'enzymes obtenus. La position adoptée par la DMS dans le site actif de CYP76F112 était tout à fait cohérente avec l'activité marmésine synthase (**Figure 48**). En effet, la DMS était positionnée juste au-dessus de l'hème, et ses carbones 2' et 3' (ceux qui sont affectés par la conversion en marmésine) étaient situés à moins de 5 Å du fer hémique, ce qui rend la réaction théoriquement possible. A l'inverse, la position adoptée par la DMS dans le site actif de CYP76F111 était très différente, et ne semblait absolument pas permettre de réaction (**Figure 50**) – ce qui était là aussi cohérent avec les résultats expérimentaux. Ces résultats de docking ont donc été considérés comme fiables, et utilisés pour identifier les acides aminés qui soient à la fois proches de la DMS, mais différents chez CYP76F111 et CYP76F112. Huit acides aminés répondaient à ces critères. Pour évaluer leur impact sur le positionnement de la DMS, j'ai réalisé toute une série de mutants « virtuels », en remplaçant successivement un ou plusieurs de ces huit acides aminés dans la séquence de CYP76F111 ou de CYP76F112 par son équivalent chez l'autre enzyme. La structure 3D associée à ces séquences mutantes a été modélisée, puis un hème et une molécule de DMS ont été dockés dans ces modèles. Les positions adoptées par la DMS dans ces différents mutants ont été comparées à celles adoptées dans CYP76F111 et CYP76F112 ; ce qui m'a permis d'identifier 4 acides aminés dont la mutation individuelle ou combinée affectait le résultat du docking. Ces 4 acides aminés, appelés A, B, C et D sont T102, S105, M117 et S305 chez CYP76F112 ; leurs équivalents respectifs chez CYP76F111sont P105, A108, W120 et P310. Ils appartiennent aux SRS1 et SRS4.

D.3. Mutagenèse dirigée, influence des acides aminés et discussion

Dans un second temps, pour tester le réel impact des acides aminés A, B, C et D sur l'activité marmésine synthase, j'ai utilisé une approche de mutagenèse dirigée. Des gènes mutants ont été synthétisés : ces

mutants correspondent aux séquences de *CYP76F112* et *CYP76F111* dans lesquels un ou plusieurs des codons correspondants aux 4 acides aminés d'intérêt avaient été remplacés par leur équivalent chez l'autre P450. Ces gènes mutants ont été exprimés de façon hétérologue en système levure, et l'activité des enzymes a été étudiée via des incubations en présence de DMS. Les résultats principaux du docking et de la caractérisation enzymatique sont les suivants.

Le remplacement de la méthionine en position C par un tryptophane est à lui seul suffisant pour empêcher totalement la métabolisation de la DMS par CYP76F112. Le résidu C joue donc un rôle clé dans la spécificité de la marmésine synthase. Son influence semble être directement liée à la taille de l'acide aminé présent dans chaque enzyme : la méthionine présente chez CYP76F112 est relativement petite et positionnée directement en dessous de la DMS. Ainsi, lorsque cette méthionine est remplacée par un tryptophane, plus massif, la DMS est physiquement repoussée et n'a plus la place de se positionner correctement dans le site actif (**Figure 61**). Cela suggère que l'acide aminé C contribue significativement à la forme générale du site actif, modulant par là même la spécificité de l'enzyme. L'acide aminé C semble être l'équivalent d'un résidu clé déjà identifié chez des P450s de bactéries, champignons et animaux, qui a également été décrit comme important pour modeler le site actif (Seifert et al. 2009; Sirim et al. 2010).

A l'inverse, le remplacement individuel ou combiné des acides aminés A, B et D dans la séquence de CYP76F112 par leur équivalent de CYP76F111 n'empêche pas la métabolisation de la DMS, mais modifie les paramètres cinétiques de l'enzyme. En particulier, les acides aminés A, B (dans une moindre mesure) et D présents chez CYP76F112 sont associés à une affinité plus élevée vis-à-vis de la DMS, permettant à l'enzyme de rester active à des concentrations plus faibles en substrat. Ces résultats peuvent être expliqués par le caractère polaire des acides aminés présents chez CYP76F112, et non polaire chez CYP76F111. Chez CYP76F112, A et B sont en effet deux résidus polaires voisins qui semblent pouvoir établir des liaisons hydrogènes avec le groupement cétone de la DMS, tandis que l'acide aminé D, polaire également, semble pouvoir établir une liaison hydrogène avec le groupement hydroxy de la DMS (**Figure 63**). Ces interactions contribueraient ainsi à positionner et stabiliser la DMS au-dessus de l'hème, sans toutefois être indispensables pour que la réaction ait lieu.

D.4. Perspectives évolutives et apparition de l'activité marmésine synthase

Les acides aminés en position A, B, C et D ont été identifiés dans les CYP76Fs proches de CYP76F112, et reportés sur l'arbre des CYP76Fs précédemment établit. Les résultats de cette analyse suggèrent fortement que les résidus présents chez CYP76F112 ne correspondent pas aux états ancestraux des acides aminés, mais à de récentes synapomorphies partagées par un sous-clade restreint de F112-like (**Figure 64**). En d'autres termes, ces acides aminés ont évolué récemment, lors de la diversification des CYP76Fs de Moracées. De plus, comme il est hautement improbable qu'un CYP76F avec un tryptophane en position C (état à priori ancestral) puisse métaboliser la DMS, l'apparition de l'activité marmésine synthase ne peut pas précéder cette évolution de séquence. L'émergence de l'activité marmésine synthase est donc récente, et n'aurait pas pu avoir lieu sans l'évolution de l'acide aminé C. L'évolution des acides aminés A, B et D a quant à elle permis d'accroitre l'affinité de l'enzyme. Cependant, d'autres évolutions de séquences, probablement au niveau du canal d'accès, semblent aussi avoir joué un rôle clé. En effet, remplacer les acides aminés A, B, C et D de CYP76F111 par leur équivalent chez CYP76F112 ne suffit pas à lui conférer une activité marmésine synthase.

E. RECONSTITUTION DE LA VOIE DES FUROCOUMARINES DANS LA TOMATE

Mon second objectif de thèse consistait à évaluer le coût métabolique lié à la production de psoralène, qui est le premier composé toxique de la voie des furocoumarines linéaires.

E.1. Approche globale

La production de métabolites spécialisés défensifs comme les furocoumarines a un coût qui peut affecter la physiologie de la plante et impacter négativement sa croissance et sa reproduction. Pour mieux comprendre ce coût, nous avons choisi d'insérer le début de la voie de biosynthèse des furocoumarines dans le génome d'une plante qui, naturellement, ne produit pas ces molécules. Des expériences préliminaires avaient été initiées au LAE par G. Galati : l'objectif de ces expérimentations était d'insérer les deux premiers gènes de la voie dans le génome du tabac (*Nicotiana tabacum*), permettant ainsi la synthèse de DMS à partir de *p*-coumaroyl CoA (un composé ubiquitaire chez les plantes). Pour cela, il a généré un plasmide portant les gènes *PsDiox* et *PsPT1*, et l'a utilisé pour initier des transformations stables de tabac (Galati 2019). Avec l'identification de CYP76F112, il devenait possible de continuer son travail. En effet, l'ajout de *CYP76F112* et de *CYP71AJ3* (psoralène synthase – Larbat et al. 2007) aux deux gènes précédant rend possible la production de psoralène à partir du *p*-coumaroyl CoA (**Figure 65**). En conséquence, mon objectif consistait à compléter le plasmide construit par G. Galati – portant les séquences codantes de PsDiox et PsPT1 – avec les séquences codantes de CYP76F112 et CYP71AJ3. Un fois construit, ce plasmide devait être utilisé pour insérer les 4 gènes dans le génome de la tomate (*Solanum lycopersicum*), une plante qui ne produit pas de furocoumarines.

E.2. Construction d'un plasmide multigénique

Dans un premier temps, je me suis donc attachée à construire un plasmide unique portant les 4 gènes permettant la synthèse de psoralène. Cette construction plasmidique devait répondre à plusieurs critères. Premièrement, le plasmide devait être utilisable pour la transformation stable de plantes. Deuxièmement, un marqueur de sélection devait être ajouté afin de permettre l'identification des plantes transgéniques : un gène de résistance à la kanamycine (*KanaR*) a été choisi pour remplir ce rôle. Enfin, pour permettre leur expression dans la tomate, les 5 séquences codantes de *PsDiox*, *PsPT1*, *CYP76F112*, *CYP71AJ3* et *KanaR* devaient être respectivement placées sous contrôle de promoteurs 35S et de terminateurs tNOS. Pour construire un tel plasmide, j'ai utilisé la technologie du GoldenBraid (Sarrion-Perdigones et al. 2011, 2013, 2014). Ce système standardisé de clonage multigénique permet de construire des plasmides binaires complexes, portant des unités de transcriptions multiples.

La construction plasmidique a commencé par la domestication des séquences codantes de *CYP76F112* et *KanaR* : des adaptateurs ont été ajoutés de part et d'autre des séquences, et des sites de restriction internes ont été supprimés. Cela a rendu les 2 séquences compatibles avec la technologie GoldenBraid. Les autres séquences étaient déjà domestiquées. Les séquences de *CYP76F112*, *CYP71AJ3* et *KanaR* ont ensuite été assemblées avec un promoteur et un terminateur, afin de former des unités de transcriptions simples. Enfin, ces unités de transcription ont été combinées ensemble et avec le plasmide de G. Galati, menant à l'assemblage du plasmide final désiré – nommé 5UT (**Figure 70**). Ce plasmide 5UT porte 5 unités de transcriptions, associées aux 4 gènes de la voie de biosynthèse du

psoralène et au marqueur de sélection. L'ensemble des 5 unités de transcription est encadré par les séquences LB et RB, définissant l'ADN-T qui peut être inséré dans la plante. Le correct assemblage du plasmide 5UT a été confirmé par PCRs, digestions, et séquençages (**Figure 71**).

E.3. Génération de tomates transgéniques

Dans un second temps, j'ai utilisé le plasmide 5UT pour insérer la voie de biosynthèse du psoralène dans le génome de la tomate. Le but était de générer des tomates transgéniques produisant du psoralène, dont la croissance et la défense puisse être comparées à des tomates témoins ne produisant pas de furocoumarines. Afin d'éviter tout biais qui pourrait être causé par le marqueur de sélection, des tomates témoins résistantes à la kanamycine devaient également être générées. Lors de la construction du plasmide 5UT, un plasmide binaire intermédiaire portant l'unité de transcription simple associée à KanaR avait été généré. Ce plasmide a donc aussi été utilisé dans des transformations, pour insérer l'unité de transcription KanaR dans le génome de la tomate.

Ainsi, j'ai successivement utilisé les plasmides 5UT et KanaR pour effectuer des transformations stables médiées par *Agrobacterium* (*A. tumefaciens*, EHA105) sur des tomates de la lignée WVa106 (*Solanum lycopersicum* var. *cerasiformae* 'West Virginia 106'). Des techniques de cultures *in vitro* m'ont alors permis de régénérer des plantes entières. Afin de confirmer – ou non – la présence des transgènes dans les génomes des tomates régénérés, des feuilles de chaque plant ont été échantillonnés. L'ADN de ces feuilles a été extrait et utilisé comme matrice pour réaliser des PCRs avec des amorces ciblant les transgènes. Dans le cas des tomates transformées avec le plasmide KanaR, 28 plants transgéniques contenant le transgène ont été identifiés (**Figure 75**). Ces plants sont montés à fleurs et ont donné des graines, qui constituent la génération T1 des témoins résistants à la kanamycine. Dans le cas des tomates transformées avec le plasmide 5UT, seul le transgène correspondant au marqueur de sélection (KanaR) a pu être identifié. Il semblerait donc que les plantes obtenues et régénérées avec ce second plasmide résultent toutes d'une insertion partielle de l'ADN-T porté par le plasmide 5UT, et ne portent pas les gènes permettant la synthèse de psoralène. Au vu du temps nécessaire pour transformer et régénérer les plants de tomates, l'expérience n'a pas pu être reconduite.

E.4. Discussion

Dans cette dernière partie de ma thèse, mon objectif était d'étudier le coût métabolique lié à la synthèse de psoralène. Des tomates productrices de psoralène n'ont pas pu être générées, mais j'ai cependant pu accomplir un certain nombre d'étapes préliminaires qui sont indispensables pour la poursuite du projet. En particulier, les gènes ont été domestiqués, les constructions plasmidiques ont été réalisées, et les témoins résistants à la kanamycine ont été générés. De plus, les insertions partielles obtenues avec le plasmide 5UT m'ont permis de mettre en évidence certaines limites liées aux choix de construction du plasmide. En conséquence, afin d'augmenter les chances de réussite d'une future transformation, j'ai établi de nombreuses recommandations permettant de reconstruire le plasmide 5UT à partir des gènes domestiqués. Les principales recommandations consistent à modifier l'ordre relatif des 5 unités de transcription sur le plasmide 5TU, à inverser le sens de l'unité de transcription KanaR, et à remplacer les promoteurs 35S associés aux 4 gènes de la voie de biosynthèse du psoralène par des promoteurs inductibles.

F. CONCLUSION GENERALE ET PERSPECTIVES

Pour rappel, mon premier objectif de thèse consistait à poursuivre l'élucidation de la voie de biosynthèse des furocoumarines chez le figuier. Cet objectif a été atteint avec l'identification des 3 premiers P450s de Moracées capables de métaboliser des coumarines et furocoumarines linéaires : CYP76F112, CYP82J18 et CYP81BN4. Le cas des enzymes CYP82J18 et CYP81BN4 ne sera pas rediscuté ici. Le résultat majeur de cette étude correspond à l'identification de CYP76F112, un P450 capable de convertir un substrat à priori unique, la DMS, en un produit unique ayant été identifié comme étant de la marmésine. Les intérêts associés à l'identification de CYP76F112 sont multiples. Premièrement, la conversion de DMS en marmésine correspondait à la dernière étape enzymatique non résolue du début de la voie des furocoumarines. CYP76F112 permet donc d'élucider cette étape, ouvrant le champ des possibles quant à l'utilisation de la voie des furocoumarines. Deuxièmement, CYP76F112 est le premier CYP76 connu impliqué dans la synthèse de furocoumarines, et une analyse phylogénétique a permis de montrer que ce gène résulte d'une diversification récente des CYP76Fs, qui a eu lieu dans la famille des Moracées. Grâce à des approches de modélisation et de mutagenèse dirigée, il est même raisonnable de penser que l'activité marmésine synthase elle-même est une acquisition récente des CYP76Fs, datant d'après la divergence entre les genres *Morus* et *Ficus*. Cela constitue un élément fort en faveur d'une apparition multiple de la voie de biosynthèse des furocoumarines, par évolution convergente, dans des taxons éloignés d'angiospermes. Troisièmement, CYP76F112 correspond au premier CYP76F caractérisé qui métabolise un substrat n'appartenant pas à la classe des terpénoïdes. Il n'est donc pas impossible que le substrat ancestral de CYP76F112 soit un terpénoïde, ce qui témoignerait de la grande adaptabilité évolutionnaire des P450s.

De nombreuses perspectives liées à l'identification de CYP76F112 sont également à considérer. D'une part, la publication très récente du génome complet de *F. carica* devrait permettre de compléter les études phylogénétiques des CYP76Fs de Moracées. D'autre part, des expériences de modélisation additionnelles ont désigné certains des CYP76Fs de *Ficus* comme étant des candidats prometteurs qui pourrait correspondre à de nouveaux gènes de la voie des furocoumarines. Cependant, des familles de P450s autres que celle des CYP76Fs pourraient être impliquées dans la synthèse de furocoumarine chez le figuier. Pour les identifier, il serait possible de se baser sur une comparaison des CYPomes de *Morus* et de *Ficus*, afin de rechercher des familles de P450s qui se seraient diversifiées après la séparation des deux genres. Enfin, CYP76F112 est une enzyme très efficace, possédant une très haute affinité et des propriétés remarquables qui pourraient être analysées plus finement et utilisées pour mieux comprendre le fonctionnement des P450s, ou même utilisées pour ingénieurer d'autres P450s.

En utilisant une approche transcriptomique similaire à celle utilisée pour les P450s candidats, j'ai également identifié des dioxygénases et des méthyltransférases candidates. Un travail préliminaire réalisé au LAE a déjà permis de montrer qu'au moins trois des candidats pouvaient métaboliser des (furo)coumarines.

Enfin, mon second objectif de thèse consistait à évaluer le coût métabolique de la production de furocoumarines. Cet objectif n'a pas pu être atteint, mais mon travail a permis de réaliser un certain nombre d'étapes préliminaires (domestication, constructions plasmidiques, génération des lignées de tomates témoins) et de réenvisager sous un nouvel angle la construction du plasmide 5UT. Ce projet doit donc être poursuivi.

References

Abegaz BM, Ngadjui BT, Folefoc GN, et al (2004) Prenylated flavonoids, monoterpenoid furanocoumarins and other constituents from the twigs of *Dorstenia elliptica* (Moraceae). Phytochemistry 65:221–226. https://doi.org/10.1016/j.phytochem.2003.10.028

Abu-Mustafa EA, El-Tawil BAH, Fayez MBE (1963) Constituents of local plants - IV. *Ficus carica* L., *F. sycomorus* L. and *F. salicifolia* L. leaves. Phytochemistry 3:701–703. https://doi.org/10.1016/S0031-9422(00)82968-4

Agrawal AA, Konno K (2009) Latex: A model for understanding mechanisms, ecology, and evolution of plant defense against herbivory. Annu Rev Ecol Evol Syst 40:311–331. https://doi.org/10.1146/annurev.ecolsys.110308.120307

Aharoni A, Gaidukov L, Khersonsky O, et al (2005) The "evolvability" of promiscuous protein functions. Nat Genet 37:73–76. https://doi.org/10.1038/ng1482

Aikens J, Sligar SG (1994) Kinetic solvent isotope effects during oxygen activation by cytochrome P-450cam. J Am Chem Soc 116:1143–1144. https://doi.org/10.1021/ja00082a051

Akashi T, Aoki T, Ayabe S (1998) CYP81E1, a cytochrome P450 cDNA of licorice (Glycyrrhiza echinata L.), encodes isoflavone 2'-hydroxylase. Biochem Biophys Res Commun 251:67–70. https://doi.org/10.1006/bbrc.1998.9414

Alam P, Siddiqui N, Basudan O, et al (2015) Comparative profiling of biomarker psoralen in antioxidant active extracts of different species of genus *Ficus* by validated HPTLC method. Afr J Tradit Complement Altern Med 12:57–67. https://doi.org/10.4314/ajtcam.v12i1.9

Alber AV, Renault H, Basilio-Lopes A, et al (2019) Evolution of coumaroyl conjugate 3-hydroxylases in land plants: lignin biosynthesis and defense. Plant J 99:924–936. https://doi.org/10.1111/tpj.14373

Al-Khdhairawi AAQ, Krishnan P, Mai C-W, et al (2017) A bis-benzopyrroloisoquinoline alkaloid incorporating a cyclobutane core and a chlorophenanthroindolizidine alkaloid with cytotoxic activity from *Ficus fistulosa* var. *tengerensis*. J Nat Prod 80:2734–2740. https://doi.org/10.1021/acs.jnatprod.7b00500

Amo L, Jansen JJ, van Dam NM, et al (2013) Birds exploit herbivore-induced plant volatiles to locate herbivorous prey. Ecol Lett 16:1348–1355. https://doi.org/10.1111/ele.12177

Anstett MC (2001) Unbeatable strategy, constraint and coevolution, or how to resolve evolutionary conflicts: the case of the fig/wasp mutualism. Oikos 95:476–484. https://doi.org/10.1034/j.1600-0706.2001.950313.x

Appel HM (1993) Phenolics in ecological interactions: The importance of oxidation. J Chem Ecol 19:1521–1552. https://doi.org/10.1007/BF00984895

Appel HM, Cocroft RB (2014) Plants respond to leaf vibrations caused by insect herbivore chewing. Oecologia 175:1257–1266. https://doi.org/10.1007/s00442-014-2995-6

Athnasios AK, El-Kholy IE-S, Soliman G, et al (1962) Constituents of leaves of *Ficus carica*, l .1. Isolation of psoralen, bergapten, psi-taraxasterol, and beta-sitosterol. J Chem Soc Resumed 10:4253–4264. https://doi.org/10.1039/jr9620004253

Atkins WM, Sligar SG (1988) The roles of active site hydrogen bonding in cytochrome P-450cam as revealed by site-directed mutagenesis. J Biol Chem 263:18842–18849

Bailey DG, Dresser G, Arnold JMO (2013) Grapefruit-medication interactions: Forbidden fruit or avoidable consequences? Can Med Assoc J 185:309–316. https://doi.org/10.1503/cmaj.120951

Bailey DG, Spence JD, Munoz C, Arnold JMO (1991) Interaction of citrus juices with felodipine and nifedipine. The lancet 337:268–269. https://doi.org/10.1016/0140-6736(91)90872-M

Barrett LG, Heil M (2012) Unifying concepts and mechanisms in the specificity of plant–enemy interactions. Trends Plant Sci 17:282–292. https://doi.org/10.1016/j.tplants.2012.02.009

Baskin JM, Ludlow CJ, Harris TM, Wolf FT (1967) Psoralen, an inhibitor in the seeds of *Psoralea subacaulis* (Leguminosae). Phytochemistry 6:1209–1213. https://doi.org/10.1016/S0031-9422(00)86083-5

Basset Y, Novotny V (1999) Species richness of insect herbivore communities on *Ficus* in Papua New Guinea. Biol J Linn Soc 67:477–499. https://doi.org/10.1111/j.1095-8312.1999.tb01943.x

Batard Y, LeRet M, Schalk M, et al (1998) Molecular cloning and functional expression in yeast of CYP76B1, a xenobiotic-inducible 7-ethoxycoumarin O-de-ethylase from *Helianthus tuberosus*. Plant J 14:111–120. https://doi.org/10.1046/j.1365-313X.1998.00099.x

Bathe U, Tissier A (2019) Cytochrome P450 enzymes: A driving force of plant diterpene diversity. Phytochemistry 161:149–162. https://doi.org/10.1016/j.phytochem.2018.12.003

Baudry J, Rupasinghe S, Schuler MA (2006) Class-dependent sequence alignment strategy improves the structural and functional modeling of P450s. Protein Eng Des Sel 19:345–353. https://doi.org/10.1093/protein/gzl012

Baumgartner B, Erdelmeier CAJ, Wright AD, et al (1990) An antimicrobial alkaloid from *Ficus septica*. Phytochemistry 29:3327–3330. https://doi.org/10.1016/0031-9422(90)80209-Y

Bayoudh C, Elair M, Labidi R, et al (2017) Efficacy of tissue culture in virus elimination from caprifig and female fig varieties (*Ficus carica* L.). Plant Pathol J 33:288–295. https://doi.org/10.5423/PPJ.OA.10.2016.0205

Beaudoin GAW, Facchini PJ (2013) Isolation and characterization of a cDNA encoding (S)-cis-N-methylstylopine 14-hydroxylase from opium poppy, a key enzyme in sanguinarine biosynthesis. Biochem Biophys Res Commun 431:597–603. https://doi.org/10.1016/j.bbrc.2012.12.129

Beier RC, Oertli EH (1983) Psoralen and other linear furocoumarins as phytoalexins in celery. Phytochemistry 22:2595–2597. https://doi.org/10.1016/0031-9422(83)80173-3

Benikhlef L, L'Haridon F, Abou-Mansour E, et al (2013) Perception of soft mechanical stress in *Arabidopsis* leaves activates disease resistance. BMC Plant Biol 13:1–12. https://doi.org/10.1186/1471-2229-13-133

Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27:343–350. https://doi.org/10.1093/bioinformatics/btq662

Benson DA, Cavanaugh M, Clark K, et al (2012) GenBank. Nucleic Acids Res 41:D36–D42. https://doi.org/10.1093/nar/gks1195

Berenbaum M (1983) Coumarins and caterpillars: A case for coevolution. Evolution 37:163–179. https://doi.org/doi:10.1111/j.1558-5646.1983.tb05524.x

Berenbaum M, Feeny P (1981) Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? Science 212:927–929. https://doi.org/10.1126/science.212.4497.927

Berenbaum MR (2002) Postgenomic chemical ecology: from genetic code to ecological interactions. J Chem Ecol 28:873–896. https://doi.org/10.1023/A:1015260931034

Berenbaum MR (1978) Toxicity of a furanocoumarin to armyworms: a case of biosynthetic escape from insect herbivores. Science 201:532–533. https://doi.org/10.1126/science.201.4355.532

Berenbaum MR, Zangerl AR (2008) Facing the future of plant-insect interaction research: le retour a la "Raison d'Etre." Plant Physiol 146:804–811. https://doi.org/10.1104/pp.107.113472

Berenbaum MR, Zangerl AR (1992) Genetics of physiological and behavioral resistance to host furanocoumarins in the parsnip webworm. Evolution 46:1373–1384. https://doi.org/10.2307/2409943

Berenbaum MR, Zangerl AR (1993) Furanocoumarin metabolism in *Papilio polyxenes*: biochemistry, genetic variability, and ecological significance. Oecologia 95:370–375. https://doi.org/10.1007/BF00320991.

Berenbaum MR, Zangerl AR (1998) Chemical phenotype matching between a plant and its insect herbivore. Proc Natl Acad Sci 95:13743–13748. https://doi.org/10.1073/pnas.95.23.13743

Berim A, Gang DR (2013) The Roles of a flavone-6-hydroxylase and 7-*O*-demethylation in the flavone biosynthetic network of sweet basil. J Biol Chem 288:1795–1805. https://doi.org/10.1074/jbc.M112.420448

Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235

Bertoni M, Kiefer F, Biasini M, et al (2017) Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep 7:1–15. https://doi.org/10.1038/s41598-017-09654-8

Bhuiyan NH, Selvaraj G, Wei Y, King J (2009) Gene expression profiling and silencing reveal that monolignol biosynthesis plays a critical role in penetration defence in wheat against powdery mildew invasion. J Exp Bot 60:509–521. https://doi.org/10.1093/jxb/ern290

Biasini M, Bienert S, Waterhouse A, et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258. https://doi.org/10.1093/nar/gku340

Bienert S, Waterhouse A, de Beer TAP, et al (2017) The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res 45:D313–D319. https://doi.org/10.1093/nar/gkw1132

Boachon B, Junker RR, Miesch L, et al (2015) CYP76C1 (Cytochrome P450)-Mediated linalool metabolism and the formation of volatile and soluble linalool oxides in *Arabidopsis* flowers: a strategy for defense against floral antagonists. Plant Cell tpc.15.00399. https://doi.org/10.1105/tpc.15.00399

Bonamonte D, Foti C, Lionetti N, et al (2010) Photoallergic contact dermatitis to 8-methoxypsoralen in *Ficus carica*. Contact Dermatitis 62:343–348. https://doi.org/10.1111/j.1600-0536.2010.01713.x

Botelho LH, Ryan DE, Levin W (1979) Amino acid compositions and partial amino acid sequences of three highly purified forms of liver microsomal cytochrome P-450 from rats treated with polychlorinated biphenyls, phenobarbital, or 3-methylcholanthrene. J Biol Chem 254:5635–5640

Bourgaud F, Allard N, Forlot P, Guckert A (1990) Study of two pharmaceutically useful *Psoralea* (Leguminosae species) : influence of inoculation on growth, grain and dry matter yield. Agronomy 10:1–8. https://doi.org/10.1051/agro:19900101

Bourgaud F, Allard N, Guckert A, Forlot P (1989) Natural sources for furocoumarins. Psoralens Past Present Future Photochem Biol Act J Libbey Eurotext Paris 301–306

Bourgaud F, Hehn A, Larbat R, et al (2006) Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. Phytochem Rev 5:293–308. https://doi.org/10.1007/s11101-006-9040-2

Bozak KR, O'Keefe DP, Christoffersen RE (1992) Expression of a ripening-related avocado (*Persea americana*) cytochrome P450 in yeast. Plant Physiol 100:1976–1981. https://doi.org/10.1104/pp.100.4.1976

Bozak KR, Yu H, Sirevag R, Christoffersen RE (1990) Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. Proc Natl Acad Sci 87:3904–3908. https://doi.org/10.1073/pnas.87.10.3904

Broadway RM, Duffey SS (1986) Plant proteinase inhibitors: Mechanism of action and effect on the growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exiqua*. J Insect Physiol 32:827–833. https://doi.org/10.1016/0022-1910(86)90097-1

Brodie BB, Axelrod J, Cooper JR, et al (1955) Detoxication of drugs and other foreign compounds by liver microsomes. Science 121:603–604. https://doi.org/10.1126/science.121.3147.603

Brown SA, Steck W (1973) 7-Demethylsuberosin and osthenol as intermediates in furanocoumarin biosynthesis. Phytochemistry 12:1315–1324. https://doi.org/10.1016/0031-9422(73)80558-8

Bruni R, Barreca D, Protti M, et al (2019) Botanical sources, chemistry, analysis, and biological activity of furanocoumarins of pharmaceutical interest. Molecules 24:2163. https://doi.org/10.3390/molecules24112163

Bryant JP, Chapin FS, Klein DR (1983) Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. Oikos 40:357–368. https://doi.org/10.2307/3544308

Burley SK, Berman HM, Bhikadiya C, et al (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res 47:D464–D474. https://doi.org/10.1093/nar/gky1004

Cabello-Hurtado F, Batard Y, Salaün J-P, et al (1998) Cloning, expression in yeast, and functional characterization of CYP81B1, a plant cytochrome P450 that catalyzes in-chain hydroxylation of fatty acids. J Biol Chem 273:7260–7267. https://doi.org/10.1074/jbc.273.13.7260

Caboni P, Saba M, Oplos C, et al (2015) Nematicidal activity of furanocoumarins from parsley against *Meloidogyne* spp.: Parsley amendments reduce root knot nematodes in tomato. Pest Manag Sci 71:1099–1105. https://doi.org/10.1002/ps.3890

Cahoon EB, Ripp KG, Hall SE, McGonigle B (2002) Transgenic production of epoxy fatty acids by expression of a cytochrome P450 enzyme from *Euphorbia lagascae* seed. Plant Physiol 128:615–624. https://doi.org/10.1104/pp.010768

Calka O, Akdeniz N, Metin A, Behcet L (2005) Phototoxic dermatitis due to *Chenopodium album* in a mother and son. Contact Dermatitis 53:58–60. https://doi.org/10.1111/j.0105-1873.2005.0456e.x

Calla B, Wu W -Y., Dean CAE, et al (2020) Substrate-specificity of cytochrome P450-mediated detoxification as an evolutionary strategy for specialization on furanocoumarin-containing hostplants: CYP6AE89 in parsnip webworms. Insect Mol Biol 29:112–123. https://doi.org/10.1111/imb.12612

Caporale G, Dall'Acqua F, Marciani S, Capozzi A (1970) Studies on the biosynthesis of psoralen and bergapten in the leaves of *Ficus carica*. Z Für Naturforschung B 25:700–703. https://doi.org/10.1515/znb-1970-0709

Caporale G, Innocenti G, Guitto A, et al (1981) Biogenesis of linear O-alkylfuranocoumarins: a new pathway involving 5-hydroxymarmesin. Phytochemistry 20:1283–1287. https://doi.org/10.1016/0031-9422(81)80022-2

Carl J, Schwarzer M, Klingelhoefer D, et al (2014) Curare - A curative poison: a scientometric analysis. PLoS ONE 9:e112026. https://doi.org/10.1371/journal.pone.0112026

Carpenter EJ, Matasci N, Ayyampalayam S, et al (2019) Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). GigaScience 8:1–7. https://doi.org/10.1093/gigascience/giz126

Chae L, Kim T, Nilo-Poyanco R, Rhee SY (2014) Genomic signatures of specialized metabolism in plants. Science 344:510–513. https://doi.org/10.1126/science.1252076

Chamont S (2014) *Ficus carica* . Les maladies et ravageurs. In: Ephytia. http://ephytia.inra.fr/fr/C/18200/Hypp-encyclopedie-en-protection-des-plantes-Les-maladies-et-ravageurs. Accessed 17 Sep 2018

Chang M-S, Yang Y-C, Kuo Y-C, et al (2005) Furocoumarin glycosides from the leaves of *Ficus ruficaulis* merr. var. antaoensis. J Nat Prod 68:11–13. https://doi.org/10.1021/np0401056

Chantarasuwan B, Baas P, van Heuven B-J, et al (2014) Leaf anatomy of *Ficus* subsection *Urostigma* (Moraceae). Bot J Linn Soc 175:259–281. https://doi.org/10.1111/boj.12165

Chauthe SK, Mahajan S, Rachamalla M, et al (2015) Synthesis and evaluation of linear furanocoumarins as potential anti-breast and anti-prostate cancer agents. Med Chem Res 24:2476–2484. https://doi.org/10.1007/s00044-014-1312-6

Chen L-W, Cheng M-J, Peng C-F, Chen I-S (2010) Secondary metabolites and antimycobacterial activities from the roots of *Ficus nervosa*. Chem Biodivers 7:1814–1821. https://doi.org/10.1002/cbdv.200900227

Chiang C-C, Cheng M-J, Peng C-F, et al (2010) A novel dimeric coumarin analog and antimycobacterial constituents from *Fatoua pilosa*. Chem Biodivers 7:1728–1736. https://doi.org/10.1002/cbdv.200900326

Cho M-J, Jiang W, Lemaux PG (1998) Transformation of recalcitrant barley cultivars through improvement of regenerability and decreased albinism. Plant Sci 138:229–244. https://doi.org/10.1016/S0168-9452(98)00162-9

Choi HW, Klessig DF (2016) DAMPs, MAMPs, and NAMPs in plant innate immunity. BMC Plant Biol 16:232–242. https://doi.org/10.1186/s12870-016-0921-2

Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5:823–826. https://doi.org/10.1002/j.1460-2075.1986.tb04288.x

Christ B, Xu C, Xu M, et al (2019) Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. Nat Commun 10:3206–3217. https://doi.org/10.1038/s41467-019-11286-7

Clement WL, Weiblen GD (2009) Morphological evolution in the mulberry family (Moraceae). Syst Bot 34:530–552. https://doi.org/10.1600/036364409789271155

Coley PD, Bryant JP, Chapin FS (1985) Resource availability and plant antiherbivore defense. Science 230:895–899. https://doi.org/10.1126/science.230.4728.895

Collu G, Unver N, Peltenburg-Looman AMG, et al (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. FEBS

Lett 508:215–220. https://doi.org/10.1016/S0014-5793(01)03045-9

Croteau R, Kutchan TM, Lewis NG (2000) Natural products (secondary metabolites). In: Biochemistry & Molecular Biology of Plants, B. Buchanan, W. Gruissem, R. Jones, Eds. American Society of Plant Physiologists, pp 1250–1319

Cruaud A, Rønsted N, Chantarasuwan B, et al (2012) An extreme case of plant–insect codiversification: figs and fig-pollinating wasps. Syst Biol 61:1029–1047. https://doi.org/10.1093/sysbio/sys068

Dall'Acqua F, Innocenti G, Caporale G, et al (1979) The role of 5-hydroxymarmesin in the biogenesis of bergapten. Z Für Naturforschung C 34:1278–1280. https://doi.org/10.1515/znc-1979-1236

Damu A, Kuo P-C, Shi L-S, et al (2009) Cytotoxic phenanthroindolizidine alkaloids from the roots of *Ficus septica*. Planta Med 75:1152–1156. https://doi.org/10.1055/s-0029-1185483

Damu AG, Kuo P-C, Shi L-S, et al (2005) Phenanthroindolizidine alkaloids from the stems of *Ficus septica*. J Nat Prod 68:1071–1075. https://doi.org/10.1021/np050095o

Dang TT, Facchini PJ (2014) CYP82Y1 is *N*-methylcanadine 1-hydroxylase, a key noscapine biosynthetic enzyme in opium poppy. J Biol Chem 289:2013–2026. https://doi.org/10.1074/jbc.M113.505099

Dardalhon M, de Massy B, Nicolas A, Averbeck D (1998) Mitotic recombination and localized DNA double-strand breaks are induced after 8-methoxypsoralen and UVA irradiation in *Saccharomyces cerevisiae*. Curr Genet 34:30–42. https://doi.org/10.1007/s002940050363

Debib A, Tir-Touil A, Mothana RA, et al (2014) Phenolic content, antioxidant and antimicrobial activities of two fruit varieties of Algerian *Ficus carica L*: Antioxdant and antimicrobial activity of Algerian figs. J Food Biochem 38:207–215. https://doi.org/10.1111/jfbc.12039

Diawara MM, Trumble JT (1997) 1 2. Linear furanocoumarins. In: Handbook of Plant and Fungal Toxicants, CRC Press. p 175

Diawara MM, Trumble JT, White KK, et al (1993) Toxicity of linear furanocoumarins to *Spodoptera exigua*: evidence for antagonistic interactions. J Chem Ecol 19:2473–2484. https://doi.org/10.1007/BF00980684

