

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

IDENTIFICATION ET CARACTÉRISATION FONCTIONNELLE
D'ORTHOLOGUES DE LA NORBELLADINE SYNTHASE CHEZ
LEUCOJUM AESTIVUM ET *NARCISSUS PAPYRACEUS*

MÉMOIRE PRÉSENTÉ
COMME EXIGENCE PARTIELLE DE LA
MAÎTRISE EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

PAR
LAURENCE TOUSIGNANT

AVRIL 2022

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

REMERCIEMENTS

Je souhaite tout d'abord exprimer ma gratitude à toutes les personnes m'ayant apporté du soutien, que ce soit de près ou de loin, durant mon cheminement à la maîtrise. J'en suis très reconnaissante. Je tiens à remercier tout particulièrement ma directrice Isabel Desgagné-Penix de m'avoir accueillie parmi son équipe de recherche et d'avoir cru en moi tout au long de mon parcours académique. Son encadrement, son soutien, son écoute et sa disponibilité ont fait en sorte que ce travail voit le jour et m'ont permis de persévérer du début jusqu'à la fin. Ces années de maîtrise m'ont été très enrichissantes et il a été un plaisir de travailler aux côtés de gens aussi inspirants et agréables.

Je tiens également à remercier le professeur Hugo Germain (Université du Québec à Trois-Rivières) pour ses nombreux conseils, son support et pour avoir partagé une quantité considérable de connaissances. J'aimerais remercier mes nombreux collègues du laboratoire d'Isabel et d'Hugo dont j'ai eu la chance de côtoyer durant ces dernières années. Je voudrais remercier plus particulièrement mes collègues Marguerite Cinq-Mars et Joëlle Rancourt, sans qui mon passage à la maîtrise n'aurait pas été le même. Nous partageons de grandes amitiés remplies de beaux souvenirs. Joëlle, merci d'avoir été autant à l'écoute et toujours présente pour moi durant mon parcours ainsi que durant la rédaction de mon mémoire.

Je souhaite remercier les organismes subventionnaires de ce projet, soit le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et également la Chaire de recherche du Canada sur le métabolisme spécialisé végétal dont Dr Isabel Desgagné-Penix est titulaire.

Finalement, je souhaite remercier ma famille, plus particulièrement mes parents, Marcel et Linda, ainsi que mon frère Joshua pour m'avoir toujours encouragée à poursuivre mes études. Vous avez tous été une source d'inspiration et de support pour moi et je n'y serais jamais arrivé sans vous. Un immense merci à mon conjoint, André,

pour avoir été un partenaire incroyable de support et d'encouragements tout au long de mon parcours universitaire.

RÉSUMÉ

Dernièrement, les plantes de la famille des Amaryllidacées ont suscité un grand intérêt en raison de leur capacité à synthétiser des métabolites spécialisés. Parmi ces métabolites spécialisés, les alcaloïdes sont des sources très prometteuses de produits naturels bioactifs aux puissantes activités pharmacologiques, pouvant être utilisés comme médicament. Cependant, comme d'autres métabolites végétaux, les alcaloïdes des Amaryllidacées (AA) sont produits en très petite quantité dans les plantes, ce qui les rend difficiles à produire à grande échelle de façon rentable. De plus, la structure des AAs est complexe et il est difficile de les synthétiser chimiquement. Il existe un intérêt à manipuler génétiquement la production d'AA afin de développer de nouveaux produits pharmaceutiques, mais le manque de connaissances sur les voies de biosynthèse des AA restreint cette possibilité. Une connaissance approfondie du métabolisme des AA est nécessaire pour la production d'alcaloïdes dans le but de (1) *générer des plateformes de microorganismes modifiés* et (2) *d'améliorer le profil en alcaloïdes des cultivars existant*. La biosynthèse d'alcaloïdes à partir de ces méthodes alternatives diminuerait grandement les coûts associés à sa production et permettrait ainsi la commercialisation de nouveaux médicaments.

Tous les AA sont synthétisés à partir d'un intermédiaire commun dans leur voie de biosynthèse : la norbelladine. Cette molécule est formée par la condensation de la tyramine avec le 3,4-dihydroxybenzaldéhyde à l'aide de la norbelladine synthase (NBS). Le projet proposé a pour but de caractériser la NBS, soit en déterminant sa localisation subcellulaire, chez les espèces de plantes *Leucojum aestivum* et *Narcissus papyraceus*. Pour ce faire, plusieurs expérimentations ont dû être réalisées. Tout d'abord, les gènes candidats codant pour la NBS ont été identifiés chez *L. aestivum* et *N. papyraceus* à partir de leur banque de données transcriptomique respective. Par la suite, un clonage des gènes candidats de la NBS dans un vecteur arborant la eGFP a été réalisé à l'aide du système Gateway. Les vecteurs ont ensuite été introduits dans la plante *Nicotiana benthamiana* par agroinfiltration afin de permettre l'expression transitoire de la NBS dans celle-ci. L'observation des feuilles agroinfiltrées au microscope confocal a démontré que la NBS est localisée dans le cytoplasme et le noyau des cellules végétales.

En conclusion, l'élucidation de la localisation nucléocytoplasmique de la NBS suggère que la biosynthèse des AA se produit dans le cytoplasme et/ou le noyau. Cette découverte permettra de faciliter la bioingénierie de microorganismes ou de plantes pour la production rentable de médicaments dans le futur.

Mots-clés : Amaryllidacée, alcaloïde, norbelladine synthase, norbelladine, *Leucojum aestivum*, *Narcissus papyraceus*, localisation subcellulaire.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
RÉSUMÉ.....	iv
LISTE DES TABLEAUX.....	viii
LISTE DES FIGURES	ix
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	x
LISTE DES SYMBOLES.....	xii
CHAPITRE I	
INTRODUCTION.....	1
1.1 Les métabolites spécialisés des plantes et leurs fonctions.....	1
1.1.1 Rôles chez les végétaux	1
1.1.2 Fonctions médicinales.....	2
1.1.3 Limitations des plantes médicinales	4
1.2 Les classes de métabolites spécialisés	6
1.2.1 Les terpènes	6
1.2.2 Les composés phénoliques.....	10
1.2.3 Les alcaloïdes.....	12
1.3 Biosynthèse, transport et accumulation des métabolites spécialisés dans la plante.....	15
1.4 Classification des alcaloïdes	17
1.4.1 Les purines.....	18
1.4.2 Les alcaloïdes tropaniques.....	19
1.4.3 Les pyrrolizidines	20
1.4.4 Les alcaloïdes quinolizidiniques.....	20
1.4.5 Les alcaloïdes indolo-monoterpéniques.....	21
1.4.6 Les alcaloïdes isoquinoléiques.....	24
1.4.6.1 Les alcaloïdes benzyloisoquinolines	24
1.4.6.2 Les alcaloïdes des Amaryllidacées	26
1.5 Problématique, objectifs et méthodologie	31

CHAPITRE II	
ANALYSE DU TRANSCRIPTOME DE <i>LEUCOJUM AESTIVUM</i> ET IDENTIFICATION DE GÈNES IMPLIQUÉS DANS LA SYNTHÈSE DE LA NORBELLADINE.....	34
2.1 Contribution des auteurs	34
2.2 Résumé de l'article	35
2.3 Article scientifique.....	36
Abstract.....	36
Introduction.....	38
Materials and methods.....	40
Plant material	40
Alkaloid extraction, TLC and HPLC analysis.....	41
RNA extraction, Illumina sequencing and transcriptome assembly.....	42
Candidate gene selection and primers design	43
Protein alignment and phylogenetic analysis.....	44
PCR amplification, cloning, and construction of the expression cassette for enzyme assay	45
Expression and purification of MBP fusion proteins in <i>E. coli</i>	45
Substrates, standards preparation and LC-MS/MS analysis of enzymatic assays.....	47
Cloning and PCR analysis for localization assay	48
Plant transformation and confocal microscopy	49
Protein extraction from <i>N. benthamiana</i> leaves and Western blotting...	50
Accession Numbers	50
Results	52
Alkaloids.....	52
RNA-Seq, de novo assembly and annotation of <i>L. aestivum</i> bulb's transcriptome	53
Identification of transcripts involved in AA metabolism	54
PCR amplification of NBS	55
Phylogenetic analysis of norbelladine synthase.....	55
Heterologous expression of <i>LaNBS</i> in <i>E. coli</i>	56
<i>LaNBS</i> is able to synthesize norbelladine in vitro.....	57

Predictions of subcellular localizations	58
Cloning of NBS into pB7FWG2 vector.....	59
Subcellular localization of GFP tagged at the C-terminus of NBS	59
Discussion.....	60
Conclusion.....	66
References.....	67
Tables	74
Figure legends.....	79
CHAPITRE III	
DISCUSSION	89
3.1 Retour sur les résultats de recherche	90
3.2 Perspectives du projet de recherche.....	96
3.3 Conclusions	98
RÉFÉRENCES BIBLIOGRAPHIQUES.....	99
ANNEXE A	
SUPPLEMENTARY DATA	115
ANNEXE B	
CLONING AND CHARACTERIZATION OF NORBELLADINE SYNTHASE CATALYZING THE FIRST COMMITTED REACTION IN AMARYLLIDACEAE ALKALOID BIOSYNTHESIS.....	123

LISTE DES TABLEAUX

Tableau		Page
1.1	Classification de différents métabolites spécialisés et leurs effets biologiques comme agent thérapeutique.....	4
1.2	Classification des terpènes selon le nombre d'unité d'isoprène (C ₅) dans leurs structures	9
1.3	Différents alcaloïdes commercialisés/testés cliniquement comme médicament et leurs propriétés thérapeutiques	14
1.4	Les 9 types d'AA ainsi que leurs structures et activités médicinales	28

LISTE DES FIGURES

Figure		Page
1.1	Facteurs biotiques et abiotiques qui influencent la biosynthèse de métabolites secondaires dans les plantes	3
1.2	Voies de biosynthèses des métabolites primaires et spécialisés	7
1.3	Structure chimique de l'isopentényl-pyrophosphate ($C_5H_{12}O_7P_2$)	8
1.4	Biosynthèse des terpénoïdes dans les plantes, montrant les étapes majeures menant à la formation des différents produits terpéniques. CoA, coenzyme A; GAP, glycéraldéhyde-3-phosphate	10
1.5	Classification des composés phénoliques	11
1.6	Représentation du mécanisme de transport entre différents tissus par les cellules du Xylème et du Phloème	17
1.7	Quelques exemples de purines dérivées de la plante ainsi que leurs structures	18
1.8	Structure du squelette tropanique et les trois groupes majeurs des tropanes .	19
1.9	Exemple de structure de différents alcaloïdes pyrrolizidiques	20
1.10	Quelques exemples d'alcaloïdes quinolizidiniques	21
1.11	Différents alcaloïdes indolo-monoterpéniques, leurs fonctions biologiques et leurs sources biosynthétiques.....	23
1.12	Exemples de diverses structures à squelettes simples des alcaloïdes des benzyloquinolines	25
1.13	Distribution géographique des plantes de la famille des Amaryllidacées	27
1.14	Représentations structurales des différents types d'AA	29
1.15	Synthèse de la norbelladine à partir de la tyramine et de 3,4-dihydroxybenzaldéhyde (3,4-DHBA), précurseur commun à tous les alcaloïdes des Amaryllidacées incluant la galanthamine, la lycorine et l'haemanthamine.....	30

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

3,4-DHBA	3,4-Dihydroxybenzaldéhyde
4-HPAA	4-Hydroxyphénylacétaldéhyde
AA	Alcaloïde des Amaryllidacées
ABI	Alcaloïde des benzylioquinolines
ACh	Acétylcholine
AChE	Acétylcholinestérase
ADN	Acide désoxyribonucléique
ADNc	Acide désoxyribonucléique complémentaire
AI	Alcaloïde isoquinoléique
AIM	Alcaloïde indolo-monoterpénique
AP	Alcaloïde pyrrolizidinique
APX	Ascorbate peroxydase
AQ	Alcaloïde quinolizidinique
ARN	Acide ribonucléique
AT	Alcaloïde tropanique
C3'H1	Coumarate 3-hydroxylase
CYP96T1	Cytochrome P450 monooxygénase 96T1
DMAPP	Diméthylallyl-pyrophosphate
DRA	Dérivé réactif de l'azote
DRO	Dérivé réactif de l'oxygène
FPP	Farnésyl-pyrophosphate
GFP	Green fluorescent protein
GGPP	Géranylgéranyl-pyrophosphate

GPP	Géranyl-pyrophosphate
IPP	Isopentényl-pyrophosphate
IPR	Protéine intracellulaire liée à la pathogénèse
LC-MS/MS	Chromatographie en phase liquide couplée à la spectrométrie de masse
MA	Maladie d'Alzheimer
MEP	Voie du méthylérythritol-phosphate
MS	Métabolite spécialisé
MVA	Voie du mévalonate
N4OMT	Norbelladine 4'- <i>O</i> -méthyltransférase
NBS	Norbelladine synthase
NCS	Norcoclaurine synthase
NR	Noroxomaritidine réductase
PAL	Phénylalanine ammonia-lyase
PR10	Protéine liée à la pathogénèse de classe 10
pb	Paire de base
RE	Réticulum endoplasmique
SLN	Signal de localisation nucléaire
TYDC	Tyrosine décarboxylase

LISTE DES SYMBOLES

%	Pourcentage
UV	Ultraviolet
min	Minute
h	Heure
mL	Millilitre
°C	Degré Celsius
rpm	Rotation par minute
g	Gramme
pH	Potentiel hydrogène
μL	Microlitre
cm	Centimètre
m/z	Masse sur charge
mg	Milligramme
mm	Millimètre
λ	Lambda
R ²	Coefficient de corrélation
nm	Nanomètre
α	Alpha
mM	Millimolaire
ppm	Partie par million
kDa	Kilo Dalton

CHAPITRE I

INTRODUCTION

1.1 Les métabolites spécialisés des plantes et leurs fonctions

Les plantes produisent une grande variété de composés subdivisés en deux groupes majeurs : les métabolites primaires et les métabolites spécialisés (également appelé métabolites secondaires). Le métabolisme primaire est impliqué dans les fonctions vitales telles que la division cellulaire, le développement, la respiration, l'entreposage de nutriments et la reproduction (Bourgaud, Gravot et al. 2001). Les métabolites primaires rassemblent les lipides, les glucides, les acides aminés et les nucléotides et sont communs à tous les organismes. À l'opposé, le métabolisme spécialisé est plutôt impliqué dans des processus essentiels tels que la croissance, le développement, la défense et l'adaptation des organismes. On les retrouve généralement chez des organismes spécifiques comme les bactéries, les champignons et les plantes. Les métabolites spécialisés (MS) sont biosynthétisés et accumulés dans des organes et/ou tissus particuliers, à un moment précis de la vie de l'organisme, afin de lui conférer un avantage sélectif (Wink 2010). Chez les organismes végétaux, un peu plus de 100 000 MS ont été identifiés jusqu'à maintenant, et leur nombre exact dans la nature est beaucoup plus élevé puisque seulement 20-30 % des plantes ont été étudiées (Wink 2010). De plus, leur distribution est restreinte, certains se trouvent chez une famille, un genre ou même une espèce spécifique.