Diaz-Chavez ML, Moniodis J, Madilao LL, et al (2013) Biosynthesis of Sandalwood oil: *Santalum album* CYP76F Cytochromes P450 Produce Santalols and Bergamotol. PLoS ONE 8:e75053. https://doi.org/10.1371/journal.pone.0075053

Dicke M, Van Beek TA, Posthumus MA, et al (1990) Isolation and identification of volatile kairomone that affects acarine predatorprey interactions. Involvement of host plant in its production. J Chem Ecol 16:381–396. https://doi.org/10.1007/BF01021772

Du H, Ran F, Dong H-L, et al (2016) Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome P450 93 family in land plants. PLOS ONE 11:e0165020. https://doi.org/10.1371/journal.pone.0165020

Du L, Dong S, Zhang X, et al (2017) Selective oxidation of aliphatic C–H bonds in alkylphenols by a chemomimetic biocatalytic system. Proc Natl Acad Sci 114:E5129–E5137. https://doi.org/10.1073/pnas.1702317114

Dueholm B, Krieger C, Drew D, et al (2015) Evolution of substrate recognition sites (SRSs) in cytochromes P450 from Apiaceae exemplified by the CYP71AJ subfamily. BMC Evol Biol 15:122–136. https://doi.org/10.1186/s12862-015-0396-z

Dugrand A, Olry A, Duval T, et al (2013) Coumarin and furanocoumarin quantitation in *Citrus* peel via Ultraperformance Liquid Chromatography coupled with Mass Spectrometry (UPLC-MS). J Agric Food Chem 61:10677–10684. https://doi.org/10.1021/jf402763t

Dugrand-Judek A (2015) Contribution à l'étude phytochimique et moléculaire de la synthèse des coumarines et furocoumarines chez diverses variétés d'agrumes du genre *Citrus*. Doctoral dissertation, Unité Mixte de Recherche Agronomie et Environnement (UMR UL-INRA)

Dugrand-Judek A, Olry A, Hehn A, et al (2015) The distribution of coumarins and furanocoumarins in *Citrus* species closely matches *Citrus* phylogeny and reflects the organization of biosynthetic pathways. PLOS ONE 10:e0142757. https://doi.org/10.1371/journal.pone.0142757

Durairaj P, Fan L, Machalz D, et al (2019) Functional characterization and mechanistic modeling of the human cytochrome P450 enzyme CYP4A22. FEBS Lett 593:2214–2225. https://doi.org/10.1002/1873-3468.13489

Dus K, Katagirit M, Yu C-A, et al (1970) Chemical characterization of cytochrome P-450 cam*. Biochem Biophys Res Commun 40:1423–1430. https://doi.org/10.1016/0006-291X(70)90026-4

Dussourd DE, Denno RF (1991) Deactivation of plant defense: correspondence between insect behavior and secretory canal architecture. Ecology 72:1383–1396. https://doi.org/10.2307/1941110

Dussourd DE, Denno RF (1994) Host range of generalist caterpillars: trenching permits feeding on plants with secretory canals. Ecology 75:69–78. https://doi.org/10.2307/1939383

Dussourd DE, Eisner T (1987) Vein-cutting behavior: insect counterploy to the latex defense of plants. Science 237:898–901. https://doi.org/10.1126/science.3616620

Edreva A (2005) Pathogenesis-related proteins: research progress in the last 15 years. Gen Appl Plant Physiol 31:105–124. https://doi.org/10.1007/s11816-017-0439-6

Ehrlich PR, Raven PH (1964) Butterflies and plants: a study in coevolution. Evolution 18:586–608. https://doi.org/10.1111/j.1558-5646.1964.tb01674.x

Elbeaino T, Digiaro M, Martelli GP (2011) Complete sequence of Fig fleck-associated virus, a novel member of the family Tymoviridae. Virus Res 161:198–202. https://doi.org/10.1016/j.virusres.2011.07.022

El-Sakhawy F, Kassem H, Abou-Hussein D, et al (2016) Phytochemical investigation of the bioactive extracts of the leaves of *Ficus cyathistipula* Warb. Z Für Naturforschung C 71:141–154. https://doi.org/10.1515/znc-2015-0274

Estévez-Braun A, González AG (1997) Coumarins. Nat Prod Rep 14:465–475. https://doi.org/10.1039/NP9971400465

FAO (2017) FAOSTAT Database. In: Food Agric. Organ. U. N. http://www.fao.org/faostat/en/#data. Accessed 24 Jan 2019

Farrell BD, Dussourd DE, Mitter C (1990) Escalation of plant defense: do latex and resin canals spur plant diversification? Am Nat 138:881–900. https://doi.org/10.1086/285258

Felton GW, Gatehouse JA (1996) Antinutritive plant defence mechanisms. In: Biology of the Insect Midgut, Lehane MJ, Billingsley PF (edn). Springer, Dordrecht, pp 373–416

Feyereisen R (2011) Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Biochim Biophys Acta BBA - Proteins Proteomics 1814:19–28. https://doi.org/10.1016/j.bbapap.2010.06.012

Field B, Osbourn AE (2008) Metabolic diversification. Independent assembly of operon-like gene clusters in different plants. Science 320:543–547. https://doi.org/10.1126/science.1154990

Floss H-G, Mothes U (1966) On the biosynthesis of furocoumarins in *Pimpinella magna*. Phytochemistry 5:161–169. https://doi.org/10.1016/S0031-9422(00)85094-3

Forman V, Bjerg-Jensen N, Dyekjær JD, et al (2018) Engineering of CYP76AH15 can improve activity and specificity towards forskolin biosynthesis in yeast. Microb Cell Factories 17:181–198. https://doi.org/10.1186/s12934-018-1027-3

Fowlks WL, Griffith DG, Oginsky EL (1958) Photosensitization of bacteria by furocoumarins and related compounds. Nature 181:571–572. https://doi.org/10.1038/181571a0.

Fracarolli L, Rodrigues GB, Pereira AC, et al (2016) Inactivation of plant-pathogenic fungus *Colletotrichum acutatum* with natural plant-produced photosensitizers under solar radiation. J Photochem Photobiol B 162:402–411. https://doi.org/10.1016/j.jphotobiol.2016.07.009

Fraenkel GS (1959) The raison d'être of secondary plant substances. Science 129:1466–1470

Franck A, Guilley H, Jonard G, et al (1980) Nucleotide sequence of cauliflower mosaic virus DNA. Cell 21:285–294. https://doi.org/10.1016/0092-8674(80)90136-1.

Franke K, Porzel A, Masaoud M, et al (2001) Furanocoumarins from *Dorstenia gigas*. Phytochemistry 56:611–621. https://doi.org/10.1016/S0031-9422(00)00419-2

Franke R, Humphreys JM, Hemm MR, et al (2002) The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J 30:33–45. https://doi.org/10.1046/j.1365-313X.2002.01266.x

Fresquet-Corrales S, Roque E, Sarrión-Perdigones A, et al (2017) Metabolic engineering to simultaneously activate anthocyanin and proanthocyanidin biosynthetic pathways in *Nicotiana* spp. PLOS ONE 12:e0184839. https://doi.org/10.1371/journal.pone.0184839

Frey M, Chomet P, Glawischnig E, et al (1997) Analysis of a chemical plant defense mechanism in grasses. Science 277:696–699. https://doi.org/10.1126/science.277.5326.696

Frost CJ, Mescher MC, Carlson JE, De Moraes CM (2008) Plant defense priming against herbivores: getting ready for a different battle. Plant Physiol 146:818–824. https://doi.org/10.1104/pp.107.113027

Fujii-Kuriyama Y, Mizukami Y, Kawajiri K, et al (1982) Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobarbital-inducible cytochrome P-450 cDNA from rat liver. Proc Natl Acad Sci 79:2793–2797. https://doi.org/10.1073/pnas.79.9.2793

Fujita M, Fujita Y, Noutoshi Y, et al (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol 9:436–442. https://doi.org/10.1016/j.pbi.2006.05.014

Fujiyama K, Hino T, Kanadani M, et al (2019) Structural insights into a key step of brassinosteroid biosynthesis and its inhibition. Nat Plants 5:589–594. https://doi.org/10.1038/s41477-019-0436-6

Galati G (2019) Etude de la synthèse des furocoumarines chez le panais par des approches d'ingénierie métabolique et de multi-omique. Doctoral dissertation, Unité Mixte de Recherche Agronomie et Environnement (UMR INPL-INRA)

Galati G, Gandin A, Jolivet Y, et al (2019) Untargeted metabolomics approach reveals diverse responses of *Pastinaca sativa* to ozone and wounding stresses. Metabolites 9:153. https://doi.org/10.3390/metabo9070153

Genovese S, Epifano F (2011) Auraptene: a natural biologically active compound with multiple targets. Curr Drug Targets 12:381–386. https://doi.org/10.2174/138945011794815248

Gierl A, Frey M (2001) Evolution of benzoxazinone biosynthesis and indole production in maize. Planta 213:493–498. https://doi.org/10.1007/s004250100594

Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteine inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem 267:83–90

Grassa CJ, Wenger JP, Dabney C, et al (2018) A complete *Cannabis* chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. BioRxiv 458083. https://doi.org/10.1101/458083

Gravot A, Larbat R, Hehn A, et al (2004) Cinnamic acid 4-hydroxylase mechanism-based inactivation by psoralen derivatives: cloning and characterization of a C4H from a psoralen producing plant—*Ruta graveolens*—exhibiting low sensitivity to psoralen inactivation. Arch Biochem Biophys 422:71–80. https://doi.org/10.1016/j.abb.2003.12.013

Gray AI, Waterman PG (1978) Coumarins in the Rutaceae. Phytochemistry 17:845–864. https://doi.org/10.1016/S0031-9422(00)88634-3

Gu M, Wang M, Guo J, et al (2019) Crystal structure of CYP76AH1 in 4-PI-bound state from *Salvia miltiorrhiza*. Biochem Biophys Res Commun 511:813–819. https://doi.org/10.1016/j.bbrc.2019.02.103

Guex N, Peitsch MC, Schwede T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30:S162–S173. https://doi.org/10.1002/elps.200900140

Guo Q, Du G, He H, et al (2016) Two nematicidal furocoumarins from *Ficus carica* L. leaves and their physiological effects on pine wood nematode (*Bursaphelenchus xylophilus*). Nat Prod Res 30:1969–1973. https://doi.org/10.1080/14786419.2015.1094804

Guo Q, Major IT, Howe GA (2018) Resolution of growth–defense conflict: mechanistic insights from jasmonate signaling. Curr Opin Plant Biol 44:72–81. https://doi.org/10.1016/j.pbi.2018.02.009

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos Trans R Soc B Biol Sci 368:20120426. https://doi.org/10.1098/rstb.2012.0426

Hamberger B, Bohlmann J (2006) Cytochrome P450 mono-oxygenases in conifer genomes: discovery of members of the terpenoid oxygenase superfamily in spruce and pine. Biochem Soc Trans 34:1209–1214. https://doi.org/10.1042/BST0341209

Hamerski D, Matern U (1988a) Elicitor-induced biosynthesis of psoralens in *Ammi majus* L. suspension cultures - Microsomal conversion of demethylsuberosin into (+)marmesin and psoralen. Eur J Biochem 171:369–375. https://doi.org/10.1111/j.1432-1033.1988.tb13800.x

Hamerski D, Matern U (1988b) Biosynthesis of psoralens. Psoralen 5-monooxygenase activity from elicitor-treated *Ammi majus* cells. FEBS Lett 239:263–265. https://doi.org/10.1016/0014-5793(88)80930-X

Hann CT, Bequette CJ, Dombrowski JE, Stratmann JW (2014) Methanol and ethanol modulate responses to danger- and microbe-associated molecular patterns. Front Plant Sci 5:550. https://doi.org/10.3389/fpls.2014.00550

Hare RS, Fulco AJ (1975) Carbon monoxide and hydroxymercuribenzoate sensitivity of a fatty acid (ω-2) hydroxylase from *Bacillus megaterium*. Biochem Biophys Res Commun 65:665–672. https://doi.org/10.1016/S0006-291X(75)80198-7

Hasemann CA, Kurumbail RG, Boddupalli SS, et al (1995) Structure and function of cytochromes P450:a comparative analysis of three crystal structures. Structure 3:41–62. https://doi.org/10.1016/S0969-2126(01)00134-4

He K, Iyer KR, Hayes RN, et al (1998) Inactivation of cytochrome P450 3A4 by bergamottin, a component of grapefruit juice. Chem Res Toxicol 11:252–259. https://doi.org/10.1021/tx970192k

Hebeish A, Fouda MMG, Hamdy IA, et al (2008) Preparation of durable insect repellent cotton fabric: Limonene as insecticide. Carbohydr Polym 74:268–273. https://doi.org/10.1016/j.carbpol.2008.02.013

Hebsgaard S (1996) Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global

sequence information. Nucleic Acids Res 24:3439–3452. https://doi.org/10.1093/nar/24.17.3439

Hehmann M, Lukačin R, Ekiert H, Matern U (2004) Furanocoumarin biosynthesis in *Ammi majus* L.: Cloning of bergaptol O-methyltransferase. Eur J Biochem 271:932–940. https://doi.org/10.1111/j.1432-1033.2004.03995.x

Heil M (2009) Damaged-self recognition in plant herbivore defence. Trends Plant Sci 14:356–363. https://doi.org/10.1016/j.tplants.2009.04.002

Heil M, Baldwin IT (2002) Fitness costs of induced resistance: emerging experimental support for a slippery concept. Trends Plant Sci 7:61–67. https://doi.org/10.1016/S1360-1385(01)02186-0

Heil M, Land WG (2014) Danger signals - damaged-self recognition across the tree of life. Front Plant Sci 5: 356–363. https://doi.org/10.3389/fpls.2014.00578

Heinke R, Franke K, Michels K, et al (2012) Analysis of furanocoumarins from *Yemenite Dorstenia* species by liquid chromatography/electrospray tandem mass spectrometry: Analysis of furanocoumarins. J Mass Spectrom 47:7–22. https://doi.org/10.1002/jms.2017

Hellens RP, Edwards EA, Leyland NR, et al (2000) pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. Plant Mol Biol 42:819–832. https://doi.org/10.1023/a:1006496308160.

Herms DA, Mattson WJ (1992) The dilemma of plants: to grow or defend. Q Rev Biol 67:283–335. https://doi.org/10.1086/417659

Hevia MA, Canessa P, Müller-Esparza H, Larrondo LF (2015) A circadian oscillator in the fungus *Botrytis cinerea* regulates virulence when infecting *Arabidopsis thaliana*. Proc Natl Acad Sci 112:8744–8749. https://doi.org/10.1073/pnas.1508432112

Hickel A, Hasslacher M, Griengl H (1996) Hydroxynitrile lyases: functions and properties. Physiol Plant 98:891–898. https://doi.org/10.1111/j.1399-3054.1996.tb06700.x

Hilder VA, Gatehouse AMR, Sheerman SE, et al (1987) A novel mechanism of insect resistance engineered into tobacco. Nature 330:160. https://doi.org/10.1038/330160a0

Hilker M, Kobs C, Varama M, Schrank K (2002) Insect egg deposition induces *Pinus sylvestris* to attract egg parasitoids. J Exp Biol 205:455–461

Höfer R, Boachon B, Renault H, et al (2014) Dual function of the cytochrome P450 CYP76 Family from *Arabidopsis thaliana* in the metabolism of monoterpenols and phenylurea herbicides. Plant Physiol 166:1149–1161. https://doi.org/10.1104/pp.114.244814

Höfer R, Dong L, André F, et al (2013) Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco)iridoid pathway. Metab Eng 20:221–232. https://doi.org/10.1016/j.ymben.2013.08.001

Hood EE, Gelvin SB, Melchers LS, Hoekema A (1993) New *Agrobacterium* helper plasmids for gene transfer to plants. Transgenic Res 2:208–218. https://doi.org/10.1007/BF01977351

Hosomi A, Miwa Y, Furukawa M, Kawaradani M (2012) Growth of fig varieties resistant to ceratocystis canker following infection with *Ceratocystis fimbriata*. J Jpn Soc Hortic Sci 81:159–165. https://doi.org/10.2503/jjshs1.81.159

Hou X, Ding L, Yu H (2013) Crosstalk between GA and JA signaling mediates plant growth and defense. Plant Cell Rep 32:1067–1074. https://doi.org/10.1007/s00299-013-1423-4

Hritz J, de Ruiter A, Oostenbrink C (2008) Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. J Med Chem 51:7469–7477. https://doi.org/10.1021/jm801005m

Hudgins JW, Krekling T, Franceschi VR (2003) Distribution of calcium oxalate crystals in the secondary phloem of conifers: a constitutive defense mechanism? New Phytol 159:677–690. https://doi.org/10.1046/j.1469-8137.2003.00839.x

Humphreys JM, Hemm MR, Chapple C (1999) New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. Proc Natl Acad Sci 96:10045–10050. https://doi.org/10.1073/pnas.96.18.10045

Hunt JW, Dean AP, Webster RE, et al (2008) A novel mechanism by which silica defends grasses against herbivory. Ann Bot 102:653–656. https://doi.org/10.1093/aob/mcn130

Huynh QK, Borgmeyer JR, Zobel JF (1992) Isolation and characterization of a 22 kDa protein with antifungal properties from maize seeds. Biochem Biophys Res Commun 182:1–5. https://doi.org/10.1016/S0006-291X(05)80103-2

Ilc T, Arista G, Tavares R, et al (2018) Annotation, classification, genomic organization and expression of the *Vitis vinifera* CYPome. PLOS ONE 13:e0199902. https://doi.org/10.1371/journal.pone.0199902

Imaishi H, Ishitobi U (2008) Molecular cloning of CYP76A3, a novel cytochrome P450 from *Petunia hybrida* catalyzing the ω-hydroxylation of myristic acid. Biol Plant 52:242–250. https://doi.org/10.1007/s10535-008-0053-0

Imaishi H, Petkova-Andonova M (2007) Molecular Cloning of CYP76B9, a Cytochrome P450 from *Petunia hybrida* , Catalyzing the ω-Hydroxylation of Capric Acid and Lauric Acid. Biosci Biotechnol Biochem 71:104–113. https://doi.org/10.1271/bbb.60396

Ingle RA, Stoker C, Stone W, et al (2015) Jasmonate signalling drives time-of-day differences in susceptibility of *Arabidopsis* to the fungal pathogen *Botrytis cinerea*. Plant J 84:937–948. https://doi.org/10.1111/tpj.13050

Innocenti G, Bettero A, Caporale G (1982) Determination of the coumarinic constituents of *Ficus carica* leaves by HPLC. Il Farm Ed Sci 37:475–485

Innocenti G, Bourgaud F, Piovan A, Favretto D (1997a) Furocoumarins and other secondary metabolites from *Psoralea canescens*. Int J Pharmacogn 35:232–236

Innocenti G, Piovan A, Filippini R, et al (1997b) Quantitative recovery of furanocoumarins from *Psoralea bituminosa*. Phytochem Anal 8:84–86. https://doi.org/10.1002/(SICI)1099-1565(199703)8:2<84::AID-PCA336>3.0.CO;2-W

Iranshahi M, Arfa P, Ramezani M, et al (2007) Sesquiterpene coumarins from *Ferula szowitsiana* and in vitro antileishmanial activity of 7-prenyloxycoumarins against promastigotes. Phytochemistry 68:554–561. https://doi.org/10.1016/j.phytochem.2006.11.002

Jaina R, Jain SC, Bhagchandani T, Yadav N (2013) New furanocoumarins and other chemical constituents from *Ficus carica* root heartwood. Z Für Naturforschung C 68:3–7. https://doi.org/10.1515/znc-2013-1-201

Janda M, Ruelland E (2015) Magical mystery tour: salicylic acid signalling. Environ Exp Bot 114:117–128. https://doi.org/10.1016/j.envexpbot.2014.07.003

Janzen DH (1966) Coevolution of mutualism between ants and acacias in central America. Evolution 20:249–275. https://doi.org/10.1111/j.1558-5646.1966.tb03364.x

Janzen DH (1979) How to be a fig. Annu Rev Ecol Syst 10:13–51. https://doi.org/10.1146/annurev.es.10.110179.000305

Jenkins C, Orsburn B (2019) The first publicly available annotated genome for Cannabis plants. BioRxiv 786186. https://doi.org/10.1101/786186

Jennewein S, Wildung MR, Chau M, et al (2004) Random sequencing of an induced *Taxus* cell cDNA library for identification of clones involved in Taxol biosynthesis. Proc Natl Acad Sci 101:9149–9154. https://doi.org/10.1073/pnas.0403009101

Jeong M-R, Kim H-Y, Cha J-D (2009) Antimicrobial activity of methanol extract from *Ficus carica* leaves against oral bacteria. J Bacteriol Virol 39:97–102. https://doi.org/10.4167/jbv.2009.39.2.97

Junttila O (1976) Allelopathic inhibitors in seeds of *Heracleum laciniatum*. Physiol Plant 36:374–378. https://doi.org/10.1111/j.1399-3054.1976.tb02259.x

Kahn RA, Bouquin RL, Pinot F, et al (2001) A conservative amino acid substitution alters the regiospecificity of CYP94A2, a fatty acid hydroxylase from the plant *Vicia sativa*. Arch Biochem Biophys 391:180–187. https://doi.org/10.1006/abbi.2001.2415

Kai K, Mizutani M, Kawamura N, et al (2008) Scopoletin is biosynthesized via *ortho*-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. Plant J 55:989–999. https://doi.org/10.1111/j.1365-313X.2008.03568.x

Karamat F, Olry A, Munakata R, et al (2014) A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in parsley. Plant J 77:627–638. https://doi.org/10.1111/tpj.12409

Karasov TL, Chae E, Herman JJ, Bergelson J (2017) Mechanisms to mitigate the trade-off between growth and defense. Plant Cell 29:666–680. https://doi.org/10.1105/tpc.16.00931

Kaspera R, Croteau R (2006) Cytochrome P450 oxygenases of taxol biosynthesis. Phytochem Rev 5:433–444. https://doi.org/10.1007/s11101-006-9006-4

Keeling CI, Bohlmann J (2006) Diterpene resin acids in conifers. Phytochemistry 67:2415–2423. https://doi.org/10.1016/j.phytochem.2006.08.019

Kemp CA, Maréchal J-D, Sutcliffe MJ (2005) Progress in cytochrome P450 active site modeling. Arch Biochem Biophys 433:361–368. https://doi.org/10.1016/j.abb.2004.08.026

Kim JS, Kim YO, Ryu HJ, et al (2003) Isolation of stress-related genes of rubber particles and latex in fig tree (*Ficus carica*) and their expressions by abiotic stress or plant hormone treatments. Plant Cell Physiol 44:412–414. https://doi.org/10.1093/pcp/pcg058

Kirton SB, Baxter CA, Sutcliffe MJ (2002) Comparative modelling of cytochromes P450. Adv Drug Deliv Rev 54:385–406. https://doi.org/10.1016/S0169-409X(02)00010-8

Kitajima S, Aoki W, Shibata D, 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. https://doi.org/10.1007/s00425-018-2880-3

Klimko M, Truchan M (2006) Morphological variability of the leaf epidermis in selected taxa of the genus *Ficus* L. (Moraceae) and its taxonomic implication. Acta Soc Bot Pol 75:309–324. https://doi.org/10.5586/asbp.2006.038

Klingenberg M (1958) Pigments of rat liver microsomes. Arch Biochem Biophys 75:376–386. https://doi.org/10.1016/0003-9861(58)90436-3

Knight CD, Sehgal A, Atwal K, et al (1995) Molecular responses to abscisic acid and stress are conserved between moss and cereals. Plant Cell 7:499–506. https://doi.org/10.1105/tpc.7.5.499

Koenigs LL, Trager WF (1998) Mechanism-based inactivation of P450 2A6 by furanocoumarins. Biochemistry 37:10047–10061. https://doi.org/10.1021/bi980003c

Kon S, Whitaker JR (1965) Separation and partial characterization of the peroxidases of *Ficus glabrata* Latex. J Food Sci 30:977–985. https://doi.org/10.1111/j.1365-2621.1965.tb01873.x

Konno K (2011) Plant latex and other exudates as plant defense systems: Roles of various defense chemicals and proteins contained therein. Phytochemistry 72:1510–1530. https://doi.org/10.1016/j.phytochem.2011.02.016

Konno K, Hirayama C, Nakamura M, et al (2004) Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. Plant J 37:370–378. https://doi.org/10.1046/j.1365-313X.2003.01968.x

Korth KL, Doege SJ, Park S-H, et al (2006) *Medicago truncatula* mutants demonstrate the role of plant calcium oxalate crystals as an effective defense against chewing insects. Plant Physiol 141:188–195. https://doi.org/10.1104/pp.106.076737

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*. Doctoral dissertation, Unité Mixte de Recherche Agronomie et Environnement (UMR UL-INRA)

Krieger C, Roselli S, Kellner-Thielmann S, et al (2018) The CYP71AZ P450 subfamily: a driving factor for the diversification of coumarin biosynthesis in apiaceous plants. Front Plant Sci 9:820. https://doi.org/10.3389/fpls.2018.00820

Kruse T, Ho K, Yoo H-D, et al (2008) In planta biocatalysis screen of P450s identifies 8-methoxypsoralen as a substrate for the CYP82C subfamily, yielding original chemical structures. Chem Biol 15:149–156. https://doi.org/10.1016/j.chembiol.2008.01.008

Kuhlman B, Bradley P (2019) Advances in protein structure prediction and design. Nat Rev Mol Cell Biol 20:681–697. https://doi.org/10.1038/s41580-019-0163-x

Kumari M, Clarke HJ, Small I, Siddique KHM (2009) Albinism in plants: a major bottleneck in wide hybridization, androgenesis and doubled haploid culture. Crit Rev Plant Sci 28:393–409. https://doi.org/10.1080/07352680903133252

Lamb DC, Follmer AH, Goldstone JV, et al (2019) On the occurrence of cytochrome P450 in viruses. Proc Natl Acad Sci 116:12343–12352. https://doi.org/10.1073/pnas.1901080116

Larbat R (2006) Contribution à l'étude des P450 impliqués dans la biosynthèse des furocoumarines. Doctoral dissertation, Unité Mixte de Recherche Agronomie et Environnement (UMR UL-INRA)

Larbat R, Hehn A, Hans J, et al (2009) Isolation and functional characterization of *CYP71AJ4* encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. J Biol Chem 284:4776–4785. https://doi.org/10.1074/jbc.M807351200

Larbat R, Kellner S, Specker S, et al (2007) Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. J Biol Chem 282:542–554. https://doi.org/10.1074/jbc.M604762200

Lawrence SD, Novak NG (2006) Expression of poplar chitinase in tomato leads to inhibition of development in Colorado Potato Beetle. Biotechnol Lett 28:593–599. https://doi.org/10.1007/s10529-006-0022-7

Le Bot J, Bénard C, Robin C, et al (2009) The 'trade-off' between synthesis of primary and secondary compounds in young tomato leaves is altered by nitrate nutrition: experimental evidence and model consistency. J Exp Bot 60:4301–4314. https://doi.org/10.1093/jxb/erp271

Lee D-S, Nioche P, Hamberg M, Raman CS (2008) Structural insights into the evolutionary paths of oxylipin biosynthetic enzymes. Nature 455:363–368. https://doi.org/10.1038/nature07307

Lee L-Y, Gelvin SB (2008) T-DNA binary vectors and systems. Plant Physiol 146:325–332. https://doi.org/10.1104/pp.107.113001

Lee S, Badieyan S, Bevan DR, et al (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. Proc Natl Acad Sci 107:21205–21210. https://doi.org/10.1073/pnas.1009975107

Lefèvre F, Fourmeau J, Pottier M, et al (2018) The *Nicotiana tabacum* ABC transporter NtPDR3 secretes O-methylated coumarins in response to iron deficiency. J Exp Bot 69:4419–4431. https://doi.org/10.1093/jxb/ery221

Leong BJ, Last RL (2017) Promiscuity, impersonation and accommodation: evolution of plant specialized metabolism. Curr Opin Struct Biol 47:105–112. https://doi.org/10.1016/j.sbi.2017.07.005

Levin DA (1973) The Role of Trichomes in Plant Defense. Q Rev Biol 48:3–15. https://doi.org/10.1086/407484

Lewinsohn TM (1991) The geographical distribution of plant latex. Chemoecology 2:64–68. https://doi.org/10.1007/BF01240668

Li D, Ma Y, Zhou Y, et al (2019) A structural and data-driven approach to engineering a plant cytochrome P450 enzyme. Sci China Life Sci 62:873–882. https://doi.org/10.1007/s11427-019-9538-3

Li H, Mei L, Urlacher VB, Schmid RD (2008a) Cytochrome P450 BM-3 evolved by random and saturation mutagenesis as an effective indole-hydroxylating catalyst. Appl Biochem Biotechnol 144:27–36. https://doi.org/10.1007/s12010-007-8002-5

Li L, Chang Z, Pan Z, et al (2008b) Modes of heme binding and substrate access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide synthase. Proc Natl Acad Sci 105:13883–13888. https://doi.org/10.1073/pnas.0804099105

Li S, Chaulagain MR, Knauff AR, et al (2009) Selective oxidation of carbolide C-H bonds by an engineered macrolide P450 mono-oxygenase. Proc Natl Acad Sci 106:18463–18468. https://doi.org/10.1073/pnas.0907203106

Li W, Schuler MA, Berenbaum MR (2003) Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. Proc Natl Acad Sci 100:14593–14598. https://doi.org/10.1073/pnas.1934643100

Limones-Mendez M, Dugrand-Judek A, Villard C, et al (2020) Convergent evolution leading to the appearance of furanocoumarins in citrus plants. Plant Sci 292:110392. https://doi.org/10.1016/j.plantsci.2019.110392

Lin H, Kenaan C, Hollenberg PF (2012) Identification of the residue in human CYP3A4 that Is covalently modified by bergamottin and the reactive intermediate that contributes to the grapefruit juice effect. Drug Metab Dispos 40:998–1006. https://doi.org/10.1124/dmd.112.044560

Liscombe DK, Facchini PJ (2008) Evolutionary and cellular webs in benzylisoquinoline alkaloid biosynthesis. Curr Opin Biotechnol 19:173–180. https://doi.org/10.1016/j.copbio.2008.02.012

Liu C-J, Huhman D, Sumner LW, Dixon RA (2003) Regiospecific hydroxylation of isoflavones by cytochrome P450 81E enzymes from *Medicago truncatula*. Plant J 36:471–484. https://doi.org/10.1046/j.1365-313X.2003.01893.x

Liu F, Yang Z, Zheng X, et al (2011) Nematicidal coumarin from *Ficus carica* L. J Asia-Pac Entomol 14:79–81. https://doi.org/10.1016/j.aspen.2010.10.006

Liu Z, Tavares R, Forsythe ES, et al (2016) Evolutionary interplay between sister cytochrome P450 genes shapes plasticity in plant metabolism. Nat Commun 7:13026. https://doi.org/10.1038/ncomms13026

Loomis WE (1932) Growth-differentiation balance vs carbohydrate-nitrogen ratio. Proc Am Soc Hortic Sci 29:240–245

Loomis WE (1953) Growth and differentiation—an introduction and summary. In: Growth and Differentiations in Plants. Ames (IA): Iowa State College Press, pp 1–17

Lord JM, Roberts LM, Robertus JD (1994) Ricin: structure, mode of action, and some current applications. FASEB J 8:201–208. https://doi.org/10.1096/fasebj.8.2.8119491

Luscombe NM, Qian J, Zhang Z, et al (2002) The dominance of the population by a selected few: power-law behaviour applies to a wide variety of genomic properties. Genome Biol 3:1–7. https://doi.org/10.1186/gb-2002-3-8-research0040

Luz RF, Vieira IJC, Braz-Filho R, Moreira VF (2015) 13C-NMR data from coumarins from Moraceae family. Am J Anal Chem 06:851–866. https://doi.org/10.4236/ajac.2015.611081

Lynch M (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155. https://doi.org/10.1126/science.290.5494.1151

Ma B, Luo Y, Jia L, et al (2014) Genome-wide identification and expression analyses of cytochrome *P450* genes in mulberry (*Morus notabilis*): Analyses of mulberry *P450* genes. J Integr Plant Biol 56:887–901. https://doi.org/10.1111/jipb.12141

Maddison WP, Maddison DR (2019) Mesquite: a modular system for evolutionary analysis. Version 3.61. In: Mesquite. http://www.mesquiteproject.org. Accessed 18 Sep 2020

Maher EA, Bate NJ, Ni W, et al (1994) Increased disease susceptibility of transgenic tobacco plants with suppressed levels of preformed phenylpropanoid products. Proc Natl Acad Sci 91:7802–7806. https://doi.org/10.1073/pnas.91.16.7802

Mamoucha S, Fokialakis N, Christodoulakis NS (2016) Leaf structure and histochemistry of *Ficus carica* (Moraceae), the fig tree. Flora - Morphol Distrib Funct Ecol Plants 218:24–34. https://doi.org/10.1016/j.flora.2015.11.003

Mandalia MR, Chalmers R, Schreuder FB (2008) Contact with fig tree sap: An unusual cause of burn injury. Burns 34:719–721. https://doi.org/10.1016/j.burns.2007.03.026

Marrelli M, Menichini F, Statti GA, et al (2012) Changes in the phenolic and lipophilic composition, in the enzyme inhibition and antiproliferative activity of *Ficus carica* L. cultivar Dottato fruits during maturation. Food Chem Toxicol 50:726–733. https://doi.org/10.1016/j.fct.2011.12.025

Marrelli M, Statti GA, Tundis R, et al (2014) Fatty acids, coumarins and polyphenolic compounds of *Ficus carica* L. cv. Dottato: variation of bioactive compounds and biological activity of aerial parts. Nat Prod Res 28:271–274. https://doi.org/10.1080/14786419.2013.841689

Massey FP, Ennos AR, Hartley SE (2007a) Grasses and the resource availability hypothesis: the importance of silica-based defences. J Ecol 95:414–424. https://doi.org/10.1111/j.1365-2745.2007.01223.x

Massey FP, Hartley SE (2009) Physical defences wear you down: progressive and irreversible impacts of silica on insect herbivores. J Anim Ecol 78:281–291. https://doi.org/10.1111/j.1365-2656.2008.01472.x

Massey FP, Roland Ennos A, Hartley SE (2007b) Herbivore specific induction of silica-based plant defences. Oecologia 152:677–683. https://doi.org/10.1007/s00442-007-0703-5

Matsui K, Shibutani M, Hase T, Kajiwara T (1996) Bell pepper fruit fatty acid hydroperoxide lyase is a cytochrome P450 (CYP74B). FEBS Lett 394:21–24. https://doi.org/10.1016/0014-5793(96)00924-6

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

Matsunaga W, Shimura H, Shirakawa S, et al (2019) Transcriptional silencing of 35S driven-transgene is differentially determined depending on promoter methylation heterogeneity at specific cytosines in both plus- and minus-sense strands. BMC Plant Biol 19:24. https://doi.org/10.1186/s12870-019-1628-y

Mauch F, Mauch-Mani B, Boller T (1988) Antifungal hydrolases in pea tissue. Plant Physiol 88:936–942. https://doi.org/10.1104/pp.88.3.936

McKenzie EHC (1986) New plant disease record in New Zealand: Fig rust (*Cerotelium fici*) on *Ficus carica*. N Z J Agric Res 29:707–710. https://doi.org/10.1080/00288233.1986.10430467

McKey D (1974) Adaptive patterns in alkaloid physiology. Am Nat 108:305–320. https://doi.org/10.1086/282909

McKey D (1979) The distribution of secondary compounds within plants. In: Herbivores-their interaction with secondary plant metabolites. pp 55–133

Mclean KJ, Girvan HM, Mason AE, et al (2011) Chapter 8. Structure, mechanism and function of cytochrome P450 enzymes. In: de Visser SP, Kumar D (eds) Iron-Containing Enzymes. Royal Society of Chemistry, Cambridge, pp 255–280

Medina-Ortega KJ, Walker GP (2015) Faba bean forisomes can function in defence against generalist aphids: Forisomes defend *V. faba* against generalist aphids. Plant Cell Environ 38:1167–1177. https://doi.org/10.1111/pce.12470

Melchers LS, Apotheker-de Groot M, van der Knaap JA, et al (1994) A new class of tobacco chitinases homologous to bacterial exo-chitinases displays antifungal activity. Plant J 5:469–480. https://doi.org/10.1046/j.1365-313X.1994.05040469.x

Melough MM, Chun OK (2018) Dietary furocoumarins and skin cancer: A review of current biological evidence. Food Chem Toxicol 122:163–171. https://doi.org/10.1016/j.fct.2018.10.027

Messer A, Raquet N, Lohr C, Schrenk D (2012) Major furocoumarins in grapefruit juice II: Phototoxicity, photogenotoxicity, and inhibitory potency vs. cytochrome P450 3A4 activity. Food Chem Toxicol 50:756–760. https://doi.org/10.1016/j.fct.2011.11.023

Mestres J (2004) Structure conservation in cytochromes P450. Proteins Struct Funct Bioinforma 58:596–609. https://doi.org/10.1002/prot.20354

Meyer K, Cusumano JC, Somerville C, Chapple CC (1996) Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases. Proc Natl Acad Sci 93:6869–6874. https://doi.org/10.1073/pnas.93.14.6869

Meyers BC, Kozik A, Griego A, et al (2003) Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. Plant Cell 15:809–834. https://doi.org/10.1105/tpc.009308

Milesi S, Massot B, Gontier E, et al (2001) *Ruta graveolens L.*: a promising species for the production of furanocoumarins. Plant Sci 161:189–199. https://doi.org/10.1016/S0168-9452(01)00413-7

Mizutani M (2012) Impacts of diversification of cytochrome P450 on plant metabolism. Biol Pharm Bull 35:824–832. https://doi.org/10.1248/bpb.35.824

Mizutani M, Ohta D (2010) Diversification of P450 genes during land plant evolution. Annu Rev Plant Biol 61:291–315. https://doi.org/10.1146/annurev-arplant-042809-112305

Mizutani M, Sato F (2011) Unusual P450 reactions in plant secondary metabolism. Arch Biochem Biophys 507:194–203. https://doi.org/10.1016/j.abb.2010.09.026

Mizutani M, Ward E, Dimaio J, et al (1993) Molecular cloning and sequencing of a cDNA encoding mung bean Cytochrome P450 (P450C4H) possessing cinnamate 4-hydroxylase activity. Biochem Biophys Res Commun 190:875–880. https://doi.org/10.1006/bbrc.1993.1130

Mlotshwa S, Pruss GJ, Gao Z, et al (2010) Transcriptional silencing induced by Arabidopsis T-DNA mutants is associated with 35S promoter siRNAs and requires genes involved in siRNA-mediated chromatin silencing: SALK line-induced transcriptional silencing. Plant J 64:699–704. https://doi.org/10.1111/j.1365-313X.2010.04358.x

Moghe G, Last RL (2015) Something old, something new: Conserved enzymes and the evolution of novelty in plant specialized metabolism. Plant Physiol 169:1512–1523. https://doi.org/10.1104/pp.15.00994

Moghe GD, Shiu S-H (2014) The causes and molecular consequences of polyploidy in flowering plants: Plant polyploidy. Ann N Y Acad Sci 1320:16–34. https://doi.org/10.1111/nyas.12466

Mohamed SA, Abdel-Aty AM, Belal Hamed M, et al (2011) *Ficus sycomorus* latex: A thermostable peroxidase. Afr J Biotechnol 10:17532–17543. https://doi.org/10.5897/AJB11.2375

Moore BD, Andrew RL, Külheim C, Foley WJ (2014) Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytol 201:733–750. https://doi.org/10.1111/nph.12526

Moore BD, Johnson SN (2017) Get tough, get toxic, or get a bodyguard: Identifying candidate traits conferring belowground resistance to herbivores in grasses. Front Plant Sci 7:1925. https://doi.org/10.3389/fpls.2016.01925

Moses T, Thevelein JM, Goossens A, Pollier J (2014) Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. Phytochemistry 108:47–56. https://doi.org/10.1016/j.phytochem.2014.10.002

Müller J, Clauss M, Codron D, et al (2014) Growth and wear of incisor and cheek teeth in domestic rabbits (*Oryctolagus cuniculus*) fed diets of different abrasiveness. J Exp Zool Part Ecol Genet Physiol 321:283–298. https://doi.org/10.1002/jez.1864

Munakata R, Kitajima S, Nuttens A, et al (2020) Convergent evolution of the UbiA prenyltransferase family underlies the independent acquisition of furanocoumarins in plants. New Phytol 225:2166–2182. https://doi.org/10.1111/nph.16277

Munakata R, Olry A, Karamat F, et al (2016) Molecular evolution of parsnip (*Pastinaca sativa*) membrane-bound prenyltransferases for linear and/or angular furanocoumarin biosynthesis. New Phytol 211:332–344. https://doi.org/10.1111/nph.13899

Murray RDH, Mendez J, Brown RA (1982) The natural coumarins: occurence, chemistry and biochemistry, Wiley and Sons. Chichester

Musajo L, Rodighie G, Colombo G, et al (1965) Photosensitizing furocoumarins - interaction with DNA and photo-inactivation of DNA containing viruses. Experientia 21:22–24. https://doi.org/10.1007/BF02136362.

Neal JJ, Wu D (1994) Inhibition of insect cytochromes P450 by furanocoumarins. Pestic Biochem Physiol 50:43–50. https://doi.org/10.1006/pest.1994.1056

Nelson D, Werck-Reichhart D (2011) A P450-centric view of plant evolution: P450-centric evolution. Plant J 66:194–211. https://doi.org/10.1111/j.1365-313X.2011.04529.x

Nelson DR (2004) Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. PLANT Physiol 135:756–772. https://doi.org/10.1104/pp.104.039826

Nelson DR (2018) Cytochrome P450 diversity in the tree of life. Biochim Biophys Acta BBA - Proteins Proteomics 1866:141–154. https://doi.org/10.1016/j.bbapap.2017.05.003

Nelson DR (2011) Progress in tracing the evolutionary paths of cytochrome P450. Biochim Biophys Acta BBA - Proteins Proteomics 1814:14–18. https://doi.org/10.1016/j.bbapap.2010.08.008

Nelson DR (1999) Cytochrome P450 and the individuality of species. Arch Biochem Biophys 369:1–10. https://doi.org/10.1006/abbi.1999.1352

Nelson DR (1998) Metazoan cytochrome P450 evolution. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol 121:15–22. https://doi.org/10.1016/S0742-8413(98)10027-0

Nelson DR (2006) Plant cytochrome P450s from moss to poplar. Phytochem Rev 5:193–204. https://doi.org/10.1007/s11101-006-9015-3

Nelson DR (2019) Cytochrome P450s in the sugarcane *Saccharum spontaneum*. Trop Plant Biol 12:150–157. https://doi.org/10.1007/s12042-019-09226-2

Nelson DR (2009) The cytochrome P450 homepage. Hum Genomics 4:59. https://doi.org/10.1186/1479-7364-4-1-59

Nelson DR, Koymans L, Kamataki T, et al (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics 6:1–42. https://doi.org/10.1097/00008571-199602000-00002

Nelson DR, Ming R, Alam M, Schuler MA (2008) Comparison of cytochrome P450 genes from six plant genomes. Trop Plant Biol 1:216–235. https://doi.org/10.1007/s12042-008-9022-1

Nelson DR, Zeldin DC, Hoffman SM, et al (2004) Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants: Pharmacogenetics 14:1–18. https://doi.org/10.1097/00008571-200401000-00001

Ngadjui BVT, Abegaz BM (2003) The chemistry and pharmacology of the genus *Dorstenia* (Moraceae). In: Studies in Natural Products Chemistry. Elsevier, pp 761–805

Nigg HN, Strandberg JO, Beier RC, et al (1997) Furanocoumarins in Florida celery varieties increased by fungicide treatment. J Agric Food Chem 45:1430–1436. https://doi.org/10.1021/jf960537p

Nitao JK, Zangerl AR (1987) Floral development and chemical defense allocation in wild parsnip (*Pastinaca sativa*). Ecology 68:521–529. https://doi.org/10.2307/1938457

Novotny V, Miller SE, Baje L, et al (2010) Guild-specific patterns of species richness and host specialization in plant-herbivore food webs from a tropical forest: Plant-herbivore food webs in tropical forest. J Anim Ecol 79:1193–1203. https://doi.org/10.1111/j.1365-2656.2010.01728.x

Nützmann H, Huang A, Osbourn A (2016) Plant metabolic clusters – from genetics to genomics. New Phytol 211:771–789. https://doi.org/10.1111/nph.13981

Nützmann H-W, Osbourn A (2014) Gene clustering in plant specialized metabolism. Curr Opin Biotechnol 26:91–99. https://doi.org/10.1016/j.copbio.2013.10.009

Ober D (2005) Seeing double: gene duplication and diversification in plant secondary metabolism. Trends Plant Sci 10:444–449. https://doi.org/10.1016/j.tplants.2005.07.007

Ogawa K, Kawasaki A, Yoshida T, et al (2000) Evaluation of auraptene content in *Citrus* fruits and their products. J Agric Food Chem 48:1763–1769. https://doi.org/10.1021/jf9905525

Ohnishi T, Yokota T, Mizutani M (2009) Insights into the function and evolution of P450s in plant steroid metabolism. Phytochemistry 70:1918–1929. https://doi.org/10.1016/j.phytochem.2009.09.015

Ohta T, Nagahashi M, Hosoi S, Tsukamoto S (2002) Dihydroxybergamottin caproate as a potent and stable CYP3A4 inhibitor. Bioorganic Med Chemistry 10:969–973. https://doi.org/10.1016/S0968-0896(01)00362-5

Ojala T, Remes S, Haansuu P, et al (2000) Antimicrobial activity of some coumarin containing herbal plants growing in Finland. J Ethnopharmacol 73:299–305. https://doi.org/10.1016/S0378-8741(00)00279-8

Ojo OA, Ajiboye BO, Ojo AB, et al (2014) Phytochemical, proximate analysis and mineral composition of aqueous crude extract of *Ficus asperifolia* Miq. J Adv Med Life Sci 1:1–4. https://doi.org/10.5281/zenodo.1000388

Okamoto M, Kuwahara A, Seo M, et al (2006) CYP707A1 and CYP707A2, which encode abscisic acid 8′-hydroxylases, are indispensable for proper control of seed dormancy and germination in *Arabidopsis*. Plant Physiol 141:97–107. https://doi.org/10.1104/pp.106.079475

Oliveira AP, Baptista P, Andrade PB, et al (2012) Characterization of *Ficus carica* L. cultivars by DNA and secondary metabolite analysis: Is genetic diversity reflected in the chemical composition? Food Res Int 49:710–719. https://doi.org/10.1016/j.foodres.2012.09.019

Oliveira AP, Silva LR, Ferreres F, et al (2010a) Chemical assessment and *in Vitro* antioxidant capacity of *Ficus carica* latex. J Agric Food Chem 58:3393–3398. https://doi.org/10.1021/jf9039759

Oliveira AP, Silva LR, Guedes de Pinho P, et al (2010b) Volatile profiling of *Ficus carica* varieties by HS-SPME and GC–IT-MS. Food Chem 123:548–557. https://doi.org/10.1016/j.foodchem.2010.04.064

Oliveira AP, Valentão P, Pereira JA, et al (2009) *Ficus carica* L.: Metabolic and biological screening. Food Chem Toxicol 47:2841–2846. https://doi.org/10.1016/j.fct.2009.09.004

Omura T, Sato R (1964a) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J Biol Chem 239:2370–2378

Omura T, Sato R (1964b) The carbon monoxide-binding pigment of liver microsomes: II. Solubilization, purification and properties. J Biol Chem 239:2379–2385

One Thousand Plant Transcriptomes Initiative (2019) One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574:679–685. https://doi.org/10.1038/s41586-019-1693-2

Ono E, Nakai M, Fukui Y, et al (2006) Formation of two methylenedioxy bridges by a Sesamum CYP81Q protein yielding a furofuran lignan, (+)-sesamin. Proc Natl Acad Sci 103:10116–10121. https://doi.org/10.1073/pnas.0603865103

Osakabe K, Tsao CC, Li L, et al (1999) Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms. Proc Natl Acad Sci 96:8955–8960. https://doi.org/10.1073/pnas.96.16.8955

Osbourn AE, Field B (2009) Operons. Cell Mol Life Sci 66:3755–3775. https://doi.org/10.1007/s00018-009-0114-3

Palú G, Palumbo M, Cusinato R, et al (1984) Antiviral properties of psoralen derivatives: A biological and physico-chemical investigation. Biochem Pharmacol 33:3451–3456. https://doi.org/10.1016/0006-2952(84)90119-9

Pan G, Zhang X, Liu K, et al (2006) Map-based cloning of a novel rice cytochrome P450 gene CYP81A6 that confers resistance to two different classes of herbicides. Plant Mol Biol 61:933–943. https://doi.org/10.1007/s11103-006-0058-z

Panchy N, Lehti-Shiu MD, Shiu S-H (2016) Evolution of gene duplication in plants. Plant Physiol 171:2294–2316. https://doi.org/10.1104/pp.16.00523

Paquette S, Møller BL, Bak S (2003) On the origin of family 1 plant glycosyltransferases. Phytochemistry 62:399–413. https://doi.org/10.1016/S0031-9422(02)00558-7

Paquette SM, Bak S, Feyereisen R (2000) Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. DNA Cell Biol 19:307–317. https://doi.org/10.1089/10445490050021221

Pateraki I, Andersen-Ranberg J, Jensen NB, et al (2017) Total biosynthesis of the cyclic AMP booster forskolin from Coleus forskohlii. eLife 6:e23001. https://doi.org/10.7554/eLife.23001

Pathak MA, Daniels Jr F, Fitzpatrick TB (1962) The presently known distribution of furocoumarins (psoralens) in plants. J Invest Dermatol 39:225–239. https://doi.org/10.1038/jid.1962.106

Patil Vikas V, Bhangale SC, Patil VR (2010) Evaluation of anti-pyretic potential of *Ficus carica* leaves. Int J Pharm Sci Rev Res 2:48–50

Peiffer M, Tooker JF, Luthe DS, Felton GW (2009) Plants on early alert: glandular trichomes as sensors for insect herbivores. New Phytol 184:644–656. https://doi.org/10.1111/j.1469-8137.2009.03002.x

Peroutka R, Schulzová V, Botek P, Hajšlová J (2007) Analysis of furanocoumarins in vegetables (Apiaceae) and citrus fruits (Rutaceae). J Sci Food Agric 87:2152–2163. https://doi.org/10.1002/jsfa.2979

Peumans WJ, Van Damme EJM (1995) Lectins as plant defense proteins. Plant Physiol 109:347–352. https://doi.org/10.1104/pp.109.2.347

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci 5:439–445. https://doi.org/10.1016/S1360-1385(00)01741-6

Pierantoni M, Tenne R, Rephael B, et al (2018) Mineral deposits in *Ficus* leaves: Morphologies and locations in relation to function. Plant Physiol 176:1751–1763. https://doi.org/10.1104/pp.17.01516

Pieterse CMJ, Van der Does D, Zamioudis C, et al (2012) Hormonal modulation of plant immunity. Annu Rev Cell Dev Biol 28:489–521. https://doi.org/10.1146/annurev-cellbio-092910-154055

Pinot F, Beisson F (2011) Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles: Cytochrome P450 metabolizing fatty acids in plants. FEBS J 278:195–205. https://doi.org/10.1111/j.1742-4658.2010.07948.x

Plettner E (2018) Preface: Cytochrome P450. Biochim Biophys Acta BBA - Proteins Proteomics 1866:1. https://doi.org/10.1016/j.bbapap.2017.11.001

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, pp 51–64

Poulos TL, Finzel BC, Gunsalus IC, et al (1985) The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. J Biol Chem 260:16122–16130

Poulos TL, Finzel BC, Howard AJ (1987) High-resolution crystal structure of cytochrome P450cam. J Mol Biol 195:687–700. https://doi.org/10.1016/0022-2836(87)90190-2

Poulos TL, Johnson ER (2015) Structures of Cytochrome P450 Enzymes. In: Cytochrome P450, Ortiz de Montellano P. (eds). Springer, Cham, pp 3–32

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3:827–837. https://doi.org/10.1038/nrg928

Qi X, Bakht S, Leggett M, et al (2004) A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. Proc Natl Acad Sci 101:8233–8238. https://doi.org/10.1073/pnas.0401301101

Rajniak J, Giehl RFH, Chang E, et al (2018) Biosynthesis of redox-active metabolites in response to iron deficiency in plants. Nat Chem Biol 14:442–450. https://doi.org/10.1038/s41589-018-0019-2

Rambaut A, Drummond AJ, Xie D, et al (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901–904. https://doi.org/10.1093/sysbio/syy032

Ravichandran K, Boddupalli S, Hasermann C, et al (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science 261:731–736. https://doi.org/10.1126/science.8342039

Reinold S, Hahlbrock K (1997) *In situ* localization of phenylpropanoid biosynthetic mRNAs and proteins in parsley (*Petroselinum crispum*). Bot Acta 110:431–443. https://doi.org/10.1111/j.1438-8677.1997.tb00660.x

Reitz SR, Karowe DN, Diawara MM, Trumble JT (1997) Effects of elevated atmospheric carbon dioxide on the growth and linear furanocoumarin content of celery. J Agric Food Chem 45:3642–3646. https://doi.org/10.1021/jf970383t

Remmert M, Biegert A, Hauser A, Söding J (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 9:173–175. https://doi.org/10.1038/nmeth.1818

Ro D-K, Arimura G-I, Lau SYW, et al (2005) Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. Proc Natl Acad Sci 102:8060–8065. https://doi.org/10.1073/pnas.0500825102

Roberts AG, Cheesman MJ, Primak A, et al (2010) Intramolecular heme ligation of the cytochrome P450 2C9 R108H mutant demonstrates pronounced conformational flexibility of the B–C loop region: implications for substrate binding. Biochemistry 49:8700–8708. https://doi.org/10.1021/bi100911q

Roberts WK, Selitrennikoff CP (1988) Plant and bacterial chitinases differ in antifungal activity. J Gen Microbiol 134:169–176. https://doi.org/10.1099/00221287-134-1-169

Robineau T, Batard Y, Nedelkina S, et al (1998) The chemically inducible plant cytochrome P450 CYP76B1 actively metabolizes phenylureas and other xenobiotics. Plant Physiol 118:1049–1056. https://doi.org/10.1104/pp.118.3.1049

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Ronquist F, Teslenko M, van der Mark P, et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Roselli S, Olry A, Vautrin S, et al (2017) A bacterial artificial chromosome (BAC) genomic approach reveals partial clustering of the furanocoumarin pathway genes in parsnip. Plant J 89:1119–1132. https://doi.org/10.1111/tpj.13450

Rovinski JM, Sneden AT (1984) Furanocoumarins from *Maquira calophylla*. J Nat Prod 47:557–557. https://doi.org/10.1021/np50033a036

Rubnov S, Kashman Y, Rabinowitz R, et al (2001) Suppressors of cancer cell proliferation from Fig (*Ficus carica*) resin: Isolation and structure elucidation. J Nat Prod 64:993–996. https://doi.org/10.1021/np000592z

Ruiz N, Ward D, Saltz D (2002) Calcium oxalate crystals in leaves of *Pancratium sickenbergeri*: constitutive or induced defence? Funct Ecol 16:99–105. https://doi.org/10.1046/j.0269-8463.2001.00594.x

Rupasinghe S, Baudry J, Schuler MA (2003) Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. Protein Eng Des Sel 16:721–731. https://doi.org/10.1093/protein/gzg094

Rupasinghe S, Schuler MA (2006) Homology modeling of plant cytochrome P450s. Phytochem Rev 5:473–505. https://doi.org/10.1007/s11101-006-9028-y

Ryan CA (1990) Protease inhibitors in plants: genes for improving defenses against insects and pathogens. Annu Rev Phytopathol 28:425–449. https://doi.org/10.1146/annurev.py.28.090190.002233

Sablok G, Wu X, Kuo J, et al (2013) Combinational effect of mutational bias and translational selection for translation efficiency in tomato (*Solanum lycopersicum*) cv. Micro-Tom. Genomics 101:290–295. https://doi.org/10.1016/j.ygeno.2013.02.008

Saeed MA, Sabir AW (2002) Irritant potential of triterpenoids from *Ficus carica* leaves. Fitoterapia 73:417–420. https://doi.org/10.1016/S0367-326X(02)00127-2

Salminen J-P, Karonen M, Sinkkonen J (2011) Chemical ecology of tannins: recent developments in tannin chemistry reveal new structures and structure-activity patterns. Chem - Eur J 17:2806–2816. https://doi.org/10.1002/chem.201002662

Sarker S, Nahar L (2004) Natural Medicine:The Genus Angelica. Curr Med Chem 11:1479–1500. https://doi.org/10.2174/0929867043365189

Sarrion-Perdigones A, Falconi EE, Zandalinas SI, et al (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. PLoS ONE 6:e21622. https://doi.org/10.1371/journal.pone.0021622

Sarrion-Perdigones A, Palaci J, Granell A, Orzaez D (2014) Design and construction of multigenic constructs for plant biotechnology using the GoldenBraid cloning strategy. In: Valla S, Lale R (eds) DNA Cloning and Assembly Methods. Humana Press, Totowa, NJ, pp 133–151

Sarrion-Perdigones A, Vazquez-Vilar M, Palaci J, et al (2013) GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. Plant Physiol 162:1618–1631. https://doi.org/10.1104/pp.113.217661

Sawada Y, Kinoshita K, Akashi T, et al (2002) Key amino acid residues required for aryl migration catalysed by the cytochrome P450 2-hydroxyisoflavanone synthase. Plant J 31:555–564. https://doi.org/10.1046/j.1365-313X.2002.01378.x

Schalk M, Croteau R (2000) A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase. Proc Natl Acad Sci 97:11948–11953. https://doi.org/10.1073/pnas.97.22.11948

Schalk M, Nedelkina S, Schoch G, et al (1999) Role of unusual amino acid residues in the proximal and distal heme regions of a plant P450, CYP73A1. Biochemistry 38:6093–6103. https://doi.org/10.1021/bi982989w

Schlumbaum A, Mauch F, Vögeli U, Boller T (1986) Plant chitinases are potent inhibitors of fungal growth. Nature 324:365–367. https://doi.org/10.1038/324365a0

Schoch GA, Attias R, Le Ret M, Werck-Reichhart D (2003) Key substrate recognition residues in the active site of a plant cytochrome P450, CYP73A1. Homology model guided site-directed mutagenesis. Eur J Biochem 270:3684–3695. https://doi.org/10.1046/j.1432-1033.2003.03739.x

Schopfer CR, Kochs G, Lottspeich F, Ebel J (1998) Molecular characterization and functional expression of dihydroxypterocarpan 6a-hydroxylase, an enzyme specific for pterocarpanoid phytoalexin biosynthesis in soybean (*Glycine max* L.). FEBS Lett 432:182–186. https://doi.org/10.1016/S0014-5793(98)00866-7

Schuler MA (1996) The role of cytochrome P450 monooxygenases in plant-insect interactions'. Plant Physiol 112:1411–1419. https://doi.org/10.1104/pp.112.4.1411

Schuler MA, Rupasinghe SG (2011) Molecular and structural perspectives on cytochrome P450s in plants. In: Advances in Botanical Research. Elsevier, pp 263–307

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annu Rev Plant Biol 54:629–667. https://doi.org/10.1146/annurev.arplant.54.031902.134840

Scott BR, Pathak MA, Mohn GR (1976) Molecular and genetic basis of furocoumarin reactions. Mutat Res Genet Toxicol 39:29–74. https://doi.org/10.1016/0165-1110(76)90012-9

Seifert A, Vomund S, Grohmann K, et al (2009) Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. ChemBioChem 10:853–861. https://doi.org/10.1002/cbic.200800799

Seigler DS (1998) Plant secondary metabolism, Springer Science&Business Media. New York, NY, USA

Sgarbieri VC, Gupte SM, Kramer DE, Whitaker JR (1964) *Ficus* enzymes I. Separation of the proteolytic enzymes of *Ficus carica* and *Ficus glabrata* latices. J Biol Chem 239:2170–2177

Shimada N, Akashi T, Aoki T, Ayabe S (2000) Induction of isoflavonoid pathway in the model legume Lotus japonicus: molecular characterization of enzymes involved in phytoalexin biosynthesis. Plant Sci 160:37–47. https://doi.org/10.1016/S0168-9452(00)00355-1

Shimura K, Okada A, Okada K, et al (2007) Identification of a biosynthetic gene cluster in rice for momilactones. J Biol Chem 282:34013–34018. https://doi.org/10.1074/jbc.M703344200

Shirasawa K, Yakushiji H, Nishimura R, et al (2019) The *Ficus erecta* genome to identify the Ceratocystis canker resistance gene for breeding programs in common fig (*F. carica*). Plant J 102:1313–1322. https://doi.org/10.1111/tpj.14703

Siminszky B, Gavilano L, Bowen SW, Dewey RE (2005) Conversion of nicotine to nornicotine in Nicotiana tabacum is mediated by CYP82E4, a cytochrome P450 monooxygenase. Proc Natl Acad Sci 102:14919–14924. https://doi.org/10.1073/pnas.0506581102

Singh A, Singh S, Singh IK (2016) Recent insights into the molecular mechanism of jasmonate signaling during insect-plant interaction. Australas Plant Pathol 45:123–133. https://doi.org/10.1007/s13313-015-0392-1

Sintupachee S, Promden W, Ngamrojanavanich N, et al (2015) Functional expression of a putative geraniol 8-hydroxylase by reconstitution of bacterially expressed plant CYP76F45 and NADPH-cytochrome P450 reductase CPR I from Croton stellatopilosus Ohba. Phytochemistry 118:204–215. https://doi.org/10.1016/j.phytochem.2015.08.005

Sirim D, Widmann M, Wagner F, Pleiss J (2010) Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. BMC Struct Biol 10:34. https://doi.org/10.1186/1472-6807-10-34

Sirisha N, Sreenivasulu M, Sangeeta K, Chetty CM (2010) Antioxidant properties of *Ficus* species–a review. Int J PharmTech Res 2:2174–2182

Siwinska J, Siatkowska K, Olry A, et al (2018) Scopoletin 8-hydroxylase: a novel enzyme involved in coumarin biosynthesis and iron-deficiency responses in *Arabidopsis*. J Exp Bot 69:1735–1748. https://doi.org/10.1093/jxb/ery005

Smyth TJP, Ramachandran V, Brooks P, Smyth WF (2012) Investigation of antibacterial phytochemicals in the bark and leaves of *Ficus coronata* by high-performance liquid chromatography-electrospray ionization-ion trap mass spectrometry (HPLC-ESI-MSn) and ESI-MSn. Electrophoresis 33:713–718. https://doi.org/10.1002/elps.201100302

Solomon A, Golubowicz S, Yablowicz Z, et al (2006) Antioxidant activities and anthocyanin content of fresh fruits of common Fig (*Ficus carica* L.). J Agric Food Chem 54:7717–7723. https://doi.org/10.1021/jf060497h

Soltani F, Mosaffa F, Iranshahi M, et al (2010) Auraptene from *Ferula szowitsiana* protects human peripheral lymphocytes against oxidative stress. Phytother Res 24:85–89. https://doi.org/10.1002/ptr.2874

Son J-H, Jin H, You H-S, et al (2017) Five cases of phytophotodermatitis caused by fig leaves and relevant literature review. Ann Dermatol 29:86–90. https://doi.org/10.5021/ad.2017.29.1.86

Sonawane PD, Pollier J, Panda S, et al (2017) Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. Nat Plants 3:16205. https://doi.org/10.1038/nplants.2016.205

Song G, Walworth A, Hancock JF (2012) Factors influencing *Agrobacterium*-mediated transformation of switchgrass cultivars. Plant Cell Tissue Organ Cult PCTOC 108:445–453. https://doi.org/10.1007/s11240-011-0056-y

Song WC, Funk CD, Brash AR (1993) Molecular cloning of an allene oxide synthase: a cytochrome P450 specialized for the metabolism of fatty acid hydroperoxides. Proc Natl Acad Sci 90:8519–8523. https://doi.org/10.1073/pnas.90.18.8519

Sosnovsky Y (2016) Sucking herbivore assemblage composition on greenhouse *Ficus* correlates with host plant leaf architecture. Arthropod-Plant Interact 10:55–69. https://doi.org/10.1007/s11829-015-9408-6

Späth E (1937) Die natürlichen cumarine. Berichte Dtsch Chem Ges B Ser 70:A83–A117. https://doi.org/10.1002/cber.19370700648

Späth E, Okahara K, Kuffner F (1937) Die identität von ficusin mit psoralen. Berichte Dtsch Chem Ges B Ser 70:73–73. https://doi.org/10.1002/cber.19370700115

Spoel SH, Dong X (2008) Making sense of hormone crosstalk during plant immune responses. Cell Host Microbe 3:348–351. https://doi.org/10.1016/j.chom.2008.05.009

Stamp N (2003) Out of the quagmire of plant defense hypotheses. Q Rev Biol 78:23–55. https://doi.org/10.1086/367580

Stamp N (2004) Can the growth-differentiation balance hypothesis be tested rigorously? Oikos 107:439–448. https://doi.org/10.1111/j.0030-1299.2004.12039.x

Stanjek V, Boland W (1998) Biosynthesis of angular furanocoumarins: Mechanism and stereochemistry of the oxidative dealkylation of columbianetin to angelicin in Heracleum mantegazzianum (Apiaceae). Helv Chim Acta 81:1596–1607. https://doi.org/10.1002/(SICI)1522-2675(19980909)81:9<1596::AID-HLCA1596>3.0.CO;2-F

Stanjek V, Miksch M, Boland W (1997) Stereoselective syntheses of deuterium labelled marmesins; valuable metabolic probes for mechanistic studies in furanocoumarin biosynthesis. Tetrahedron 53:17699–17710. https://doi.org/10.1016/S0040-4020(97)10237-X

Steemans P, Herisse AL, Melvin J, et al (2009) Origin and radiation of the earliest vascular land plants. Science 324:353–353. https://doi.org/10.1126/science.1169659

Stevens PF (2001) Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since]. http://www.mobot.org/MOBOT/Research/APweb/. Accessed 13 May 2020

Stone BA, Clarke AE (1992) Chemistry and Biology of (1→3)-β-glucans, Bundoora: La Trobe University Press.

Studer G, Rempfer C, Waterhouse AM, et al (2020) QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics 36:1765–1771. https://doi.org/10.1093/bioinformatics/btz828

Sun W, Xue H, Liu H, et al (2020) Controlling chemo- and regioselectivity of a plant P450 in yeast cell toward rare licorice triterpenoid biosynthesis. ACS Catal 10:4253–4260. https://doi.org/10.1021/acscatal.0c00128

Sung P-H, Huang F-C, Do Y-Y, Huang P-L (2011) Functional expression of geraniol 10-hydroxylase reveals its dual function in the biosynthesis of terpenoid and phenylpropanoid. J Agric Food Chem 59:4637–4643. https://doi.org/10.1021/jf200259n

Swaminathan S, Morrone D, Wang Q, et al (2009) CYP76M7 is an ent-cassadiene C11 -hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. Plant Cell 21:3315–3325. https://doi.org/10.1105/tpc.108.063677

Taira T, Ohdomari A, Nakama N, et al (2005) Characterization and antifungal activity of Gazyumaru (*Ficus microcarpa*) latex chitinases: Both the chitin-binding and the antifungal activities of class I chitinase are reinforced with increasing ionic strength. Biosci Biotechnol Biochem 69:811–818. https://doi.org/10.1271/bbb.69.811

Takahashi S, Zhao Y, O'Maille PE, et al (2005) Kinetic and molecular analysis of 5-epiaristolochene 1,3-dihydroxylase, a cytochrome P450 enzyme catalyzing successive hydroxylations of sesquiterpenes. J Biol Chem 280:3686–3696. https://doi.org/10.1074/jbc.M411870200

Takahashi T, Okiura A, Kohno M (2017) Phenylpropanoid composition in fig (*Ficus carica* L.) leaves. J Nat Med 71:770–775. https://doi.org/10.1007/s11418-017-1093-6

Takahashi T, Okiura A, Saito K, Kohno M (2014) Identification of phenylpropanoids in fig (*Ficus carica* L.) leaves. J Agric Food Chem 62:10076–10083. https://doi.org/10.1021/jf5025938

Takei K, Yamaya T, Sakakibara H (2004) Arabidopsis CYP735A1 and CYP735A2 encode cytokinin hydroxylases that catalyze the biosynthesis of trans-zeatin. J Biol Chem 279:41866–41872. https://doi.org/10.1074/jbc.M406337200

Tanaka K, Choi J, Cao Y, Stacey G (2014) Extracellular ATP acts as a damage-associated molecular pattern (DAMP) signal in plants. Front Plant Sci 5:. https://doi.org/10.3389/fpls.2014.00446

Terras Frg, Schoofs Hme, Thevissen K, et al (1993) Synergistic enhancement of the antifungal activity of wheat and barley thionins by radish and oilseed rape 2S albumins and by barley trypsin inhibitors. Plant Physiol 103:1311–1319. https://doi.org/10.1104/pp.103.4.1311

Teutsch HG, Hasenfratz MP, Lesot A, et al (1993) Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. Proc Natl Acad Sci 90:4102–4106. https://doi.org/10.1073/pnas.90.9.4102

The Angiosperm Phylogeny Group (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 181:1–20. https://doi.org/10.1111/boj.12385

Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. Science 324:203–207. https://doi.org/10.1126/science.1169375

Tollrian R, Harvell CD (eds) (1999) The ecology and evolution of inducible defenses, Princeton University Press. Princeton, New Jersey

Trott O, Olson AJ (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461. https://doi.org/10.1002/jcc.21334

Trumble JT, Millar JG, Ott DE, Carson WC (1992) Seasonal patterns and pesticidal effects on the phototoxic linear furanocoumarins in celery, Apium graveolens L. J Agric Food Chem 40:1501–1506. https://doi.org/10.1021/jf00021a006

Tuomi J (1992) Toward integration of plant defence theories. Trends Ecol Evol 7:365–367. https://doi.org/10.1016/0169-5347(92)90005-V

Umezawa T, Okamoto M, Kushiro T, et al (2006) CYP707A3, a major ABA 8'-hydroxylase involved in dehydration and rehydration response in *Arabidopsis thaliana*. Plant J 46:171–182. https://doi.org/10.1111/j.1365-313X.2006.02683.x

Uno Y, Iwasaki K, Yamazaki H, Nelson DR (2011) Macaque cytochromes P450: nomenclature, transcript, gene, genomic structure, and function. Drug Metab Rev 43:346–361. https://doi.org/10.3109/03602532.2010.549492

Urban P, Cullin C, Pompon D (1990) Maximizing the expression of mammalian cytochrome P-450 monooxygenase activities in yeast cells. Biochimie 72:463–472. https://doi.org/10.1016/0300-9084(90)90070-W

Urban P, Mignotte C, Kazmaier M, et al (1997) Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. J Biol Chem 272:19176–19186. https://doi.org/10.1074/jbc.272.31.19176

Urlacher VB (2006) Biotransformation of ionones by engineered cytochrome P450 BM-3. Appl Microbiol Biotechnol 70:53–59. https://doi.org/10.1007/s00253-005-0028-4

Usai G, Mascagni F, Giordani T, et al (2020) Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. Plant J 102:600–614. https://doi.org/10.1111/tpj.14635

van Velzen R, Doyle JJ, Geurts R (2019) A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. Trends Plant Sci 24:49–57. https://doi.org/10.1016/j.tplants.2018.10.005

Vazquez-Albacete D, Montefiori M, Kol S, et al (2017) The CYP79A1 catalyzed conversion of tyrosine to (E)-p-hydroxyphenylacetaldoxime unravelled using an improved method for homology modeling. Phytochemistry 135:8–17. https://doi.org/10.1016/j.phytochem.2016.11.013

Venugopala KN, Rashmi V, Odhav B (2013) Review on natural coumarin lead compounds for their pharmacological activity. BioMed Res Int 2013:1–14. https://doi.org/10.1155/2013/963248

Vialart G, Hehn A, Olry A, et al (2012) A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants: C2'H involved in umbelliferone synthesis. Plant J 70:460–470. https://doi.org/10.1111/j.1365-313X.2011.04879.x

Vieira IJC, Mathias L, Monteiro VDFF, et al (1999) A new coumarin from *Brosimum gaudichaudii* trecul. Nat Prod Lett 13:47–52. https://doi.org/10.1080/10575639908048490

Villard C, Larbat R, Munakata R, Hehn A (2019) Defence mechanisms of *Ficus*: pyramiding strategies to cope with pests and pathogens. Planta 249:617–633. https://doi.org/10.1007/s00425-019-03098-2

Vogel A (1820) Darstellung von benzoesäure aus der Tonka-bohne und aus den meliloten - oder steinklee - blumen. Ann Phys Phys Chem 64:161–166. https://doi.org/10.1002/andp.18200640205

Voigt CA (2014) Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. Front Plant Sci 5:1–7. https://doi.org/10.3389/fpls.2014.00168

Volf M, Segar ST, Miller SE, et al (2018) Community structure of insect herbivores is driven by conservatism, escalation and divergence of defensive traits in *Ficus*. Ecol Lett 21:83–92. https://doi.org/10.1111/ele.12875