1.1.1 Rôles chez les végétaux

Chez les organismes végétaux, les MS ont plusieurs fonctions importantes. Tout d'abord, certains jouent un rôle d'attraction de pollinisateurs biotiques dans le but de transférer le pollen d'une fleur à l'autre ou afin de disperser les graines des plantes dans l'environnement. Les pigments, les arômes et les parfums volatils, engendrés par la production des MS, permettent d'attirer les insectes et les animaux tels que les oiseaux,

chauve-souris et mammifères pour la pollinisation et la dispersion des graines, respectivement (Böttger, Vothknecht et al. 2018). En outre, la biosynthèse de MS volatils serait également impliquée dans les interactions allélopathiques entre les organismes végétaux en réponse à un stress. Par exemple, il a été démontré par plusieurs études que les plantes ont une plus grande résistance aux herbivores lorsqu'elles sont exposées aux composés volatils produits par une plante voisine endommagée (Karban, Yang et al. 2014). Des substances allélochimiques (des métabolites secondaires) permettent également d'inhiber l'invasion et la germination des autres plantes (Bourgaud, Gravot et al. 2001). Néanmoins, la plupart des MS connus aujourd'hui sont majoritairement associés au mécanisme de défense des plantes, qui leur permet de s'adapter sous des conditions de stress biotiques et abiotiques (Croteau, Kutchan et al. 2000, Akula et Ravishankar 2011). Par conséquent, l'absence de MS dans les plantes ne mène pas à la mort de celles-ci, mais elles deviennent plus susceptibles aux maladies et aux prédateurs, ce qui affecte leur survie à long terme. Les MS confèrent une protection contre une variété de stress environnementaux comme la température, l'humidité, la lumière, les polluants, le vent, etc. (Akula et Ravishankar 2011, Yang, Wen et al. 2018, Isah 2019). De plus, ils jouent un rôle clé dans la défense contre les agents pathogènes microbiens (virus, bactéries et champignons) et les organismes herbivores (Piasecka, Jedrzejczak-Rey et al. 2015, Pusztahelyi, Holb et al. 2015, Yang, Wen et al. 2018). La figure 1.1 présente les différents facteurs pouvant induire la production de métabolites secondaires chez les plantes.

1.1.2 Fonctions médicinales

Les MS sont connus depuis des siècles pour avoir des effets biologiques importants permettant de traiter une grande variété de maladies et cancers. Les plus vieilles indications écrites de l'utilisation de plantes médicinales en médecine traditionnelle remontent à il y a approximativement 5000 ans, par les Sumériens en Mésopotamie antique (Falodun 2010, Dias, Urban et al. 2012). Par exemple, leurs travaux documentaient l'utilisation d'huiles essentielles provenant de *Cupressus sempervirens* (Cyprès) et du genre *Commiphora* (Myrrhe), traitant le rhume, la grippe et l'inflammation

(Dias, Urban et al. 2012). L'ancien traité médical papyrus Ebers, écrit par les Égyptiens durant l'Égypte antique, présente environ 800 prescriptions faisant référence à plus de 700 espèces végétales et médicaments comme l'aloès, l'ail, le saule, la coriandre, etc. (Dias, Urban et al. 2012, Petrovska 2012). De façon similaire, les civilisations telles que la Chine, l'Inde, la Grèce antique ont documenté dans leurs archives une grande variété de plantes médicinales (Petrovska 2012).

Encore aujourd'hui, la médecine moderne emploie une grande quantité de médicaments dont la source principale est les plantes. Par exemple, l'ingrédient actif de l'aspirine (acide salicylique), qui est couramment utilisé de nos jours et autrefois prescrit par Hippocrate durant la Grèce antique, provenait de l'écorce de saule (Falodun 2010). Un autre exemple est la morphine, un des principaux composés bioactifs synthétisés par le pavot à opium (*Papaver somniferum*), est utilisé depuis des siècles pour ses propriétés narcotiques et analgésiques (Stefano, Pilonis et al. 2017).

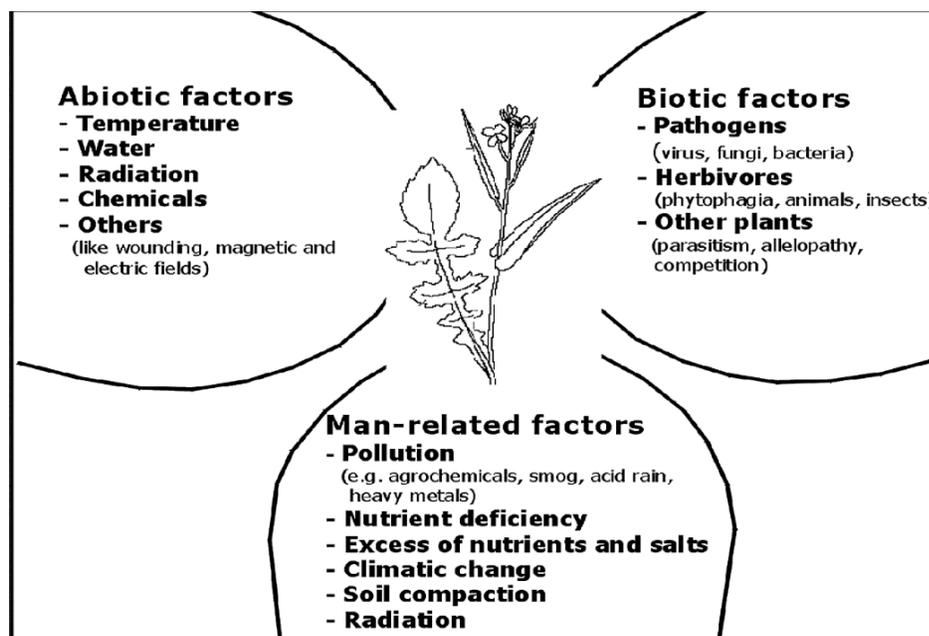


Figure 1.1 Facteurs biotiques et abiotiques qui influencent la biosynthèse de métabolites secondaires dans les plantes. (Tirée de Mazid, Khan et al. 2011.)

1.1.3 Limitations des plantes médicinales

En raison de leurs fonctions biologiques importantes comme agents thérapeutiques, les MS sont couramment extraits à partir de sources naturelles par les compagnies pharmaceutiques. Celles-ci cultivent les plantes produisant des composés de valeur afin de les commercialiser et les rendre accessibles pour des applications thérapeutiques. Près de 25 % des médicaments modernes utilisés aujourd'hui sont dérivés de plantes médicinales, et plus de 60 % des médicaments anticancéreux proviennent de sources naturelles telles que les plantes et les microorganismes (Cragg et Newman 2005, Isah 2019). Le tableau 1.1 montre quelques exemples de MS, de leurs applications en pharmaceutique et de leurs effets biologiques.

Tableau 1.1

Classification de différents métabolites spécialisés et leurs effets biologiques comme agent thérapeutique

Classification	Métabolite spécialisé	Source	Effets biologiques
Alcaloïde	Nicotine	Plante de tabac	Bloque l'action enzymatique, interfère avec la neurotransmission
Monoterpène	Menthol	Menthe	Interfère avec la neurotransmission, bloque le transport des ions, anesthésique
Diterpène	Gossypol	Cotonnier	Bloque la phosphorylation
Triterpène	Digitaline	Digitale	Stimule les muscles du cœur, altère le transport d'ions
Tétraterpène	Carotène	De nombreuses plantes	Antioxydant
Acide phénolique	Acide caféique	Toutes les plantes	Cause du dommage oxydatif (anticancéreux)
Coumarine	Ombelliférone	Carotte et panais	Bloque la division cellulaire, Réticulation de l'ADN
Flavonoïde	Anthocyane	Mûre, cerise, orange, maïs, framboise, raisin	Anti-allergique, anti-inflammatoire et antioxydant

Classification	Métabolite spécialisé	Source	Effets biologiques
Flavonoïde	Apigénine	Pomme, artichaut, céleri, basilic, cerise, raisin, noix, persil	Anti-inflammatoire, antioxydant, antispasmodique, chimiopréventif, induit l'apoptose, inhibe le cancer du sein et de l'ovaire
Acide phénolique	Acide caféique	Artichaut, poire, basilic, origan	Antioxydant et anti-inflammatoire
Polyphénol	Gallotanins	Chêne, légumes, pruche, lotier corniculé	Antioxydant, se lie aux protéines et enzymes, bloque la digestion

(Tiré de www.biologyreference.com.)

Bien que les effets que procurent les MS chez l'Homme sont une grande source d'intérêt pour les chercheurs et les entreprises pharmaceutiques, la culture de ces composés à partir de la plante a ses limitations. Entre autres, l'une des plus importantes est la minime quantité de composés bioactifs synthétisés par celle-ci. En effet, la biosynthèse de MS produit généralement un rendement de moins de 1 % du poids sec de la plante (Akula et Ravishankar 2011). En plus de rapporter un rendement très faible, la quantité produite est également variable d'un organisme à l'autre, en raison de facteurs environnementaux et de stress biotiques (Figure 1.1), du stade physiologique et du développement de la plante. Pour commercialiser un médicament, des quantités extrêmement importantes sont nécessaires afin de réaliser des essais cliniques et ensuite pour son usage général par la société. En conséquence, le processus d'extraction des MS à partir de sources végétales est excessivement coûteux et rend ainsi inaccessible une quantité considérable de médicaments. De plus, la synthèse chimique de ces composés est souvent difficile puisque les MS ont des structures complexes et les coûts qui y sont associés sont économiquement non viables (Matsuura, Malik et al. 2018). Les défis que procure la production industrielle des MS ont développé l'intérêt des chercheurs afin de produire des méthodes de biosynthèse alternative. Plusieurs techniques de cultivation sont considérées, comme modifier génétiquement les cultivars pour un meilleur profil en MS ou la génération de plateformes de microorganismes modifiés par ingénierie métabolique afin de produire des composés d'intérêts.

1.2 Les classes de métabolites spécialisés

La biosynthèse des MS est dérivée en grande partie du métabolisme primaire. En effet, les métabolites primaires tels que les sucres, les acides aminés, les acides gras et les acides nucléiques sont très importants puisqu'ils maintiennent l'homéostasie de l'organisme et sont également les molécules précurseurs des MS (Maeda 2019). La figure 1.2 montre la relation entre les différentes voies métaboliques primaires et secondaires. Selon leurs origines biosynthétiques, les métabolites spécialisés sont regroupés en trois grandes classes : les terpènes, les composés phénoliques et les alcaloïdes. Ils sont dérivés d'intermédiaires communs comme l'acide shikimique, l'acétyl-coenzyme A, l'acide mévalonique et le 1-deoxyxylulose-5-phosphate (Dias, Urban et al. 2012).

1.2.1 Les terpènes

Les terpènes et les terpénoïdes constituent la plus grande classe de métabolites spécialisés et la plus diversifiée. Les terpénoïdes (ou isoprénoïdes) sont définis comme étant des molécules dérivées des terpènes, qui contiennent des groupements fonctionnels additionnels. Cependant, de nos jours les termes terpènes, terpénoïdes et isoprénoïdes sont tous utilisés de façon interchangeable, bien que cela soit erroné. À ce jour, plus de 40 000 structures terpéniques ont été rapportées par la littérature (Cho, Lim et al. 2017). Généralement, on les retrouve chez les plantes, mais certaines formes de terpènes tel que le squalène se retrouvent chez les mammifères également (Cox-Georgian, Ramadoss et al. 2019). Ces composés naturels sont responsables des couleurs, des arômes et des odeurs chez les organismes végétaux. Plus particulièrement, ils jouent un rôle de défense contre les animaux herbivores ainsi que les microorganismes pathogènes. Les terpènes volatils jouent également un rôle dans la reproduction des plantes en attirant les organismes pollinisateurs et les animaux dispersant les graines, et inhibent la germination et le développement de plantes voisines (Ashour, Wink et al. 2010, Pichersky et Raguso 2018). Tous ces composés naturels terpéniques ont en commun des origines provenant d'intermédiaires tels que

l'acétyl-CoA ainsi que des composés glycolytiques comme le montre la figure 1.2 (Mazid, Khan et al. 2011).

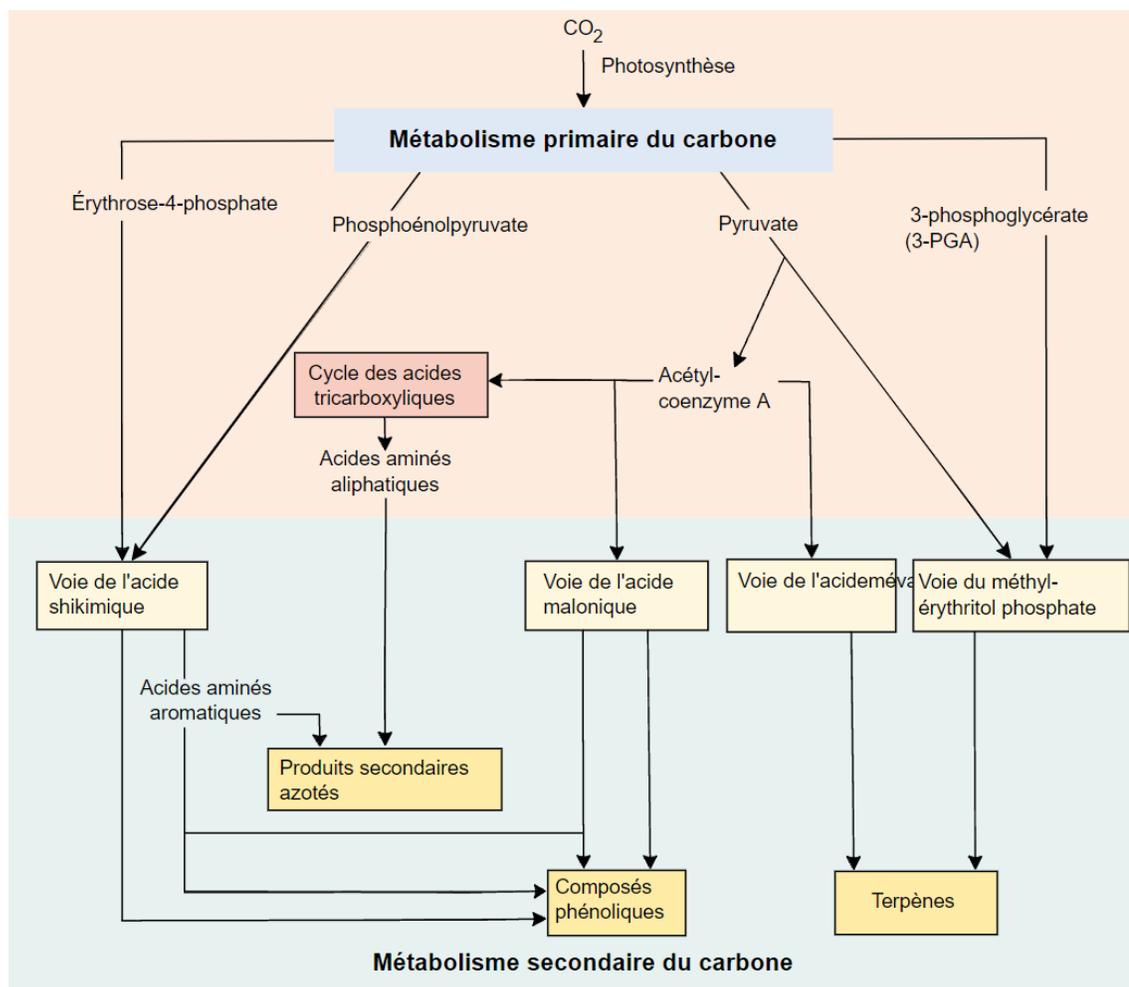


Figure 1.2 Voies de biosynthèses des métabolites primaires et spécialisés. (Tirée de Wikipédia.org et reproduite de Taiz et Zeiger 2002.)

La structure chimique de base des terpènes et des terpénoïdes est constituée d'une unité à cinq atomes de carbones. Cette unité se nomme l'isopentényl pyrophosphate (IPP), soit l'unité fonctionnelle d'isoprène dans les organismes vivants, incluant les végétaux (González-Burgos et Gómez-Serranillos 2012). La figure 1.3 présente la structure chimique de la molécule d'IPP. L'IPP et son isomère, diméthylallyl-pyrophosphate (DMAPP), sont synthétisés par deux voies métaboliques distinctes. La première est la voie du mévalonate (MVA) qui est cytoplasmique, et la seconde est la voie du méthylérythritol-phosphate (MEP) se produisant au niveau des plastides, incluant les chloroplastes,

les phényloplastes et autres plastes spécialisés des plantes. Ces deux voies permettent donc la formation de blocs de construction menant à la biosynthèse des différents terpènes.