Vries J, Evers JB, Dicke M, Poelman EH (2019) Ecological interactions shape the adaptive value of plant defence: Herbivore attack versus competition for light. Funct Ecol 33:129–138. https://doi.org/10.1111/1365-2435.13234

Wagner GJ (1991) Secreting glandular trichomes: more than just hairs. Plant Physiol 96:675–679. https://doi.org/10.1104/pp.96.3.675

Wahler D, Gronover CS, Richter C, et al (2009) Polyphenoloxidase silencing affects latex coagulation in *Taraxacum* species. Plant Physiol 151:334–346. https://doi.org/10.1104/pp.109.138743

Walters D (2017) Fortress Plant: How to survive when everything wants to eat you. Oxford University Press

Walters D (2011) Plant defense: warding off attack by pathogens, herbivores and parasitic plants, Oxford: Wiley-Blackwell

Wamer WG, Timmer WC, Wei RR, et al (1995) Furocoumarin-photosensitized hydroxylation of guanosine in RNA and DNA. Photochem Photobiol 61:336–340. https://doi.org/10.1111/j.1751-1097.1995.tb08618.x

Wang J, Liu Y, Cai Y, et al (2010) Cloning and Functional Analysis of Geraniol 10-Hydroxylase, a Cytochrome P450 from *Swertia mussotii* Franch. Biosci Biotechnol Biochem 74:1583–1590. https://doi.org/10.1271/bbb.100175

Wang K, Herrera-Estrella L, Van Montagu M, Zambryski P (1984) Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from Agrobacterium to the plant genome. Cell 38:455–462

Wang M, Yuan J, Qin L, et al (2020) *TaCYP 81D5*, one member in a wheat cytochrome P450 gene cluster, confers salinity tolerance via reactive oxygen species scavenging. Plant Biotechnol J 18:791–804. https://doi.org/10.1111/pbi.13247

Wang Q, Hillwig ML, Okada K, et al (2012) Characterization of CYP76M5–8 indicates metabolic plasticity within a plant biosynthetic gene cluster. J Biol Chem 287:6159–6168. https://doi.org/10.1074/jbc.M111.305599

Wang S, Wang R, Liu T, et al (2018) CYP76B74 catalyzes the 3''-hydroxylation of geranylhydroquinone in shikonin biosynthesis. Plant Physiol pp.01056.2018. https://doi.org/10.1104/pp.18.01056

Wang T, Jiao J, Gai Q-Y, et al (2017) Enhanced and green extraction polyphenols and furanocoumarins from Fig (*Ficus carica* L.) leaves using deep eutectic solvents. J Pharm Biomed Anal 145:339–345. https://doi.org/10.1016/j.jpba.2017.07.002

Ward D, Spiegel M, Saltz D (1997) Gazelle herbivory and interpopulation differences in calcium oxalate content of leaves of a desert lily. J Chem Ecol 23:333–346. https://doi.org/10.1023/B:JOEC.0000006363.34360.9d

Waterhouse A, Bertoni M, Bienert S, et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303. https://doi.org/10.1093/nar/gky427

Waxman DJ, Walsh C (1982) Phenobarbital-induced Rat Liver Cytochrome P-450. J Biol Chem 257:10446–10457

Weigel D, Ahn JH, Blázquez MA, et al (2000) Activation tagging in Arabidopsis. Plant Physiol 122:1003–1014. https://doi.org/10.1104/pp.122.4.1003

Weng J-K (2014) The evolutionary paths towards complexity: a metabolic perspective. New Phytol 201:1141–1149. https://doi.org/10.1111/nph.12416

Weng J-K, Li X, Stout J, Chapple C (2008) Independent origins of syringyl lignin in vascular plants. Proc Natl Acad Sci 105:7887–7892. https://doi.org/10.1073/pnas.0801696105

Weng J-K, Noel JP (2012) The remarkable pliability and promiscuity of specialized metabolism. Cold Spring Harb Symp Quant Biol 77:309–320. https://doi.org/10.1101/sqb.2012.77.014787

Weng J-K, Philippe RN, Noel JP (2012) The rise of chemodiversity in plants. Science 336:1667–1670. https://doi.org/10.1126/science.1217411

Werck-Reichhart D, Bak S, Paquette S (2002) Cytochromes P450. Arab Book e0028. https://doi.org/10.1199/tab.0028

Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: a success story. Genome Biol 1:reviews3003–1. https://doi.org/10.1186/gb-2000-1-6-reviews3003

Werck-Reichhart D, Hehn A, Didierjean L (2000) Cytochromes P450 for engineering herbicide tolerance. Trends Plant Sci 5:116–123. https://doi.org/10.1016/S1360-1385(00)01567-3

Weryszko-Chmielewska E, Chwil M (2017) Localisation of furanocoumarins in the tissues and on the surface of shoots of *Heracleum sosnowskyi*. Botany 95:1057–1070. https://doi.org/10.1139/cjb-2017-0043

White RF (1979) Acetylsalicylic acid (aspirin) induces resistance to tobacco mosaic virus in tobacco. Virology 99:410–412. https://doi.org/10.1016/0042-6822(79)90019-9

Wilkinson GR (2005) Drug metabolism and variability among patients in drug response. N Engl J Med 352:2211–2221. https://doi.org/10.1056/NEJMra032424

Williams PA, Cosme J, Sridhar V, et al (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Mol Cell 5:121–131. https://doi.org/10.1016/S1097-2765(00)80408-6

Wink M (2008) Plant secondary metabolism: diversity, function and its evolution. Nat Prod Commun 3:1934578X0800300. https://doi.org/10.1177/1934578X0800300801

Wu CC, Kuo-Huang LL (1997) Calcium crystals in the leaves of some species of Moraceae. Bot Bull Acad Sin 38:97–104

Wu P-L, V. Rao K, Su C-H, et al (2002) Phenanthroindolizidine alkaloids and their cytotoxicity from the leaves of *Ficus septica*. Heterocycles 57:2401–2408. https://doi.org/10.3987/COM-02-9615

Xiang H, Chen J (2004) Interspecific variation of plant traits associated with resistance to herbivory among four species of *Ficus* (Moraceae). Ann Bot 94:377–384. https://doi.org/10.1093/aob/mch153

Yamada T, Kambara Y, Imaishi H, Ohkawa H (2000) Molecular cloning of novel cytochrome P450 species induced by chemical treatments in cultured tobacco cells. Pestic Biochem Physiol 68:11–25. https://doi.org/10.1006/pest.2000.2496

Yan R, Wang Z, Ren Y, et al (2019) Establishment of efficient genetic transformation systems and application of CRISPR/Cas9 genome editing technology in Lilium pumilum DC. Fisch. and Lilium longiflorum white heaven. Int J Mol Sci 20:2920. https://doi.org/10.3390/ijms20122920

Yap VA, Loong B-J, Ting K-N, et al (2015) Hispidacine, an unusual 8,4'-oxyneolignan-alkaloid with vasorelaxant activity, and hispiloscine, an antiproliferative phenanthroindolizidine alkaloid, from *Ficus hispida* Linn. Phytochemistry 109:96–102. https://doi.org/10.1016/j.phytochem.2014.10.032

Yap VA, Qazzaz ME, Raja VJ, et al (2016) Fistulopsines A and B antiproliferative septicine-type alkaloids from *Ficus fistulosa*. Phytochem Lett 15:136–141. https://doi.org/10.1016/j.phytol.2015.12.007

Yonekura-Sakakibara K, Higashi Y, Nakabayashi R (2019) The origin and evolution of plant flavonoid metabolism. Front Plant Sci 10:943. https://doi.org/10.3389/fpls.2019.00943

Yoo SY, Bomblies K, Yoo SK, et al (2005) The 35S promoter used in a selectable marker gene of a plant transformation vector affects the expression of the transgene. Planta 221:523–530. https://doi.org/10.1007/s00425-004-1466-4

Yoshihara T, Sogawa K, Pathak MD, et al (1980) Oxalic acid as a sucking inhibitor of the brown planthopper in rice (Delphacidae, Homoptera). Entomol Exp Appl 27:149–155. https://doi.org/10.1111/j.1570-7458.1980.tb02959.x

Zabolinejad N, Maleki M, Salehi M, et al (2020) Psoralen and narrowband UVB combination provides higher efficacy in treating vitiligo compared with narrowband UVB alone: A randomised clinical trial. Australas J Dermatol 61:. https://doi.org/10.1111/ajd.13184

Zangar R, Davydov DR, Verma S (2004) Mechanisms that regulate production of reactive oxygen species by cytochrome P450. Toxicol Appl Pharmacol 199:316–331. https://doi.org/10.1016/j.taap.2004.01.018

Zangerl AR, Arntz AM, Berenbaum MR (1997) Physiological price of an induced chemical defense: photosynthesis, respiration, biosynthesis, and growth. Oecologia 109:433–441. https://doi.org/10.1007/s004420050103

Zangerl AR, Berenbaum MR (1990) Furanocoumarin induction in wild parsnip - Genetics and populational variation. Ecology 71:1933–1940. https://doi.org/10.2307/1937601

Zangerl AR, Nitao JK (1998) Optimal defence, kin conflict and the distribution of furanocoumarins among offspring of wild parsnip. Evol Ecol 12:443–457. https://doi.org/10.1023/A:1006572805289

Zangerl AR, Rutledge CE (1996) The probability of attack and patterns of constitutive and induced defense: A test of optimal defense theory. Am Nat 147:599–608. https://doi.org/10.1086/285868

Zaynoun ST, Aftimos BG, Ali LA, et al (1984) *Ficus carica*; isolation and quantification of the photoactive components. Contact Dermatitis 11:21–25. https://doi.org/10.1111/j.1600-0536.1984.tb00164.x

Zerega NJC, Gardner EM (2019) Delimitation of the new tribe Parartocarpeae (Moraceae) is supported by a 333-gene phylogeny and resolves tribal level Moraceae taxonomy. Phytotaxa 388:253. https://doi.org/10.11646/phytotaxa.388.4.1

Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18:292–298. https://doi.org/10.1016/S0169-5347(03)00033-8

Zhang L, Lu Q, Chen H, et al (2006) Identification of a cytochrome P450 hydroxylase, CYP81A6, as the candidate for the bentazon and sulfonylurea herbicide resistance gene, Bel, in rice. Mol Breed 19:59–68. https://doi.org/10.1007/s11032-006-9044-z

Zhang L, Rybczynski JJ, Langenberg WG, et al (2000) An efficient wheat transformation procedure: transformed calli with long-term morphogenic potential for plant regeneration. Plant Cell Rep 19:241–250. https://doi.org/10.1007/s002990050006

Zhu-Salzman K, Luthe DS, Felton GW (2008) Arthropod-inducible proteins: broad spectrum defenses against multiple herbivores. Plant Physiol 146:852–858. https://doi.org/10.1104/pp.107.112177

Ziegler J, Schmidt S, Strehmel N, et al (2017) *Arabidopsis* transporter ABCG37/PDR9 contributes primarily highly oxygenated coumarins to root exudation. Sci Rep 7:3704. https://doi.org/10.1038/s41598-017-03250-6

Zobel AM, Brown SA (1988) Determination of furanocoumarins on the leaf surface of *Ruta graveolens* with an improved extraction technique. J Nat Prod 51:941–946. https://doi.org/10.1021/np50059a021

Zobel AM, Brown SA (1989) Histological localization of furanocoumarins in *Ruta graveolens* shoots. Can J Bot 67:915–921. https://doi.org/10.1139/b89-120

Zobel AM, Brown SA (1991a) Psoralens on the surface of seeds of rutaceae and fruits of umbelliferae and leguminosae. Can J Bot 69:485–488. https://doi.org/10.1139/b91-065

Zobel AM, Brown SA (1991b) Psoralens in senescing leaves of *Ruta graveolens*. J Chem Ecol 17:1801–1810. https://doi.org/10.1007/BF00993729

Zobel AM, Brown SA (1993) Influence of low-intensity ultraviolet radiation on extrusion of furanocoumarins to the leaf surface. J Chem Ecol 19:939–952. https://doi.org/10.1007/BF00992529

Zobel AM, Brown SA, Glowniak K (1990) Localization of furanocoumarins in leaves, fruits, and seeds of plants causing contact photodermatitis. Planta Med 56:571–572. https://doi.org/10.1055/s-2006-961167

Zumwalt JG, Neal JJ (1993) Cytochromes P450 from *Papilio polyxenes*: adaptations to host plant allelochemicals. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol 106:111–118. https://doi.org/10.1016/0742-8413(93)90261-I

Züst T, Agrawal AA (2017) Trade-offs between plant growth and defense against insect herbivory: An emerging mechanistic synthesis. Annu Rev Plant Biol 68:513–534. https://doi.org/10.1146/annurev-arplant-042916-040856

Popular references

Popular references

~

Popular references were hidden in almost all the main titles of this document.
Did you find them all?

How to produce your furanocoumarins: the hidden pathway

How to Train Your Dragon: The Hidden World. Dean DeBlois (2019). DreamWorks Animation, Dentsu, Fuji Television. Based on the book series of the same name by Cressida Cowell.

💗 My favourite animated film of the list!

A la croisée des voies, ou comment produire des furocoumarines

His Dark Materials (*À la Croisée des Mondes*)*.* Philip Pullman (1995-2000). Scholastic. The series includes 3 books: *Northern Lights*, *The Subtle Knife*, *The Amber Spyglass*.

Chapter I – State of art (No hidden reference)

I. Journey to the centre of plant defence
Voyage au centre de la Terre (*Journey to the Center of the Earth*)*.* Jules Verne (1864). P. J. Hetzel.

A. Through the looking-glass: open your eyes, don't be plant blind!
Through the Looking-Glass, and What Alice Found There. Lewis Carroll (1871). Macmillan.

B. The defence in our plants: from martial strategies to the cost of war
The Fault in Our Stars. John Green (2012). Dutton Books.

C. Feed me kill me: agronomical interest and defence of *Ficus carica*
Heal Me Kill Me. Shaka Ponk (2014). Album: *The White Pixel Ape (Smoking Isolate to Keep in Shape).* Label: Tôt ou tard.

💗 Shaka Ponk, completely crazy, highly talented, my favourite music group ever!
They are Shaka Ponk, we are little Monkeyz!

D. Smells like toxic furanocoumarins
Smells Like Teen Spirit. Nirvana (1991). Album: *Nevermind.* Label: DGC.

E. The furanocoumarin pathway must go on
The Show Must Go On. Queen (1991). Album: *Innuendo.* Labels: Parlophone, Hollywood.

II. Endless P450s most beautiful (Triple reference)

"Endless forms most beautiful" quote from *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Charles Darwin (1859) J. Murray.

Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom. Sean B. Carroll (2005). W. W. Norton.

Endless Forms Most Beautiful. Nightwish (2015). Album: *Endless Forms Most Beautiful.* Label: Nuclear Blast.

A. Introduction: all you need is a P450
All You Need Is Love. The Beatles (1967). Non-album single. Labels: Parlophone, Capitol.

B. A brief history of P450s: researches and discoveries over time
A Brief History of Time: From the Big Bang to Black Holes. Stephen Hawking (1988). Bantam Dell Publishing Group.

C. One classification to name them all
"One ring to rule them all" quote from *The Fellowship of the Ring*, first book of *The Lord of the Rings* trilogy. J. R. R. Tolkien (1954). George Allen & Unwin.

D. If you please draw me a functional P450: from structure to activity
« *S'il vous plaît… dessine-moi un mouton !* » (*"If you please… draw me a sheep!")* quote from *Le Petit Prince* (*The Little Prince*). Antoine de Saint-Exupéry (1943). Reynal & Hitchcock.

E. Highway to phytochemistry: P450s' importance in the plant kingdom
Highway to Hell. AC/DC (1979). Album: *Highway to Hell.* Label: Albert Productions.

F. On the origin of P450s genes: evolutive story of P450s in land plants
On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. Charles Darwin (1859). J. Murray.

III. Objective and approach of the PhD (No hidden reference)

CHAPTER II – WHOLE NEW GENES
A Whole New World. Brad Kane, Lea Salonga (1992). From Walt Disney's animated film *Aladdin*. Album: *Aladdin: Original Motion Picture Soundtrack.* Label: Walt Disney.

A. Introduction and strategy: in search of the lost gene
À la Recherche du Temps Perdu (*In Search of Lost Time*). Marcel Proust (1913-1927). B. Grasset, Gallimard.

B. Candidate genes and how to find them
Fantastic Beasts and Where to Find Them. David Yates (2016). Warner Bros. Pictures, Heyday Films. Based on the book of the same name by J. K. Rowling.

C. I'll make an enzyme out of you: cloning and expression of the P450s
I'll Make a Man Out of You. Donny Osmond (1998). From Walt Disney's animated film *Mulan*. Album: *Mulan: An Original Walt Disney Records Soundtrack.* Label: Walt Disney.

D. Convert this and I'll love you: enzyme assays and characterisation
Hate This and I'll Love You. Muse (1999). Album: Showbiz*. Labels: Mushroom, Taste.

E. Discussion: another P450 in the pathway
Another Brick in the Wall. Pink Floyd (1979). Album: The Wall*. Labels: Harvest, Columbia.

CHAPTER III – ONCE UPON A P450 (Multiple references)

Once Upon a Time. Edward Kitsis, Adam Horowitz (2011-2018). ABC Studios.

Once Upon a December. Liz Callaway (1997). From Fox Animation Studios animated film *Anastasia*. Album: *Anastasia: Music from the Motion Picture.* Label: Atlantic.

And, basically, any fairy tale starting with "*Once upon a time*".

A. Back to the past: introduction and strategy
Back to the Future. Robert Zemeckis (1985). Universal Pictures, Amblin Entertainment.

B. Data mining: the P450s coming out of the nitrogen fixing clade
Fire Coming Out Of The Monkey's Head. Gorillaz (2005). Album: Demon Days*. Labels: Parlophone, Virgin Records.

C. Inferring phylogenies: the realm of the elder genes
The Realm of the Elderlings. Robin Hobb (1995-2017). It includes 16 major books from 5 series: *The Farseer Trilogy* (1995-1997), *The Liveship Traders Trilogy* (1998-2000), *The Tawny Man Trilogy* (2001-2003), *The Rain Wilds Chronicles* (2009-2013), *The Fitz and the Fool Trilogy* (2014-2017).

> These are by far the best books I have ever read. Amazing story. Wonderfully written. Nothing left to chance. More than books; this is an adventure that must be lived.

D. Discussion: the story o' my P450s
Story O' my LF. Shaka Ponk (2014). Album: *The White Pixel Ape (Smoking Isolate to Keep in Shape).* Label: Tôt ou tard.

CHAPTER IV – IN THE ACTIVE SITE OF THE MARMESIN SYNTHASES

In the Dark of the Night. Jim Cummings (1997). From Fox Animation Studios animated film *Anastasia*. Album: *Anastasia: Music from the Motion Picture.* Label: Atlantic.

A. Rise of the marmesin synthases: introduction and strategy
Rise of the Guardians. Peter Ramsey (2012). DreamWorks Animation. Based on *The Guardians of Childhood and The Man in the Moon* by William Joyce.

B. The molecular shape of your P450: 3D modelling and docking (Double reference)

Shape of you. Ed Sheeran (2017). Album: ÷. Labels: Asylum, Atlantic.

The Molecular Shape of You. A Capella Science (2017). Scientific parody of the above song.

C. Finding key amino acids influencing the docking of the DMS
Finding Nemo. Andrew Stanton (2003). Walt Disney Pictures, Pixar Animation Studios.

Well, sorry about it, but I was out of imagination for this one.

D. Site-directed mutagenesis: 4 amino acids, and nothing else matters?
Nothing Else Matters. Metallica (1992). Album: *Metallica.* Label: Elektra.

E. Gotta dock them all: additional dockings with the F112-like
"*Gotta catch 'em all!*" lyrics / alternative title of the *Pokémon Theme* song. Jason Paige (1999). Album: *Pokémon 2.B.A. Master*. Theme song for the first season of the Pokémon anime: Kunihiko Yuyama, 4Kids Entertainment, The Pokémon Company International.

F. Discussion: a single amino acid is missing, and all begins anew
« Un seul être vous manque et tout est dépeuplé » ("*A single entity is missing, and all becomes a barren waste*") quote from the poem *L'isolement* in *Méditations poétiques* (*Isolation* in *Poetic meditation)*. Alphonse de Lamartine (1820).

CHAPTER V – THE COST OF FURANOCOUMARINS

A Resurrection of Magic. Kathleen Duey (2008-2009). It includes two books: Skin Hunger (2008) and Sacred Scars (2009). In French, *A Resurrection of Magic* has been translated into *Le Prix de la magie*, which means "the cost of magic".

I read the books in French, as a teenager. They were initially supposed to be part of a trilogy, but there will be no third book. Sadly, the author passed away this year. Because of the "bad" translation, this reference does not really make sense. Yet, for me, this is a tribute to a talented author and her powerful story which can now only end in the reader's imagination.

A. Introduction: do tomatoes dream of toxic furanocoumarins?
Do Androids Dream of Electric Sheep? Philip K. Dick (1968). Doubleday.

B. Brick by brick: the GoldenBraid multi-genic constructions
Brick by Brick. Arctic Monkeys (2011). Album: *Suck It and See.* Label: Domino.

C. The tomatoes of evil: tomato transformation and regeneration
Les Fleurs du mal (*Flowers of Evil).* Charles Baudelaire (1857). A. Poulet-Malassis.

D. Discussion: get a better plasmid, don't give up the transformations
"*Get up, stand up, don't give up the fight*" lyrics from the song *Get Up, Stand Up*. Bob Marley and The Wailers (1973). Album: Burnin'. Labels: Tuff, GongIsland.

CHAPTER VI – GENERAL CONCLUSION AND PERSPECTIVES (No hidden reference)

A. Into the unknown steps of the furanocoumarin pathway
Into the Unknown. Idina Menzel (2019). From Walt Disney's animated film *Frozen II.* Album: *Frozen II (Original Motion Picture Soundtrack).* Label: Walt Disney.

B. Too much furanocoumarins will cost you
Too Much Love Will Kill You. Brian May (1992). Album: *Back to the Light.* Label: Parlophone, Hollywood.

Supplemental data

Supp. Table 1 Details of the candidate genes. Expression levels of the different contigs from the RNA-seq library corresponding to every candidate (Kitajima et al. 2018). When the genetic data from F. religiosa (1KP) have been used as a frame to assemble the contigs, or to fill the gap between the contigs, the corresponding accession number have also been specified.

	Candidate	Contigs from the RNA-seq library	Expression patterns in the three latexes						Acc. num. of the *F. religiosa* sequences used to assemble the contigs	Length of the final amino acid sequence
			rpkm			Percentage				
			Fruit	Petiole	Trunk	Fruit	Petiole	Trunk		
P450s	1	TRINITY_DN39161_c0_g1_i1	0,01	1,47	0,27	0,7%	83,7%	15,6%	---	507
	2	TRINITY_DN40663_c2_g1_i1	0,12	10,87	2,54	0,9%	80,4%	18,8%	---	504
	3	TRINITY_DN34694_c0_g1_i1	1,88	7,34	2,63	15,9%	62,0%	22,2%	---	497
		TRINITY_DN36863_c1_g1_i1	1,94	6,64	2,87	16,9%	58,0%	25,1%		
	4	TRINITY_DN36440_c0_g1_i1	1,80	16,17	1,15	9,4%	84,6%	6,0%	---	497
		TRINITY_DN37681_c0_g1_i1	1,77	15,88	1,02	9,5%	85,1%	5,5%		
	5	TRINITY_DN20434_c0_g1_i1	0,01	15,62	1,83	0,0%	89,5%	10,5%	---	506
		TRINITY_DN24530_c0_g1_i1	0,01	15,16	1,82	0,0%	89,3%	10,7%		
	6	TRINITY_DN10219_c0_g1_i1	1,35	2,79	2,11	21,6%	44,7%	33,7%	---	502
		TRINITY_DN16915_c0_g1_i1	1,43	2,98	1,97	22,4%	46,7%	30,8%		
	7	TRINITY_DN18087_c0_g1_i1	0,00	12,85	1,58	0,0%	89,1%	10,9%	EDHN-2011625	504
		TRINITY_DN19124_c0_g1_i1	0,02	8,17	0,96	0,2%	89,2%	10,5%		
		TRINITY_DN24253_c0_g1_i1	0,03	5,10	0,59	0,5%	89,3%	10,2%		
		TRINITY_DN23815_c0_g1_i1	0,01	4,59	0,51	0,2%	89,8%	10,0%		
		TRINITY_DN56115_c0_g1_i1	0,00	13,26	1,18	0,0%	91,8%	8,2%		
	8	TRINITY_DN19583_c0_g1_i1	0,02	2,23	0,35	0,6%	85,9%	13,4%	EDHN-2013627 EDHN-2013626 EDHN-2013625	506
		TRINITY_DN20865_c0_g1_i1	0,05	1,36	0,32	2,8%	78,4%	18,8%		
	9	TRINITY_DN11096_c0_g1_i1	0,29	0,59	0,57	20,1%	40,7%	39,1%	---	520
		TRINITY_DN11533_c0_g1_i1	0,37	0,74	0,69	20,7%	40,9%	38,4%		
		TRINITY_DN63824_c0_g1_i1	0,22	1,05	0,71	10,9%	53,1%	36,0%		
	10	TRINITY_DN31066_c0_g1_i1	0,80	1,14	0,63	31,0%	44,3%	24,6%	---	499
		TRINITY_DN31617_c0_g1_i1	0,57	1,12	0,60	24,8%	48,8%	26,4%		
Dioxygenases	1	TRINITY_DN40955_c4_g10_i1	224,44	459,31	243,38	24,2%	49,5%	26,3%	EDHN-2003239	357
	2	TRINITY_DN40957_c0_g3_i1	36,54	103,31	51,95	19,1%	53,9%	27,1%	EDHN-2003240	349
	3	TRINITY_DN56000_c0_g1_i1	0,12	0,43	0,04	20,8%	72,6%	6,6%	EDHN-2052579	370
	4	TRINITY_DN311_c0_g1_i1	0,20	1,34	0,62	9,3%	62,1%	28,6%	---	363
	5	TRINITY_DN69424_c0_g1_i1	0,61	1,80	1,02	17,8%	52,5%	29,6%	---	356
	6	TRINITY_DN16212_c0_g1_i1	2,39	21,65	2,14	9,1%	82,7%	8,2%	---	344
	7	TRINITY_DN69009_c0_g1_i1	0,02	0,74	0,06	2,6%	90,3%	7,1%	EDHN-2010973	314
		TRINITY_DN62753_c0_g1_i1	0,02	0,72	0,17	2,2%	79,5%	18,3%		
Methyltransferases	1	TRINITY_DN34340_c0_g2_i1	0,30	156,30	8,46	0,2%	94,7%	5,1%	---	368
		TRINITY_DN69326_c0_g1_i1	0,38	150,40	7,41	0,2%	95,1%	4,7%		
	2	TRINITY_DN16866_c0_g1_i1	0,00	2,45	0,01	0,0%	99,4%	0,6%	---	358
	3	TRINITY_DN29383_c0_g1_i1	0,01	2,72	0,30	0,2%	90,0%	9,8%	---	363
		TRINITY_DN30809_c0_g2_i1	0,01	1,71	0,41	0,5%	80,4%	19,2%		
	4	TRINITY_DN12105_c0_g1_i1	0,00	1,88	0,26	0,0%	87,8%	12,2%	EDHN-2002064 EDHN-2051022	347
		TRINITY_DN59562_c0_g1_i1	0,14	0,99	0,52	8,5%	59,9%	31,6%		
		TRINITY_DN15646_c0_g1_i1	0,01	1,32	0,49	0,8%	72,3%	26,9%		
	5	TRINITY_DN76471_c0_g1_i1	17,15	46,99	28,25	18,6%	50,9%	30,6%	---	367
		TRINITY_DN49045_c0_g1_i1	0,90	1,93	2,53	16,8%	36,1%	47,1%		
		TRINITY_DN6118_c0_g1_i1	1,26	2,10	2,92	20,0%	33,4%	46,5%		
	6	TRINITY_DN35326_c0_g1_i1	8,14	11,47	9,18	28,3%	39,8%	31,9%	---	357
	7	TRINITY_DN40899_c10_g1_i3	280,87	232,61	323,16	33,6%	27,8%	38,6%	---	368

Supp. Table 2 Summary of the dioxygenase and methyltransferase candidate genes. The relative expression levels of every candidate correspond to the expression patterns of the contigs from the RNA-seq library (Kitajima et al. 2018). The candidates have been arbitrarily numbered. More information such as the accession numbers and expression patterns can be found in **Supp. Table 1**.

Candidates		Relative expression levels in the three latexes			Length of the amino acid sequence
		Petiole	Trunk	Fruit	
Dioxygenases	1	49,5%	26,3%	24,2%	357
	2	53,9%	27,1%	19,1%	349
	3	72,6%	6,6%	20,8%	370
	4	62,1%	28,6%	9,3%	363
	5	52,5%	29,6%	17,8%	356
	6	82,7%	8,2%	9,1%	344
	7	90,3%	7,1%	2,6%	314
Methyltransferases	1	94,7%	5,1%	0,2%	368
	2	99,4%	0,6%	0,0%	358
	3	90,0%	9,8%	0,2%	363
	4	87,8%	12,2%	0,0%	347
	5	50,9%	30,6%	18,6%	367
	6	39,8%	31,9%	28,3%	357
	7	27,8%	38,6%	33,6%	368

Supp. Table 3 **Pairs of primers used to amplify the P450 candidates.** "Rev-HIS" refers to the reverse orientation of the His-tagged version of the primers.

Candidate	Orientation	Sequence
1	Direct	ATGGCTATGGATGCCCAACAAG
	Reverse	CTATGCAAATAGGTGCTTCTTTGGG
2	Direct	ATGAAATATGTCCTTTGGAGTTGCAG
	Reverse	CTAAGAAATTATCGCCACGGGAC
	Rev-HIS	CTAATGATGATGATGATGATGAGAAATTATCGCCACGGGAC
3	Direct	ATGTTCATCTACTTGCTGCTGTTTATAG
	Reverse	TCAAATTTGAGAAAGAACATTCAACATGC
	Rev-HIS	TCAATGATGATGATGATGATGAATTTGAGAAAGAACATTCAACATGC
4	Direct	ATGGAAGAACCCATATTGTACACATC
	Reverse	TTAAGAGAGAACAACATTCATAATGGGG
	Rev-HIS	TTAATGATGATGATGATGATGAGAGAGAACAACATTCATAATGGGG
5	Direct	ATGGATTTGATCACCTCTATATTGTGTTTTG
	Reverse	CTAAGCGATTGCTATAGGGACAAC
	Rev-HIS	CTAATGATGATGATGATGATGAGCGATTGCTATAGGGACAAC
6	Direct	ATGGAAGTCTTCTACCTTTACCTGTC
	Reverse	TCAAAGTTGAGATAGGAGGCGAG
	Rev-HIS	TCAATGATGATGATGATGATGAAGTTGAGATAGGAGGCGAG
7	Direct	ATGGATATTTTCACCTCCTTACTGTATC
	Reverse	CTAATGCTTTGTCGGCACGGG
	Rev-HIS	CTAATGATGATGATGATGATGATGCTTTGTCGGCACGGG
8	Direct	ATGACTGCCTTCCTTATCTTTCTTTTC
	Reverse	TTAAGAGATCTGGTCATAAGGTGTTG
9	Direct	ATGGACTTTCTCATCTCTCATTCTC
	Reverse	TTAATGTTGATAAAATTCAACAGAGAGGC
	Rev-HIS	TTAATGATGATGATGATGATGATGTTGATAAAATTCAACAGAGAGGC
10	Direct	ATGGGTAGTTGGGGGGCG
	Reverse	TCATGAAGATGTGCATTTCTCAACTTTTG
	Rev-HIS	TCAATGATGATGATGATGATGTGAAGATGTGCATTTCTCAACTTTTG

Supp. Table 4 Determination of the quantity of marmesin formed by CYP76F112 incubated in presence of variable DMS concentrations. 0.54 pmol of P450 were incubated in 1mL reaction mixtures containing various concentration of DMS. 5 to 15 reaction mixture of 1mL, performed with the same modalities, were pooled together, concentrated 50 to 150 times, and analysed by UHPLC-MS. The quantity of marmesin was quantified using the peak area obtained with the MS data. The quantity of marmesin (mol) formed in one incubation of 1mL is obtained by dividing the peak area by 1.41×10^{14} (see the materials and methods section, **Chapter VII**, **B.2.f.3**) and by the concentration factor. The specific activity (mol of marmesin per min per mol of P450) is obtained by dividing the quantity of marmesin by the duration of the incubation (1min) and the quantity of functional CYP76F112 present in the reaction mix (5.41×10^{13} mol).

		CYP76F112 Quantity of P450 for 1mL of reaction: 0.54 pmol – Incubation time: 1min				
		Concentration of the reaction mix in 100µL		**UHPLC-MS analysis: determination of the quantity of formed marmesin**		**Specific activity (mol marm. / min / mol P450)**
Rep	**[DMS]$_{ini}$ (nM)**	**Number of pooled reactions (1mL each)**	**Concentration factor**	**Quantity of marm. in the concentrated solution (Peak Area)**	**Quantity of marm. for 1mL of reaction mix (mol)**	
A1	**10**	15	150	68638	3,25E-12	**6,001**
A2	**10**	15	150	78201	3,70E-12	**6,837**
B1	**15**	15	150	Lost sample – Problem during the concentration		
B2	**15**	15	150	83547	3,96E-12	**7,305**
C1	**20**	15	150	99258	4,70E-12	**8,678**
C2	**20**	15	150	Lost sample – Problem during the analysis		
D1	**30**	15	150	131204	6,21E-12	**11,472**
D2	**30**	15	150	108447	5,13E-12	**9,482**
E1	**50**	15	150	141088	6,68E-12	**12,336**
E2	**50**	15	150	135072	6,39E-12	**11,810**
F1	**100**	5	50	68476	9,73E-12	**17,961**
F2	**100**	5	50	63289	8,99E-12	**16,601**
G1	**300**	5	50	77331	1,10E-11	**20,284**
G2	**300**	5	50	76602	1,09E-11	**20,093**

Supp. Table 5 CYP76F dataset. The sequences in red have not been included in any phylogenetic tree (pseudogene, or outside the CYP71 clan). The sequences in white and blue have been used for the first global tree. The sequences in blue have also been included in the second tree, that is a zoom of the first one. Pg stands for "PseudoGene", FS for "FrameShift", SC for "Stop Codon", and m.r. for "manually removed".

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
CYP76F110_F.carica	CYP76F110	---	---	---	---	*Ficus carica* (Moraceae)
CYP76F111_F.carica	CYP76F111	---	---	---	---	
CYP76F112_F.carica	CYP76F112	---	---	---	---	
CYP81B114_F.carica	CYP81B114	---	---	---	---	
CYP81BN4_F.carica	CYP81BN4	---	---	---	---	
CYP81CA1_F.carica	CYP81CA1	---	---	---	---	
CYP82J18_F.carica	CYP82J18	---	---	---	---	
CYP88A103_F.carica	CYP88A103	---	---	---	---	
G01_76F_F.religiosa	76F	Gene01	EDHN_scaffold_2011625	---	---	*Ficus religiosa* (Moraceae)
G02_76F_F.religiosa	76F	Gene02	EDHN_scaffold_2011626	---	---	
Pg01_76F_F.religiosa	76F	Pg01 (FS)	EDHN_scaffold_2011627	---	---	
G03_76F_F.religiosa	76F	Gene03	EDHN_scaffold_2012455	---	---	
G04_76F_F.religiosa	76F	Gene04	EDHN_scaffold_2006849	---	---	
G01_76F_F.erecta	76F	Gene01	BKCH01000037.1	Chr10; 1467776 to 1469539	m.r.	*Ficus erecta* (Moraceae)
G02_76F_F.erecta	76F	Gene02	BKCH01000037.1	Chr10; 1498959 to 1500758	m.r.	
G03_76F_F.erecta	76F	Gene03	BKCH01000037.1	Chr10; 1485033 to 1486832	m.r.	
Pg01_76F_F.erecta	76F	Pg01 (FS)	BKCH01000037.1	Chr10; 1452680 to 1454461	m.r.	
Pg02_76F_F.erecta	76F	Pg02 (1SC)	BKCH01000037.1	Chr10; 1657943 to 1660156	m.r.	
G04_76F_F.erecta	76F	Gene04	BKCH01000037.1	Chr10; 1532919 to 1535139	m.r.	
G05_76F_F.erecta	76F	Gene05	BKCH01000037.1	Chr10; 1524881 to 1526757	m.r.	
G06_76F_F.erecta	76F	Gene06	BKCH01000037.1	Chr10; 1478177 to 1480688	m.r.	
G07_76F_F.erecta	76F	Gene07	BKCH01000037.1	Chr10; 1491273 to 1494633	m.r.	
G08_76F_F.erecta	76F	Gene08	BKCH01000037.1	Chr10; 1505528 to 1508292	m.r.	
G09_76F_F.erecta	76F	Gene09	BKCH01000037.1	Chr10; 1393102 to 1395292	m.r.	
G10_76F_F.erecta	76F	Gene10	BKCH01000037.1	Chr10; 1444970 to 1447299	m.r.	
Pg03_76F_F.erecta	76F	Pg03 (1SC)	BKCH01001365.1	Chr10; 1438994 to 1441189	m.r.	
G11_76F_F.erecta	76F	Gene11	BKCH01001369.1	Chr10; 1664773 to 1666685	m.r.	
G12_76F_F.erecta	76F	Gene12	BKCH01000017.1	Chr09; 12950690 to 12952902	m.r.	
G13_76F_F.erecta	76F	Gene13	BKCH01000017.1	Chr09; 12940242 to 12942122	m.r.	
G01_76F_M.notabilis	76F	Gene01	XM_024171830.1	---	---	*Morus notabilis* (Moraceae)
G02_76F_M.notabilis	76F	Gene02	XM_010104907.2	---	---	
G03_76F_M.notabilis	76F	Gene03	XM_024171829.1	---	---	
G04_76F_M.notabilis	76F	Gene04	XM_010107479.2	---	---	
G05_76F_M.notabilis	76F	Gene05	XM_010109776.2	---	---	
G06_76G_M.notabilis	76G	Gene06	XM_010114207.1	---	---	
G01_76F_B.nivea	76F	Gene01	ACFP_scaffold_2005551	---	---	*Boehmeria nivea* (Urticaceae)
G02_76F_B.nivea	76F	Gene02	PHNS01007790.1	931823 to 934791	m.r.	
G03_76F_B.nivea	76F	Gene03	PHNS01007790.1	925948 to 927804	m.r.	
G04_76F_B.nivea	76F	Gene04	PHNS01007790.1	938908 to 942366	m.r.	
Pg01_76F_B.nivea	76F	Pg01 (FS)	PHNS01007790.1	944667to 947205	m.r.	
Pg02_76F_B.nivea	76F	Pg02 (FS)	PHNS01007790.1	949078 to 952208	m.r.	
G05_76F_B.nivea	76F	Gene05	PHNS01008384.1	466226 to 468675	m.r.	
G06_76F_B.nivea	76F	Gene06	PHNS01006891.1	3121336 to 3124016	m.r.	
G07_76F_B.nivea	76F	Gene07	PHNS01007284.1	506313 to 509244	m.r.	

Supp. Table 5 – **Continuation**

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
G01_76F_C.sativa	76F	Gene01	XM_030644098.1	---	---	*Cannabis sativa* (Cannabaceae)
G02_76F_C.sativa	76F	Gene02	XM_030643420.1	---	---	
G03_76F_C.sativa	76F	Gene03	XM_030654999.1	---	---	
G04_76F_C.sativa	76F	Gene04	XM_030633170.1	---	---	
G05_76F_C.sativa	76F	Gene05	XM_030644317.1	---	---	
G06_76F_C.sativa	76F	Gene06	UZAU01000228.1	2170135 to 2171750	m.r.	
G01_76F_P.andersonii	76F	Gene01	JXTB01000218.1	---	---	*Parasponia andersonii* (Cannabaceae)
G02_76F_P.andersonii	76F	Gene02	JXTB01000159.1	---	---	
G03_76F_P.andersonii	76F	Gene03	JXTB01000105.1	---	---	
G01_76F_T.orientale	76F	Gene01	JXTC01000188.1	---	---	*Trema orientale* (Cannabaceae)
G02_76F_T.orientale	76F	Gene02	JXTC01000331.1	---	---	
G03_76F_T.orientale	76F	Gene03	JXTC01000049.1	---	---	
G04_76F_T.orientale	76F	Gene04	JXTC01001031.1	2504 to 4445	m.r.	
G05_76E_T.orientale	76E	Gene05	JXTC01000558.1	57196 to 58814	m.r.	
G01_76F_H.lupulus	76F	Gene01	LD175939.1	3168 to 6366	m.r.	*Humulus lupulus* (Cannabaceae)
Pg01_76F_H.lupulus	76F	Pg01 (FS)	LD170049.1	10369 to 12936	m.r.	
G01_76F_Z.jujuba	76F	Gene01	XM_016038877.2	---	---	*Ziziphus jujuba* (Rhamnaceae)
G02_76F_Z.jujuba	76F	Gene02	XM_016038862.2	---	---	
G03_76F_Z.jujuba	76F	Gene03	XM_016026492.2	---	---	
G04_76G_Z.jujuba	76G	Gene04	XM_025066845.1	---	---	
G01_76F_P.avium	76F	Gene01	XM_021973699.1	---	---	*Prunus avium* (Rosaceae)
G02_76F_P.avium	76F	Gene02	XM_021974251.1	---	---	
G03_76F_P.avium	76F	Gene03	XM_021953954.1	---	---	
G04_76F_P.avium	76F	Gene04	XM_021946941.1	---	---	
G05_76F_P.avium	76F	Gene05	XM_021953913.1	---	---	
G01_76F_F.vesca	76F	Gene01	XM_011459248.1	---	---	*Fragaria vesca* subsp. Vesca (Rosaceae)
G02_76F_F.vesca	76F	Gene02	XM_004310009.2	---	---	
G03_76F_F.vesca	76F	Gene03	XM_011459247.1	---	---	
G04_76F_F.vesca	76F	Gene04	XM_011459138.1	---	---	
G05_76F_F.vesca	76F	Gene05	XM_004310160.2	---	---	
G01_76F_R.chinensis	76F	Gene01	XM_024314281.1	---	---	*Rosa chinensis* (Rosaceae)
G02_76F_R.chinensis	76F	Gene02	XM_024315574.1	---	---	
G03_76F_R.chinensis	76F	Gene03	XM_024339061.1	---	---	
G04_76F_R.chinensis	76F	Gene04	XM_024337524.1	---	---	
G05_76F_R.chinensis	76F	Gene05	XM_024339675.1	---	---	
G06_76F_R.chinensis	76F	Gene06	XM_024318077.1	---	---	
G07_76F_R.chinensis	76F	Gene07	XM_024340369.1	---	---	
G08_76F_R.chinensis	76F	Gene08	XM_024337656.1	---	---	
G09_76F_R.chinensis	76F	Gene09	XM_024312917.1	---	---	
G10_76F_R.chinensis	76F	Gene10	XM_024312916.1	---	---	
G11_76F_R.chinensis	76F	Gene11	XM_024340593.1	---	---	
G12_76F_R.chinensis	76F	Gene12	XM_024336913.1	---	---	
G01_76F_M.domestica	76F	Gene01	XM_008391597.3	---	---	*Malus domestica* (Rosaceae)
G02_76F_M.domestica	76F	Gene02	XM_008391617.3	---	---	
G03_76F_M.domestica	76F	Gene03	XM_029095914.1	---	---	
G04_76F_M.domestica	76F	Gene04	XM_029105024.1	---	---	
G05_76F_M.domestica	76F	Gene05	XM_008391592.3	---	---	
G06_76F_M.domestica	76F	Gene06	XM_029095915.1	---	---	
G07_76F_M.domestica	76F	Gene07	XM_008391905.3	---	---	

Supp. Table 5 – **Continuation and ending**

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
G01_76F_P.bretschneideri	76F	Gene01	XM_009364544.1	---	---	*Pyrus x Bretschneideri* (Rosaceae)
G02_76F_P.bretschneideri	76F	Gene02	XM_018650951.1	---	---	
G03_76F_P.bretschneideri	76F	Gene03	XM_009374042.2	---	---	
G04_76F_P.bretschneideri	76F	Gene04	XM_009374045.2	---	---	
G05_76F_P.bretschneideri	76F	Gene05	XM_018650950.1	---	---	
G06_76F_P.bretschneideri	76F	Gene06	XM_018650949.1	---	---	
G07_76F_P.bretschneideri	76F	Gene07	XM_009374039.2	---	---	
G01_76F_C.arietinum	76F	Gene01	NM_001364785.1	---	---	*Cicer arietinum* (Fabaceae)
G02_76F_C.arietinum	76F	Gene02	XM_027336245.1	---	---	
G03_76F_C.arietinum	76F	Gene03	XM_027335581.1	---	---	
G04_76F_C.arietinum	76F	Gene04	XM_004511396.3	---	---	
G05_76F_C.arietinum	76F	Gene05	NM_001364785.1	---	---	
G06_76F_C.arietinum	76F	Gene06	XM_004505535.3	---	---	
G07_76F_C.arietinum	76F	Gene07	XM_004496664.3	---	---	
G08_76F_C.arietinum	76F	Gene08	XM_004496792.3	---	---	
G09_76F_C.arietinum	76F	Gene09	XM_027336245.1	---	---	
G01_76F_C.cajan	76F	Gene01	XM_020357766.2	---	---	*Cajanus cajan* (Fabaceae)
G02_76F_C.cajan	76F	Gene02	XM_020350453.2	---	---	
G01_76F_M.truncatula	76F	Gene01	XM_003628196.3	---	---	*Medicago truncatula* (Fabaceae)
G02_76F_M.truncatula	76F	Gene02	XM_013611413.2	---	---	
G03_76F_M.truncatula	76F	Gene03	XM_003610780.2	---	---	
G04_76F_M.truncatula	76F	Gene04	XM_003628199.3	---	---	
G05_76F_M.truncatula	76F	Gene05	XM_024782607.1	---	---	
G06_76F_M.truncatula	76F	Gene06	XM_024782607.1	---	---	
G01_76F_C.sativus	76F	Gene01	XM_004147550.2	---	---	*Cucumis sativus* (Cucurbitaceae)
G02_76F_C.sativus	76F	Gene02	XM_011656478.1	---	---	
G03_76F_C.sativus	76F	Gene03	XM_004147552.2	---	---	
G04_76F_C.sativus	76F	Gene04	XM_004152021.2	---	---	
G05_76F_C.sativus	76F	Gene05	XM_004147554.2	---	---	
G01_76F_J.regia	76F	Gene01	XM_018964949.1	---	---	*Juglans regia* (Juglandaceae)
G02_76F_J.regia	76F	Gene02	XM_018964948.1	---	---	
G03_76F_J.regia	76F	Gene03	XM_018964947.1	---	---	
G04_76F_J.regia	76F	Gene04	XM_018964971.1	---	---	
G05_76F_J.regia	76F	Gene05	XM_018985782.1	---	---	
G06_76F_J.regia	76F	Gene06	XM_018976414.1	---	---	
G07_76F_J.regia	76F	Gene07	XM_018976415.1	---	---	
G08_76F_J.regia	76F	Gene08	XM_018964950.1	---	---	
G09_76F_J.regia	76F	Gene09	XM_018964214.1	---	---	
CYP71AJ1_A.majus	CYP71AJ1	---	AY532370.2	---	---	*Ammi majus* (Apiaceae)
CYP71AZ1_A.majus	CYP71AZ1	---	EF127863.1	---	---	
CYP71AZ4_P.sativa	CYP71AZ4	---	MH000219.1	---	---	*Pastinaca sativa* (Apiaceae)
CYP71AZ6_P.sativa	CYP71AZ6	---	MH000221.1	---	---	
CYP71AJ3_P.sativa	CYP71AJ3	---	EF191020	---	---	
CYP71AJ4_P.sativa	CYP71AJ4	---	EF191021	---	---	
CYP71AJ2_A.graveolens	CYP71AJ2	---	EF191022		---	*Apium graveolens* (Apiaceae)
CYP82C2_A.thaliana	CYP82C2	---	NM_119348.2	---	---	*Arabidopsis thaliana* (Brassicaceae)
CYP82C4_A.thaliana	CYP82C4	---	NM_119345.3	---	---	

Supp. Table 6 **CYP81BN dataset.** The sequences in red have not been included in any phylogenetic tree (pseudogene, or outside the CYP71 clan). The sequences in white and blue have been used for the first global tree. The sequences in blue have also been included in the second tree, that is a zoom of the first one. Pg stands for "PseudoGene", FS for "FrameShift", SC for "Stop Codon", and m.r. for "manually removed".

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
CYP81BN4_F.carica	CYP81BN4	---	---	---	---	*Ficus carica (Moraceae)*
CYP76F110_F.carica	CYP76F110	---	---	---	---	
CYP76F111_F.carica	CYP76F111	---	---	---	---	
CYP76F112_F.carica	CYP76F112	---	---	---	---	
CYP81B114_F.carica	CYP81B114	---	---	---	---	
CYP81CA1_F.carica	CYP81CA1	---	---	---	---	
CYP82J18_F.carica	CYP82J18	---	---	---	---	
CYP88A103_F.carica	CYP88A103	---	---	---	---	
G01_81_F.erecta	81	Gene01	BKCH01000005.1	Chr04; 24413323 to 24415673	m.r.	*Ficus erecta (Moraceae)*
G02_81_F.erecta	81	Gene02	BKCH01000005.1	Chr04; 24429766 to 24432599	m.r.	
G03_81_F.erecta	81	Gene03	BKCH01000005.1	Chr04; 24444234 to 24447907	m.r.	
G04_81_F.erecta	81	Gene04	BKCH01000005.1	Chr04; 24434462 to 24436786	m.r.	
G05_81_F.erecta	81	Gene05	BKCH01000005.1	Chr04; 24459571 to 24461474	m.r.	
G06_81_F.erecta	81	Gene06	BKCH01000005.1	Chr04; 24482470 to 24484885	m.r.	
G07_81_F.erecta	81	Gene07	BKCH01000005.1	Chr04; 24494446 to 24497276	m.r.	
G01_81_F.religiosa	81	Gene01	EDHN_scaffold_2011790		---	*Ficus religiosa (Moraceae)*
G02_81_F.religiosa	81	Gene02	EDHN_scaffold_2011790	---	---	
G01_81_M.notabilis	81	Gene01	XM_010097029.2	---	---	*Morus notabilis (Moraceae)*
G02_81_M.notabilis	81	Gene02	XM_024164884.1	---	---	
G03_81_M.notabilis	81	Gene03	XM_010097030.2	---	---	
Pg01_81_M.notabilis	81	Pg01 (1SC)	XM_024164879.1	---	---	
G04_81_M.notabilis	81	Gene04	XM_010097033.2	---	---	
G05_81_M.notabilis	81	Gene05	XM_010097031.2	---	---	
G01_81_B.nivea	81	Gene01	PHNS01011820.1	2275276 to 2277464	m.r.	*Boehmeria nivea (Urticaceae)*
G02_81_B.nivea	81	Gene02	PHNS01001376.1	276295 to 279142	m.r.	
G03_81_B.nivea	81	Gene03	PHNS01001376.1	269427 to 271858	m.r.	
G04_81_B.nivea	81	Gene04	PHNS01001376.1	262397 to 264935	m.r.	
G05_81_B.nivea	81	Gene05	PHNS01005842.1	3767277 to 3769657	m.r.	
G01_81_C.sativa	81	Gene01	XM_030630113.1	---	---	*Cannabis sativa (Cannabaceae)*
G02_81_C.sativa	81	Gene02	XM_030630192.1	---	---	
G03_81_C.sativa	81	Gene03	XM_030630110.1	---	---	
G04_81_C.sativa	81	Gene04	XM_030640327.1	---	---	
G05_81_C.sativa	81	Gene05	XM_030630198.1	---	---	
G06_81_C.sativa	81	Gene06	XM_030630201.1	---	---	
G07_81_C.sativa	81	Gene07	XM_030630200.1	---	---	
G08_81_C.sativa	81	Gene08	XM_030630196.1	---	---	
G09_81_C.sativa	81	Gene09	XM_030630195.1	---	---	
G10_81_C.sativa	81	Gene10	XM_030643149.1	---	---	
G11_81_C.sativa	81	Gene11	XM_030630194.1	---	---	
G12_81_C.sativa	81	Gene12	XM_030630193.1	---	---	
G13_81_C.sativa	81	Gene13	XM_030630115.1	---	---	
G14_81_C.sativa	81	Gene14	XM_030630199.1	---	---	

Supp. Table 6 – **Continuation**

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
G01_81_P.andersonii	81	Gene01	JXTB01000251.1	55329 to 57338	m.r.	*Parasponia andersonii* (Cannabaceae)
G01_81_T.orientale	81	Gene01	JXTC01000555.1	173979 to 176049	m.r.	
G02_81_T.orientale	81	Gene02	JXTC01000555.1	181429 to 183692	m.r.	*Trema orientale* (Cannabaceae)
G03_81_T.orientale	81	Gene03	JXTC01000555.1	155341 to 157863	m.r.	
G04_81_T.orientale	81	Gene04	JXTC01000555.1	146183 to 148067	m.r.	
G05_81_T.orientale	81	Gene05	JXTC01000555.1	132412 to 134506	m.r.	
Pg01_81_H.lupulus	81	Pg01 (FS)	LD150659.1		m.r.	*Humulus lupulus* (Cannabaceae)
G01_81_Z.jujuba	81	Gene01	XM_016034769.2	---	---	
G02_81_Z.jujuba	81	Gene02	XM_016034757.2	---	---	
G03_81_Z.jujuba	81	Gene03	XM_016034761.2	---	---	
G04_81_Z.jujuba	81	Gene04	XM_016012829.2	---	---	
G05_81_Z.jujuba	81	Gene05	XM_016011379.2	---	---	
G06_81_Z.jujuba	81	Gene06	XM_016012830.2	---	---	
G07_81_Z.jujuba	81	Gene07	XM_016012827.2	---	---	
G08_81_Z.jujuba	81	Gene08	XM_016034756.2	---	---	*Ziziphus jujuba* (Rhamnaceae)
G09_81_Z.jujuba	81	Gene09	XM_025069204.1	---	---	
G10_81_Z.jujuba	81	Gene10	XM_016011376.2	---	---	
G11_81_Z.jujuba	81	Gene11	XM_016034765.2	---	---	
G12_81_Z.jujuba	81	Gene12	XM_016034755.1	---	---	
G13_81_Z.jujuba	81	Gene13	XM_016012831.2	---	---	
G14_81_Z.jujuba	81	Gene14	XM_016011378.2	---	---	
G15_81_Z.jujuba	81	Gene15	XM_016045026.2	---	---	
G16_81_Z.jujuba	81	Gene16	XM_016029493.2	---	---	
G01_81_P.avium	81	Gene01	XM_021975948.1	---	---	
G02_81_P.avium	81	Gene02	XM_021975949.1	---	---	
G03_81_P.avium	81	Gene03	XM_021961664.1	---	---	*Prunus avium* (Rosaceae)
G04_81_P.avium	81	Gene04	XM_021975946.1	---	---	
Pg01_81_P.avium	81	Pg01 (1SC)	XM_021947513.1	---	---	
G05_81_P.avium	81	Gene05	XM_021964542.1	---	---	
G01_81_F.vesca	81	Gene01	XM_004287941.2	---	---	
G02_81_F.vesca	81	Gene02	XM_004287937.2	---	---	*Fragaria vesca subsp. Vesca* (Rosaceae)
G03_81_F.vesca	81	Gene03	XM_004290888.2	---	---	
G04_81_F.vesca	81	Gene04	XM_004288272.2	---	---	
G01_81_R.chinensis	81	Gene01	XM_024329949.1	---	---	
G02_81_R.chinensis	81	Gene02	XM_024327652.1	---	---	
G03_81_R.chinensis	81	Gene03	XM_024325431.1	---	---	
G04_81_R.chinensis	81	Gene04	XM_024326829.1	---	---	
G05_81_R.chinensis	81	Gene05	XM_024326816.1	---	---	*Rosa chinensis* (Rosaceae)
G06_81_R.chinensis	81	Gene06	XM_024322634.1	---	---	
G07_81_R.chinensis	81	Gene07	XM_024303826.1	---	---	
G08_81_R.chinensis	81	Gene08	XM_024326583.1	---	---	
G09_81_R.chinensis	81	Gene09	XM_024311045.1	---	---	

Supp. Table 6 – **Continuation and ending**

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
G01_81_M.domestica	81	Gene01	XM_029094682.1	---	---	*Malus domestica* (Rosaceae)
G02_81_M.domestica	81	Gene02	XM_008370905.3	---	---	
G03_81_M.domestica	81	Gene03	XM_029093829.1	---	---	
G04_81_M.domestica	81	Gene04	XM_008369429.3	---	---	
G01_76F_P.bretschneideri	81	Gene01	XM_009374181.2	---	---	*Pyrus x bretschneideri* (Rosaceae)
G02_76F_P.bretschneideri	81	Gene02	XM_009374182.2	---	---	
G03_76F_P.bretschneideri	81	Gene03	XM_009374179.2	---	---	
CYP81E8_M.truncatula	CYP81E8	---	AY278229.1	---	---	*Medicago truncatula* (Fabaceae)
CYP81Q32_C.roseus	CYP81Q32	---	KF302070.1	---	---	*Catharanthus roseus* (Apocynaceae)
CYP81D1_A.thaliana	CYP81D1	---	NM_123013.4	---	---	*Arabidopsis thaliana* (Brassicaceae)
CYP81B1c_H.tuberosus	CYP81B1c	---	AJ000477.1	---	---	*Helianthus tuberosus* (Asteraceae)
G01_81_C.arietinum	81	Gene01	NM_001282318.1	---	---	*Cicer arietinum* (Fabaceae)
G02_81_C.arietinum	81	Gene02	XM_004511824.2	---	---	
G01_81_C.cajan	81	Gene01	XM_020374753.2	---	---	*Cajanus cajan* (Fabaceae)
G02_81_C.cajan	81	Gene02	XM_020369244.2	---	---	
G01_81_M.truncatula	81	Gene01	CM001218.2	14956186 to 14960642	m.r.	*Medicago truncatula* (Fabaceae)
G02_81_M.truncatula	81	Gene02	CM001220.2	38953847 to 38956118	m.r.	
G01_81_C.sativus	81	Gene01	XM_011659578.2	---	---	*Cucumis sativus* (Cucurbitaceae)
G02_81_C.sativus	81	Gene02	XM_004141948.3	---	---	
G01_81_J.regia	81	Gene01	XM_018986507.1	---	---	*Juglans regia* (Juglandaceae)
G02_81_J.regia	81	Gene02	XM_018986508.1	---	---	
CYP71AJ1_A.majus	CYP71AJ1	---	AY532370.2	---	---	*Ammi majus* (Apiaceae)
CYP71AZ1_A.majus	CYP71AZ1	---	EF127863.1	---	---	
CYP71AZ4_P.sativa	CYP71AZ4	---	MH000219.1	---	---	*Pastinaca sativa* (Apiaceae)
CYP71AZ6_P.sativa	CYP71AZ6	---	MH000221.1	---	---	
CYP71AJ2_A.graveolens	CYP71AJ2	---	EF191022	---	---	*Apium graveolens* (Apiaceae)
CYP71AJ3_P.sativa	CYP71AJ3	---	EF191020	---	---	*Pastinaca sativa* (Apiaceae)
CYP71AJ4_P.sativa	CYP71AJ4	---	EF191021	---	---	
CYP82C2_A.thaliana	CYP82C2	---	NM_119348.2	---	---	*Arabidopsis thaliana* (Brassicaceae)
CYP82C4_A.thaliana	CYP82C4	---	NM_119345.3	---	---	

Supp. Table 7 CYP82J dataset. The sequences in red have not been included in any phylogenetic tree (pseudogene, or outside the CYP71 clan). The sequences in white and blue have been used for the first global tree. The sequences in blue have also been included in the second tree, that is a zoom of the first one. Pg stands for "PseudoGene", FS for "FrameShift", SC for "Stop Codon", and m.r. for "manually removed".

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
CYP82J18_F.carica	CYP82J18	---	---	---	---	*Ficus carica (Moraceae)*
CYP76F110_F.carica	CYP76F110	---	---	---	---	
CYP76F111_F.carica	CYP76F111	---	---	---	---	
CYP76F112_F.carica	CYP76F112	---	---	---	---	
CYP81B114_F.carica	CYP81B114	---	---	---	---	
CYP81BN4_F.carica	CYP81BN4	---	---	---	---	
CYP81CA1_F.carica	CYP81CA1	---	---	---	---	
CYP88A103_F.carica	CYP88A103	---	---	---	---	
G01_82J_F.erecta	82J	Gene01	BKCH01000715.1	Chr09; 5501120 to 5498611	m.r.	*Ficus erecta (Moraceae)*
G02_82_F.erecta	82	Gene02	BKCH01000004.1	Chr09; 5470922 to 5472821	m.r.	
G03_82_F.erecta	82	Gene03	BKCH01000004.1	Chr09; 5513451 to 5515949	m.r.	
G04_82_F.erecta	82	Gene04	BKCH01000252.1	Chr12; 12092087 to 12094687	m.r.	
G01_82J_M.notabilis	82J	Gene01	XM_024170626.1	---	---	*Morus notabilis (Moraceae)*
G02_82_M.notabilis	82	Gene02	XM_010107781.1	---	---	
G03_82_M.notabilis	82	Gene03	XM_010092993.2	---	---	
G04_82_M.notabilis	82	Gene04	XM_010094055.2	---	---	
G05_82_M.notabilis	82	Gene05	XM_010107777.2	---	---	
G01_82J_B.nivea	82J	Gene01	PHNS01008662.1	1718481 to 1720801	m.r.	*Boehmeria nivea (Urticaceae)*
G02_82J_B.nivea	82J	Gene02	NHTU01003778.1	15572 to 17908	m.r.	
G03_82_B.nivea	82	Gene03	PHNS01007953.1	3836708 to 3838451	m.r.	
G01_82_C.sativa	82	Gene01	XM_030625418.1	---	---	*Cannabis sativa (Cannabaceae)*
G02_82_C.sativa	82	Gene02	XM_030655318.1	---	---	
G01_82J_P.andersonii	82J	Gene01	JXTB01000586.1	224751 to 226892	m.r.	*Parasponia andersonii (Cannabaceae)*
G01_82J_T.orientale	82J	Gene01	JXTC01000376.1	170166 to 172286	m.r.	*Trema orientale (Cannabaceae)*
G02_82_T.orientale	82	Gene02	JXTC01000376.1	158427 to 160548	m.r.	
G03_82_T.orientale	82	Gene03	JXTC01000376.1	126925 to 129008	m.r.	
G01_82J_H.lupulus	82J	Gene01	LD145570.1	36590 to 38901	m.r.	*Humulus lupulus (Cannabaceae)*
G01_82J_Z.jujuba	82J	Gene01	XM_016036730.2	---	---	*Ziziphus jujuba (Rhamnaceae)*
G02_82_Z.jujuba	82	Gene02	XM_016036850.2	---	---	
G03_82_Z.jujuba	82	Gene03	XM_016047182.2	---	---	
G04_82_Z.jujuba	82	Gene04	XM_016021164.2	---	---	
G01_82J_P.avium	82J	Gene01	XM_021958574.1	---	---	*Prunus avium (Rosaceae)*
G02_82J_P.avium	82J	Gene02	XM_021958467.1	---	---	
G03_82_P.avium	82	Gene03	XM_021973829.1	---	---	
G04_82_P.avium	82	Gene04	XM_021965212.1	---	---	
G01_82J_F.vesca	82J	Gene01	XM_011461424.1	---	---	*Fragaria vesca subsp. Vesca (Rosaceae)*
G02_82J_F.vesca	82J	Gene02	XM_011459856.1	---	---	
G03_82_F.vesca	82	Gene03	XM_004308804.2	---	---	

Supp. Table 7 – **Continuation and ending**

Identification	P450 name or family	Gene number	Accession number	Range	Intron	Species
G01_82J_R.chinensis	82J	Gene01	XM_024312696.1	---	---	*Rosa chinensis* (Rosaceae)
G02_82J_R.chinensis	82J	Gene02	XM_024306539.1	---	---	
G03_82_R.chinensis	82	Gene03	XM_024320429.1	---	---	
G04_82_R.chinensis	82	Gene04	XM_024320430.1	---	---	
G01_82J_M.domestica	82J	Gene01	XM_008374293.3	---	---	*Malus domestica* (Rosaceae)
G02_82_M.domestica	82	Gene02	XM_008380179.3	---	---	
G03_82_M.domestica	82	Gene03	XM_008380187.3	---	---	
G01_82J_P.bretschneideri	82J	Gene01	XM_009350098.2	---	---	*Pyrus x bretschneideri* (Rosaceae)
G02_82J_P.bretschneideri	82J	Gene02	XM_009350099.1	---	---	
G03_82_P.bretschneideri	82	Gene03	XM_009365887.2	---	---	
G01_82J_C.arietinum	82J	Gene01	XM_012718550.2	---	---	*Cicer arietinum* (Fabaceae)
G02_82_C.arietinum	82	Gene02	XM_012718553.2	---	---	
G03_82_C.arietinum	82	Gene03	XM_004488138.3	---	---	
G01_82J_M.truncatula	82J	Gene01	CM001224.2	3657106 to 3659102	m.r.	*Medicago truncatula* (Fabaceae)
G02_82_M.truncatula	82	Gene02	CM001224.2	3651766 to 3655164	m.r.	
G03_82J_M.truncatula	82J	Gene03	CM001222.2	7841391 to 7843353	m.r.	
G04_82_M.truncatula	82	Gene04	CM001222.2	14753204 to 14755390	m.r.	
G01_82_C.sativus	82	Gene01	XM_004147772.3	---	---	*Cucumis sativus* (Cucurbitaceae)
G02_82_C.sativus	82	Gene02	XM_004147754.3	---	---	
G01_82J_J.regia	82J	Gene01	XM_018994625.1	---	---	*Juglans regia* (Juglandaceae)
G02_82J_J.regia	82J	Gene02	XM_018988640.1	---	---	
G03_82J_J.regia	82J	Gene03	XM_018965476.1	---	---	
G04_82_J.regia	82	Gene04	XM_018972244.1	---	---	
G05_82_J.regia	82	Gene05	XM_018966110.1	---	---	
CYP82G1_A.thaliana	CYP82G1	---	NM_113423.4	---	---	*Arabidopsis thaliana* (Brassicaceae)
CYP82B1_E.califorica	CYP82B1	---	AF014802.1	---	---	*Eschscholzia californica* (Papaveraceae)
CYP82C1_G.max	CYP82C1	---	NM_001249250.3	---	---	*Glycine max* (Fabaceae)
CYP82A4_G.max	CYP82A4	---	NM_001317501.1	---	---	
CYP82D3_G.hirsutum	CYP82D3	---	KJ704111.1	---	---	*Gossypium hirsutum* (Malvaceae)
CYP82H1_A.majus	CYP82H1	---	AY532373.1	---	---	*Ammi majus* (Apiaceae)
CYP71AJ1_A.majus	CYP71AJ1	---	AY532370.2	---	---	
CYP71AZ1_A.majus	CYP71AZ1	---	EF127863.1	---	---	
CYP71AZ4_P.sativa	CYP71AZ4	---	MH000219.1	---	---	*Pastinaca sativa* (Apiaceae)
CYP71AZ6_P.sativa	CYP71AZ6	---	MH000221.1	---	---	
CYP71AJ2_A.graveolens	CYP71AJ2	---	EF191022	---	---	*Apium graveolens* (Apiaceae)
CYP71AJ3_P.sativa	CYP71AJ3	---	EF191020	---	---	*Pastinaca sativa* (Apiaceae)
CYP71AJ4_P.sativa	CYP71AJ4	---	EF191021	---	---	
CYP82C2_A.thaliana	CYP82C2	---	NM_119348.2	---	---	*Arabidopsis thaliana* (Brassicaceae)
CYP82C4_A.thaliana	CYP82C4	---	NM_119345.3	---	---	

Supp. Table 8 **Set of 15 flexible amino acids used for the docking of the DMS within CYP76F111 and CYP76F112.** The amino acids described in a single row are equivalent in CYP76F111 and CYP76F112. The distance between the amino acids and the DMS have been measured after the docking.

SRS	Amino acid in...		Distance between to the DMS	
	CYP76F112	**CYP76F111**	**CYP76F112**	**CYP76F111**
SRS1	T102 – Polar	P105 – Other	3Å	4Å
	S105 – Polar	A108 – Hydrophobic	3Å	4Å
	H110 – Polar	H113 – Polar	5Å	5Å
	V116 – Hydrophobic	A119 – Hydrophobic	4Å	5Å
	M117 – Hydrophobic	W120 – Hydrophobic	4Å	3Å
SRS2	M211 – Hydrophobic	I214 – Hydrophobic	6Å	8Å
	L215 – Hydrophobic	L218 – Hydrophobic	5Å	5Å
SRS4	D302 – Negative	D307 – Negative	3Å	3Å
	S305 – Polar	P310 – Other	4Å	5Å
	A306 – Hydrophobic	A311 – Hydrophobic	4Å	4Å
	T310 – Polar	T315 – Polar	4Å	4Å
SRS5	T370 – Polar	V375 – Hydrophobic	6Å	7Å
	A371 – Hydrophobic	A376 – Hydrophobic	5Å	5Å
	I375 – Hydrophobic	V380 – Hydrophobic	4Å	4Å
SRS6	L487 – Hydrophobic	L490 – Hydrophobic	5Å	5Å

Supp. Table 9 **Summary of the docking of the DMS within CYP76F112 and CYP76F111.** The modes in which the DMS was positioned in a way that seemed to allow its conversion into marmesin are highlighted in blue. The ones in which the reaction would be clearly impossible are in orange. The ones in which the reaction seemed unlikely, but maybe not impossible, are in grey. The best modes are highlighted with entirely coloured lines. The docking has been performed with Autodock Vina: see Trott and Olson (2009) for more information about the definition, the calculation and the interpretation of the affinity and root-mean-square deviation (rmsd).

Mutant	Mode	Affinity (kcal/ mol)	Distance from the best mode		Distance to the heme iron (Å)		Remark
			rmsd l.b.	rmsd u.b.	C2'	C3'	
CYP76F112	1	-8.9	0.000	0.000	4.3	4.6	Correct orientation
	2	-8.7	1.454	2.104	6.0	4.9	Bad orientation / too far
	3	-8.4	1.565	3.646	11.0	12.1	Bad orientation / too far
	4	-8.3	1.762	3.588	10.5	11.7	Bad orientation / too far
	5	-8.2	1.719	2.689	6.1	4.9	Bad Bad orientation / too far
	6	-7.7	1.979	3.853	10.7	10.9	Bad orientation / too far
	7	-7.7	2.004	2.962	6.8	6.5	Bad orientation / too far
	8	-7.3	1.584	2.205	6.1	4.9	Bad orientation / too far
	9	-7.0	9.093	11.068	9.1	9.5	Bad orientation / too far
CYP76F111	1	-8.2	0.000	0.000	9.3	10.3	Bad orientation / too far
	2	-8.1	1.536	3.446	4.9	4.4	Bad orientation?
	3	-8.0	1.560	2.546	9.7	11.0	Bad orientation / too far
	4	-7.2	1.660	3.465	9.5	10.4	Bad orientation / too far
	5	-6.9	2.481	3.975	10.9	11.9	Bad orientation / too far
	6	-6.9	2.406	4.225	6.3	5.0	Bad orientation / too far
	7	-6.8	1.577	3.756	5.7	5.8	Too far?
	8	-6.7	3.647	5.022	12.8	11.7	Bad orientation / too far
	9	-6.7	2.114	2.795	7.7	8.2	Bad orientation / too far

Supp. Table 10 Description of the main CYP76F111/112 mutants. The protein sequence of every mutant corresponds to the sequence of the associated "original enzyme", in which the amino acids in positions A, B, C or/and D have been replaced by their equivalent from the other "original" enzyme. The amino acids corresponding to CYP76F112 are coloured in blue, those corresponding to CYP76F111 are in purple. The amino acids that have been modified in every sequence are in bold.

Original enzyme		Original amino acids in positions…			
		A	B	C	D
CYP76F111		P105	A108	W120	P310
CYP76F112		T102	S105	M117	S305

Mutant	Original enzyme	Amino acids in positions…			
		A	B	C	D
CYP76F111-ABCD	CYP76F111	**T105**	**S108**	**M120**	**S310**
CYP76F112-ABCD	CYP76F112	**P102**	**A105**	**W117**	**P305**
CYP76F112-A	CYP76F112	**P102**	S105	M117	S305
CYP76F112-B	CYP76F112	T102	**A105**	M117	S305
CYP76F112-C	CYP76F112	T102	S105	**W117**	S305
CYP76F112-D	CYP76F112	T102	S105	M117	**P305**
CYP76F112-AB	CYP76F112	**P102**	**A105**	M117	S305

Supp. Table 11 **Summary of the docking of the DMS within the mutants.** The modes in which the DMS was positioned in a way that seemed to allow its conversion into marmesin are highlighted in blue. The ones in which the reaction would be clearly impossible are in orange. The ones in which the reaction seemed unlikely, but maybe not impossible, are in grey. The best modes are highlighted with entirely coloured lines. The docking has been performed with Autodock Vina.