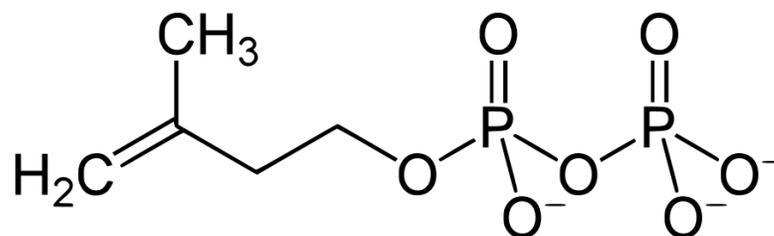


Figure 1.3 Structure chimique de l'isopentényl-pyrophosphate ($C_5H_{12}O_7P_2$).
(Tirée de Wikipédia.org.)

Une fois les unités à cinq carbones d'IPP synthétisé par les voies MVA et MEP, ceux-ci sont ensuite condensés afin d'élaborer le squelette précurseur aux terpènes. En effet, la polymérisation de deux ou plus IPP mène à la formation de différents squelettes, ce qui permet de diviser ces MS en différentes classes basées sur le nombre d'unité d'isoprène composant leur structure. La classification des terpénoïdes est présentée dans le tableau 1.2.

La voie biosynthétique menant à la formation de composés naturels terpéniques comporte quatre étapes distinctes. Tout d'abord, la première étape est la synthèse des molécules d'IPP par les voies MVA et MEP comme mentionné dans le paragraphe précédent. La deuxième étape consiste à la condensation des unités IPP afin de générer des prényl-pyrophosphates beaucoup plus larges, soit le géranyl-pyrophosphate (GPP, C_{10}), le farnésyl-pyrophosphate (FPP, C_{15}) et le géranyl-géranyl-pyrophosphate (GGPP, C_{20}). À la troisième étape, les pyrophosphates (C_{10} - C_{20}) subissent une grande variété de cyclisations et réarrangements afin de former les squelettes précurseurs selon la classe de terpènes (Figure 1.4). Par exemple, GPP est converti en monoterpène, FPP en sesquiterpène et GGPP en diterpène. FPP et GGPP peuvent également se dimériser afin de former les squelettes précurseurs triterpènes et tétraterpènes, respectivement. Dans la dernière étape, les molécules terpéniques subissent plusieurs modifications secondaires comme des oxydations/réductions, isomérisations, conjugaisons et plusieurs autres

transformations par lesquelles les squelettes parents de chaque classe sont convertis en milliers de métabolites distincts.

Tableau 1.2

Classification des terpènes selon le nombre d'unité d'isoprène (C₅) dans leurs structures

Nombre d'unité d'isoprène n	Nombre d'atome de carbone n	Nom
1	5	Hemiterpène
2	10	Monoterpène
3	15	Sesquiterpène
4	20	Diterpène
5	25	Sesterpène
6	30	Triterpène
8	40	Tétraterpène
9–30 000	> 40	Polyterpène

(Tiré de Ashour, Wink et al. 2010.)

En plus de leurs fonctions biologiques dans la plante leurs, les terpènes possèdent diverses utilisations chez l'humain. Par exemple, l'artémisinine qui est un sesquiterpène, dérivé de la plante médicinale chinoise *Artemisia annua*, est connu pour ses différentes propriétés. En effet, l'artémisinine et ses dérivés ont une activité anti-malariale, antimicrobienne et des effets cytotoxiques contre plusieurs types de cellules cancéreuses (Efferth, Romero et al. 2008). Une grande quantité de terpènes sont également utilisés dans l'industrie des cosmétiques, mais aussi dans la nourriture et les boissons en raison de leurs fragrances, arômes et saveurs.

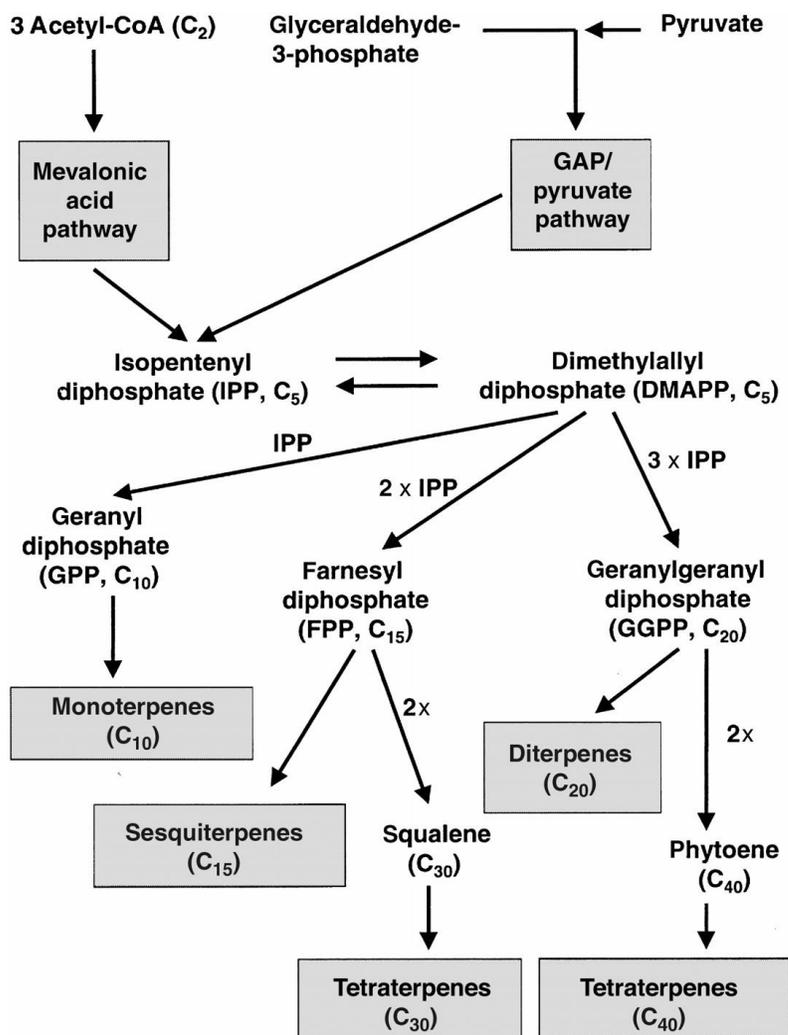


Figure 1.4 Biosynthèse des terpénoïdes dans les plantes, montrant les étapes majeures menant à la formation des différents produits terpéniques. CoA, coenzyme A; GAP, glycéraldéhyde-3-phosphate. (Tirée de Ashour, Wink et al. 2010.)

1.2.2 Les composés phénoliques

Les composés phénoliques ou polyphénols représentent une des trois grandes classes de MS et sont largement répandus dans le règne végétal. Plus de 8000 structures phénoliques très variées ont été rapportées dans la littérature dans les dernières années (Carocho et Ferreira 2013, Velderrain-Rodríguez, Palafox-Carlos et al. 2014). Ces composés sont caractérisés par la présence d'au moins un groupement phénol, qui est composé d'un anneau benzénique (cycle aromatique à 6 carbones) avec un ou plusieurs groupements hydroxyles (OH) (Cosme, Rodríguez et al. 2020). Dérivé d'acides aminés

aromatiques, leur biosynthèse découle de la voie du shikimate ou de la voie des phénylpropanoïdes (Vuolo, Lima et al. 2019). Selon leur structure, les polyphénols peuvent être divisés en plusieurs groupes, allant des acides phénoliques simples à des composés hautement polymérisés comme les tannins (Velderrain-Rodríguez, Palafox-Carlos et al. 2014). Il existe donc de nombreux groupes de composés phénoliques résultant de différentes structures chimiques : les acides phénoliques, les flavonoïdes, les tannins, les stilbènes, les lignanes, etc. (Dai et Mumper 2010). Généralement, leur classification est basée sur leur diversité structurale (Figure 1.5). Cette grande diversité entraîne un large éventail de propriétés physico-chimiques et biologiques. Ces composés sont impliqués dans l'adaptation de la plante lorsqu'elle est exposée aux conditions stressantes de l'environnement, telles que les infections microbiennes, l'exposition aux rayons UV et les attaques par les herbivores (Dai et Mumper 2010). D'autant plus, les polyphénols jouent un rôle important dans le développement de la plante grâce à la biosynthèse de la lignine composant les parois cellulaires, ainsi que de pigments signalant la maturation des fruits et parfums attirant les pollinisateurs (Bhattacharya, Sood et al. 2010).

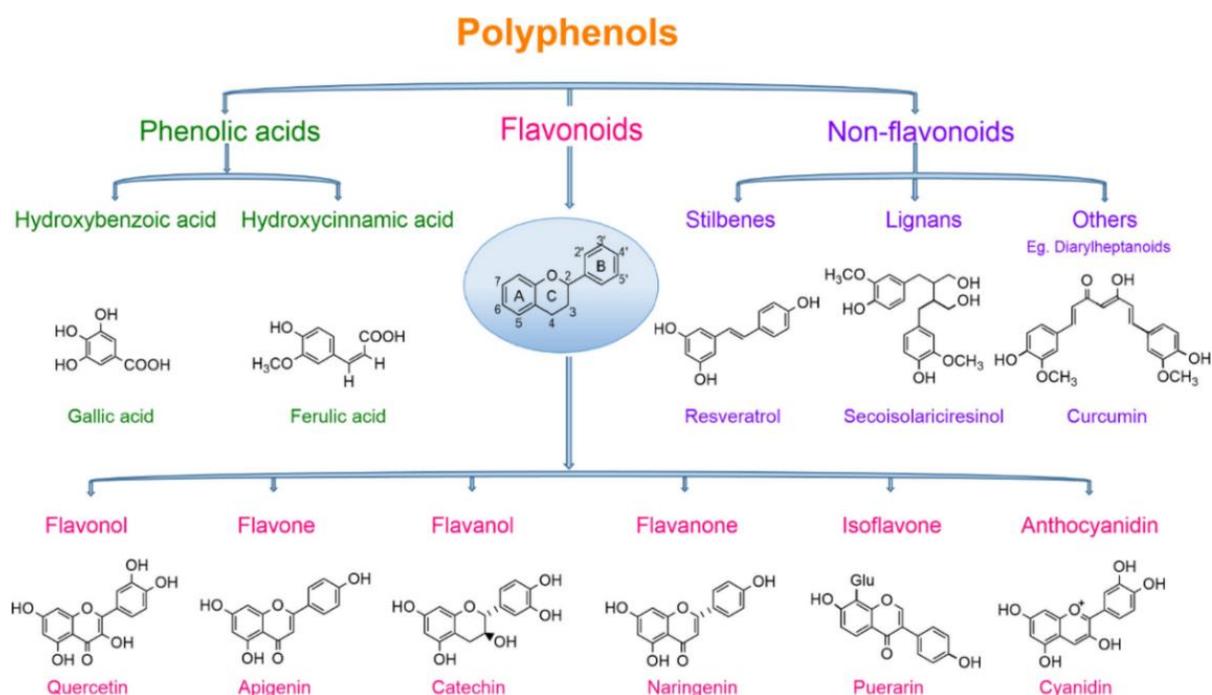


Figure 1.5 Classification des composés phénoliques.
(Tirée de Rambaran 2020.)

Ces MS sont bien connus pour être une source abondante de molécules antioxydantes naturelles sous forme de composés phénoliques (Zhao, Zhang et al. 2014). En effet, les groupements hydroxyles que contiennent les phénols sont de bons donneurs d'hydrogènes et peuvent interagir avec les dérivés réactifs de l'oxygène (DRO) et les dérivés réactifs de l'azote (DRA) (Pereira, Valentão et al. 2009). Cette réaction permet donc de rompre le cycle menant à la génération de nouveaux radicaux libres. Les composés phénoliques sont bénéfiques pour la santé puisqu'ils permettraient de se protéger contre le dommage oxydatif. En plus d'avoir des propriétés antioxydantes, leur utilisation à des fins thérapeutiques permettrait de prévenir le cancer et autres maladies, telles que les complications reliées au diabète et auraient des effets préventifs sur le développement de la maladie d'Alzheimer (MA) (Yamada, Ono et al. 2015, Lin, Xiao et al. 2016). Ces molécules ont également une activité anti-inflammatoire, antimicrobienne et antiproliférative (Lin, Xiao et al. 2016).

1.2.3 Les alcaloïdes

Les alcaloïdes sont des composés naturels provenant d'origine majoritairement végétale. Ces molécules organiques de faible poids moléculaire sont présentes sous forme hétérocyclique et contiennent généralement une ou plusieurs bases azotées. Principalement dérivés des acides aminés, les alcaloïdes ont des propriétés basiques et leurs structures sont hautement variées et complexes. Le terme « alcaloïde » a été introduit pour la première fois par le chimiste allemand Carl F. W. Meissner en 1819 et est dérivé du nom arabe *al-qali* (Dey, Kundu et al. 2020). Selon ses observations, ces composés agissent comme un alcali, en raison de leurs propriétés basiques. Jusqu'à aujourd'hui, plus de 27 000 alcaloïdes ont été caractérisés, dont 21 000 provenant d'organismes végétaux représentant plus de 20 % des MS chez les plantes (Dewick 2009, Kaur 2015). D'après une perspective structurale, les alcaloïdes sont classifiés sur la base de leurs structures chimiques et leurs origines biosynthétiques (Dey, Kundu et al. 2020). Ils peuvent également être classés selon leurs activités biologiques et écologiques. La classification des différentes classes d'alcaloïdes sera discutée plus en détails à la section 1.4.

Il est proposé que les alcaloïdes jouent un rôle majeur dans le mécanisme de défense des plantes. En effet, la biosynthèse de ces molécules permettrait aux organismes végétaux de se protéger contre les agressions par les microorganismes et les prédateurs tels que les animaux herbivores et les insectes. Les alcaloïdes peuvent également jouer un rôle d'herbicide afin de compétitionner contre les autres plantes l'environnement (Wink 2007). Dans la plupart des cas, les alcaloïdes présentent plusieurs fonctions biologiques. Par le fait même, leur synthèse étant plutôt coûteuse pour la cellule, ces molécules ont évolué de façon à ce qu'elles contiennent habituellement plusieurs groupements fonctionnels actifs afin d'interagir avec différentes cibles moléculaires (Wink 2007).

Ces composés sont reconnus pour être synthétisés en petite quantité *in planta*, puisque ce sont des molécules toxiques. Afin d'empêcher les organismes végétaux de s'intoxiquer, les alcaloïdes sont accumulés dans des cellules spécialisées à l'intérieur de la plante se nommant les laticifères (De Luca et St Pierre 2000). Leur accumulation se produit dans les parties végétales couramment attaquées par les herbivores et pathogènes. Par exemple, l'alcaloïde nicotine est synthétisé dans les racines de la plante de *Nicotiana tabacum*, alors que 85 % de la celle-ci se retrouve dans les feuilles par la suite afin de conférer une protection contre les herbivores (Pallardy 2008). En outre, les alcaloïdes quinolizidiniques et pyrrolizidiniques s'accumulent dans les tissus aériens qui sont plus sujets aux attaques par les pathogènes (Lee, Pate et al. 2007). Même si les alcaloïdes n'ont pas un rôle vital dans la croissance des plantes, beaucoup d'entre eux sont connus pour leurs propriétés pharmaceutiques et sont utilisés comme agent thérapeutique depuis des milliers d'années. Le tableau 1.3 montre différents alcaloïdes ayant été commercialisés ou testés cliniquement en date de 2020 ainsi que leurs propriétés bioactives (Heinrich, Mah et al. 2021).

Tableau 1.3

Différents alcaloïdes commercialisés/testés cliniquement
comme médicament et leurs propriétés thérapeutiques

Alkaloid	Therapeutic Indications	Source; Other Uses of the Source
Aconitine	Rheumatism, neuralgia, sciatica	<i>Aconitum napellus</i> L. and others; reduce fever, pneumonia, laryngitis
Atropine	Antispasmodic, anti--parkinson, cycloplegic drug	<i>Atropa bella-donna</i> L. and others; to treat peptic ulcers, relives intestinal colic, pupil dilation agents
Caffeine	Neonatal apnea, atopic dermatitis	<i>Theobroma cacao</i> L. and others; food additive, emollient, angina treatment, hypertension treatment
Cocaine	Local anaesthetic	<i>Erythroxylum coca</i> Lam. and others; GI symptoms treatment, altitude sickness treatment
Codeine	Antitussive, analgesic	<i>Papaver somniferum</i> L. and others; antioxidant, antimutagenic, and anticarcinogenic effects
Emetine	Intestinal amoebiasis, expectorant drug	<i>Alangium lamarckii</i> Thwaites and others; emetic, anthelmintic, purgative
Ergotamine	Migraine treatment	<i>Claviceps purpurea</i> var. <i>purpurea</i> (Fr.) Tul. and others; migraine, Parkinson's diseases. Antitumor
Galanthamine	Muscle relaxant, Alzheimer's	<i>Galanthus woronowii</i> Losinsk. and many other species of the Amarylidaceae; emmenagogue, treatment of traumatic injuries to nervous system
Hydrastine	Gastrointestinal disorders	<i>Corydalis fimbriifera</i> Korsh. and others; treatment for depression, hypertension, intestinal spasms
Lobeline	Anti-smoking, asthma, cough	<i>Lobelia inflata</i> L. and others; antispasmodic, respiratory stimulant
Morphine	Pain relief, diarrhoea	<i>Papaver somniferum</i> L. and others; antispasmodic, expectorant, antitussive
Narceine	Cough suppressant	<i>Papaver somniferum</i> L. and others; antispasmodic, expectorant, antitussive
Nicotine	Anti-smoking	<i>Nicotiana tabacum</i> L. and others; relaxant, antispasmodic, discutient, diuretic
Noscapine	Cough suppressant	<i>P. somniferum</i> and others; antispasmodic, expectorant, antitussive
Papaverine	Vasodilator, gastrointestinal disorders	<i>P. somniferum</i> . and others; antispasmodic, expectorant, antitussive

Alkaloid	Therapeutic Indications	Source; Other Uses of the Source
Quinine	Malaria, babesiosis, myotonic disorders	<i>Cinchona officinalis</i> L. and others; fever, spasm relax, neuralgia
Reserpine	Hypertension, psychoses	<i>R. serpentine</i> and others; hypnotic, increase urine contractions, treat wounds and itches
Sanguinarine	Antiplatelet agent	<i>Sanguinaria canadensis</i> L. and others; fever, rheumatism, expectorant
Sparteine	Uterine contractions, cardiac arrhythmias	<i>Lupinus pusillus</i> Pursh var. <i>pusillus</i> and others; treatment for haemostatic, ears and eyes disorders
Strychnine	Eye disorders	<i>Strychnos wallichiana</i> Steud. ex A.DC. and others; leprosy, antidote for rabies, ulcers treatment, rheumatism treatment
Taxol	Mamma and ovary carcinoma	<i>Taxus brevifolia</i> Nutt. and others; treatment for diabetes, cancer treatment
Theobromine	Asthma, diuretic agent	<i>Theobroma cacao</i> L. and others; food, emollient, angina treatment, hypertension treatment
Turbocuranine	Muscle relaxant	<i>Chondrodendron tomentosum</i> Ruiz and Pav.; oedema, kidney stones, persistent urinary tract infections
Vindesine	Chemotherapy	<i>C. roseus</i> and others; anticancer, hypoglycaemic agent, emetic

(Tiré de Heinrich, Mah et al. 2021.)

1.3 Biosynthèse, transport et accumulation des métabolites spécialisés dans la plante

Lorsque les MS sont biosynthétisés par les plantes, ils sont souvent sécrétés en direction d'autres types cellulaires et/ou entreposés par celles-ci dans des compartiments spécialisés. Cela implique une organisation complexe de plusieurs types de cellules participant à un processus de transport intracellulaire et extracellulaire des enzymes et des métabolites (Samanani, Park et al. 2005, Ziegler et Facchini 2008). En outre, l'accumulation des MS dans des compartiments subcellulaires particuliers et dans des types de cellules spécifiques permettrait aux plantes de se protéger contre la cytotoxicité de ces molécules, d'assurer la réponse de défense et de maintenir la biosynthèse de ces produits afin d'améliorer les interactions avec d'autres organismes (Bidlack, Omaye et al. 2000, Ziegler et Facchini 2008). Différentes structures telles que les trichomes

glandulaires, les cellules idioblastiques, les laticifères, les canaux résinifères, l'épiderme, l'endoderme, le péricycle et le cortex sont tous connus pour être impliqués dans leur processus de biosynthèse et/ou accumulation à l'intérieur de la plante (Facchini et St-Pierre 2005).

Une grande variété d'études sur la séquestration des MS a été réalisée. Plusieurs ont montré qu'il y a accumulation des alcaloïdes de la classe des benzyloquinolines de *Papaver somniferum* dans les cellules laticifères, près des cellules du phloème (Alcantara, Bird et al. 2005). Le cytoplasme des cellules laticifères, qui produit du latex, contient des alcaloïdes entreposés dans de petites vésicules. Également, les alcaloïdes de la classe des tropanes et la nicotine sont biosynthétisés dans les racines des plantes telles que *Nicotiana*, *Hyoscyamus*, *Atropa* et *Datura* et leur transport suite à la synthèse a été rapporté dans les cellules du xylème pour ensuite être accumulé dans les parties aériennes des plantes (Wink 2007). La plupart des plantes produisent les MS dans un compartiment, type de cellule ou tissu particuliers, pour ensuite être transportés par les cellules du phloème et du xylème vers d'autres tissus de la plante afin d'être entreposés pour ultérieurement participer au mécanisme de défense ou pour signaler (Figure 1.6).

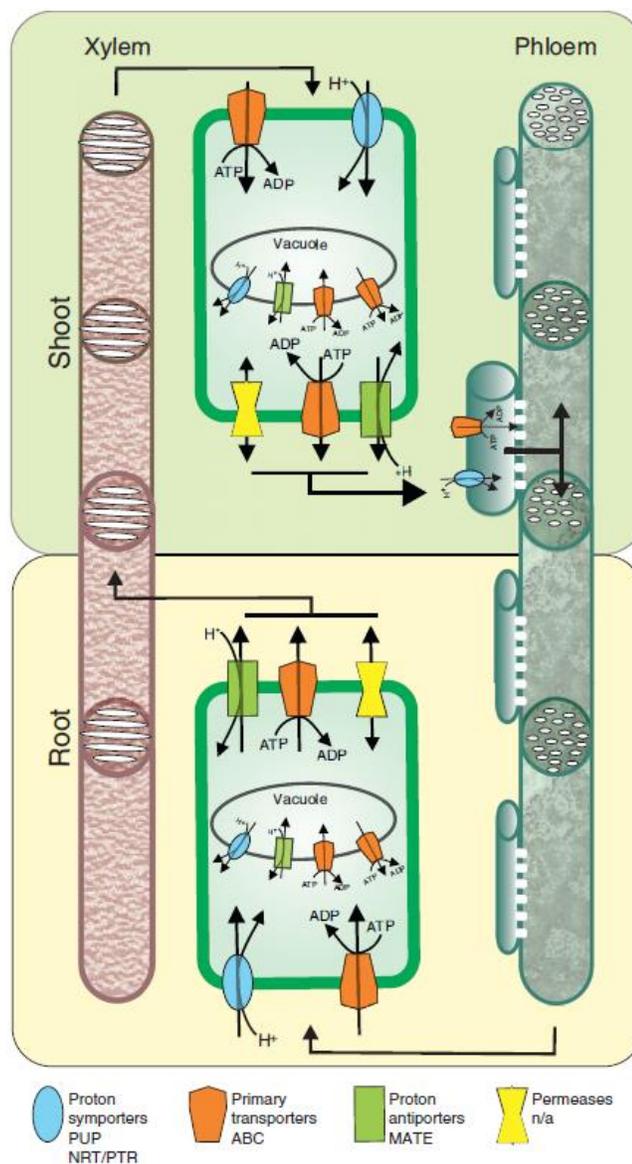


Figure 1.6 Représentation du mécanisme de transport entre différents tissus par les cellules du Xylème et du Phloème.
(Tirée de Nour-Eldin et Halkier 2013.)

1.4 Classification des alcaloïdes

En raison du grand nombre d'alcaloïdes existant, ils ont été divisés en différents groupes principaux selon leurs structures chimiques (les alcaloïdes purines, tropanes, pyrrolizidines, quinolizidines, benzylisoquinolines, indolo-monoterpènes et les amaryllidacées) (Dewick 2011).

1.4.1 Les purines

Les purines sont des alcaloïdes principalement dérivés de la xanthosine et de la xanthine (Smith et Atkins 2002). Le noyau de la purine est un hétérocycle azoté et ne se retrouve jamais sous la forme non-modifié dans la nature, mais plutôt substitué sous forme aminée, hydroxylée, méthylée ou de nucléosides (Rosemeyer 2004). Ces composés sont donc regroupés sous le terme général de purines. Ces molécules sont synthétisées chez plusieurs espèces de plantes de taxonomies différentes. Entre autres, on les retrouve chez les espèces *Coffea arabica*, *Camellia sinensis*, *Theobroma cacao*, *Ilex paraguariensis*, *Paullinia cupana*, et *Cola nitida* (Springob et Kutchan 2009). La purine la plus abondante dans la nature est la caféine. En raison de son accumulation en grande quantité dans les plantes (2-3 % du contenu des feuilles), la caféine ainsi que d'autres purines joueraient un rôle au niveau de la défense contre les herbivores (Hollingsworth, Armstrong et al. 2002). De plus, elles démontrent des effets allélopathiques en inhibant la germination des graines (Springob et Kutchan 2009). Étant un stimulant très puissant, la caféine est largement utilisée dans les boissons telles que le café, le thé et le soda, mais est également utilisée dans certains médicaments contre le rhume et les analgésiques. Depuis quelques années, les chercheurs ont développé un intérêt pour la caféine puisque des études sur les consommateurs ont montré que le risque de développement de la maladie de Parkinson est réduit (Ascherio, Weisskopf et al. 2004). En général, les purines sont intensément étudiées puisqu'elles ont la capacité de stimuler le système nerveux central. La figure 1.7 présente quelques exemples de purines dérivées de plantes.

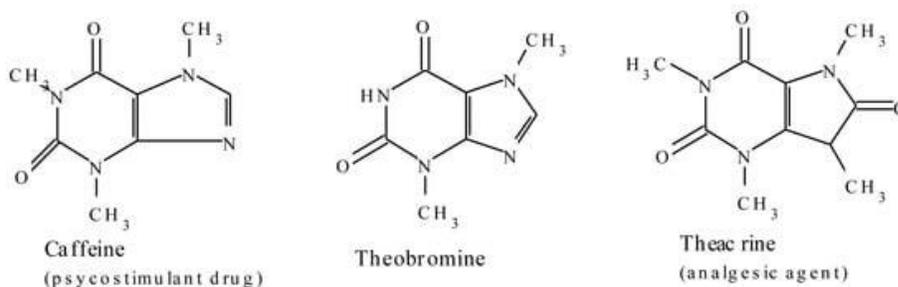


Figure 1.7 Quelques exemples de purines dérivées de la plante ainsi que leurs structures.
(Tirée de Ncube et Van Staden 2015.)

1.4.2 Les alcaloïdes tropaniques

Les alcaloïdes tropaniques (AT) sont des composés caractérisés par un noyau cyclique de tropane. Ces MS sont majoritairement retrouvés en concentration élevée chez les plantes de la famille des Solanaceae et des Erythroxylaceae (Kohnen-Johannsen et Kayser 2019). Trois groupes d'AT majeurs découlent du squelette tropanique : cocaïne, hyoscyamine et scopolamine et finalement calystégine (Figure 1.8). Ces composés ont la capacité de traverser la barrière hémato-encéphalique pouvant provoquer des hallucinations et des effets psychoactifs, plus particulièrement les tropanes des groupes de la cocaïne et des hyoscyamines et scopolamines (Kohnen-Johannsen et Kayser 2019). Les calystégines n'ont pas la capacité d'être absorbés par le système nerveux central en raison de son hydrophilie, et ne causent donc pas d'effets psychoactifs chez les humains (Dräger 1995).

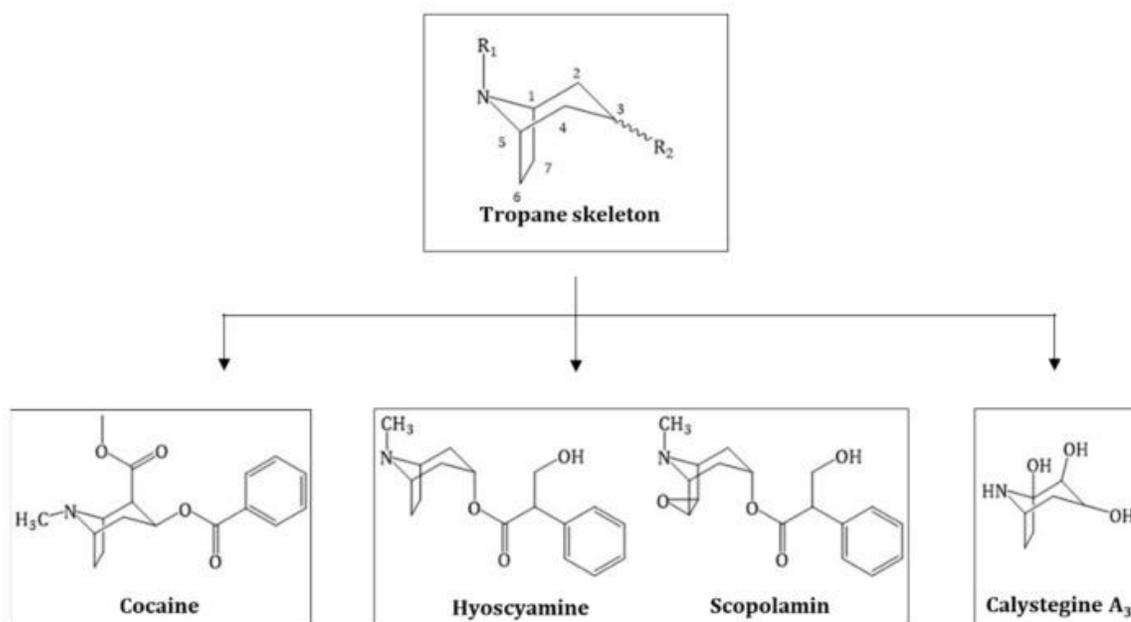


Figure 1.8 Structure du squelette tropanique et les trois groupes majeurs des tropanes. (Tirée de Kohnen-Johannsen et Kayser 2019.)

1.4.3 Les pyrrolizidines

Les alcaloïdes pyrrolizidiniques (AP) sont des MS caractérisés par une structure pyrrolizidine (Figure 1.9). Les AP dérivant de cette structure sont nommés les nécines puisque le cycle de pyrrolizidine est substitué par un groupement hydroxyméthyle et comportent parfois également une fonction alcool secondaire (Bruneton 2009). Les AP sont généralement biosynthétisés chez les familles de plantes des Asteraceae, Fabaceae et Boraginaceae (Wiedenfeld 2011). Leur rôle est de défendre la plante contre les attaques par les animaux herbivores et les insectes (Hartmann 1999). Chez l'humain, leurs effets sont mutagènes et très toxiques.