Mutant	Mode	Affinity (kcal/mol)	Distance from the best mode		Distance to the heme iron (Å)		Remark
			rmsd l.b.	rmsd u.b.	C2'	C3'	
CYP76F112-ABCD	1	-8.6	0.000	0.000	11.2	12.1	Bad orientation / too far
	2	-8.6	1.499	3.630	5.9	4.8	Bad orientation
	3	-8.5	1.989	4.099	6.3	5.2	Too far
	4	-8.4	2.171	3.632	6.4	6.3	Bad orientation / too far
	5	-8.0	2.063	3.729	10.4	11.5	Bad orientation / too far
	6	-7.9	2.030	3.159	12.2	13.0	Bad orientation / too far
	7	-7.6	1.865	3.467	12.2	11.6	Bad orientation / too far
	8	-7.6	2.315	3.912	5.5	6.4	Bad orientation / too far?
	9	-7.4	2.194	3.881	6.0	4.9	Bad orientation?
CYP76F111-ABCD	1	-8.0	0.000	0.000	4.4	4.6	Correct orientation
	2	-7.5	1.986	3.505	6.6	5.5	Bad orientation
	3	-6.8	1.555	3.717	9.1	10.3	Bad orientation / too far
	4	-6.8	2.090	4.068	10.6	11.0	Bad orientation / too far
	5	-6.8	2.037	2.941	5.2	6.0	Too far?
	6	-6.6	1.830	2.711	6.0	5.3	Bad orientation
	7	-6.5	2.644	3.491	6.4	5.1	Bad orientation / too far
	8	-6.2	4.448	6.847	17.5	18.6	Bad orientation / too far
	9	-5.9	9.678	13.826	13.9	14.1	Bad orientation / too far
CYP76F112-A	1	-8.8	0.000	0.000	4.2	4.5	Correct orientation
	2	-8.5	1.654	3.432	10.0	10.9	Bad orientation / too far
	3	-8.5	1.530	3.745	10.9	11.9	Bad orientation / too far
	4	-8.2	2.027	3.026	6.4	6.3	Bad orientation / too far
	5	-8.2	1.846	3.577	9.9	11.2	Bad orientation / too far
	6	-7.8	1.739	3.971	11.1	12.3	Bad orientation / too far
	7	-7.3	6.776	9.712	19.9	18.9	Bad orientation / too far
	8	-7.0	3.888	5.418	10.4	9.2	Bad orientation / too far
	9	-6.9	4.448	7.068	15.8	14.5	Bad orientation / too far

Mutant	Mode	Affinity (kcal/mol)	Distance from the best mode		Distance to the heme iron (Å)		Remark
			rmsd l.b.	rmsd u.b.	C2'	C3'	
CYP76F112-B	1	-8.7	0.000	0.000	4.3	4.5	Correct orientation
	2	-8.4	1.685	2.706	5.8	5.0	Bad orientation
	3	-8.1	1.697	3.749	11.0	12.0	Bad orientation / too far
	4	-7.9	1.961	3.611	10.5	11.6	Bad orientation / too far
	5	-7.4	1.718	3.693	9.1	10.4	Bad orientation / too far
	6	-7.4	1.894	4.109	9.8	10.9	Bad orientation / too far
	7	-6.8	9.169	11.119	9.0	9.4	Bad orientation / too far
	8	-6.6	2.774	3.907	7.5	7.9	Too far
	9	-6.6	2.687	4.476	9.0	8.2	Too far
CYP76F112-C	1	-8.6	0.000	0.000	5.7	4.8	Bad orientation
	2	-8.5	1.858	3.470	10.4	11.6	Bad orientation / too far
	3	-8.5	1.524	3.646	11.0	12.1	Bad orientation / too far
	4	-8.5	1.207	1.895	5.9	4.8	Bad orientation
	5	-8.2	1.945	2.984	6.4	6.2	Bad orientation / too far
	6	-7.8	1.770	3.272	10.3	11.5	Bad orientation / too far
	7	-7.5	1.827	3.934	10.3	11.5	Bad orientation / too far
	8	-7.5	1.457	2.094	6.2	4.9	Bad orientation
	9	-7.1	2.390	3.416	7.4	7.9	Too far
CYP76F112-D	1	-8.2	0.000	0.000	6.1	5.0	Bad orientation
	2	-8.2	1.444	3.736	10.9	11.8	Bad orientation / too far
	3	-8.0	1.658	2.413	4.3	4.5	Correct orientation
	4	-7.6	1.530	2.417	6.5	5.3	Bad orientation
	5	-7.6	1.953	2.966	10.5	11.0	Bad orientation / too far
	6	-7.3	2.151	3.441	6.7	6.5	Bad orientation / too far
	7	-6.6	1.959	3.933	11.0	12.1	Bad orientation / too far
	8	-6.4	1.914	3.164	11.2	10.7	Bad orientation / too far
	9	-6.3	8.992	10.970	9.4	8.9	Bad orientation / too far
CYP76F112-AB	1	-8.1	0.000	0.000	11.0	12.0	Bad orientation / too far
	2	-7.9	2.049	3.674	10.5	11.6	Bad orientation / too far
	3	-7.6	2.311	3.778	6.6	6.3	Bad orientation
	4	-7.1	2.076	2.878	11.3	12.3	Bad orientation / too far
	5	-6.9	7.778	10.823	19.9	18.8	Bad orientation / too far
	6	-6.7	3.194	4.513	7.4	7.7	Too far
	7	-6.7	5.644	7.574	18.2	19.1	Bad orientation / too far
	8	-6.6	6.113	8.214	12.2	12.8	Too far
	9	-6.5	3.777	5.785	9.4	8.4	Too far

Supp. Table 12 **Determination of the quantity of marmesin formed by the 4 CYP76F112 mutants incubated in presence of variable DMS concentrations.** The mutants were incubated in 1mL reaction mixtures containing various concentration of DMS. 5 to 15 reaction mixture of 1mL, performed with the same modalities, were pooled together, concentrated 50 to 150 times, and analysed by UHPLC-MS. The quantity of marmesin was quantified using the peak area obtained with the MS data. The quantity of marmesin (mol) formed in one incubation of 1mL is obtained by dividing the peak area by 1.41×10^{14} (see the materials and methods section, **Chapter VII**, **B.2.f.3**) and by the concentration factor. The specific activity (mol of marmesin per min per mol of P450) is obtained by dividing the quantity of marmesin by the duration of the incubation and the quantity of functional P450s present in the reaction mix.

CYP76F112-A – Quantity of P450 for 1mL of reaction: 2.08 pmol – Incubation time: 2min						
Rep	[DMS]$_{ini}$ (nM)	Concentration of the reaction mix in 100µL		UHPLC-MS analysis: determination of the quantity of formed marmesin		Specific activity (mol marm. / min / mol P450)
		Number of pulled reactions (1mL each)	Concentration factor	Quantity of marm. in the concentrated solution (Peak Area – UA)	Quantity of marm. for 1mL of reaction mix (mol)	
A1	10	15	150	43951	2,08E-12	0,454
A2	10	15	150	37800	1,79E-12	0,391
B1	15	15	150	54994	2,60E-12	0,569
B2	15	15	150	52938	2,51E-12	0,547
C1	20	15	150	67497	3,20E-12	0,698
C2	20	15	150	66958	3,17E-12	0,692
D1	30	15	150	87076	4,12E-12	0,900
D2	30	15	150	96837	4,58E-12	1,001
E1	50	15	150	114947	5,44E-12	1,189
E2	50	15	150	116738	5,53E-12	1,207
F1	100	5	50	81363	1,16E-11	2,524
F2	100	5	50	70916	1,01E-11	2,200
G1	300	5	50	101388	1,44E-11	3,145
G2	300	5	50	89137	1,27E-11	2,765

		Concentration of the reaction mix in 100µL		UHPLC-MS analysis: determination of the quantity of formed marmesin		Specific activity (mol marm. / min / mol P450)
Rep	[DMS]$_{ini}$ (nM)	Number of pulled reactions (1mL each)	Concentration factor	Quantity of marm. in the concentrated solution (Peak Area – UA)	Quantity of marm. for 1mL of reaction mix (mol)	
A1	10	15	150	44163	2,09E-12	8,644
A2	10	15	150	30022	1,42E-12	5,876
B1	15	15	150	46825	2,22E-12	9,165
B2	15	15	150	41965	1,99E-12	8,214
C1	20	14	140	56832	2,88E-12	11,919
C2	20	15	150	53342	2,53E-12	10,441
D1	30	15	150	65190	3,09E-12	12,760
D2	30	15	150	62808	2,97E-12	12,294
E1	50	15	150	81627	3,86E-12	15,978
E2	50	15	150	71065	3,36E-12	13,910
F1	100	5	50	34938	4,96E-12	20,516
F2	100	5	50	Lost sample – Problem during the concentration		
G1	300	5	50	42774	6,08E-12	25,118
G2	300	5	50	44375	6,30E-12	26,058

CYP76F112-B – Quantity of P450 for 1mL of reaction: 0.322 pmol – Incubation time: 45s

		Concentration of the reaction mix in 100µL		UHPLC-MS analysis: determination of the quantity of formed marmesin		Specific activity (mol marm. / min / mol P450)
Rep	[DMS]$_{ini}$ (nM)	Number of pulled reactions (1mL each)	Concentration factor	Quantity of marm. in the concentrated solution (Peak Area – UA)	Quantity of marm. for 1mL of reaction mix (mol)	
A1	10	15	150	25237	1,19E-12	2,247
A2	10	14	140	22314	1,13E-12	2,129
B1	15	15	150	31879	1,51E-12	2,839
B2	15	15	150	36978	1,75E-12	3,293
C1	20	15	150	46096	2,18E-12	4,105
C2	20	15	150	35959	1,70E-12	3,202
D1	30	15	150	66924	3,17E-12	5,959
D2	30	15	150	66682	3,16E-12	5,938
E1	50	15	150	88538	4,19E-12	7,884
E2	50	15	150	83633	3,96E-12	7,447
F1	100	5	50	57393	8,15E-12	15,332
F2	100	5	50	51594	7,33E-12	13,783
G1	300	5	50	72170	1,03E-11	19,280
G2	300	5	50	79608	1,13E-11	21,267

CYP76F112-AB – Quantity of P450 for 1mL of reaction: 0.266 pmol – Incubation time: 2min

Rep	[DMS]$_{ini}$ (nM)	Concentration of the reaction mix in 100µL		UHPLC-MS analysis: determination of the quantity of formed marmesin		Specific activity (mol marm. / min / mol P450)
		Number of pulled reactions (1mL each)	Concentration factor	Quantity of marm. in the concentrated solution (Peak Area – UA)	Quantity of marm. for 1mL of reaction mix (mol)	
A1	10	15	150	46518	2,20E-12	21,007
A2	10	15	150	51822	2,45E-12	23,402
B1	25	15	150	95262	4,51E-12	43,020
B2	25	15	150	106494	5,04E-12	48,092
C1	50	10	100	116908	8,30E-12	79,192
C2	50	10	100	130091	9,24E-12	88,122
D1	75	10	100	140477	9,98E-12	95,158
D2	75	10	100	150244	1,07E-11	101,774
E1	100	5	50	100784	1,43E-11	136,540
E2	100	5	50	102902	1,46E-11	139,409
F1	200	5	50	139070	1,98E-11	188,409
F2	200	5	50	131070	1,86E-11	177,571
G1	500	5	50	160530	2,28E-11	217,482
G2	500	5	50	161241	2,29E-11	218,446
H1	1000	5	50	182387	2,59E-11	247,094
H2	1000	5	50	184727	2,62E-11	250,264

Table title: **CYP76F112-D** — Quantity of P450 for 1mL of reaction: 0.0349 pmol – Incubation time: 3min

Supp. Table 13 **Summary of the docking of the DMS within the 4 F112-like.** In the best models highlighted in blue, the DMS was docked in a way that should allow its conversion into marmesin. In the best models highlighted in orange, it was badly positioned. The docking has been performed with Autodock Vina (Trott and Olson, 2009).

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
GO2_F.erecta + DMS	1	-8.7	0.000	0.000
	2	-8.3	1.463	2.206
	3	-8.1	1.642	3.737
	4	-8.0	1.888	3.491
	5	-7.5	1.488	3.591
	6	-7.5	1.887	4.180
	7	-7.5	1.686	3.859
	8	-7.4	2.017	3.037
	9	-7.0	2.257	4.127

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
G02_F.religiosa +DMS	1	-8.3	0.000	0.000
	2	-8.2	1.878	2.596
	3	-7.7	1.612	2.337
	4	-7.4	1.902	3.821
	5	-7.3	2.626	4.297
	6	-7.3	2.065	4.229
	7	-7.2	1.430	2.525
	8	-7.2	2.138	3.037
	9	-6.7	1.824	2.942

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
GO4_F.erecta + DMS	1	-8.3	0.000	0.000
	2	-8.0	1.580	2.789
	3	-8.0	1.768	2.372
	4	-7.7	1.258	3.711
	5	-7.5	1.562	2.245
	6	-7.3	2.874	5.011
	7	-7.2	1.784	2.828
	8	-7.2	1.903	3.887
	9	-7.0	5.679	7.492

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
G11_F.erecta + DMS	1	-8.9	0.000	0.000
	2	-8.6	2.162	3.959
	3	-8.5	1.650	2.473
	4	-8.3	2.966	3.894
	5	-8.2	2.054	3.608
	6	-8.1	2.520	3.671
	7	-8.0	2.369	4.501
	8	-7.7	2.141	3.030
	9	-7.6	2.848	4.755

Supp. Table 14 Pairs of primers used to validate the 5-TUs and KanaR plasmids, and to test the transgenic tomatoes. The primers associated to the PsDiox, PsPT1, CYP76F112, CYP71AJ3 and KanaR genes (in red and black) overlap the end of the 35 promoter and the beginning of the coding sequence (direct), or the end of the coding sequence and the beginning of the tNOS terminator (reverse). The primers "GB-PsDiox-PsPT1" have been design to amplify a fragment that overlap both PsDiox and PsPT1 genes on the 5-TU plasmid. The primers "MYB8", used as a control, allow the amplification of the MYB8 gene.

Gene	Orientation	Sequence
GB-PsDiox	Direct	CTAGTCGAAATGGCTCCATCTCC
GB-PsDiox	Reverse	CGAAGATCCATTCCAAGCTCACATTTG
GB-PsPT1	Direct	CTAGTCGAAATGGCTCAAACAATTATGC
GB-PsPT1	Reverse	GAAGATCCATTCCAAGCTCACCTC
GB-CYP76F112	Direct	CTATTCTAGTCGAAATGGATATTTTCACCTC
GB-CYP76F112	Reverse	CGAAGATCCATTCCAAGCCTAATGATG
GB-CYP71AJ3	Direct	CTATTCTAGTCGAAATGAAGATGCTTGAG
GB-CYP71AJ3	Reverse	GATCGAAGATCCATTCCAAGCTCATC
GB-KanaR	Direct	CAATTTACTATTCTAGTCGAAATGGTTGAACAAG
GB-KanaR	Reverse	GAAGATCCATTCCAAGCTCAGAAGAAC
GB-PsDiox-PsPT1	Direct	ATGAAGTGGCTTTTGAAAGCTCTTATGAAGGGAT
GB-PsDiox-PsPT1	Reverse	CATCAATCCACTTGCTTTGAAGACGTGGT
MYB13	Direct	CAGCTCCAAGTTCCAAGTGA
MYB13	Reverse	TCACTAATCTGTCGGCCATT

Supp. Table 15 **Summary of the docking of the psoralen within CYP76F112 and the 4 F112-like.** In the best models highlighted in blue, the psoralen was docked in a way that should allow its conversion into bergaptol. In the best model highlighted in orange, it was badly positioned. The docking has been performed with Autodock Vina (Trott and Olson, 2009).

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
CYP76F112 + psoralen	1	-7.9	0.000	0.000
	2	-7.1	1.273	1.777
	3	-7.0	1.243	3.141
	4	-6.9	1.693	3.078
	5	-6.8	1.790	3.230
	6	-6.7	1.740	3.037
	7	-6.3	7.938	8.750
	8	-6.3	5.203	8.443
	9	-6.1	7.959	9.15

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
GO2_F.erecta + psoralen	1	-8.1	0.000	0.000
	2	-6.9	1.512	2.955
	3	-6.6	1.743	2.741
	4	-6.5	1.739	3.229
	5	-6.4	2.223	3.176
	6	-6.4	8.069	8.881
	7	-6.2	1.666	2.333
	8	-5.9	6.453	8.752
	9	-5.9	5.054	7.416

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
GO2_F.religiosa + psoralen	1	-8.1	0.000	0.000
	2	-6.9	1.512	2.955
	3	-6.6	1.743	2.741
	4	-6.5	1.739	3.229
	5	-6.4	2.223	3.176
	6	-6.4	8.069	8.881
	7	-6.2	1.666	2.333
	8	-5.9	6.453	8.752
	9	-5.9	5.054	7.416

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
G11_F.erecta + psoralen	1	-8.2	0.000	0.000
	2	-7.6	0.471	2.823
	3	-7.3	1.220	2.943
	4	-7.1	1.834	2.561
	5	-6.7	4.867	7.335
	6	-6.2	5.048	8.726
	7	-6.2	8.873	10.657
	8	-6.2	8.903	10.731
	9	-6.1	4.895	7.287

Mutant	Mode	Affinity (kcal/mol)	Dist. from best mode	
			rmsd l.b.	rmsd u.b.
G04_F.erecta + psoralen	1	-7.6	0.000	0.000
	2	-7.2	0.835	2.922
	3	-7.1	1.960	3.205
	4	-7.0	2.014	2.894
	5	-6.4	1.523	2.251
	6	-6.4	1.577	2.670
	7	-6.2	2.091	3.381
	8	-6.0	5.280	7.814
	9	-5.8	5.106	7.663

Supp. Figure 1 Sequence of the P450 candidates amplified from *F. carica* cDNA.

>CYP76F110
ATGAAATATGTCCTTTGGAGTTGCAGTTTCTTGTGTGTTATCTCATCCTCCAGCTTTTAATCCAAACTTTGTCCATTTCTAG
GTCAAGAAGTAGCAAAATATTGCCACCGGGACCAAGACCATATCCCATCATAGGAAGCCTCTTGGCTTTGGGAGACAAACCTC
ACAAATCCTTGGCCACACTTTCCCACACTTTTGGCCCCATCATGCATTTGAAGCTCGGCCAAATCACCACAATAGTAATTTCC
TCCTCATCCACGGCCAAACAAGTTCTCCAAACCGACGAGAAGTTCTTCGCCAACCGACCGTCATGCGATGCAGCCAACGCCTA
CAACCATAGCCAGTTCGGCATGCCTTGGTTGCCCGCTTCATCCTTGTGGAGAAATCTTCGAAAATAAGCAACTCCCACTTGT
TTTCCGCCATGGTTCTCGATGCCAACATGAACTTGAGGCACAACAAGGTCCGAGACCTCCTAAACCGTGTCCATAAAAGCGCG
GAAACCGGTGAGGCAGTTGACATCGGTAGGGCCGCTTTCGAAACGATCCTGAATCTGTTGTCCACCGCCTTTTTCTCCGTCGA
TTTGGCTGATTCAACTTCCGACATCGCGAGAGAGCTCAAGGAGACCGTGTGGAATCTAATGGAAGAGGTGGGGAAACCGAACT
TGGCGGATTATTTCCCGTTGCTTAGGAAGCTTGACCCGCAAGGAATAAGGAGACGCGCCGCTATTTATATCCGGAAATTCATG
GATTTATTTGATCGAATCGTCGAAGAAAGAGTACGACTTAGATCGGAAATTAGTACTTCCGATTCGACAAAAGATGATGATAT
GTTAGACGCTCTTCTCAACATGATGCCGATCAGTGAATATCAGCTTGACAAAAACCAAATTGAACATTTGTTCTTGGATCTAC
TTGTTGCGGGTATTGAAACGACTTCATCCGCACTCGAGTGGTCGATAGCCGAGCTACTCAAAAGCCCGGAGATCATGTCGAAA
GCCAATGCAGAGCTTGAGCAAGTTATAGGAAAAGGAAACCAAGTGAAGGAATCAGACATAACTCAACTCCCTTACTTGCAAGC
TATTGTCAAAGAGACCTTACGCTTGTACCCATCAATTCCGCTAATTCCCCGTGAAGCGGAAAGGGATGTGCAAGTATGCGGCT
ATGTCATCCCAAAGGGTGCACAAGTTCTAGTTAATGCATGGGCTATTGGGAGAGATCCAACCATTTGGGACAATCCCAACAAG
TTTCTGCCGGAGAGGTTTCTTGAATCAAATATTGATATCGGAGGTCGGAGTTTCGAGTTCATTCCCTTTGGTGGCGGCCGGCG
ACTATGTCCCGGCTTGCCGTTGGCAACTAGGATGTTGCACTTGATATTGGCTTCTTTGCTTCACTCCTTTACTTGGAAGTTAG
AAGATGGGATTTCACCAGAAAATATGAACATGGAGGACAAGTTTGGCCTCGCCACGCAGATGGCTCAGCCTTTGAGAGCTCGT
CCCGTGGCGATAATTTCTTAG

>CYP81B114
ATGTTCATCTACTTGCTGCTGTTTATAGCTCTCTATGCCATCACCAATCACTTGCTCGGGAAAATCAGAAACTACCCACCAAG
CCCTTTCCCGGCCCTCCCGATCATCGGCCATCTCTACCTCTTCAAGAAGCCTCTCTACCGAGCTCTCTCCAAGATCTCAAACC
GGCACGGTCCGGTTCTTTTCCTGCAGTTCGGGTCCCGCCGCATCCTCGTCGTTTCGTCCCCTTCGGCGGCGGAGGACTGCCTG
TCCAAACACGACGTCATCTTCGCGAACCGCCCTCGTCTTCTGGCCGGGAAACACTTCGGGTCCAACTACACCAGCCTCGTCTA
CTCCCCCTACGGCGACCACTGGCGGAACCTCCGCCGGATCTCGACCCTCGAGCTCCTCTCCACCCCCACGCCTCCAGATCCAGT
CCCAAATTCGCTCCGACGAGGCCCTGTCTCTCGTCGGCCGCCTCCTCCGGTCGCCGGATCAGACGGTGGACATGAAGACGCTG
CTGTTCGAGCTGACGCTGAACGTGATGATGAGGACGATCGCCGGGAAAAGGTACTACGGGGAAAGCGTGGCGGAAGTTGAGGA
AGCCAGGACGTTTCAGGACATCGTGAAAGAAACGTTTCGGATGGGCCCGGCTTCGAACATCAGTGATTTTCTGCCGTTTATGA
AGTTGTTGGGGACTAGTGATTACGAAAAGAAACTGATTGAGTTGCAGAGAAAGAGAGATGATTTCATGCAATATCTGATTGAG
GAGCACAGGAAAAATATGGGAAAAATTGGGGAAGATAAAATAAAGACCATGATTGAAGTTTTGTTGAAGCTGCAAGAGTCGGA
GCCGGATTACTATACGGACGACATGATCAGAGGCCTCTTGCTGGTCTTATTTTCAGCAGGAACTGATACATCAGCTGGAACTA
TGGAGTGGGCACTGTCTCTTCTACTAAACCATCCAGAAGTCATGGAAAAAGCACGTGCCGAGTTAGAAAATTGCGTGGGACAT
GATCGTCTAGTGGATGAGTCAGATTTAAACAATCTTCCTTATCTCCAATGCATCATAAAAGAGACCATGCGATTGTTCCCAGC
GGGTCCATTACTAATACCCCATGAATCATCAGAAGAGTGCAAGGTGAGTGGGTTCCACATACCACGTGGCACAATGCTAATGG
TCAACGTGTGGGCAATACAACAAGACCCCAAAATTTGGGAAGATCCTACTAGTTTCAAGCCAGAGAGGTTTAAGGACGTGGAA
GGGATGAAAGATGGATTTAAGTACATTCCTTTTGGGACAGGGAGAAGGGGCTGCCCGGGCGAAAACTTCGCCGTTCGAATTGT
TGGGCTCAATTTGGGCCTGCTGATTCAGTGCTTTGAGTGGGAAAGGCTTGGCAAGGAGTTGGTTGACTTAAGTGAAGGGACTG
GGCTCACCCTGCCCAAGGCCCACCCGTTGCTCGCTAAGTGCAGACCTCGTCCAAGCATGTTGAATGTTCTTTCTCAAATTTGA

>CYP81BN4
ATGGAAGAACCCATATTGTACACATCCCTTTGCCTCATCATCCTCCTAATAGCTCTAAAGCTCTTTCTCCAAACCAGAAAACC
ACATAATAAGAACCTGCCACCAAGCCCACTGGCTCTTCCGGTCATCGGTCACCTCCACCTCCTCAAGGAGCCGCTCCACCGCA
CTTTACACCGCCTCTCGAAAACATACGGCGATGTTTTCTCCCTCAGGTTTGGCTCTCGCCATGTCGTCGTCATCTCCTCGCCT
TCAGCAGTCGAGGAATGCTTCACCAAAAACGACGTCGTCCTCGCCAACCGCCCTCCCCTCATCATGGGCAAGCACGTTCACTA
CAACCATACTACCCTTGCCCTGGCCCCCCACGGCGACCACTGGCGCAACCTCCGCCGCATCGGCAGTCTTCAGATCTTCTCCT
CGGCCCGACTCAACGCTTTCAACGACACCCGAAGGGACGAGGTTAAGCGCATGCTTCGACGAATCTCTACCAATTCATTCCAT
GGTGCGGGGAAGGTACAGCTGACGTCGATGTTCGCTGACCTGACGATAAACGTCATAATGAGAATGGCCACCGGGAAGAGGTA
CTACGGTGATGATGTGGCGGACGTCGACGAGGCAAGGAAGTTTAGAGAGATAGTGACGGAGGTCTTTGAGAATGCTGGGGCGG
GAAATCCCGCAGATTTCTTGCCTGTTTTGAATTGGATTCCTGGCAGTTATGAAACGAAAATCAAGAGGCTCGCAGAAAAGACA
GATGGTTTTCTGCAACAGTTGGTCGATGAGCATCGGAATGCGAAAGAGAGTAAGAACACCATGATTGATCATTTGCTTTCCTT
GCAGGAGTCGCAGCCCGAGTATTACACTGACGAAATTATCAAGGGCTTCATAATGATGATATTATTGGGTGCGACTGATACAA
TAGTCGTTAATCTAGAATGGACAATGTCGAATCTCCTTAACCATCCCAACATCCTTAAAAAGGTGAAAGCAGAATTGGATGCT
CAAGTTGGTGAACAGCAGTTGTTGGAGGAATCAGATCTTTCTAATCTAAATTACCTCAAAAACGTCATCTCTGAGACCCTTCG
ACTGTACCCGTCAGCTCCACTGCTTTGTACCCCACTACTCATCTGATGATTGCACCATTGGCGGAAACCACGTGCCACGTGACA
CGATATTATTGGTCAATGCGTGGGCCATACAAAGAGATGCCAAATTGTGGGATGATGCAGAGAGTTTTAAGCCCGAGAGGTTT

GAGAATGGCGAGAGTGAGAGTTACAAGCTAATTCCATTTGGGGTTGGAAGAAGGATTTGTCCCGGAATAGGCTTGGCCAACCG
TGTTGTGGGCTTGGCTTTGGGAACATTAATTCAGTGCTTTGAGTGGGAGAGGATTGGCGAAGAAGAAGTTGATATGGCTGAAG
GTAAAGGCCTCACTATGCCTAGATTGGTGCCATTGGAGGCAAAGTGTAAAGCACTCCCCATTATGAATGTTGTTCTCTCTTAA

>CYP76F111
ATGGATTTGATCACCTCTATATTGTGTTTTGTTCTAATCTTATGTATCTCGATTAAAGCCATCCGTTCATATTCGAATCCCAA
ATCAAGCAAAAAACTCAATATTCCACCTGGTCCAAAACCCCTTCCCATCATAGGAAACCTCTTTGATCTCGGCAACAAACCGC
ACCAGTCCTTGGCCAGGCTTGCCAAGTCCTATGGCCCTCTTATGAGTTTGAAGCTTGGCCAAGTCACGACGATTGTCGTCTCT
TCGCCGGCCCTGGCTAAGGAAGTCATAATGGGACAGAGCCAGTTCTTCTCCAACCGATCAGTCCCGGACGCCGCCCACGCCCT
GGGACACTGCCAGTACAGCATGGCTTGGTTGCCTGTCTCAAACTACTGGAGAAACCTCCGAAGGCTAAGCAACTCCCAGTTGC
TCGCCGCTTCGGTTCTCGACGCCAACATGAACCTGAGGCACAACAAGGTTGAAGACCTCTTACGTGAGGTCCAGAAAAGTGCC
GAGGCTGGTCAGGCCATGGAGGTTGGGAAAGCCGGTTTTTTAACGACTCTAAATCTCTTGTCCACCACCTTCTTCTCTATGGA
TTTGTTAGCGGATCCGAATTCGGAGACTGTTAGACAGTTCAAGAAGGACGTGGTTGGAATCTTTGAAGAGTTGGGAAAATCCA
ACATGGCTGACCATTTCCCGTTCCTTAAGAAGTTTGATCCCCAAGGAATTCGGAGGCGAACCGAATTGTACTTCCAAAAGATG
TTGGACTTGTTCAATCGCATTATCGACGAAAGGCTGCAGCGGCTCAAGGAAGCGAATGGTTCTGTGAAAGAGAATGACATATT
GGACAATCTCATCAACATGATGTTGGCTGGTGAAGATAATAGAGGATACGATCCAGTAGACAGAGCCGCAATTTCACACTTTT
TACTGGATCTATTTCCTGCCGGGACTGACACGACATCAGCTACGTTAGAATGGGCAATGGCTGAGCTAGTAAAAGCCCCGGAG
ATCATGGCAAAAGCCCGAAAAGAACTAGAGCAAGTTATTGGCAAAGGAAACCAAGTGAAGGAATCGGACATAAACCAACTCCC
TTATTTGCAAGCCATTGTCAAAGAAACCTTCCGCTTGCACCCCGTTGCTCCATTATTGGTTCCTCGCATAGCCGATGGAGACG
TCGAACTCTACGGCTATAGCGTCCCGAAGGATGCTAAAGTGCTAATAAACATATGGGCCATTGGTAGAGACTCAAACGTTTGG
GAAAATTTCGAAGAGTTTATACCGGAGAGGTTTTTGAAAAGTACTATTGAGACCGGAGGTCGGGATTTCGAGCTCATTCCCTT
TGGTGCGGGGAGGAGAATATGTCCTGGGATACTGTTGGCGAACCGAATGTTGCACTTGATGTTGGGTTCTTTGCTTCACTCGT
TTGATTGGAAATTGGAAGACGGGGTTGAGACTGTGAACTTGGATGACAAGTTTGGCCTGACCTTGCAGCTGGCTCAGCCTTTG
CGAGTTGTCCCTATAGCAATCGCTTAG

>CYP81CA1
ATGGAAGTCTTCTACCTTTACCTGTCTCTCTTCCTCTCCATTTTTCTTCTGATCACTAAATTTCTACTTCACCAAACCAAAAA
TTATCCACCAAAACCTCCTTCTCTTCCAATAATCGGCCACCTTCGCCTTCTAAAGCCTCCTCTTCATCAAACACTTCAATCTC
TCTCTCAAAAATATGGTCCAATCTTCTCTCTTCGAGTTGGTCGCCAATCCTTCTTGGTCCTATCTTCTCCTTCCGCCGTCGAA
GACTGCTTCACGAAAAACGACATCCTTTTTGCTAACCGGCCGCGTAGCACGGCCGGTGATCTTATGTCATACGATCACACTTC
CATCATGACGGCTCCTTATGGCCATCTCTGGCGCAGCCTCCGGCGCCTCACCACCGTCGAGATCTTCTCTCAGATCAGCCTTA
ATAAGTCCTCTTCCGCACGAGAAGAGGTAGTGCGTTCGCTTGTTCGTCAAGTAATTAAGGGACCGGAAAAGGCGGATTTGATT
TATGCGTTTGGTGTTCTGGTTTTCAATATGACGACGAGGATGGCTTGTGGGAAGGACTATATTGAAGATGAGATATTTGCAAC
GGATCTGGGGAGACAACGTCTGAAAGAAGCGAAGGAGACGCTCTCCCCTCTAGCGATATTAAATCTGTGTGATTATTTCCCAG
TTTTGAGGTTGTTTGGCTTTAGAGATTTGGAAAAGAGGCTGATAAGGTTGCAGAGGAAGAGAGACGAGTTCATCAGTAGTCTG
ATCGATGAGTTTCGGCAGAACAGAGTTGCCCCTTTTAATTTCAAGAAAGGAAAAACGCTAGTTGAAACTTTGCTCTCAATTCA
GGAATCAGAACCTGAATTCTACTCGGACAACGTTATTAAAAGCATTATTCAGATCATGTTTGTGGGTGGAACCGACACATCCG
CAAGCACCATGGAATGGACAATGTCACTCCTCCTAAGTCATCCGGAAGAACTACAAAAGCTAAGAAGACAGATTGATGATCAA
GTCGGACATGATCGGTTGCTCAAGGACTCCGATCTCCCCATGCTTCCCTACCTTCGATGCATCATCAGCGAGGCGCTTCGGTT
ATATCCCGTAGCTCCCCTTTTAATACCTCATTTTTCATCACAAGATTGTACCATCAAGGGATATCACATACCAAGAGGCACCA
CTTTGATGGTCAACGCCTGGGCAATTCATAGGGATCCAAATGTTTGGGACGAGCCGACAAAGTTCAAGCCAGAGCGATTTGAA
GTGACGGTAGTGGACCGTCAGAAGGAAGGGTTCAAATATCTACCATTTGGAGTGGGGAGGAGGGCTTGTCCTGGCTCTCCCTT
GGCCATCCGGGCGGTGTCGTTGGTCATTGGAACGCTGGTTCATTGCTTTAATTGGGAGAGGGTTGGGGAGGAACCTGTGGATA
CGAGTCTTGGCTCTGGAGCTTCTTTTTCCAAGAGTAAGCCTCTTGAAGCTGTATGGACTCCTCGTCCTGTCATGTCTCGCCTC
CTATCTCAACTTTGA

>CYP76F112
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC
ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCACCGACGCAAGCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTT
CAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC

CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
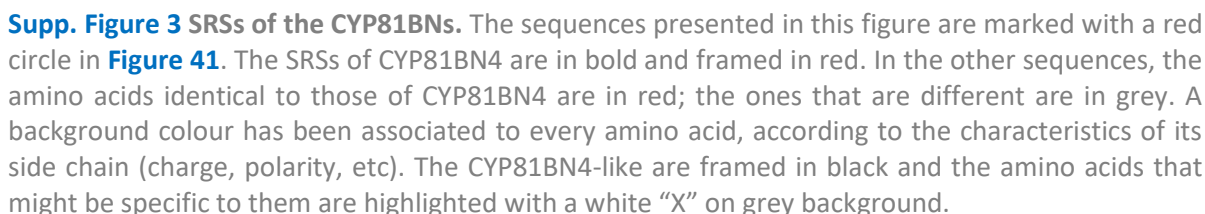CCCGTGCCGACAAAGCATTAG