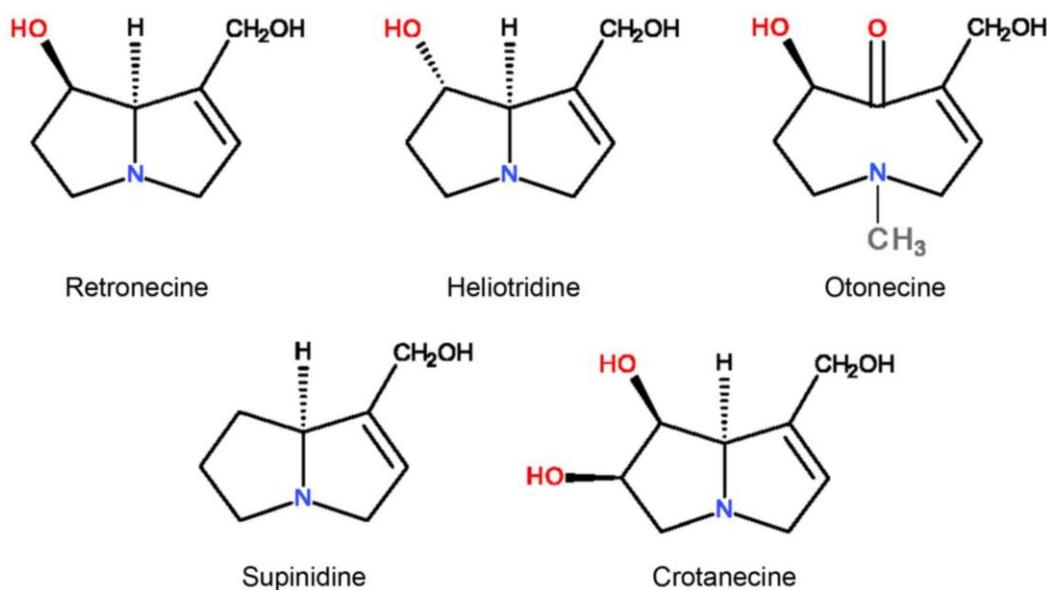


Figure 1.9 Exemple de structure de différents alcaloïdes pyrrolizidiniques. (Tirée de Seremet, Olaru et al. 2018.)

1.4.4 Les alcaloïdes quinolizidiniques

Les alcaloïdes quinolizidiniques (AQ) sont des MS caractérisés par un cycle quinolizidinique ou un cycle pipéridique et sont dérivés de l'acide aminé lysine (Saito et Murakoshi 1995, Wink 2019). Les AQ sont plus particulièrement biosynthétisés chez la famille des Leguminosae (Osborn et Lanzotti 2009, Frick, Kamphuis et al. 2017). On les retrouve également chez les familles Chenopodiaceae, Berberidaceae,

Ranunculaceae, Scrophulariaceae et Solanaceae (Wink 2003). L'accumulation des AQ est toxique pour grand nombre d'insectes, de vers et d'animaux vertébrés. Ils procurent également une protection contre les bactéries, virus et champignons et peuvent être utilisés comme un pesticide naturel (Wink 2019). Quelques exemples de AQ sont représentés dans la figure 1.10.

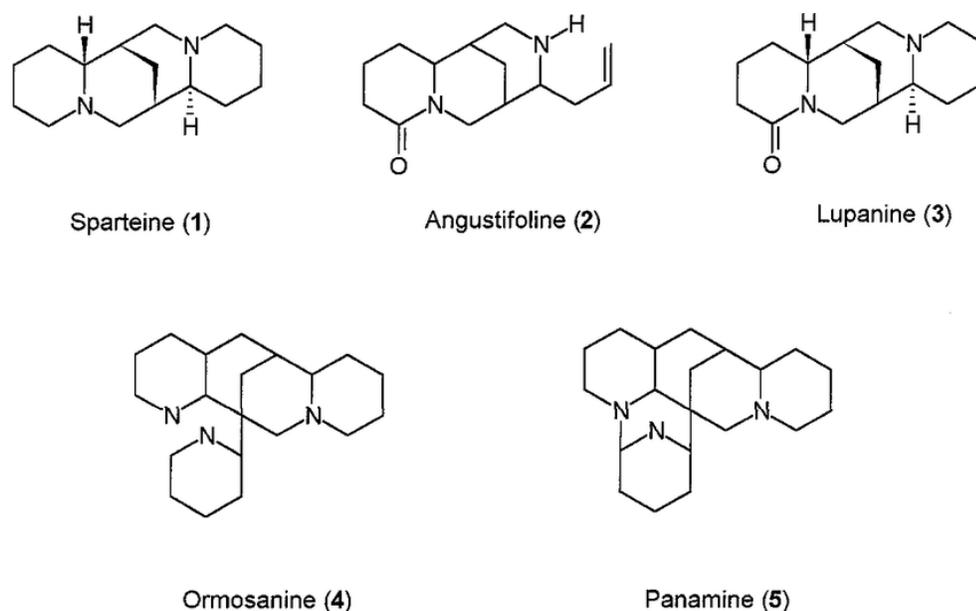


Figure 1.10 Quelques exemples d'alcaloïdes quinolizidiniques.
(Tirée de Guimarães, José et al. 2003.)

1.4.5 Les alcaloïdes indolo-monoterpéniques

Les alcaloïdes indolo-monoterpéniques (AIM) représentent un groupe large et divers de MS comptant plus de 3000 composés connus jusqu'à ce jour (Brown, Clastre et al. 2015). On les retrouve plus particulièrement chez les familles de plantes des Apocynaceae, Loganiaceae, Rubiaceae, Icacinaceae, Nyssaceae et Alangiaceae (O'Connor et Maresh 2006, De Luca, Salim et al. 2014). Ces composés sont caractérisés par une structure bicyclique comprenant un cycle benzénique et un cycle pyrrole (Mohammed, Abdul-Hameed et al. 2021). La figure 1.11 présente quelques exemples d'alcaloïdes indolo-monoterpéniques. Depuis quelques décennies, les AIM sont devenus une source d'intérêt en raison de leurs activités biologiques. En effet, certains de ces métabolites possèdent une activité anticancéreuse, antispasmodique, antipaludique et

anti-inflammatoire (Ferreira et Paterna 2019, Mohammed, Abdul-Hameed et al. 2021). Leurs structures étant complexes, les voies de biosynthèse des AIM sont plus particulièrement étudiées afin de développer des systèmes de biologie synthétique. Récemment, Brown et al. ont reproduit la voie de biosynthèse menant à la production de l'intermédiaire commun strictosidine. En effet, ils ont introduit 21 gènes dans le génome de la levure *Saccharomyces cerevisiae*, ce qui a mené à l'expression hétérologue de l'intermédiaire commun à tous les AIM connus jusqu'à ce jour (Brown, Clastre et al. 2015).

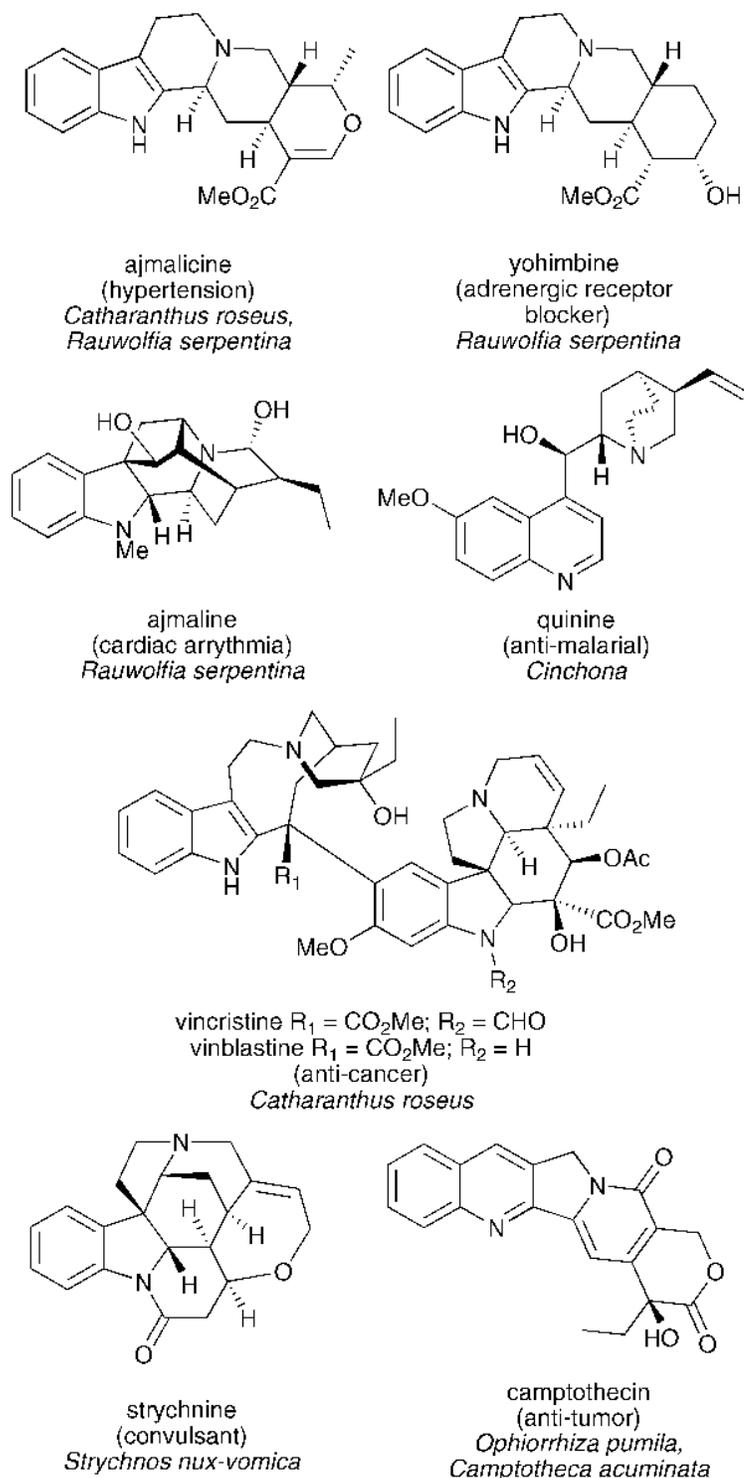


Figure 1.11 Différents alcaloïdes indolo-monoterpéniques, leurs fonctions biologiques et leurs sources biosynthétiques.
(Tirée de O'Connor et Maresh 2006.)

1.4.6 Les alcaloïdes isoquinoléiques

Les alcaloïdes isoquinoléiques (AI) constituent l'un des groupes de MS les plus larges. Leur biosynthèse est prédominante chez les familles des Papaveraceae, Berberidaceae et Ranunculaceae (Khan et Suresh Kumar 2015). Ces composés sont synthétisés à partir de précurseurs d'acides aminés, soit de la phénylalanine et de la tyrosine, et leur squelette de base est caractérisé par un cycle isoquinoléique ou un cycle tétrahydroisoquinoléique (Grycová, Dostál et al. 2007). La distribution taxonomique des AI est limitée et représente plutôt des molécules signatures de certaines espèces ou familles de plantes (Diamond et Desgagné-Penix 2016). Selon la taxonomie, une molécule précurseur intermédiaire spécifique est biosynthétisée avant de subir des modifications menant à un large spectre d'alcaloïdes isoquinoléiques. Parmi les AI, les sous-groupes des alcaloïdes des benzyloquinolines (ABI) et les alcaloïdes des Amaryllidacées (AA) seront décrits plus en détail.

1.4.6.1 Les alcaloïdes benzyloquinolines

Les alcaloïdes des benzyloquinolines (ABI) sont un large groupe d'environ 2500 composés azotés structurellement divers (Figure 1.12) (Desgagné-Penix et Facchini 2011, Hagel et Facchini 2013). Malgré la grande diversité structurale, les ABI partagent tous les mêmes origines biosynthétiques. En effet, ils sont tous synthétisés à partir du précurseur central (S)-norcoclaurine, qui est la première étape engagée dans la voie de biosynthèse des alcaloïdes des benzyloquinolines (Samanani, Liscombe et al. 2004, Lee et Facchini 2010, Hagel et Facchini 2013). Leur biosynthèse débute par la décarboxylation, *ortho*-hydroxylation et déamination de deux L-tyrosines, et sont converties en 3,4-dihydroxyphénylamine (dopamine) et en 4-hydroxyphénylacétaldéhyde (4-HPAA) (Rueffer et Zenk 1987, Samanani, Liscombe et al. 2004). Par la suite, la dopamine et le 4-HPAA subissent une condensation de Pictet-Spengler catalysée par l'enzyme norcoclaurine synthase (NCS), afin de former le précurseur intermédiaire commun (S)-norcoclaurine qui contient un noyau benzyloquinoléique. Les ABI sont principalement retrouvés chez les familles de plantes des Papaveraceae, Fumariaceae, Ranunculaceae, Berberidaceae et Mernispermaceae (Samanani et Facchini 2001).

Beaucoup des ABI possèdent une activité pharmacologique, incluant la morphine et la codéine qui ont des propriétés analgésiques et narcotiques, la sanguinarine et la berbérine qui ont une activité antimicrobienne, la papavérine et le (+)-tubocurarine qui ont des propriétés relaxantes des muscles ainsi que la noscapine ayant une activité anticancéreuse (Hagel et Facchini 2013). Certains des médicaments les plus anciens de l'humanité sont des ABI dérivés de plantes. Par exemple, le pavot à opium (*Papaver somniferum*) a été l'une des premières plantes dont la culture a été domestiquée et est également une source traditionnelle d'analgésique. En effet, le latex séché aussi connu sous le nom de « opium », a été utilisé parmi les premières drogues par l'humanité puisque celui-ci contient de la morphine et de la codéine (Desgagné-Penix et Facchini 2011).

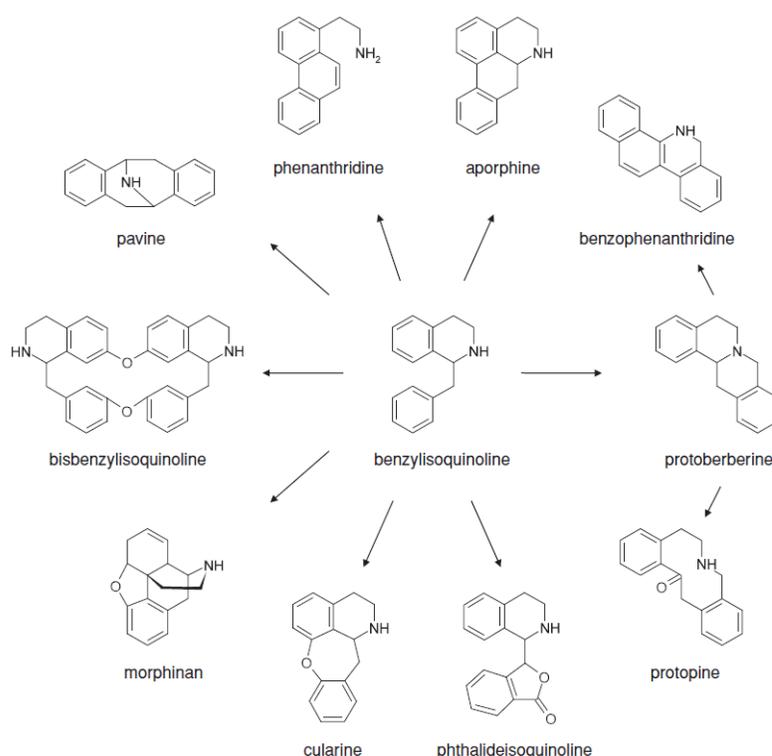


Figure 1.12 Exemples de diverses structures à squelette simple des alcaloïdes des benzylisoquinolines.
(Tirée de Desgagné-Penix et Facchini 2011.)

Les ABI ont également été étudiés considérablement sur leur synthèse et entreposage des alcaloïdes dans des cellules spécialisées ou dans des tissus spécifiques pour séquestrer les constituants cytotoxiques. Plusieurs études ont démontré que les gènes biosynthétiques et les enzymes impliquées dans la biosynthèse des ABI sont localisés dans les cellules compagnes et les tubes criblés dans le phloème. Une fois synthétisés, ces composés toxiques sont ensuite transportés dans les cellules laticifères adjacentes pour être entreposés (Bird, Franceschi et al. 2003, Samanani, Alcantara et al. 2006). Cette distribution spatiale complexe des gènes, enzymes et métabolites requiert plusieurs niveaux de régulation dans la voie de biosynthèse des benzyloquinolines (Desgagné-Penix et Facchini 2011).

1.4.6.2 Les alcaloïdes des Amaryllidacées

Les alcaloïdes des Amaryllidacées (AA) proviennent essentiellement des plantes bulbeuses, vivaces à fleurs appartenant à l'ordre des *Asparagales*. Cette famille de plantes contient 75 genres, 1600 espèces et est largement distribuée dans les régions tropicales et subtropicales (Christenhusz et Byng 2016, Desgagné-Penix 2021). Depuis la première isolation de la lycorine de *Narcissus pseudonarcissus* en 1877, plus de 650 AA aux structures diverses ont été reportés jusqu'à maintenant, et ce nombre continue d'augmenter chaque année (Gerrard 1877, Jin et Yao 2019, Ka, Koirala et al. 2020). Dans la figure 1.13, la distribution géographique des AA est représentée. Ces plants sont plus particulièrement retrouvés en Eurasie, Australie, Amérique et Afrique et se retrouvent parmi les 20 meilleures familles de plantes médicinales largement considérées à ce jour (Jin et Xu 2013, Jin et Yao 2019).



Figure 1.13 Distribution géographique des plantes de la famille des Amaryllidacées. (Tirée de Jin et Xu 2013.)

Activités biologiques

Dernièrement, les alcaloïdes de la famille des Amaryllidacées ont suscité l'intérêt en raison de leurs sources très prometteuses de produits naturels bioactifs à large spectre d'actions thérapeutiques. Plus particulièrement, les Amaryllidacées produisent des métabolites spécialisés aux puissantes activités pharmacologiques pouvant avoir une activité anticancéreuse, antivirale, antibactérienne, antihistaminique, anti-inflammatoire, antiparasitaire, antiproliférative, cytotoxique et également un potentiel d'inhibition de l'acétylcholinestérase (Kornienko et Evidente 2008, Nair et van Staden 2013, Ncube et Van Staden 2015, Kilgore et Kutchan 2016, Hotchandani et Desgagné-Penix 2017, Desgagné-Penix 2021). La plus vieille évidence de l'utilisation des AA remonte à 400 ans avant J-C, lorsque Hippocrate de Cos utilisait l'huile produite par les Narcisses (*Narcissus poeticus L.*) pour le traitement des tumeurs utérines (Kornienko et Evidente 2008, He, Qu et al. 2015).

En raison du nombre élevé d'AA et de leurs structures diverses, ceux-ci ont été classifiés en différents groupes basés sur leurs squelettes moléculaires. Selon leurs squelettes et leurs origines biosynthétiques, les différents types d'alcaloïdes sont la norbelladine, cherylline, galanthamine, lycorine, lycorénine, crinine, narciclasine,

tazettine et montanine (Desgagné-Penix 2021). Le tableau 1.4 et la figure 1.14 montre tous les groupes de AA mentionnés plus haut ainsi que leurs fonctions médicinales.

Tableau 1.4

Les 9 types d'AA ainsi que leurs structures et activités médicinales

Type de AA	Anneau structural	Orientation couplage de phénols	Activités thérapeutiques
Norbelladine	N-(3,4-Dioxybenzyl)-4-oxyphenethylamine		Inhibiteur de la butyrylcholinestérase
Cherylline	Tétrahydroisoquinoline		Antiviral, cytotoxique
Galanthamine	6H-Benzofuro[3a,3,2-e,f]-2-benzazépine	<i>para-ortho'</i>	Inhibiteur de l'acétylcholinestérase, traitement de la maladie d'Alzheimer
Lycorine	Pyrrolo[d,e]phénanthridine	<i>ortho-para'</i>	Cytotoxique, antimicrobial, analgésique, anti-malarial, anti-inflammatoire
Homolycorine	2-Benzopyrano-[3,4-g]indole	<i>ortho-para'</i>	Anti-inflammatoire, cytotoxique
Crinine	5,10b-Éthanophénanthridine	<i>para-para'</i>	Cytotoxique, antimicrobial, anti-inflammatoire, anticancéreux
Narciclasine	Lycoricidine	<i>para-para'</i>	Cytotoxique, antitumoral
Pretazettine	2-Benzopyrano[3,4-c]indole	<i>para-para'</i>	Antiviral, antiparasite, cytotoxique, analgésique
Montanine	5,11-Méthanomorphanthridine	<i>para-para'</i>	Antidépresseur, anxiolytique

Sources : He, Qu et al. 2015, Ka, Koirala et al. 2020, Desgagné-Penix 2021.

Un AA bien connu pour ses propriétés inhibitrices de l'acétylcholinestérase (AChE) est la galanthamine, originalement isolé de la plante *Galanthus nivalis L.* en 1947 (He, Qu et al. 2015). En effet, ce métabolite permet aux patients atteints de la maladie d'Alzheimer (MA) de soulager leurs symptômes, sans toutefois empêcher la progression de la maladie. L'enzyme AChE a pour fonction de dégrader les neurotransmetteurs d'acétylcholines (ACh), ceux-ci jouant un rôle très important dans la transmission de l'influx nerveux d'un neurone à l'autre. Un déficit en ACh mène à des problèmes neurodégénératifs tels que la MA. La galanthamine se lie donc à l'enzyme AChE afin

d'inhiber la dégradation de l'ACh, ce qui mène à une augmentation de la transmission nerveuse causant une stimulation en continu des muscles, des glandes et du système nerveux central.

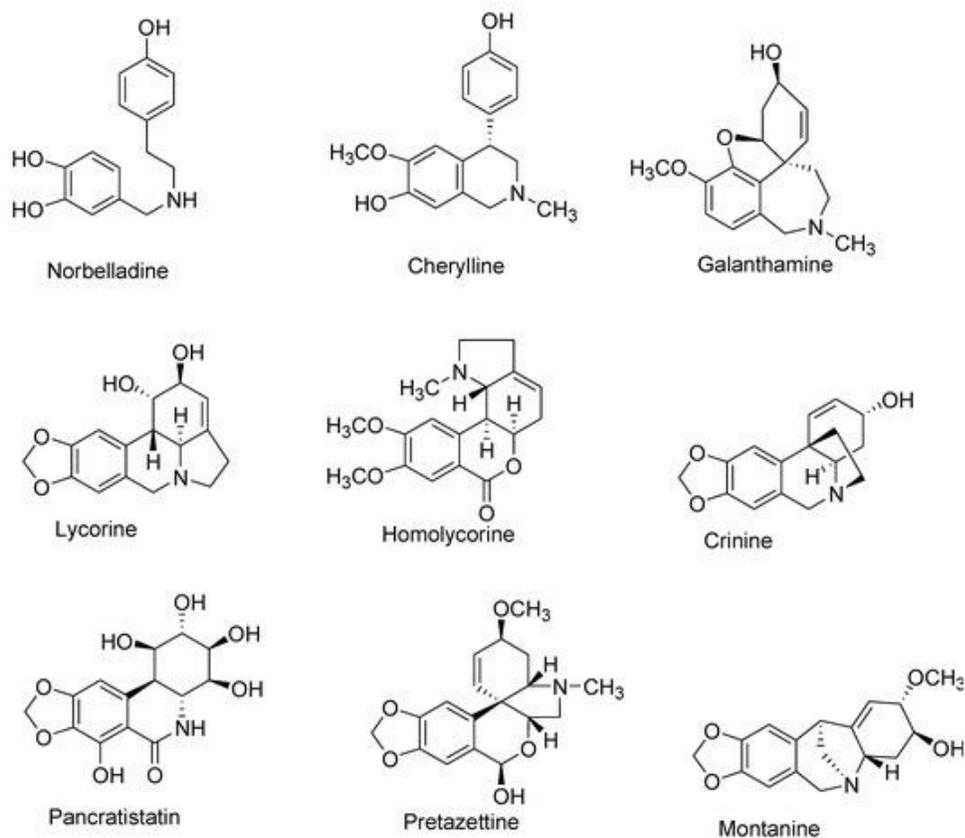


Figure 1.14 Représentations structurales des différents types d'AA.
(Tirée de Ka, Koirala et al. 2020.)

Leur voie de biosynthèse

Contrairement à la littérature sur les effets pharmacologiques des AA, peu d'information est connue sur leur voie de biosynthèse jusqu'à ce jour. Beaucoup d'enzymes sont à identifier et caractériser afin de compléter les connaissances sur la voie métabolique des AA. La biosynthèse des AA débute initialement avec des précurseurs d'acides aminés aromatiques comme substrats, soit la L-tyrosine et L-phénylalanine, provenant de la voie du shikimate impliquée dans le métabolisme primaire (Desgagné-Penix 2021). Le précurseur L-phénylalanine subit alors plusieurs réactions

chimiques par la voie des phénylpropanoïdes, afin de former le précurseur aldéhyde 3,4-dihydroxybenzaldéhyde (3,4-DHBA). La séquence de réactions menant à la formation du 3,4-DHBA est inconnue jusqu'à ce jour. Par contre, la première réaction de la séquence de modifications chimiques est bien connue et est catalysée par l'enzyme phénylalanine ammonia-lyase (PAL), une enzyme clé permettant la déamination de la phénylalanine. Pour ce qui est du précurseur L-tyrosine, celui-ci est converti en tyramine par une réaction de décarboxylation catalysée par l'enzyme tyrosine décarboxylase (TYDC). Donc, les molécules PAL et TYDC sont deux enzymes régulatrices uniques permettant la conversion de métabolites primaires en métabolites secondaires (Singh et Desgagné-Penix 2014). Par la suite, les métabolites précurseurs tyramine et 3,4-DHBA subissent une réaction de condensation à l'aide de l'enzyme nouvellement identifiée, la norbelladine synthase (NBS) pour former le précurseur intermédiaire commun à tous les AA : la norbelladine (Figure 1.15) (Singh, Massicotte et al. 2018).

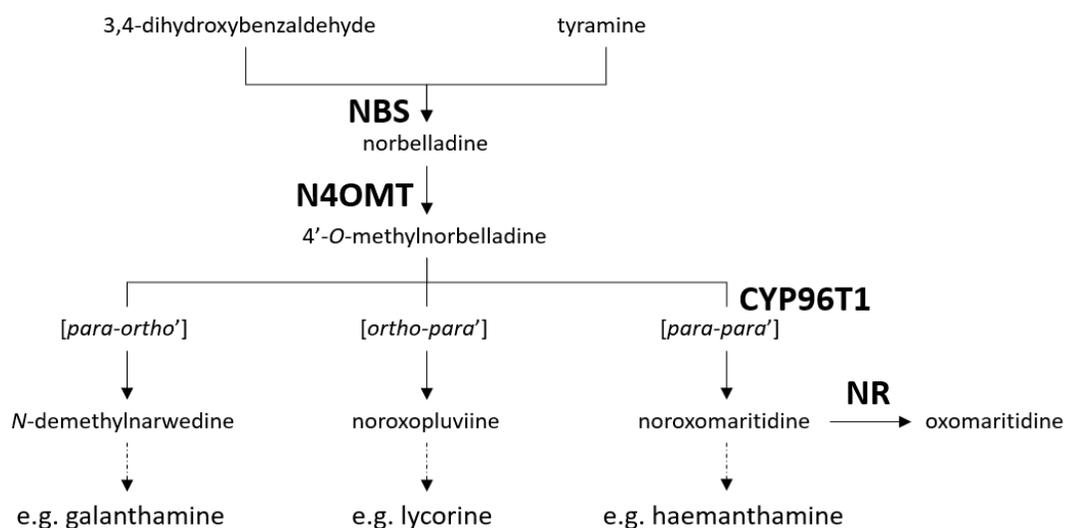


Figure 1.15 Synthèse de la norbelladine à partir de la tyramine et de 3,4-dihydroxybenzaldéhyde (3,4-DHBA), précurseur commun à tous les alcaloïdes des Amaryllidacées incluant la galanthamine, la lycorine et l'haemanthamine.

D'autres réactions de condensations similaires impliquant une amine et un aldéhyde ont été décrites. Par exemple, les réactions initiales dans la synthèse des ABI impliquent entre autres la condensation entre la dopamine et le 4-HPAA pour former la

(S)-norcoclaurine, le précurseur intermédiaire à tous les ABI (Hagel et Facchini 2013). Cette réaction est catalysée par la NCS et est la première étape engagée dans la voie de biosynthèse des benzylisoquinolines (Samanani, Liscombe et al. 2004, Lee et Facchini 2010).

La suite de la voie de biosynthèse des AA implique la méthylation de la norbelladine par l'enzyme norbelladine 4'-*O*-méthyltransférase (N4OMT) (Kilgore, Augustin et al. 2014). Le 4'-*O*-méthylnorbelladine formé par la N4OMT subit ensuite une cyclisation via trois modes de couplages oxydatifs phénoliques C-C, soit en *ortho-ortho*', *ortho-para*' et *para-ortho*'. L'enzyme CYP96T1 a été caractérisée à partir des plantes *Narcissus* sp. *aff. pseudonarcissus*, *Ganlanthus* sp. et *Galanthus elwesii*. Cette enzyme catalyse la transformation du 4'-*O*-méthylnorbelladine en (10b*R*,4a*S*)-noroxomaritidine et (10b*S*,4a*R*)-noroxomaritidine (Kilgore, Augustin et al. 2016). La noroxomaritidine est subséquentement transformée en oxomaritinamine par la noxomaritidine réductase (NR). Curieusement, la protéine NR pourrait également produire de petites quantités de norbelladine avec les substrats tyramine et 3,4-DHBA, mais avec une activité 400 fois moins élevée que celle de la conversion de la noroxomaritidine (Kilgore, Holland et al. 2016). En conclusion, seulement 4 gènes ont été identifiés jusqu'à maintenant dans la voie de biosynthèse des AA. Il est donc nécessaire d'élargir les connaissances sur les gènes et enzymes biosynthétiques des AA, afin de fournir des outils pour des applications biotechnologiques ou de biologie synthétique.

1.5 Problématique, objectifs et méthodologie

Comme mentionné précédemment, beaucoup de connaissances ont été acquises sur les propriétés pharmacologiques des AA, mais peu est connu sur leur métabolisme. Le manque d'information sur les voies de biosynthèse a suscité l'intérêt des chercheurs ces dernières années, afin d'identifier et caractériser de nouveaux gènes des AA. De façon semblable aux autres métabolites végétaux, les AA sont synthétisés en très petite quantité dans la plante et celle-ci est variable, ce qui les rend difficiles à produire à grande échelle de façon rentable. De plus, leur structure est très complexe, ce qui rend la synthèse

chimique laborieuse. Actuellement, la galanthamine est le seul AA utilisé cliniquement pour soulager les symptômes de la MA et est habituellement extraite de cultivars de *Leucojum* et *Narcissus* (Tako et Rook 2013). Il a été démontré que la galanthamine contenue dans la plante peut varier de quantité de traces, allant jusqu'à 0,5 % du poids sec des feuilles de *Leucojum aestivum*, tout dépendant la localisation géographique de celle-ci (Berkov, Georgieva et al. 2009). Il existe un grand intérêt à manipuler génétiquement la production des AA afin d'améliorer leur biosynthèse pour pouvoir commercialiser de nouveaux médicaments, mais le manque de connaissances sur les voies de biosynthèse et le métabolisme des AA restreignent cette possibilité. Donc, une meilleure compréhension au niveau moléculaire de la voie de biosynthèse des AA bénéficierait grandement à l'optimisation des cultivars pour la production de composés pharmaceutiques et au développement de nouvelles approches basées sur la biotechnologie. Pour le moment, la plupart des AA sont inaccessibles et ne peuvent pas être utilisés à des fins thérapeutiques.

Récemment, la NBS catalysant la réaction de condensation entre la tyramine et le 3,4-DHBA, a été identifiée chez la plante *Narcissus pseudonarcissus* et a été caractérisée. Malgré l'élucidation de la NBS à un niveau moléculaire, aucun autre orthologue a été identifié chez les autres espèces d'Amaryllidacées. De plus, la localisation subcellulaire de la NBS *in planta* n'a pas été explorée. Le manque d'information sur la localisation subcellulaire de la NBS contribue au manque de connaissances concernant les compartiments intracellulaires dans lesquels les AA sont biosynthétisés. La détermination de la localisation de la première étape engagée des AA où la NBS synthétise la norbelladine permettrait de faciliter la bio-ingénierie de la voie des AA dans les plantes ainsi que dans les microorganismes.