>CYP82J18
ATGGACTTTCTCATCTCTCATTCTCTAATAGCAATTTTTGCAGGCCTTTTGGGGCTGGTCTCTCTGTACCTATGGAGGACAAA
AACAAGCCACAAAACCAAGCACAAGTTTCTAGCCCCACCAGAGCCCTCCGGTGCGCTACCGATTATCGGCCACCTCCACAAGC
TAAGCTCCCACAACCCGATCTTCCGATCCCTGGCCGACATCGCCGACAAGTACGGCCACATCTTCATGTTCCGCCTCGGCAAG
AACCCCGTGCTCGTGGTCAGCAACTACGACGCGGTCAAGGAGTGTTTCACCACAAACGACAAGAACTTCGCCTCCCGACCGAG
GTCCACGCACGGCACACACCTCGGCTACGACTACGCCGCGTTCGGGTTCGCGGCGTACGGGCCGTTCTGGCAGCACATGAGAA
AGCTGGTCATGGTCGAGCTGCTCTCGGCCCGCCGGCTCGAGACATTAAGACACGTGCAGATTTCAGAGGTTGAGGCTTTGGTC
AACGGATTGTACGGACAGTGCAAGGCCGGCAACGGGTTTGGTCCGACCAAGGTGGTGATCAGCGAGTGGATTGAGCACTTCAC
TTTGAACATCATAACGGAGATGATTGCCCGGAAGAGGTATTTCAAGAGTGGTCGGGAATTCGATGGAGAGGCGGAGAGGATTG
GGAAGGTTATAAAGGACTATATGCTCATTTCCGGGGCTTTTGTACTTTCCGATTTGTTCCCGATTCCGACATGGCTTGATCTT
AATGGGAATCTGAAAGAAATGAAGCGGATTGGGAGGGAAATTGATGTTCTGATTCAGAGTTGGATTGACGAACACAAGATGAG
TGAGTCGAGTAAAAAGCAGGATTTCATTGATGTCATGTTGTCCTTGATTGAGGATAGTCCCGACCTTGGCTATACGCGCGACA
CCATTATTAAAGGAACAGTTTCGACTCTAATAATAGCAGGCTCGGACACCACAGCCATAAACTTGACATGGATTCTATCACTA
TTACTAAACAACAGACAATCTTTAAACCATGCCCAAGAAGAGCTAGACCATGTAGTTGGAAGAGAAAGATGGGTCGAAGACTC
CGACATCAAAAACTTAGTTTACCTTCAAGCCATTGTCAAGGAGACACTGCGGCTATACCCTCCGGGTCCTCTAGCAGTTCCAC
ACGAGGCTAGGGAAGACTGCACAGTTTGTGGCTACCATGTAAAGAAAGGCACTCGTTTGTTTGTCAATCTTTGGAAGCTGCAC
CGAGACCCGATCGAGTGGCCCGACCCGGGGGAGTTTATGCCGGAGAGGTTTTTGACGAGCCATGCAGGGCTAGATGCATCGGG
TAAGCATTTTGAGTTCATCCCATTTGGTTCCGGTAGAAGATCATGTCCCGGGATGACTTTTGCCTTGCAAGTGACGCATTTGA
CTCTCGCACGCCTTCTCCAAGGGTTTGAGTTGGCAACACCGTTGGATGATCCGGTGGATATGACTGAAGGATTGGGAATTACC
TTGCCTAGAGCCACTCCACTAGAAGCGGTCCTAACCCCTCGCCTCTCTGTTGAATTTTATCAACATTAA

>CYP88A103
ATGGGTAGTTGGGGGGCGGTCGTGGTGATGGTGATGGTGATGGTGAGCAGTTTGGGGATGGGAGTGGTCCTGAAATGGGTTCT
GAAGAACGTGAATGGGTGGATGTATGAGACAAAGTTGGGTCATGAGAAGCAATTCTCTCTTCCACCAGGTGATTTGGGATGGC
CTTTCATTGGCAACATGTGGTCCTTCCTCGGAGCCTTCAAGTCCTCTAACCCTGACTCTTTCATCTCTTCCTTTGTTTCCAGG
TACGGACGTACAGGAATGTACAAGGCCTTCATGTTCGGTAACCCGAGTATCATTGTCACGACGCCAGAAACCTGCAGGCGAGT
GTTGACAGATGACGATTCGTTTAAGCCAGGTTGGCCTCGTTCCACTGTGGAACTCATCGGGAGGAAATCCTTCATTGGGATTT
CTTTCGAAGAACACAAGCGTCTAAGGAAGTTAACGGCAGCTCCTGTGAATGGCCATGAAGCATTGTCCATGTACATTCAATAC
ATTGAAGAGAATGTAGTCGCTTCCTTGGAAAAGTGGGCTACCATGGGACAAATAGAGTTCTTGACTGAAGTCCGGAAACTCAC
TTTTAGGATCATCATGTACATTTTTCTAAGCTCAGAAAGTGAGCCAGTAATGGAGGCTTTGGAGAATGAATATACCGCACTTA
ATTATGGAGTCAGAGCCATGGCAATTAACATTCCGGGATTCGCTTACCATAGAGCATTCAAGGCTAGGAAAAATCTTGTATCC
ATGTTTCAATCTATTGTGGATGAGCGACGATCTCAGAGGAGGAAGAGCAACATAAGCTCGACAAAGAAGAAAGATATGATGGA
TGCTTTGCTCGATGTTGAAGATGAAAATGGCAGAAAATTGAGTGATGAGGAGATCGTCGATGTTTTGTTAATGTACTTGAATG
CCGGCCATGAGTCTTCTGGTCATATCATGATGTGGGCTGTCGTTTTCCTACAAAAACACCCGGAGATTTTCCAAATAGCCAAG
GCTGAGCAAGAAGAGATTTTGAAGAGGAGGCCACCTACAAAGAAAAGCTTGAGCCTTAAAGAAATTCGAGAAATGAACTATCT
ATCAAAGGTAGTTGACGAAACACTTCGTGTGGTAACTTTCTCGCTCACAGTGTTTCGAGAAGCAAAACGAGATGTCCAAATAG
CTGGTTACACCATCCCGAAGGGTTGGAAAGTTCTTGTCTGGTTTAGAAGTGTTCATTTGGATCCAGAAATCTATCCAAATCCA
CTTGAATTCGATCCTTCAAGATGGGATAACCACACACCGAAAGCAGGATCTTTCCTCCCCTTTGGAGCAGGAAGCAGGTTGTG
CCCGGGAAACGATCTAGCTAAGATGGAAATCTCGATTTTCCTTCATCACTTTCTCCTCAATTATCGGCTGAAACGGACTAATC
CTGATTGTCCATTGATGTACTTGCCACACTCGCGGCCTAAAGACAATTGCTTGGCAAAAGTTGAGAAATGCACATCTTCATGA
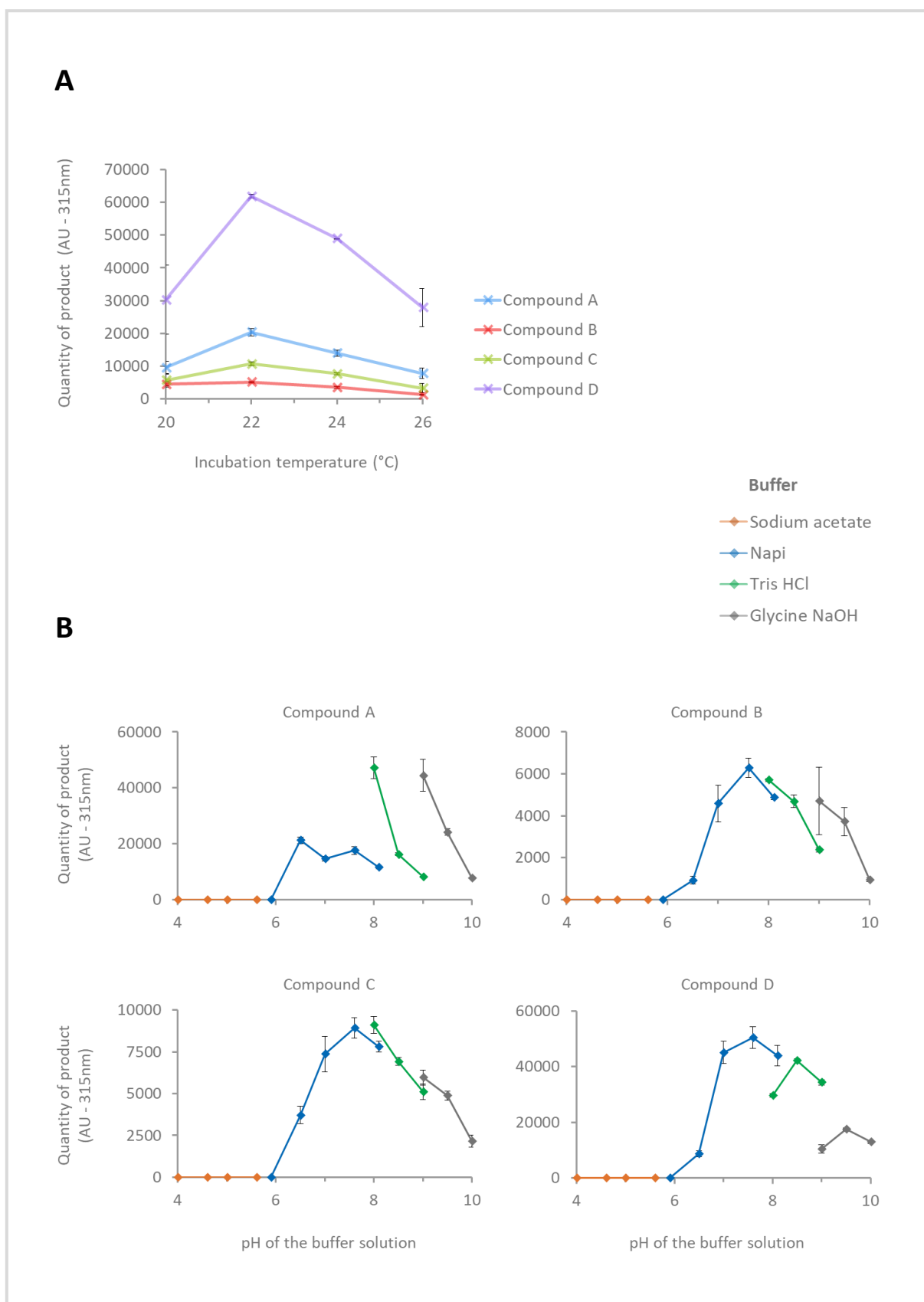
Supp. Figure 2 Determination of the optimal temperature and pH for the different products formed by CYP82J18 when incubated in presence of auraptene and NADPH. All incubations were repeated in triplicates and analysed by UHPLC-MS. **A** Optimal temperature. **B**. Optimal pH.

Supp. Figure 3 SRSs of the CYP81BNs. The sequences presented in this figure are marked with a red circle in **Figure 41**. The SRSs of CYP81BN4 are in bold and framed in red. In the other sequences, the amino acids identical to those of CYP81BN4 are in red; the ones that are different are in grey. A background colour has been associated to every amino acid, according to the characteristics of its side chain (charge, polarity, etc). The CYP81BN4-like are framed in black and the amino acids that might be specific to them are highlighted with a white "X" on grey background.

Supp. Figure 4 SRSs of the CYP82Js. The sequences presented in this figure are highlighted in blue in **Figure 43**. The SRSs of CYP82J18 are in bold and are framed in red. In the other sequences, the amino acids identical to CYP76F112 are in red; the ones that are different are in grey. A background colour has been associated to every amino acid, according to the characteristics of its side chain.

Supp. Figure 5 **Sequence of the mutants synthesised and inserted into the expression vector pYeDP60.** The codons corresponding to the four amino acids of interest have been highlighted according to their original enzyme: blue for CYP76F112, green for CYP76F111.

>CYP76F112
```
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC
ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCACCGACGCAAGCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTT
CAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC
CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
CCCGTGCCGACAAAGCATCATCATCATCATCATTAG
```

>CYP76F112-ABCD
```
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC
ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCCCGGACGCAGCCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTTGGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTC
CTGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC
CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
CCCGTGCCGACAAAGCATCATCATCATCATCATTAG
```

>CYP76F112-A
```
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC
ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCCCGGACGCAAGCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTT
CAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC
CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
CCCGTGCCGACAAAGCATCATCATCATCATCATTAG
```

>CYP76F112-B
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC
ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCACCGACGCAGCCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTT
CAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC
CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
CCCGTGCCGACAAAGCATCATCATCATCATCATCATTAG

>CYP76F112-C
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC
ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCACCGACGCAAGCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTTGGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTT
CAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC
CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
CCCGTGCCGACAAAGCATCATCATCATCATCATCATTAG

>CYP76F112-D
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC
ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCACCGACGCAAGCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTC
CTGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC
CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
CCCGTGCCGACAAAGCATCATCATCATCATCATCATTAG

>CYP76F112-AB
ATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAA
ACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGACCAACCCCACAGGTCCT
TGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACC

ATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTCTCCAGCCGGACAATCCCGGACGCAGCCCGCGCCCACAGACACAG
CGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCACACTTGCTTTCCTCCA
AGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAAAAGCGTCCAGGCCGGG
GAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCTCCATGGACATGGCGGA
TGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCG
ATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTG
TTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGACACTCTTAT
CAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTACTGGATTTATTTT
CAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCC
CGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCCCTTACTTACAAGCCAT
TGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGACATCGAAATCTCCGACT
ATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAAATCCCGACAAG
TTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGG
AAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATC
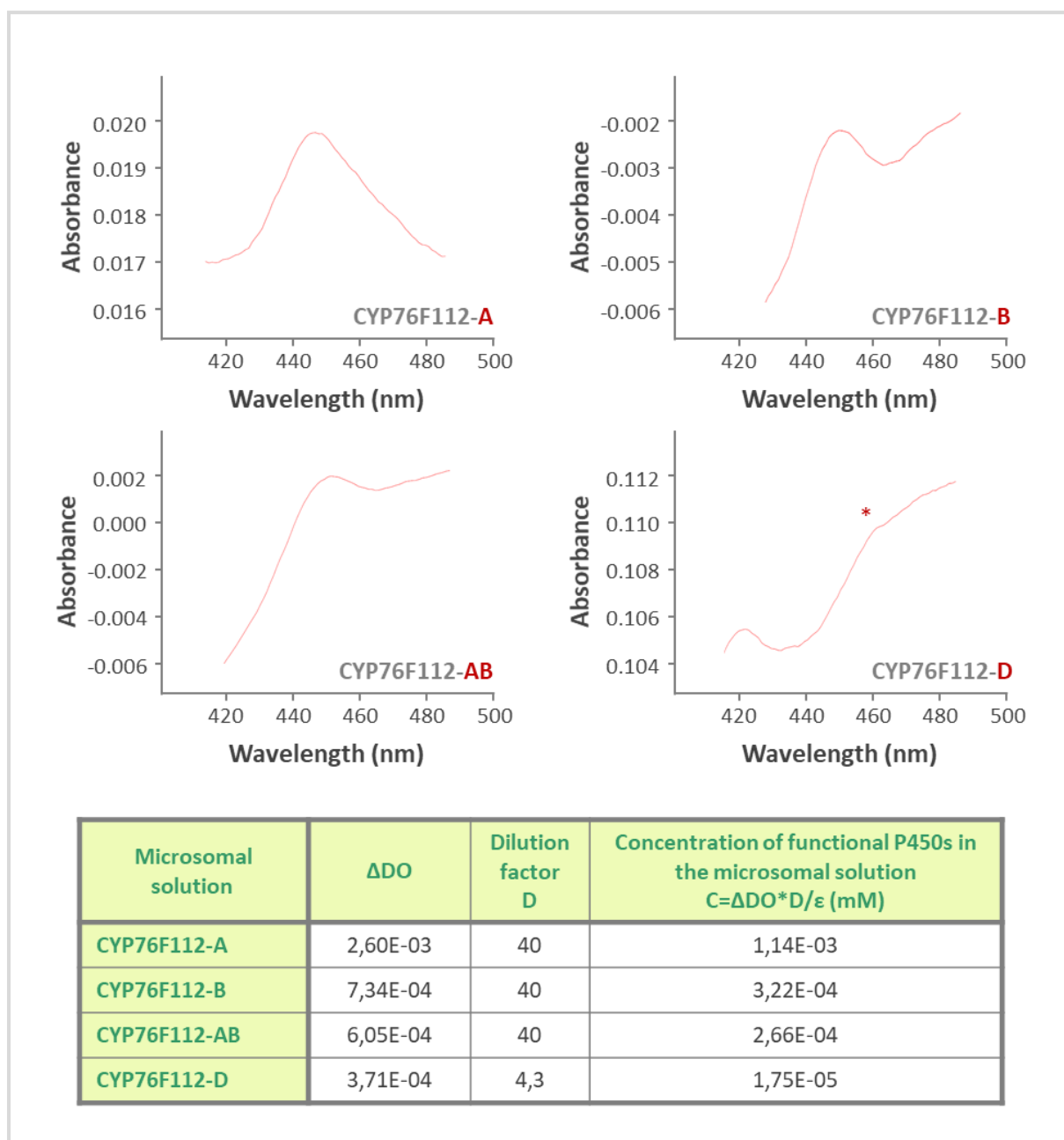CCCGTGCCGACAAAGCATCATCATCATCATCATTAG

>CYP76F111
ATGGATTTGATCACCTCTATATTGTGTTTTGTTCTAATCTTATGTATCTCGATTAAAGCCATCCGTTCATATTCGAATCCCAA
ATCAAGCAAAAAACTCAATATTCCACCTGGTCCAAAACCCCTTCCCATCATAGGAAACCTCTTTGATCTCGGCAACAAACCGC
ACCAGTCCTTGGCCAGGCTTGCCAAGTCCTATGGCCCTCTTATGAGTTTGAAGCTTGGCCAAGTCACGACGATTGTCGTCTCT
TCGCCGGCCCTGGCTAAGGAAGTCATAATGGGACAGAGCCAGTTCTTCTCCAACCGATCAGTCCCGGACGCCGCCCACGCCCT
GGGACACTGCCAGTACAGCATGGCTTGGTTGCCTGTCTCAAACTACTGGAGAAACCTCCGAAGGCTAAGCAACTCCCAGTTGC
TCGCCGCTTCGGTTCTCGACGCCAACATGAACCTGAGGCACAACAAGGTTGAAGACCTCTTACGTGAGGTCCAGAAAAGTGCC
GAGGCTGGTCAGGCCATGGAGGTTGGGAAAGCCGGTTTTTTAACGACTCTAAATCTCTTGTCCACCACCTTCTTCTCTATGGA
TTTGTTAGCGGACCCGAACTCGGAGACTGTTAGACAGTTCAAGAAGGACGTGGTTGGAATCTTTGAAGAGTTGGGAAAATCCA
ACATGGCTGACCATTTCCCGTTCCTTAAGAAGTTTGATCCCCAAGGAATCCGGAGGCGAACCGAATTGTACTTCCAAAAGATG
TTGGACTTGTTCAATCGCATTATCGACGAAAGGCTGCAGCGGCTCAAGGAAGCGAATGGTTCTGTGAAAGAGAATGACATATT
GGACAATCTCATCAACATGATGTTGGCTGGTGAAGATAATAGAGGATACGATCCAGTAGACAGAGCCGCAATTTCACACTTTT
TACTGGATCTATTTCCTGCCGGGACTGACACGACATCAGCTACGTTAGAATGGGCAATGGCTGAGCTAGTAAAAGCCCCGGAG
ATCATGGCAAAAGCCCGAAAAGAACTAGAGCAAGTTATTGGCAAAGGAAACCAAGTGAAGGAATCGGACATAAACCAACTCCC
TTATTTGCAAGCCATTGTCAAAGAAACCTTCCGCTTGCACCCCGTTGCTCCATTATTGGTTCCTCGCATAGCCGATGGAGACG
TCGAACTCTACGGCTATAGCGTCCCGAAGGATGCTAAAGTGCTAATAAACATATGGGCCATTGGTAGAGACTCAAACGTTTGG
GAAAATTTCGAAGAGTTTATACCGGAGAGGTTTTTGAAAAGTACTATTGAGACCGGAGGTCGGGATTTCGAGCTCATTCCCTT
TGGTGCGGGGAGGAGAATATGTCCTGGGATACTGTTGGCGAACCGAATGTTGCACTTGATGTTGGGTTCTTTGCTTCACTCGT
TTGATTGGAAATTGGAAGACGGGGTTGAGACTGTGAACTTGGATGACAAGTTTGGCCTGACCTTGCAGCTGGCTCAGCCTTTG
CGAGTTGTCCCTATAGCAATCGCTCATCATCATCATCATCATTAG

>CYP76F111-ABCD
ATGGATTTGATCACCTCTATATTGTGTTTTGTTCTAATCTTATGTATCTCGATTAAAGCCATCCGTTCATATTCGAATCCCAA
ATCAAGCAAAAAACTCAATATTCCACCTGGTCCAAAACCCCTTCCCATCATAGGAAACCTCTTTGATCTCGGCAACAAACCGC
ACCAGTCCTTGGCCAGGCTTGCCAAGTCCTATGGCCCTCTTATGAGTTTGAAGCTTGGCCAAGTCACGACGATTGTCGTCTCT
TCGCCGGCCCTGGCTAAGGAAGTCATAATGGGACAGAGCCAGTTCTTCTCCAACCGATCAGTCACCGACGCCAGCCACGCCCT
GGGACACTGCCAGTACAGCATGGCTATGTTGCCTGTCTCAAACTACTGGAGAAACCTCCGAAGGCTAAGCAACTCCCAGTTGC
TCGCCGCTTCGGTTCTCGACGCCAACATGAACCTGAGGCACAACAAGGTTGAAGACCTCTTACGTGAGGTCCAGAAAAGTGCC
GAGGCTGGTCAGGCCATGGAGGTTGGGAAAGCCGGTTTTTTAACGACTCTAAATCTCTTGTCCACCACCTTCTTCTCTATGGA
TTTGTTAGCGGACCCGAACTCGGAGACTGTTAGACAGTTCAAGAAGGACGTGGTTGGAATCTTTGAAGAGTTGGGAAAATCCA
ACATGGCTGACCATTTCCCGTTCCTTAAGAAGTTTGATCCCCAAGGAATCCGGAGGCGAACCGAATTGTACTTCCAAAAGATG
TTGGACTTGTTCAATCGCATTATCGACGAAAGGCTGCAGCGGCTCAAGGAAGCGAATGGTTCTGTGAAAGAGAATGACATATT
GGACAATCTCATCAACATGATGTTGGCTGGTGAAGATAATAGAGGATACGATCCAGTAGACAGAGCCGCAATTTCACACTTTT
TACTGGATCTATTTTCAGCCGGGACTGACACGACATCAGCTACGTTAGAATGGGCAATGGCTGAGCTAGTAAAAGCCCCGGAG
ATCATGGCAAAAGCCCGAAAAGAACTAGAGCAAGTTATTGGCAAAGGAAACCAAGTGAAGGAATCGGACATAAACCAACTCCC
TTATTTGCAAGCCATTGTCAAAGAAACCTTCCGCTTGCACCCCGTTGCTCCATTATTGGTTCCTCGCATAGCCGATGGAGACG
TCGAACTCTACGGCTATAGCGTCCCGAAGGATGCTAAAGTGCTAATAAACATATGGGCCATTGGTAGAGACTCAAACGTTTGG
GAAAATTTCGAAGAGTTTATACCGGAGAGGTTTTTGAAAAGTACTATTGAGACCGGAGGTCGGGATTTCGAGCTCATTCCCTT
TGGTGCGGGGAGGAGAATATGTCCTGGGATACTGTTGGCGAACCGAATGTTGCACTTGATGTTGGGTTCTTTGCTTCACTCGT
TTGATTGGAAATTGGAAGACGGGGTTGAGACTGTGAACTTGGATGACAAGTTTGGCCTGACCTTGCAGCTGGCTCAGCCTTTG
CGAGTTGTCCCTATAGCAATCGCTCATCATCATCATCATCATTAG

Microsomal solution	ΔDO	Dilution factor D	Concentration of functional P450s in the microsomal solution C=ΔDO*D/ε (mM)
CYP76F112-A	2,60E-03	40	1,14E-03
CYP76F112-B	7,34E-04	40	3,22E-04
CYP76F112-AB	6,05E-04	40	2,66E-04
CYP76F112-D	3,71E-04	4,3	1,75E-05

Supp. Figure 6 Differential CO spectrums recorded with the microsomes collected from the yeast producing CYP76F112-A, CYP76F112-B, CYP76F112-AB and CYP76F112-D. The raw spectrums have been smoothed prior to measuring the peak heights, which correspond to the ΔDO used to calculate the concentration of functional P450s in the microsomal solution. On the spectrum associated to the mutant D, the peak was small but significant. The dilution factor D corresponds to the dilution of the initial microsomal solution used to record the CO spectrums. $\varepsilon = 91mM^{-1}.cm^{-1}$ (see **Chapter VII**, **B.2.e**).
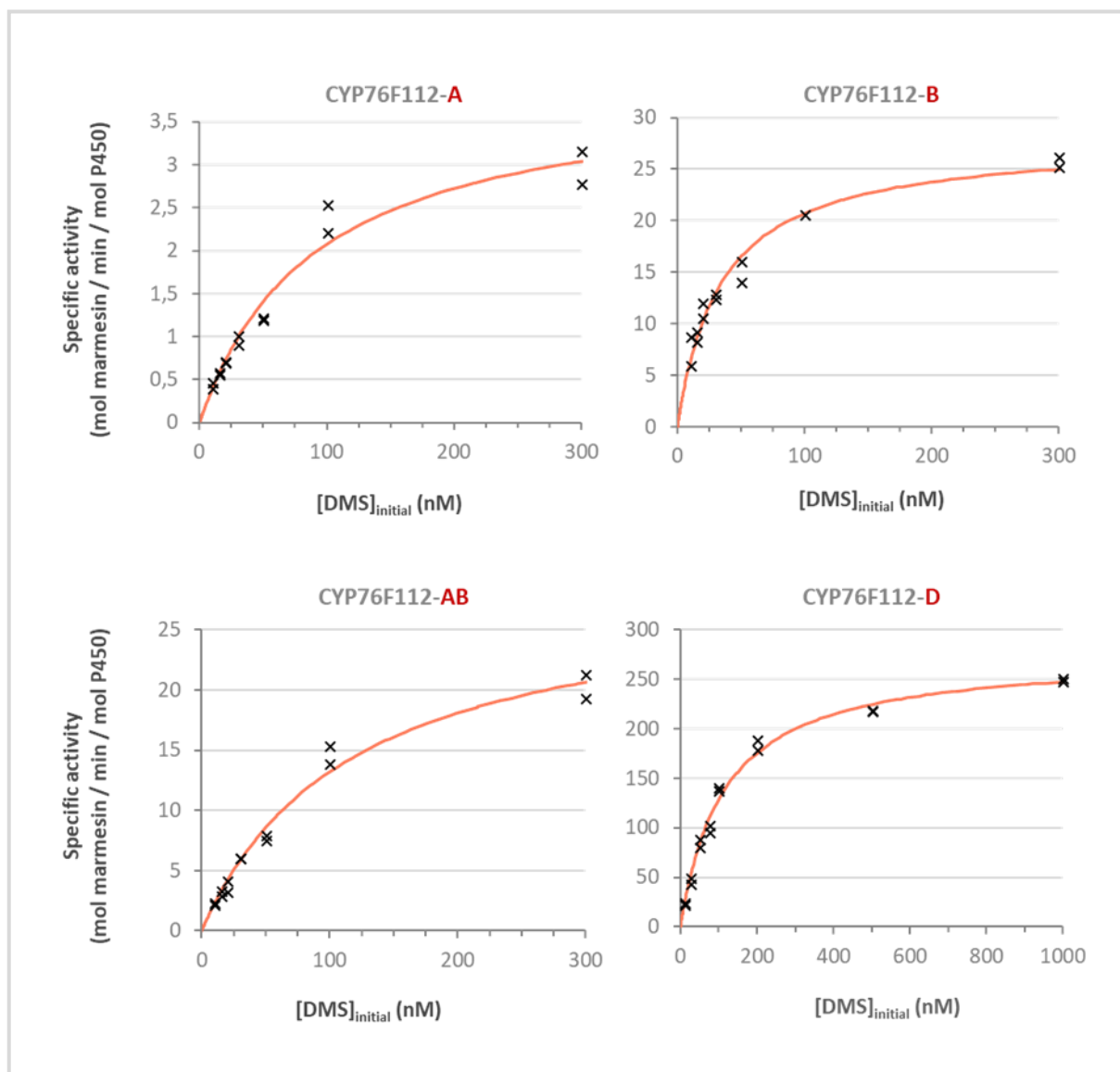
Supp. Figure 7 Determination of the kinetic parameters associated to the conversion of DMS into marmesin by the CYP76F112 mutants. Specific activity of the mutants in the presence of various initial substrate concentrations. The experimental points are marked by black crosses (**Supp. Table 12**) The Michaelis-Menten model curve, plotted on SigmaPlot to fit the experimental data, is in light red.
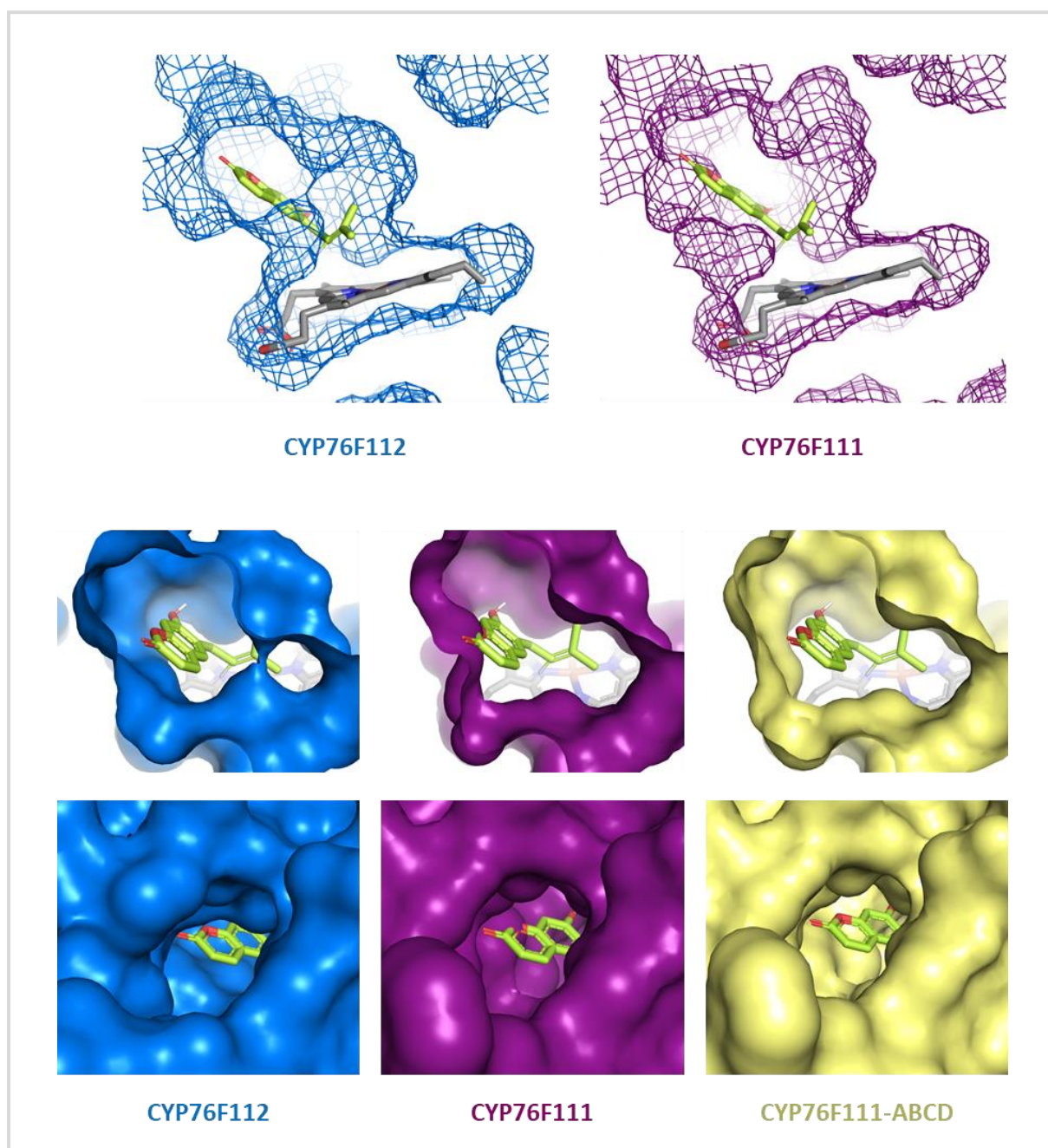
Supp. Figure 8 Global architecture of the substrate-binding site and access channel of CYP76F111, CYP76F112 and the mutant CYP76F111-ABCD. For an easier understanding, the DMS (green) and the heme (grey) are shown in all models as docked within CYP76F112 ("correct" position).
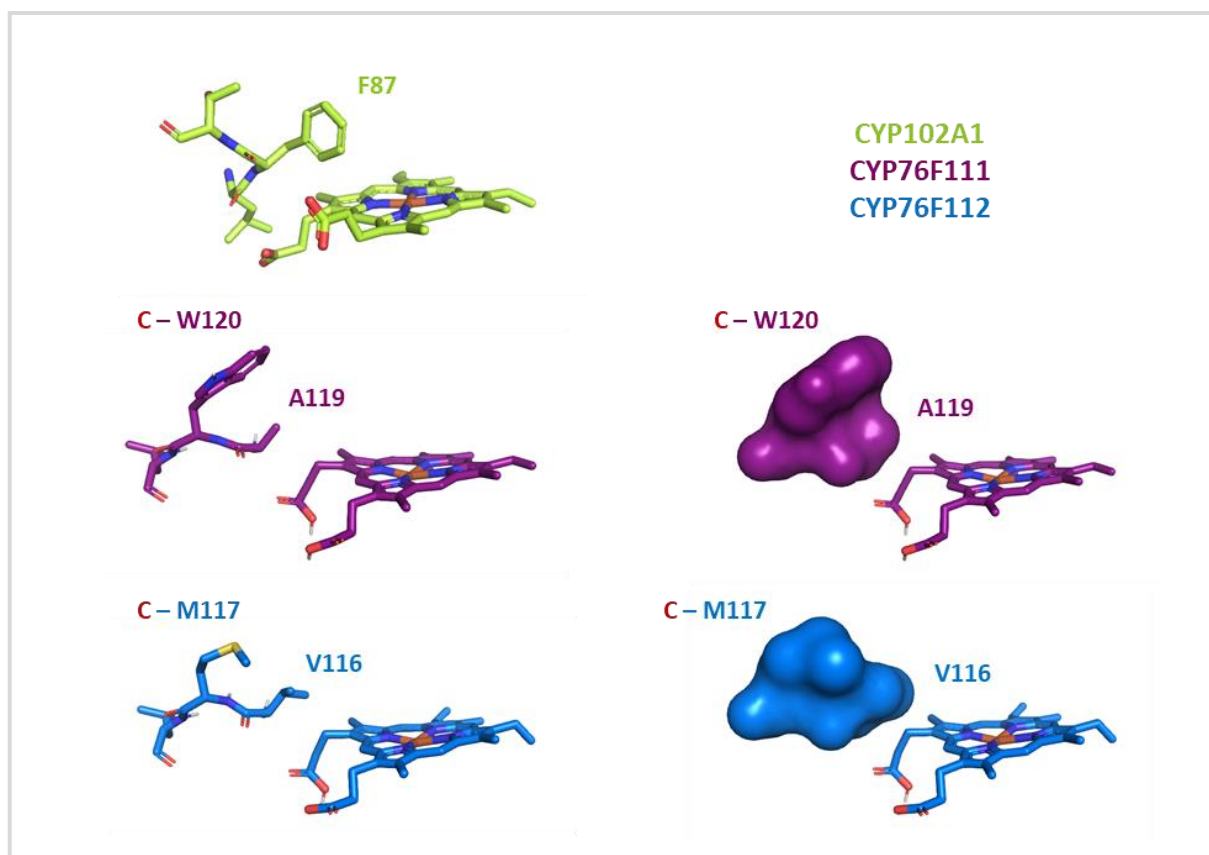
Supp. Figure 9 Identification of the residues from CYP76F111 and CYP76F112 that are equivalent to F87 from CYP102A1.

Supp. Figure 10 Pairs of primers used for the domestication of CYP76F112 and KanaR, and domesticated sequence of these genes. The initial coding sequences are in grey, the GB extremities are in blue, and the modified nucleotide is in red. The primers 1, 2, 3 and 4 refers to **Figure 88**.