L'objectif principal du projet de recherche est d'identifier et caractériser de nouveaux gènes codant pour la NBS chez les espèces *Leucojum aestivum* et *Narcissus papyraceus*. Les espèces *Leucojum* et *Narcissus* sont bien connues pour synthétiser des AA, ce qui font d'elles des cibles intéressantes pour l'identification et caractérisation de nouveaux gènes impliqués dans la biosynthèse d'alcaloïdes. Le premier

objectif spécifique est d'identifier des orthologues de la NBS chez *L. aestivum* et *N. papyraceus*. Le second objectif spécifique du projet de recherche est de caractériser la fonction de la NBS en déterminant sa localisation subcellulaire et en étudiant l'histoire évolutive des différents orthologues identifiés.

Dans un premier temps, l'identification de gènes codant pour la NBS a été réalisée à l'aide de la transcriptomique. En effet, une recherche de nos banques de données transcriptomiques des espèces à l'étude a permis l'identification de séquences ADNc complètes nommées *LaNBS*, *NpNBS1* et *NpNBS2*. Dans un deuxième temps, les clones de NBS en fusion avec la GFP ont été agroinfiltrés dans la plante *Nicotiana benthamiana*, ce qui a permis de déterminer les compartiments intracellulaires dans lesquels la NBS exerce ses fonctions. De plus, une étude de l'histoire évolutive entre les NBS, les NCS et les protéines liées à la pathogenèse de classe 10 (PR10) a été réalisée montrant l'homologie entre ces protéines étant toutes impliquées dans le mécanisme de défense des plantes.

Au final, ce projet a permis de faire avancer les connaissances sur la voie de biosynthèse des AA. Ces nouvelles connaissances pourront faciliter le développement de méthodes de production biotechnologiques alternatives pour l'obtention d'AA biosourcés qui augmenteront et diversifieront la production de façon compétitive, efficace et respectueuse de l'environnement.

CHAPITRE II

ANALYSE DU TRANSCRIPTOME DE *LEUCOJUM AESTIVUM* ET IDENTIFICATION DE GÈNES IMPLIQUÉS DANS LA SYNTHÈSE DE LA NORBELLADINE

Le contenu de ce chapitre a fait l'objet d'une publication en anglais dans la revue scientifique *Planta* le 3 janvier 2022. La référence de cet article est la suivante :

Tousignant L, Diaz-Garza AM, Majhi BB, Gélinas SE, Singh A, Desgagne-Penix I. Transcriptome analysis of *Leucojum aestivum* and identification of genes involved in norbelladine biosynthesis. *Planta*. 2022 Jan 3;255(2):30. doi: 10.1007/s00425-021-03741-x. PMID: 34981205.

2.1 Contribution des auteurs

L'étude a été conçue par Laurence Tousignant et Isabel Desgagné-Penix, supervisée par Isabel Desgagné-Penix (directrice de recherche) et Bharat Bhusan Majhi (stagiaire postdoctoral en biologie cellulaire et moléculaire). La première ébauche du manuscrit a été réalisée par Laurence Tousignant. Les expériences à la base de cet article ont été accomplies par Laurence Tousignant, soit l'identification des trois orthologues de la NBS, l'étude bioinformatique sur les relations évolutives des orthologues *LaNBS*, *NpNBS1* et *NpNBS2*, et la détermination de la localisation subcellulaire. Isabel Desgagné-Penix a identifié 50 nouveaux gènes des AA à l'aide du transcriptome de *Leucojum aestivum*. L'assemblage du transcriptome a été réalisé par Aracely Maribel Diaz-Garza. Bharat Bhusan Majhi et Sarah-Eve Gélinas ont réalisé les essais enzymatiques de *LaNBS* par LC-MS/MS. Aparna Singh a réalisé les extractions d'ARN du bulbe de *Narcissus papyraceus* et *Leucojum aestivum*. Laurence Tousignant, Aracely Maribel Diaz-Garza, Bharat Bhusan Majhi et Isabel Desgagné-Penix ont participé à la rédaction finale du manuscrit. De plus, tous les auteurs ont révisé et approuvé la version finale de l'article.

2.2 Résumé de l'article

Les alcaloïdes des Amaryllidacées (AA) sont un large groupe de métabolites spécialisés des plantes et sont connus pour leurs propriétés pharmaceutiques. Bien que les réactions chimiques générales de la voie de biosynthèse des AA ont été proposées, la plupart des gènes et enzymes impliqués dans la voie restent inconnus à ce jour. Tous les AA sont synthétisés à partir d'un précurseur commun, soit la norbelladine, qui est formée par la condensation de la tyramine et du 3,4-dihydroxybenzaldéhyde (3,4-DHBA). La norbelladine synthase (NBS), catalysant la réaction de condensation, a récemment été identifiée et caractérisée à un niveau moléculaire chez *Narcissus pseudonarcissus*, mais aucun autre orthologue n'a été identifié et les compartiments intracellulaires dans lesquels la NBS exerce ses fonctions n'ont toujours pas été explorés. Dans cette étude, une première analyse complète du transcriptome de *Leucojum aestivum* a été réalisée, afin d'identifier des gènes clés impliqués dans la biosynthèse des AA. Cinquante gènes orthologues ont été identifiés et déposés dans GenBank. De plus, la NBS a été identifiée et caractérisée à partir du transcriptome de *L. aestivum* et du transcriptome de *Narcissus papyraceus* précédemment rapporté. Les études phylogénétiques ont démontré que *LaNBS*, *NpNBS1* et *NpNBS2* partagent un pourcentage d'identité élevé de leurs séquences en acides aminés. L'expression hétérologue de la protéine recombinante *LaNBS* a permis de confirmer l'activité enzymatique de celle-ci en présence de ses substrats. Les prédictions bioinformatiques et l'expression transitoire de la NBS en fusion avec GFP dans *Nicotiana benthamiana* agroinfiltrée ont montré que *LaNBS*, *NpNBS1* et *NpNBS2* sont localisés dans le cytosol, ce qui suggère que la biosynthèse des AA commence dans le cytosol. Cette étude fournit un transcriptome des Amaryllidacées qui sera très utile pour identifier de nouveaux gènes et mener des recherches de caractérisation du métabolisme des AA dans la plante ou dans un système hétérologue. Finalement, l'étude permettra de faciliter la bio-ingénierie de la voie de biosynthèse des AA dans les plantes ou dans les microorganismes.

2.3 Article scientifique

Transcriptome analysis of *Leucojum aestivum* and identification of genes involved in norbelladine biosynthesis

Laurence Tousignant¹, Aracely Maribel Diaz-Garza¹, Bharat Bhusan Majhi¹,
Sarah-Eve Gélinas¹, Aparna Singh¹, Isabel Desgagne-Penix^{1,2*}

¹Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

²Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

*Correspondence: Isabel.Desgagne-Penix@uqtr.ca, Telephone: 1-819-376-5011

Abstract

***Main conclusion* Transcriptome analysis of *Leucojum aestivum* led to the identification of 50 key genes associated with Amaryllidaceae alkaloid biosynthesis including norbelladine synthase which localized in the cytosol and catalyzed norbelladine formation.**

Abstract

The Amaryllidaceae alkaloids (AAs) are a large group of plant specialized metabolites, which are known for their biological activities. Although the general chemical reactions in the AA biosynthetic pathway have been proposed, the genes and enzymes of the pathway remain largely unstudied. All AAs are synthesized from a common precursor, norbelladine, by the condensation of tyramine and 3,4-dihydroxybenzaldehyde. The enzyme norbelladine synthase (NBS) which catalyzes the condensation reaction has only been characterized at a molecular level from one species, and the subcellular localizations have not been explored. Hence, the intracellular compartments wherein the

AAs are biosynthesized remain unknown. In this study, a first comprehensive transcriptomic analysis of summer snowflake (*Leucojum aestivum*) was done to identify key genes associated with AA biosynthesis. Fifty orthologous genes were identified and deposited into GenBank. In addition, we identified and further characterized NBS from the transcriptome of *L. aestivum* and previously reported *Narcissus papyraceus*. Phylogenetic analysis showed that *La*NBS, *Np*NBS1 and *Np*NBS2 shared high amino acid identity. The heterologous expression of *La*NBS produced a recombinant protein with NBS activity. Bioinformatic prediction and C-terminal GFP tagging in transiently transformed *Nicotiana benthamiana* showed that *La*NBS, *Np*NBS1 and *Np*NBS2 were likely localized to the cytosol which suggests that the AA biosynthesis starts in the cytosol. This study provides an Amaryllidaceae transcriptome that will be very helpful to identify genes for characterization studies in AA metabolism *in planta* or using heterologous systems. In addition, our study will facilitate the bioengineering of AA biosynthetic pathway in plants or in microorganisms.

Keywords: Amaryllidaceae alkaloid, transcriptomic, *Leucojum aestivum*, Norbelladine synthase, specialized metabolism, subcellular localization.

Abbreviations

3,4-DHBA	3,4-dihydroxybenzaldehyde
4CL	4-hydroxycinnamoyl CoA ligase
AA	Amaryllidaceae alkaloid
C3H	Coumarate 3-hydroxylase
C4H	Cinnamate 4-hydroxylase
CYP96T1	Cytochrome P450 monooxygenase 96T1
FPKM	Fragments per kilobase of transcript per million mapped reads
HCT	Hydroxycinnamoyl transferase
N4OMT	Norbelladine 4'- <i>O</i> -methyltransferase
NBS	Norbelladine synthase
NorRed	Noroxomaritidine reductase
PAL	Phenylalanine ammonia lyase
Rt	Retention time
TPM	Transcript per million mapped reads
TYDC	Tyrosine decarboxylase

Introduction

The Amaryllidaceae plant family produces a wide range of chemical compounds through the primary and specialized metabolic pathways. One class of specialized metabolites specific to this family that possess multiple biological activities is the Amaryllidaceae Alkaloids (AAs). To date, AAs are distributed among 1100 species and more than 650 structurally diverse alkaloids have been isolated from Amaryllidaceae plants (Jin and Yao 2019, Ka, Koirala et al. 2020). Alkaloids are known to play a major role in plant defense against several biotic and abiotic stresses (Croteau, Kutchan et al. 2000, Akula and Ravishankar 2011).

Recently, AAs have been a source of interest because of their potential exploitation as therapeutic agents for various ailments and diseases. The structurally diverse groups of alkaloids from Amaryllidaceae family have been shown to display antimicrobial, anticancer and anti-inflammatory activities (Nair and van Staden 2013, Hotchandani and Desgagne-Penix 2017, Ka, Koirala et al. 2020). For example, both lycorine and narciclasine exhibit potent anti-tumor and anti-inflammatory properties (Kornienko and Evidente 2008, Lamoral-Theys, Andolfi et al. 2009, Bastida, Berkov et al. 2011, Furst 2016). Despite their powerful pharmaceutical activities, only galanthamine, which is currently used as acetylcholinesterase inhibitor to treat symptoms of Alzheimer's disease, has been commercialized (Heinrich and Lee Teoh 2004). Like the other AAs, galanthamine is produced in variable quantities in the plants. Indeed, Berkov, S. et al. (2009) showed that galanthamine content vary from trace amounts to 0.5% (generally 0.1-0.3%) of dry weight in the leaves of *Leucojum aestivum* (summer snowflake) depending on the geographical location (Berkov, Georgieva et al. 2009). Similar results were obtained from bulbs and *in vitro* cultures from *L. aestivum* (Georgieva, Berkov et al. 2007, Berkov, Georgieva et al. 2013). The variable and low yield of AAs makes them difficult to extract from plants for large-scale production, since it is an expensive process and the amount extracted can vary. In addition, the complex structure of AAs makes the chemical synthesis process difficult and costly. Therefore, most of the AAs are currently inaccessible and cannot be used for therapeutic purposes.

There is an interest to genetically modify cultivars or do *in vitro* cultures to produce higher concentrations of AAs and this way generate a synthetic biology platform using metabolic engineering. However, it is limited by the lack of genomic and transcriptomic data available of Amaryllidaceae plants, since only few enzymes involved in AA biosynthesis have been identified while several remain unknown (Kilgore and Kutchan 2016, Desgagne-Penix 2021).

AAs are assembled from a common precursor, named norbelladine, which is derived from the aromatic amino acids tyrosine and phenylalanine via the shikimate/chorismate and phenylpropanoid pathways (**Fig. 1**). Recently, the enzyme that catalyzes the first committed step in AAs biosynthesis, namely norbelladine synthase (NBS), has been identified and characterized from *Narcissus pseudonarcissus* ‘King Alfred’ (Singh, Massicotte et al. 2018). NBS catalyzes the condensation of tyramine and 3,4-dihydroxybenzaldehyde (3,4-DHBA) to form norbelladine (**Fig. 1**). Despite the recent elucidation of NBS at the molecular level, no orthologs have been reported from other Amaryllidaceae species. Additionally, NBS subcellular localization *in planta* has not been explored. Thus, there is still no information about the subcellular localization of NBS that contributes to the lack of knowledge regarding the intracellular compartments in which the AAs are biosynthesized. The determination of wherein the NBS synthesizes norbelladine will facilitate the bioengineering of the AA biosynthetic pathway in non-Amaryllidaceae plants and microorganisms.

Besides, while *L. aestivum* is also a primary source of galanthamine, the lack of transcriptomic data has not made possible to identify the AA pathway. In this study, we analyzed the alkaloid profiles from different tissues of *L. aestivum* and quantified AAs from bulb extracts by high-performance liquid chromatography. Next, we generated the transcriptome of *L. aestivum* bulbs facilitating identification of key genes encoding AA biosynthetic enzymes. BLAST searches were performed to find *norbelladine synthase* (NBS) candidate transcripts in our generated *de novo* transcriptome from *L. aestivum* and the available transcriptome of *Narcissus papyraceus* (Hotchandani, de Villers et al. 2019). Candidate NBS sequences were selected based on their homology with *NpKANBS* (Singh,

Massicotte et al. 2018). Phylogenetic analysis and alignment of amino acids sequences were performed to show similarity between the NBSs. The heterologous expression of *LaNBS* produced a recombinant protein with NBS activity confirmed using LC-MS/MS analysis. Next, we investigated the subcellular localization of NBSs in *Nicotiana benthamiana* plants. Thus, in this study, we report for the first time on the transcriptome of *L. aestivum* and the identification of 50 AA genes including the key enzyme NBS, which catalyzes the first committed step in the AAs biosynthesis. We also report on the cytosolic subcellular localization of norbelladine synthase.

Materials and methods

Plant material

Summer snowflake (*Leucojum aestivum*) bulbs and Paperwhite narcissus (*Narcissus papyraceus*) bulbs were both purchased from Vesey's (York, PE, Canada). Several *L. aestivum* bulbs were kept aside (for alkaloid and RNA extraction) and others were planted using AGRO MIX G6 potting soil (Fafard, QC, Canada), which had been autoclaved for 40 min. The plants were grown at room temperature with exposure to tube lighting for 16 h daily until being harvested. The plants were watered when necessary to keep the soil moist. Tissues (stems, flowers, leaves) for alkaloid analysis were harvested 2 days after flowering. The gateway-adapted *attP*-flanked pDONR 221 vector used for BP recombinase reaction was purchased from Invitrogen (CA, USA). The gateway expression vector pB7FWG2 with C-terminal GFP tag was obtained from VIB-UGent Center for Plant Systems Biology (Ghent, Belgium) (Karimi, Inze et al. 2002). Fungus effector MLP124466 kindly provided by Prof. Hugo Germain (Université du Québec à Trois-Rivières, QC, Canada) (Germain, Joly et al. 2018). MLP124466-GFP, well-known to have a nucleocytoplasmic localization, was used as positive control.