Gene	Primer		
	Name	**Purpose**	**Sequence**
CYP76F112	**1**	Addition of the GB extensions	**GCGCCGTCTCGCTCGA**ATGGATATTTTCACCTCCTTACTGTATC
	4		**GCGCCGTCTCGCTCGAAGC**CTAATGATGATGATGATGATGATGCTTTG
	3	Removal of the recognition site	CCAAGTCGT**A**TCCAGCCGGACAATC
	2		CCGGCTGGA**T**ACGACTTGGCTG
KanaR	Dir.	Addition of the GB extensions	**GCGCCGTCTCGCTCGA**ATGGTTGAACAAGATGGATTGCACG
	Rev.		**GCGCCGTCTCGCTCGAAGC**TCAGAAGAACTCGTCAAGAAGGC

>Domesticated_CYP76F112
GCGCCGTCTCGCTCGAATGGATATTTTCACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTT
CCTTTGCGTTTCCTAAACACAAAAGGCTTCCACCTGGTCCAAAACCTCGTCCCATCATCGGAAGCCTCTTGGAGCTCGGCGAC
CAACCCCACAGGTCCTTGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTATGCATTTGAAGCTCGGCCAAGTCACGACGGTTGT
CATTTCCTCCACCACCATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGT**A**TCCAGCCGGACAATCACCGACGCAAGCC
GCGCCCACAGACACAGCGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAATAAGCAACTCA
CACTTGCTTTCCTCCAAGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAA
AAGCGTCCAGGCCGGGGAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTCT
CCATGGACATGGCGGATGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGG
AAGCCTAACTTGGCCGATTATTTCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCG
GAAACTGATCAACTTGTTTGGGCGTATCATCGACCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATA
TTTTAGACACTCTTATCAACATGATGGTGGTGGATCAGGAGAAGAAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTT
TTACTGGATTTATTTTCAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGGCAATGGCTGAGCTAGTAAAAGCGCCAGA
GATTATGTCAAAAGCCCGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATCGGACGTATCTCGACTCC
CTTACTTACAAGCCATTGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGACAGTGAC
ATCGAAATCTCCGACTATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATG
GGAAAATCCCGACAAGTTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGT
TCGGTGCTGGTCGGAGAATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCG
TTTGATTGGAAGTTGGAAGATGGGGTTAGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCA
GCCTTTGCGAGCTATCCCCGTGCCGACAAAGCATCATCATCATCATCATCATTAG**GCTTCGAGCGAGACGGCGC**

>Domesticated_KanaR
GCGCCGTCTCGCTCGAATGGTTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATG
ACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAG
ACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGC
AGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTC
ACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTC
GACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGA
GCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACTCATG
GCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGAC
CGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTA
CGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGA**GCTTCGAGCGAGACGGCGC**

Supp. Figure 11 Annotated sequence of the 5-TUs plasmid harbouring the psoralen biosynthesis pathway. SpecR refers to the gene conferring spectinomycin resistance to the bacteria. RB and LB refers to the right and left borders delimiting the T-DNA. The different parts of the plasmid have been coloured as follow, with the same colour code than in **Figure 71** and the other GB-figures:

```
(16658…3)     LB
(22…1048)     35S        (1050…2102)    PsDiox       (2107…2598)   tNOS
(2611…3637)   35S        (3639…4844)    PsPT1        (4849…5340)   tNOS
(5353…6379)   35S        (6381…7913)    CYP76F112    (7918…8409)   tNOS
(8422…9448)   35S        (9450…10937)   CYP71AJ3     (10942…11433) tNOS
(11446…12472) 35S        (12474…13268)  KanaR        (13273…13764) tNOS
(13786…13809) RB
(14660…15670) SpecR
```

Scars left by the repeated restriction-ligations: `GGAG` `AATG` `GCTT` `CGCT` `GTCA`

>5-TUs plasmid: psoralen biosynthesis pathway in PDGB1Ω1 (16677nt)

```
TAACGGATCCGGTCTCAGGAGACTAGAGCCAAGCTGATCTCCTTTGCCCCGGAGATCACCATGGACGACTTTCTCTATCTCTACGATCTAGGAA
GAAAGTTCGACGGAGAAGGTGACGATACCATGTTCACCACCGATAATGAGAAGATTAGCCTCTTCAATTTCAGAAAGAATGCTGACCCACAGAT
GGTTAGAGAGGCCTACGCGGCAGGTCTGATCAAGACGATCTACCCGAGTAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGAT
GCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAG
GCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAGAAAGTAGTTCCTACTGAATCAAAGGCCATGGAGTCAAAAATTCAGAT
CGAGGATCTAACAGAACTCGCCGTGAAGACTGGCGAACAGTTCATACAGAGTCTTTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATG
GTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATAT
CGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTG
CGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGAC
GTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTT
CCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACTCCGGTATTTTTACAACAATACCACAACAAAACAAACAACAAACAACATTACAATTT
ACTATTCTAGTCGAAATGGCTCCATCTCCTACTGATGATTACAGAAACTTCGCTGTTACAAAGGGACATGGTGTTAAAGGACTTTCAGATCTTA
AATTGGATGCTTTGCCCAGAGCAATACATTCAACCTGTTGAAGAGAGGCTTGATATGACTAAGGTTTTGAAGAAAGAATCTATCCCAGTTATTGA
TATGTCAAATCTTGATGATCCTAAAGTTGCTGATCAAATTGCTGCTGCTGCTGAAAAGTGGGGATTTTTCCAAATTATTAATCATGGTGTTCCA
ATCGAAGTTCTTGAGAATGTTAAAGAGGCTACTAGAAGATTTTTCGCTTTGCCTGTTGAAGAGAAGATTAAATACACTCAAGAACAATCTCCAA
CAAACTCAGTTAGACTTACTACATCTTTTCTTCCTAAGGTTGATAAGGTTTTGGAGTGGAAGGATTATCTTTCTATCTTGGTTTCAGATAAGAA
ATCTTCAGAATTTTGGCCATCTACTTGTAAGAATGATGTTAAAGAATACGTTGAGAAGTCAGAATTCGTTATGAAGTGGCTTTTGAAAGCTCTT
ATGAAGGGATTGAATGTTGATATGGATTCTAAGGAGTCAATCCTTATGGGTTCAACAAGAATTAATTTGAATTATTACCCTGTTTGCCCAAATC
CTGAACTTGCTATTGGAGTTGGTAGGCATTCTGATGTTTCAACTCTTACATTTCTTTTGCAAGATAACGTTGGAGGTTTGCATATTAGAAAGAT
GGATACTGATACATGGATTTTTGTTCCACCTGTTGAGGGAGCTATTGTTATTAATATCGGAGATGCTCTTCAAATCTTGTCTAATGGTAAATAT
AAGTCAGCTGAACATTGTGTTGCTGCTAATGGTTCTAACGATAGAATCTCAGTTCCAATTTTTACTAACCCATCTCCTGATGATATTATTGGAC
CACTTCCTGAGCTTTTGAAGAATGGTGAAAAGCCTATATATAAGCATGTTTTGTACTCTGATTACGTTAAGCATTTCTACAGGAAATCACATGA
TGGAAAGCATACATTGGATTTTTGCTCAAATGTGAGCTTGGAATGGATCTTCGATCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGAT
TGAATCCTGTTGCCGGTCTTGCGACGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATT
TATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAACAAAATATAGCGCGCAAACTAGGATAAATTATCG
CGCGCGGTGTCATCTATGTTACTAGATCGGGAATTGCCAAGCTAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAAT
GTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAA
ATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATGGGACCGACTCGCGCTGTCAGGAGACTAGAGCCAAGCTGATCTCCT
TTGCCCCGGAGATCACCATGGACGACTTTCTCTATCTCTACGATCTAGGAAGAAAGTTCGACGGAGAAGGTGACGATACCATGTTCACCACCGA
TAATGAGAAGATTAGCCTCTTCAATTTCAGAAAGAATGCTGACCCACAGATGGTTAGAGAGGCCTACGCGGCAGGTCTGATCAAGACGATCTAC
CCGAGTAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGA
AAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTC
TAAGAAAGTAGTTCCTACTGAATCAAAGGCCATGGAGTCAAAAATTCAGATCGAGGATCTAACAGAACTCGCCGTGAAGACTGGCGAACAGTTC
ATACAGAGTCTTTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATA
CAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTT
CATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGT
GGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCT
CCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACTCCGGT
ATTTTTACAACAATACCACAACAAAACAAACAACAAACAACATTACAATTTACTATTCTAGTCGAAATGGCTCAAACAATTATGCATTCAAGAC
TTTCTTCAGGATTTTTGCATCTTCAAAGAGATAAGGGTTTTAGGACTTTGCCAACACAAAGAAGGCATGCTAAAGTTGTTAACGGAGATCAAGA
ATTTGCTTTTAGGGTTGTTTCATGTGATAAGAATCTTGATTCTACTAAGAACTTCTCAGGTTCTTGCGAGAAGCCTATCAGAACTCATACAAAT
AAGCTTTTGCAAACTATTTCAGCTACATCTGATAGGGAAGCTATTATTCAACCAAAGGATGATTATGAGGCTCCTTGGCAAAATACTCTTAGAA
GGAAATGGGATGCTTTTTTGTACATTTGGAAGACCATACTCAGCTATCTGCACTATTATTGGTATCTCTTCAGTTTCTCTTTTGCCATTGACATC
AGTTAAGGATTTTTCTGCTCCTTATTTTGTTGGACTTTTGCAAGCTTTGATCCCTTTTCTTTGTGCTAACATCTATACTTCAGGTATTAATCAA
CTTGTTGATGTTGATATCGATAAGATTAATAAGCCATATTTGCCTCTTGTTTCAGGAGAATTTTCTTTGGGAGAGGGTAGAGCTATTGTTTCTG
CTTTTGGCTTTTATGTGCCTTGCTGTTGGTATTTTGTCACATTCTACACCACTTTTCGTTGGAGTTTTGGTTTATTTTCTTATCGGTACAGCTTA
CTCAGTTGAATTGCCCACTTTTGAGGTGGAAGACTAAACCTGCTATGGCTGCTTTTTCTATGGCTGGATTGATGGGTCTTACTATTCAACCAGCT
```

GTTTTCTATCATATCCAAAACGCTCTTGGAAAGCCTATGGTTTTCTCTAAGACTGTTGCTTTCGCTACAATTTTCTTTTCTGTTTTCGCTGCTG
TTTTGGGTGCTATTAAGGATGTTCCTGATGTTGAAGGAGATACTGCTTTCGGTAACAGAACATTTTCAGTTAGGTATGGTCAAGAGAAGGTTTT
CTCTGTTTGTCTTAACATCCTTTTGCTTGCTTACGGATTTGCTGTTGTTGTTGGTGCTTCTTCATCTTTTCTTCTTTGCAAGATTGTTTCTGTT
ATGGGTCATACTACATTGGCTTCATTGCTTTTGCTTAGGGCTAAGTCTACTAATCCAAAGGATCCTGAATCAACACAATCTTTCTACATGTTCC
TTTTTAAATTGCTTTATGCTGAATATGTTTTGATTCATTTTATGAGGTGAGCTTGGAATGGATCTTCGATCCCGATCGTTCAAACATTTGGCAA
TAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGACGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGT
AATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAA
CTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTGCCAAGCTAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCT
ATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTT
TTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATGGGACCGACTCGCGCTGTCAGGAGACTAGA
GCCAAGCTGATCTCCTTTGCCCCGGAGATCACCATGGACGACTTTCTCTATCTCTACGATCTAGGAAGAAAGTTCGACGGAGAAGGTGACGATA
CCATGTTCACCACCGATAATGAGAAGATTAGCCTCTTCAATTTCAGAAAGAATGCTGACCCACAGATGGTTAGAGAGGCCTACGCGGCAGGTCT
GATCAAGACGATCTACCCGAGTAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAACTGC
ATCAAGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGCTTCATAAACCAAGGCAAGTAA
TAGAGATTGGAGTCTCTAAGAAAGTAGTTCCTACTGAATCAAAGGCCATGGAGTCAAAAATTCAGATCGAGGATCTAACAGAACTCGCCGTGAA
GACTGGCGAACAGTTCATACAGAGTCTTTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACTCTCGTCTACTCC
AAGAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCC
CAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGA
TGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTG
GATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATT
TGGAGAGGACTCCGGTATTTTTACAACAATACCACAACAAAACAAACAACAAACAACATTACAATTTACTATTCTAGTCGA-ATGGATATTTTC
ACCTCCTTACTGTATCTTGCTCTCATTCTTTTCTTTTCTCTTCAAGTCTTCCGTTCCTTTGCGTTTCCTAAACACAAAAGGCTTCCACCTGGTC
CAAAACCTCGTCCCATCATCGGAAGCCTCTTGGGAGCTCGGCGACCAACCCCACAGGTCCTTGGCCAGGCTTTCCGAGTCTTACGGCCCGTTTAT
GCATTTGAAGCTCGGCCAAGTCACGACGGTTGTCATTTCCTCCACCACCATGGCTAAAGAAGTCCTCCAGGCAAACAGCCAAGTCGTATCCAGC
CGGACAATCACCGACGCAAGCCGCGCCCACAGACACAGCGATTTTAGCATGGTTATGTTGCCCGTATCCCCTCTGTGGCGAAACCTTCGGAAAA
TAAGCAACTCACACTTGCTTTCCTCCAAGGCTCTTGATGGCAACATGGAGCTGAGAAACAAAAAGGTGCAAGAGCTCCTAAATGATGTCCACAA
AAGCGTCCAGGCCGGGGAGGCGGTGGAGATCGCGAGCCTTTCTTTCAGAGCTACTCTGAATCTCTTGTCCACCACATTTTTTCTCCATGGACATG
GCGGATGACACAAATTCCGTCACTCTAAAAGAGCTCAAGGAGGCTATGTCGCACATGATGGAAGAGTTGGGGAAGCCTAACTTGGCCGATTATT
TCCCGTTTCTACAAAAGATTGACCCCCAAGGCATTAGGCGGCGCAACACGGTTACTTTCCGGAAACTGATCAACTTGTTTGGGCGTATCATCGA
CCAAAGATTGAAAGTGAGAGAAGCGAGTGGTTCTTTGAAAGATGATGATATTTTAGCACTCTTATCAACATGATGGTGGTGGATCAGGAGAAG
AAAGAGGATCAGCTTGACAAAACCATAATTGAACATTTTTTTACTGGATTTATTTTCAGCGGGGACTGAAACGACTTCAACCACGTTGGAGTGGG
CAATGGCTGAGCTAGTAAAAGCGCCAGAGATTATGTCAAAAGCCCGAGCAGAGCTAGATCAAGTTATAGGCAAAGGAAACCAAGTGAAGGAATC
GGACGTATCTCGACTCCCTTACTTACAAGCCATTGTTAAAGAAACCTTCCGCATGCACCCTACAGCTCCATTATTGATTCCTCGCAAAGCCGAC
AGTGACATCGAAATCTCCGACTATATCATCCCGAAGGATGCTCAGGTGATTGTCAATGTATGGGCCATTGGTAGAGACTCAAGCACATGGGAAA
ATCCCGACAAGTTTATACCGGAGAGGTTTTTGGACATCGATATAGATGTCGGAGGCCGGGATTTTAAGCTCATTCCGTTCGGTGCTGGTCGGAG
AATATGTCCCGGATTCCCATTGGCGATGCGAATGTTGCACTTGATGTTGGGGTCTTTGCTTCACTCGTTTGATTGGAAGTTGGAAGATGGGGTT
AGACCTGATGCTCTAAACATGGATGAAAAGTTTGGCCTCACCTTGCAAATGGCTCAGCCTTTGCGAGCTATCCCCGTGCCGACAAAGCATCATC
ATCATCATCATCATTAGGCTTGGAATGGATCTTCGATCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGT
CTTGCGACGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTA
TGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTAT
GTTACTAGATCGGGAATTGCCAAGCTAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGT
TTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATG
AGACAATAACCCTGATAAATGCTTCAATAATGGGACCGACTCGCGCTGTCAGGAGACTAGAGCCAAGCTGATCTCCTTTGCCCCGGAGATCACC
ATGGACGACTTTCTCTATCTCTACGATCTAGGAAGAAAGTTCGACGGAGAAGGTGACGATACCATGTTCACCACCGATAATGAGAAGATTAGCC
TCTTCAATTTCAGAAAGAATGCTGACCCACAGATGGTTAGAGAGGCCTACGCGGCAGGTCTGATCAAGACGATCTACCCGAGTAATAATCTCCA
GGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGAAAGATATATTTCTCAAG
ATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAGAAAGTAGTTCCTA
CTGAATCAAAGGCCATGGAGTCAAAAATTCAGATCGAGGATCTAACAGAACTCGCCGTGAAGACTGGCGAACAGTTCATACAGAGTCTTTTACG
ACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAA
AGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAG
AAAAGGAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACC
CCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGAT
GACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACTCCGGTATTTTTACAACAATACC
ACAACAAAACAAACAACAAACAACATTACAATTTACTATTCTAGTCGA-ATGAAGATGCTTGAGCAAAACCCACTTTATTTGTATTTCTTTCCT
CTTTTTCTTGTTACAATTTTTCTTTATAAGTGGTTGGTTAAGAAAACTCCATCTAAAAATTTGCCACCTTCACCACCTAGACTTCCTATTATTG
GAAATTTGCATCAAATTGGTCCAGATCCTCAAATTTCTCTTAGAGATTTGGCTAGGAAGTACGGACCAGTTATGCATCTTAAATTTGGTTCAGT
TCCTGTTTTGGTTGTTTCTTCAGCTGATGGAGCTAGAGAGATTTTTAAGACACATGATCTTGCTTTTGCTGATAGGCCATACTCTTCAGTTGCT
AACAGAATTTTCTATAATGGTAGGGATATGGTTTTCGCTAGATACACTGAATACTGGAGACAGGTTAAGTCTACTTGTGTTACACAACTTTTGT
CTGTTAAAAGAGTTCAATCATTCCATAACGTTAGGGAAGAGGAAGTTGCTCTTTTGCTTGATAACATCGAGAACTCTAAGTCAAAAGTTATTAA
TCTTTCTGAGATGCTTATTGAATTGACAGGAAATGTTGTTTGCAGAGCTGCTTTGGGATCAGGTTACAACGTTGATTCTTACAAGTCATTGCTT
TTGCAAATTATGGATATGTTGGGATACTCTAGGTCAATTGAAGATTTCTTTCCATCTCTTGGTTGGGTTGATTGGATTACTGGATTGAAGGGTA
AAGTTGAGAAAGCTGCTAATGGTGTTGATGCTTTTCTTGAAGGTGTTTTGAAGAACCATACAAACCCTTCTACTTCTTCAGCTAATAAGGATTT
TGTTTCAATCCTTTTGGAGATCCAAGAAGCTGATGCTGGTTCTTCTATGGATAAGGAATGTATTAAATCACTTATTTGGGATATGTTGGGAGCT
GGTACTGAGACAATTGCTACAGCTCTTGAGTGGACTATTGGAGCTTTGATTAAGTCTCCAGATGCTATGTCAAAGCTTCAAAAAGAGGTTAGAG
AAATTGGAAAGGGTAAATCTAGGATTGAGGAAGGAGATCTTGTTAAGATGGATTATTTGAAGGCTGTTATGAAAGAGTCAATGAGATTGTACTT
CACAGCTCCACTTTTGGTTCCTAGAGAAGCTAGGCAAGATGTTAAGTTTATGGGATACGATATTAAATCTGGTACTCAAGTTCTTATTAATGCT
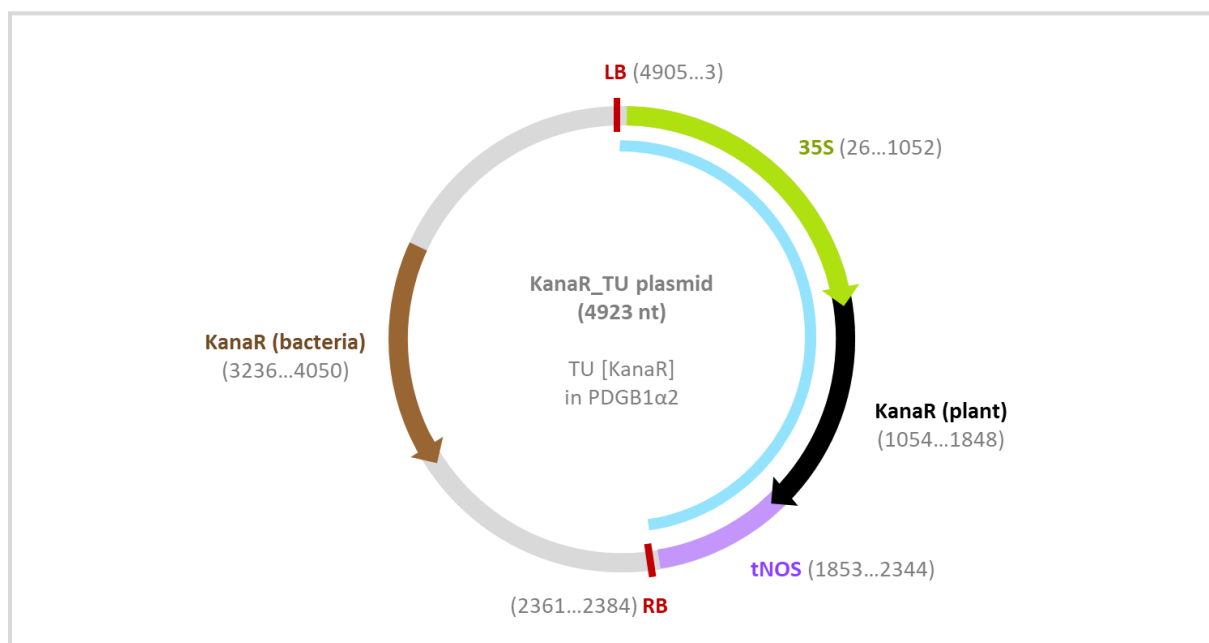TGGGCTATTGCTAGGGATCCATCTTCATGGGATAATCCAGAGGAGTTTAGACCTGAGAGGTTTTTGAACTCTCCTATCGATTACAAGGGTTTTA

ATTACGAATACATCCCATTCGGAGCTGGTAGAAGGGGATGCCCTGGTATCCAATTCGCTATCTCAGTTAATGAGCTTGTTGTTGCTAACGTTGT
TAATAAGTTTAATTTCGAACTTCCTGATGGAAAAAGATTGGAGGAGATGGATATGACTGCTTCTACAGGTATCACTTTCCATAAGAAATCACCA
CTTTTGGTTGTTGCTAGGCCTCATGTTTGATGAGCTTGGAATGGATCTTCGATCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATT
GAATCCTGTTGCCGGTCTTGCGACGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTT
ATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGC
GCGCGGTGTCATCTATGTTACTAGATCGGGAATTGCCAAGCTAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATG
TCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAA
TATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATGGGACCGACTCGCGCTGTCAGGAGACTAGAGCCAAGCTGATCTCCTT
TGCCCCGGAGATCACCATGGACGACTTTCTCTATCTCTACGATCTAGGAAGAAAGTTCGACGGAGAAGGTGACGATACCATGTTCACCACCGAT
AATGAGAAGATTAGCCTCTTCAATTTCAGAAAGAATGCTGACCCACAGATGGTTAGAGAGGCCTACGCGGCAGGTCTGATCAAGACGATCTACC
CGAGTAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGAA
AGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCAAGGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTCT
AAGAAAGTAGTTCCTACTGAATCAAAGGCCATGGAGTCAAAAATTCAGATCGAGGATCTAACAGAACTCGCCGTGAAGACTGGCGAACAGTTCA
TACAGAGTCTTTTACGACTCAATGACAAGAAGAAAATCTTCGTCAACATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATAC
AGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTC
ATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTG
GTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTC
CACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACTCCGGTA
TTTTTACAACAATACCACAACAAAACAAACAACAAACAACATTACAATTTACTATTCTAGTCGA ATG GTTGAACAAGATGGATTGCACGCAGG
TTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCG
CAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGA
CGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTC
ATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCAC
CAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC
CAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACTCATGGCGATGCCTGCTTGCCGAATATCATGGT
GGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCT
GAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTG
ACGAGTTCTTCTGAGCTTGGAATGGATCTTCGATCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTT
GCGACGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGA
TTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTT
ACTAGATCGGGAATTGCCAAGCTAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTC
TTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGA
CAATAACCCTGATAAATGCTTCAATAATGGGACCGACTCGCGCTGTCATGAGACCGGATCCTGACAGGATATATTGGCGGGTAAACTAAGTCGC
TGTATGTGTTTGTTTGAGATCTCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT
CCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTA
GGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCT
TCGGAAGAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTA
TCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGTGTAACATTGGTCTAGTGA**TTATT**
TGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAG
ATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTG
CCGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCACCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGG
TTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGC
TTTTGTCAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGT
TCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAG
CCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGG
CTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGT
TGAGTCGATACTTCGGCGATCACCGCTTCCCTCATGATGTTTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGCTGCTCCATA
ACATCAAACATCGACCCACGGCGTAACGCGCTTGCTGCTTGGATGCCCGAGGCATAGACTGTACCCAAAAAAACAGTCATAACAAGCCATGAA
AACCGCCACTGCGCCGTTACCACCGCTGCGTTCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTACAGCTTACGAACC
GAACAGGCTTATGTCCACTGGGTTCGTGCCTTCATCCGTTTCCACGGTGTGCGTCACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGCATTTC
TGTCCTGGCTGGAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACAT
CAGAGATTTTGAGACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCTTCCCGACAACGCAGACC
GTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAACTGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTCTGGCTGGATGA
TGGGGCGATTCAGGCGATCCCCATCCAACAGCCCGCCGTCGAGCGGGCTTTTTATCCCCGGAAGCCTGTGGATAGAGGGTAGTTATCCACGTG
AAACCGCTAATGCCCCGCAAAGCCTTGATTCACGGGGCTTTCCGGCCCGCTCCAAAAACTATCCACGTGAAATCGCTAATCAGGGTACGTGAAA
TCGCTAATCGGAGTACGTGAAATCGCTAATAAGGTCACGTGAAATCGCTAATCAAAAAGGCACGTGAGAACGCTAATAGCCCTTTCAGATCAAC
AGCTTGCAAACACCCCTCGCTCCGGCAAGTAGTTACAGCAAGTAGTATGTTCAATTAGCTTTTCAATTATGAATATATATATCAATTATTGGTC
GCCCTTGGCTTGTGGACAATGCGCTACGCGCACCGGCTCCGCCCGTGGACAACCGCAAGCGGTTGCCCACCGTCGAGCGCCTTTGCCCACAACC
CGGCGGCCGGCCGCAACAGATCGTTTTATAAATTTTTTTTTTTTGAAAAAGAAAAAGCCCGAAAGGCGGCAACCTCTCGGGCTTCTGGATTTCCG
ATCCCCGGAATTAGAGATCTTGGCAGGATATATTGTGGT

Supp. Figure 12 Annotated sequence of the KanaR plasmid. The KanaR gene in black refer to the kanamycin resistance gene included in the T-DNA (for plant); the one in brown refers to the gene conferring kanamycin resistance to the bacteria. The different parts of the plasmid have been coloured as follow, with the same colour code than in **Supp. Figure 11**, **Supp. Figure 13**, and the other GB-figures:

```
(4905…3)      LB          (26…1052)      35S          (1054…1848)    KanaR (plant)
(1853…2344)   tNOS        (2361…2384)    RB           (3236…4050)    KanaR (bacteria)
Scars left by the repeated restriction-ligations:     GGAG   AATG   GCTT   CGCT   GTCA
```

>KanaR plasmid: TU [KanaR] in PDGB1α2 (4923nt)

```
TAACAAGCTTCGTCTCAGTCAGGAGACTAGAGCCAAGCTGATCTCCTTTGCCCCGGAGATCACCATGGACGACTTTCTCTATCTCTACGATCTA
GGAAGAAAGTTCGACGGAGAAGGTGACGATACCATGTTCACCACCGATAATGAGAAGATTAGCCTCTTCAATTTCAGAAAGAATGCTGACCCAC
AGATGGTTAGAGAGGCCTACGCGGCAGGTCTGATCAAGACGATCTACCCGAGTAATAATCTCCAGGAGATCAAATACCTTCCCAAGAAGGTTAA
AGATGCAGTCAAAAGATTCAGGACTAACTGCATCAAGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATT
CAAGGCTTGCTTCATAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAGAAAGTAGTTCCTACTGAATCAAAGGCCATGGAGTCAAAAATTC
AGATCGAGGATCTAACAGAACTCGCCGTGAAGACTGGCGAACAGTTCATACAGAGTCTTTTACGACTCAATGACAAGAAGAAAATCTTCGTCAA
CATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTA
ATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAATGCCATC
ATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGA
AGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGAC
CCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACTCCGGTATTTTTACAACAATACCACAACAAAACAAACAACAAACAACATTACA
ATTTACTATTCTAGTCGA ATG GTTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCA
CAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCC
TGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGG
AAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCA
ATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCG
GTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGA
GGATCTCGTCGTGACTCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGT
GTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACG
GTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCTTGGAATGGATCTTCGATCCCGATCGTTCA
AACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGACGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAA
TAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAAT
ATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTGCCAAGCTAATTCTTGAAGACGAAAGGGCCT
CGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCT
ATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATGGGACCGACTCGCGCTTG
AGACGAAGCTTGACAGGATATATTGGCGGGTAAACTAAGTCGCTGTATGTGTTTGTTTGAGATCTCATGTGAGCAAAAGGCCAGCAAAAGGCCA
GGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCG
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTG
TCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACT
GGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGA
AGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTA
GCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA
GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAA
TCAATCTAAAGTATATATGTGTAACATTGGTCTAGTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTAT
CAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGC
GATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAA
TCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCA
AACCGTTATTCATTCGTGATTGCGCCTGAGCGAGTCGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCG
CAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCTGGGATCGCAGTGGTG
AGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTG
TAACAACATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGGTAGATTGTCGCACCTGATTG
CCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTTGAGCAAGACGTTTCCCGTTGAATA
TGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGA
GATTTTGAGACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCTTCCCGACAACGCAGACCGTTC
CGTGGCAAAGCAAAAGTTCAAAATCACCAACTGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTCTGGCTGGATGATGGG
GCGATTCAGGCGATCCCCATCCAACAGCCCGCCGTCGAGCGGGCTTTTTTATCCCCGGAAGCCTGTGGATAGAGGGTAGTTATCCACGTGAAAC
CGCTAATGCCCCGCAAAGCCTTGATTCACGGGCTTTCCGGCCCGCTCCAAAAACTATCCACGTGAAATCGCTAATCAGGGTACGTGAAATCGC
TAATCGGAGTACGTGAAATCGCTAATAAGGTCACGTGAAATCGCTAATCAAAAAGGCACGTGAGAACGCTAATAGCCCTTTCAGATCAACAGCT
TGCAAACACCCCTCGCTCCGGCAAGTAGTTACAGCAAGTAGTATGTTCAATTAGCTTTTCAATTATGAATATATATATCAATTATTGGTGCCC
TTGGCTTGTGGACAATGCGCTACGCGCACCGGCTCCGCCCGTGGACAACCGCAAGCGGTTGCCCACCGTCGAGCGCCTTTGCCCACAACCGGC
GGCCGGCCGCAACAGATCGTTTTATAAATTTTTTTTTTTGAAAAAGAAAAAGCCCGAAAGGCGGCAACCTCTCGGGCTTCTGGATTTCCGATCC
CCGGAATTAGAGATCTTGGCAGGATATATTGTGGT
```

Supp. Figure 13 **KanaR plasmid map**. RB and LB refers to the right and left borders delimiting the T-DNA to be inserted in the plant, highlighted in blue. The KanaR gene in black is the one that is inserted in the plant. The KanaR gene in brown refers to the gene conferring kanamycin resistance to the bacteria.

Abstract / Résumé

How to produce your furanocoumarins: the hidden pathway ~ From the characterisation of new P450s to the evolution of the furanocoumarin pathway and the development of tools allowing the study of furanocoumarins' metabolic cost.

To cope with pests and pathogens, plants have evolved many defence mechanisms such as the production of various specialised metabolites. For instance, some species including the fig tree (*Ficus carica*) produce toxic molecules called furanocoumarins. As the use of pesticides is being restricted, these age-old adaptations represent an interesting source of inspiration to find crop protection alternatives. Therefore, improving our understanding of plant defences becomes essential to rethink our crop management strategies. The first objective of this project was to pursue the molecular elucidation of the furanocoumarin biosynthesis pathway and to gain evolutionary insights through the characterisation of new genes in *F. carica*. The second objective was to assess the metabolic cost of furanocoumarin production by inserting the associated pathway in the genome of tomato (*Solanum lycopersicum*), a plant that does not naturally produce these molecules. To identify new genes involved in the furanocoumarin biosynthesis, we combined transcriptomic and metabolomic approaches: we identified candidate genes in a *F. carica* differential RNAseq database, cloned their coding sequences, heterologously expressed them, and performed enzymatic assays. This led us to identify three enzymes with original activities. Among these, CYP76F112 plays a key role in the furanocoumarin biosynthesis by converting demethylsuberosin into marmesin with a very high affinity. A phylogenetic gene-family analysis strongly suggests that CYP76F112 evolved recently in a restricted taxon of the Moraceae family, through an expansion of the CYP76Fs. Modelling and site-directed mutagenesis experiments permitted to highlight four amino acids that impact CYP76F112 specificity and affinity. Moreover, the recent evolution of these amino acids has most certainly been critical for the emergence of the marmesin synthase activity. Therefore, CYP76F112 significantly improves our understanding of furanocoumarin production in higher plants, since its recent evolution supports the hypothesis that furanocoumarins have appeared by convergent evolution in distant plant families. In addition, as CYP76F112 completes the set of four enzymes that allow the conversion of coumaric acid (a common molecule) into psoralen (a toxic furanocoumarin), it opens new horizons for the use of furanocoumarins in the study of plant defence. For instance, by generating psoralen-producing tomatoes, it is now possible to evaluate the metabolic costs and defensive profits linked to psoralen production, which might lead us to better understand the trade-offs between growth and defence. Consequently, we used a multi-gene cloning technology called GoldenBraid to construct a plasmid harbouring the four genes of the psoralen biosynthesis pathway. Using transgenesis methods along with *in vitro* culture, we used this plasmid to initiate plant transformation, in order to generate psoralen-producing tomatoes. These transformations were not conclusive, but our study constitutes a pioneer work that will be continued. In particular, it allowed to the identification of some weaknesses in the initial strategy and the establishment of recommendations that will be essential to overcome these limits.

Keywords: Furanocoumarins, *Ficus carica*, marmesin synthase, convergent evolution, multigenic cloning, trade-off.

A la croisée des voies, ou comment produire des furocoumarines ~ De la caractérisation de P450s à l'évolution de la voie des furocoumarines et au développement d'outils permettant l'étude du coût métabolique des furocoumarines.

Au cours de l'évolution, les plantes ont développé de nombreux mécanismes de défense contre les stress biotiques, telles que la production de métabolites secondaires variés. Par exemple, certaines espèces comme le figuier (*Ficus carica*) produisent des molécules toxiques appelées furocoumarines. A l'heure où l'utilisation des pesticides est de plus en plus restreinte, comprendre les mécanismes naturels de défense des plantes devient de plus en plus important pour imaginer des méthodes de protection alternatives et repenser la gestion de nos cultures. Le premier objectif de ce projet consistait à poursuivre l'élucidation moléculaire de la voie de biosynthèse des furocoumarines ainsi qu'à mieux comprendre son évolution, à travers la caractérisation de nouveaux gènes de *F. carica*. Le second objectif consistait à évaluer le coût métabolique de la production de furocoumarines, en insérant la voie associée dans le génome de la tomate (*Solanum lycopersicum*), une plante qui, naturellement, ne produit pas ces molécules. Pour identifier de nouveaux gènes impliqués dans la synthèse de furocoumarines, nous avons combiné des approches de transcriptomique et de métabolomique : nous avons identifié des candidats dans une banque différentielle d'ARNseq de *F. carica*, avons cloné leurs séquences codantes, les avons exprimés dans un système hétérologue, et avons mené des tests d'activité enzymatiques. Nous avons ainsi identifié trois enzymes possédant des activités originales. Parmi elles, CYP76F112 joue un rôle clé dans la synthèse de furocoumarines, en convertissant la déméthylsubérosine en marmésine. Une analyse phylogénétique de gènes a été réalisée, suggérant fortement que CYP76F112 a évolué récemment dans un taxon restreint de Moracées, via une expansion de CYP76Fs. Des approches de modélisation et de mutagenèse dirigée ont également permis d'identifier quatre acides aminés impactant la spécificité et l'affinité de CYP76F112. Ces acides aminés semblent eux aussi avoir évolué récemment, et leur évolution a certainement été critique pour l'apparition de l'activité marmésine synthase. Ainsi, CYP76F112 améliore significativement notre compréhension de la production de furocoumarines dans le règne végétal, dans la mesure où son apparition récente supporte l'hypothèse selon laquelle les furocoumarines seraient apparues par évolution convergente dans des familles de plantes éloignées. De plus, CYP76F112 complète le jeu de quatre enzymes permettant la synthèse de psoralène (une furocoumarine toxique) à partir d'acide coumarique (un composé commun), ce qui ouvre de nouveaux horizons quant à l'utilisation des furocoumarines dans l'étude de la défense des plantes. Par exemple, en générant des tomates productrices de psoralène, il est désormais possible d'évaluer le coût métabolique et les bénéfices défensifs liés à la production de psoralène, ce qui pourrait conduire à une meilleure compréhension des compromis croissance-défense. En conséquence, nous avons utilisé la technique de clonage multigénique du GoldenBraid pour construire un plasmide portant les 4 gènes de la voie de biosynthèse du psoralène. En utilisant des techniques de transgenèse et de culture *in vitro*, nous avons utilisé ce plasmide pour initier des transformations de plantes, dont le but était de générer des tomates productrices de psoralène. Ces transformations n'ont pas été concluantes, mais notre étude constitue un travail pionnier qui sera poursuivi. En particulier, il a mené à l'identification de failles dans notre stratégie initiale, ainsi qu'à l'établissement de recommandations essentielles pour dépasser ces limites.

Mots-clés : Furocoumarines, *Ficus carica*, marmésine synthase, évolution convergente, clonage multigénique, trade-off.