Alkaloid extraction, TLC and HPLC analysis

Alkaloids were extracted from plant tissues of *L. aestivum* with methanol and purified following an acid-base extraction protocol described previously (De Andrade, Pigni et al. 2012, Hotchandani, de Villers et al. 2019). For a given sample, 2 g/each in triplicates of fresh plant material were ground to a fine powder in liquid nitrogen with a mortar and pestle. The powder was transferred to a 15 mL centrifuge tube. Five mL of methanol were added to the powder, which was then kept at 50 °C for 2 h on a shaker (200 rpm). The tubes were vortexed then centrifuged at 7000 g for 2 min to collect supernatant without debris and left for complete evaporation. The supernatant (crude extract of alkaloids) was collected and allowed to evaporate to dryness. The dry crude extracts of alkaloids were further subjected to a modified acid-base extraction method (Singh and Desgagne-Penix 2017), where they were resuspended in methanol 100% pH-8 (adjusted with NH₃) and H₂SO₄ (2% v/v). Organic impurities were removed by washing twice with chloroform. Alcalization was done using NH₃. The purified alkaloid extracts obtained were dried under N₂ gas and finally solubilized in 300 µL methanol. Qualitative analysis of alkaloids was performed on TLC silica gel 60 F₂₅₄ aluminum sheets 20 x 20 cm, (Merck, Darmstadt, Germany). Fifteen µl of samples were loaded on TLC and visualized under 254 nm and 365 nm (**Fig. S1**). For quantitative analyses of alkaloids, prior to extraction for HPLC extracts, ten microliters of papaverine (0.03 mg/mL in water) were used as an internal standard for extraction yield determination and data quantification. High-Performance Liquid Chromatography (HPLC) analysis was performed on the alkaloid extract samples by reversed-phase HPLC with photodiode array (PDA) detector using the Prominence-i LC-2030C system (Shimadzu, www.shimadzu.com). Separations were performed at a flow rate of 0.5 mL/min using a Kinetex C₁₈ column (150 × 4.6 mm, 5 µm particle size; Phenomenex, www.phenomenex.com). HPLC oven temperature was set at 40 °C. Gradient elution was carried out using a 1% aqueous ammonium acetate solution with pH 5 and 100% acetonitrile. The ratio of the ammonium acetate solution to acetonitrile was: 90:10 for 11 min, 69:31 for 4 min, 30:70 for 1 min, 10:90 for 5 min and 90:10 for 2 min. Compounds were monitored at 280 nm. The average retention time (R_t) and absorption maxima (λ max) were determined for the AAs lycorine, galanthamine and narciclasine serving as standards, and for the alkaloid papaverine serving as internal

standard for the relative quantification of the concentration of detected compounds as described previously (Singh and Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019). Hundred microliters of the alkaloid extract of each plant sample were injected into the HPLC column. The area under the peak value of each compound detected was normalized to the area value of the internal standard papaverine in that sample. The relative concentration of alkaloid was determined using the papaverine standard curve ($y = 24731x - 15295$; $R^2 = 0.991$) and reported into equivalent mg of papaverine/g tissues.

RNA extraction, Illumina sequencing and transcriptome assembly

The transcriptome of *N. papyraceus* was generated in a previously published study by (Hotchandani, de Villers et al. 2019). Similarly, biological triplicate (2g/each) of bulbs (stage 1 (dormancy) (Hotchandani, de Villers et al. 2019)) of *L. aestivum* were used to generate the corresponding transcriptome, the same method was used for its generation. Replicates were pooled for the transcriptome generation. Briefly, total RNA from bulbs of *L. aestivum* was extracted using CTAB (cetyltrimethylammonium bromide) method as described by (Singh and Desgagne-Penix 2017). The integrity of RNA extracted from the bulb was verified on NanoVue spectrophotometer (GE Healthcare Life Sciences, www.gelifesciences.com) and on a bioanalyzer (G enome Qu ebec, QC, Canada). For example for one biological replicate, Nanodrop quantification yielded a total RNA concentration of 1022.8 ng/ μ l with a ratio 260 nm/230 nm of 2.46 and 260 nm/280 nm of 2.06. Bioanalysis of *L. aestivum* RNA gave an RNA integrity number (RIN) of 8.6 with a 28s/18s of 1.49, which confirmed its suitable quality to proceed for next generation sequencing. Biological triplicates RNA samples of suitable quality were sent to Genome Qu ebec for transcriptome generation.

The mRNA was converted into a cDNA library and sequenced through Illumina HiSeq 2000 sequencing system, PE 100 paired ends, at Genome Quebec Innovation Centre (Montr eal, QC, Canada). The transcriptomic pipeline for normalization, assembly, and annotation, is summarized in **Fig. S2**. Raw paired reads were analysed using FastQC (version 0.11.8) for quality control (Conesa, Madrigal et al. 2016). Afterwards, Trinity

in silico normalization was performed to eliminate redundant reads in datasets without affecting its k-mer content. No quality trimming was done prior to *in silico* normalization since the per-base quality plots showed high quality reads and extremely low adaptor content, in addition significant information can be lost when performing quality trimming (Mbandi, Hesse et al. 2014). *De novo* assembly of normalized reads was done using the Trinity assembler (version 2.9.0) (<https://github.com/trinityrnaseq/trinityrnaseq/wiki>) (Grabherr et al. 2011) with a 32 k-mer size. Additionally, the completeness of the assembled transcriptome was assessed with BUSCO (version 3.0.2) using liliopsida odb10 database (Waterhouse, Seppey et al. 2017). Coding regions were predicted using TransDecoder (version 5.5.0) (<http://transdecoder.github.io/>). The longest open reading frames (ORFs) obtained were functionally annotated using Trinotate pipeline (version 2.0) (<http://trinotate.github.io/>) (Brown et al. 2012). The gene ontology (GO) of the annotated transcripts was extracted to illustrate the ten most abundant GO terms of each category (molecular function, cellular component and biological process). To quantify the gene transcript abundance, the raw RNA-Seq reads were mapped to the assembled transcripts applying Bowtie2 using default parameters (Langmead et al. 2009). The gene transcript abundance was calculated as transcripts per million transcripts (TPM) using the RSEM package (Li and Dewey 2011).

Candidate gene selection and primers design

The norbelladine synthase candidate genes were selected by screening both *Leucojum aestivum* and *Narcissus papyraceus* transcriptomic databases for ortholog norbelladine synthase (NBS)/norcoclaurine synthase (NCS). Transcript sequences associated with NCS from *Thalictrum flavum* were selected for each plant species. The orthologous full-length transcripts obtained, with an open reading frame of 480 bp for both species, were selected for amplification *i.e.*, were used as templates in order to design specific primers, which contained gateway *attB* segments (**Table S1**). PCR amplification yielded one NBS ortholog from *L. aestivum* cDNA (*La*NBS; Genbank MW971977) and two from *N. papyraceus* (*Np*NBS1; Genbank MZ054103 and *Np*NBS2; Genbank MZ054104).

Protein alignment and phylogenetic analysis

Studied *Np*NBS1, *Np*NBS2 and *La*NBS sequences were aligned with *N. pseudonarcissus* ‘King Alfred’ norbelladine synthase (*Np*KANBS; GenBank: AYV96792.1) and *Thalictrum flavum* norcoclaurine synthase (*Tf*NCS; Genbank: ACO90248.1) using CLUSTAL W algorithm in MEGA X software at default parameters. Sequences alignment were edited using the BioEdit program version 7.2.6.1 (TA. 1999). The evolutionary history between norbelladine synthases, norcoclaurine synthases and pathogenesis-related proteins 10 (PR10) was inferred using the Neighbor-joining statistic method (Saitou and Nei 1987). Phylogenetic analysis was conducted on MEGA X using bootstrap method for the phylogeny test (Felsenstein 1985, Kumar, Stecher et al. 2018). Amino acid sequences of *La*NBS, *Np*NBS1 and *Np*NBS2 were aligned with norbelladine synthase *Np*KANBS, and with norcoclaurine synthase sequences from *Papaver somniferum* (*Ps*NCS1 and *Ps*NCS2), *Thalictrum flavum* (*Tf*NCS), *Coptis japonica* (*Cj*PR10A), *Sinopodophyllum hexandrum* (*Ph*NCS), *Papaver bracteatum* (*Pb*NCS), and were aligned with PR10 sequences from *Pinus monticola* (*Pm*PR10-1 and *Pm*PR10-2), *Triticum turgidum* subsp. *Durum* (*Td*PR10), *Triticum aestivum* (*Ta*PR10), *Zea mays* (*Zm*PR10 and *Zm*PR10.1), *Gossypium barbadense* (*Gb*PR10), *Arachis hypogaea* (*Ah*PR10), *Capsicum annuum* (*Ca*PR10), *Lupinus albus* (*La*PR10), *Hyperitum perforatum* (HYP-1) and *Solanum virginianum* (*Sv*PR10). The alignment was done using the CLUSTAL W algorithm.

For bioinformatics predictions of signal peptides and subcellular localizations, the entire amino acid sequences of *La*NBS, *Np*NBS1 and *Np*NBS2 were analyzed using the online tool signal peptide prediction programs TargetP2.0, SignalP5.0 and ChloroP1.1 (**Table S2**). The subcellular localization prediction programs used were Loctree3, PSLpred and WoLF PSORT. For all the programs, the defaults plants or eukaryotes were selected, prioritizing plants over eukaryotes.

PCR amplification, cloning, and construction of the expression cassette for enzyme assay

The open reading frame (ORF) of full length *LaNBS* was amplified from *L. aestivum* bulbs cDNA using PrimeStar GXL premix (TaKaRa Bio) in 50 μ L reaction with 0.2 μ M forward (LaNBS-FP-BamHI, 5'-AACGGGATCCATGAAGGGAAGTCTCTCCCATGAG-3') and reverse (LaNBS-RP-HindIII, 5'-ACGCAAGCTTCTACGCTACAATAGCTTTTTGCTCC-3') primers. PCR program parameters: 2 min 98 °C 1 cycle, 10 s 98 °C, 20 s 55 °C, 1 min 72 °C for 35 cycles, 5 min 72 °C 1 cycle, and final infinite hold at 4 °C. For protein expression in *Escherichia. coli*, *LaNBS* gene was fused to the C-terminus of the maltose binding protein (MBP) in the pMAL-c2x vector (New England Biolabs). Precisely, the full-length ORF was amplified by PCR using primers reported above with the restriction enzyme sites underlined. The PCR product was cleaned using GenepHlow Gel/PCR kit (Geneaid). The purified PCR product was digested with *Bam*HI and *Hind*III and ligated into pMAL-c2x vector digested with *Bam*HI and *Hind*III using T₄ DNA ligase (New England Biolabs). The recombinant plasmids were transformed into chemically competent *E. coli* DH5 α cells by heat shock transformation and colonies were selected on ampicillin (100 μ g/mL) LB agar plates. The positive clones were identified by colony PCR in a 20 μ L reaction using Taq DNA polymerase with ThermoPol buffer (New England Biolabs) with PCR parameters: 5 min 95 °C 1 cycle, 30 s 95 °C, 40 s 55 °C, 1 min 68 °C for 30 cycles, 5 min 68 °C 1 cycle, and final infinite hold at 4 °C. The resulting plasmids were verified by DNA Sanger sequencing at Genome Quebec Innovation Centre (Montréal, QC, Canada) to ensure the correct sequence and exclude undesired mutations.

Expression and purification of MBP fusion proteins in *E. coli*

LaNBS gene was cloned into the pMAL-c2x vector. The purified plasmids were transformed using the heat shock transformation into chemically competent *E. coli* Rosetta (DE3) pLysS strain for protein expression. Transformed cells were selected on LB plates with 100 μ g/mL ampicillin, and 34 μ g/mL chloramphenicol overnight at 37 °C. The positive colonies were screened by colony PCR using PCR parameters: 5 min 95 °C

1 cycle, 30 s 95 °C, 40 s 55 °C, 1 min 68 °C for 30 cycles, 5 min 68 °C 1 cycle, and final infinite hold at 4 °C. A PCR positive single colony was picked and grown overnight at 37 °C at 220 rpm in 12.5 mL LB broth containing 100 µg/mL ampicillin, and 34 µg/mL chloramphenicol. The overnight grown pre-culture was added into fresh 250 mL LB broth containing ampicillin (100 µg/mL), and chloramphenicol (34 µg/mL) and grown at 220 rpm at 37 °C to an OD₆₀₀ = 0.5 to 0.6. The cultures were brought to room temperature and Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.25 mM to induce the protein expression. The cultures were incubated for 20 h at 18 °C at 150 rpm. Bacterial cultures were pelleted at 10,000 rpm for 15 min at 4 °C, resuspended in 25 ml column binding buffer (25 mM Tris-HCl [pH 7.5], 150 mM NaCl, and 1 mM EDTA) and frozen at -80 °C overnight. The cultures were thawed on ice water and 1 mM phenylmethylsulfonyl fluoride [PMSF], and 0.1x protease inhibitor cocktail (Cell Signaling Technology) were added and lysed using a sonicator at 41% amplitude for a total of 8 min with 15 s run and 35 s cooling time in ice. The lysates were centrifuged twice at 14,000 rpm for 20 min at 4 °C to pellet the cell debris and the clear supernatants were collected. Supernatants were mixed with 500 µL of amylose resin beads (New England Biolabs) (prewashed with column binding buffer and resuspended to 50% slurry) and incubated for 1 h at 4 °C with constant rocking. The mixture was passed twice through the filter columns (Thermo Scientific) with gravitational flow to retain the beads with bound protein in the column matrix. The beads were washed with gravitational flow in the columns three times with 30 mL column binding buffer. The bead slurry was transferred to microcentrifuge tube and centrifuged at 1000 rpm for 1 min and the supernatant was removed. Finally, the bound protein from the bead pellet was eluted twice (elution 1 and 2) each time in 500 µL elution buffer (15 mM maltose in column binding buffer) by centrifugation at 1000 rpm for 1 min and supernatant/elute was collected. The protein samples were flash frozen in liquid nitrogen and stored at -80 °C. Protein quantification was done using DC protein assay kit (Bio-Rad) according to the manufacturer's instructions with bovine serum albumin (BSA) as standard, and protein samples were fractionated by 10% (v/v) SDS-PAGE and stained with Coomassie Blue.

Substrates, standards preparation and LC-MS/MS analysis of enzymatic assays

Norbelladine was synthesized as previously described (Singh, Massicotte et al. 2018). Standard solutions of 3,4-dihydroxybenzaldehyde (3,4-DHBA) (Fisher Scientific), tyramine (Sigma-Aldrich), and papaverine (Sigma-Aldrich) were prepared at 1000 ppm in LC-MS grade methanol (Sigma-Aldrich). Standard solution of norbelladine was prepared as previously described (Singh, Massicotte et al. 2018). From these standard solutions, dilutions were performed to obtain working solutions of 100 ppm in methanol, and 1 ppm in the mobile phase (10 mM ammonium acetate (Sigma-Aldrich) (pH 5.0), and acetonitrile (Sigma-Aldrich) [60:40]). The substrates and standard solutions were stored in the dark at $-20\text{ }^{\circ}\text{C}$.

The catalytic activity of the purified *La*NBS enzyme was analyzed following the method of (Singh, Massicotte et al. 2018) with some modifications. Reactions were conducted using 80 μg of purified protein in 100 mM HEPES buffer (pH 6.0), with 10 μM tyramine and 300 μM 3,4-DHBA, in a total volume of 100 μL . Reaction components were equilibrated at $35\text{ }^{\circ}\text{C}$ and the reaction was started by the addition of enzyme to the substrate and buffer mixture. The reaction was incubated at $35\text{ }^{\circ}\text{C}$ for 2 h and terminated by addition of 10 μL of 20% trichloroacetic acid (TCA). Negative controls were without substrate and without enzyme containing assays. Following the reaction termination, papaverine (1000 PPM) was added to all the reaction serving as an internal standard to assure equal and efficient extraction of detected compounds among the replicates. All reactions were performed in triplicates. The reaction samples were diluted 100x with mobile phase (10 mM ammonium acetate [pH 5.0], and acetonitrile [60:40]) and analysis of the enzymatic product (norbelladine) using a high-performance liquid chromatography (HPLC) system coupled with a tandem mass spectrometer (MS/MS) (Agilent, QC, Canada) equipped with an Agilent Jet Stream ionization source, a Kinetex EVO C₁₈ column (150 x 4.6 mm, 5 μm , 100 \AA ; Phenomenex, Torrance, USA), a binary pump, an autosampler and a column compartment were used for the analysis. The LC-MS/MS analysis was carried out as described in **Table S3**, **Table S4** and by (Singh, Massicotte et al. 2018).

Cloning and PCR analysis for localization assay

The open reading frame (ORF) of full length *LaNBS* and *NpNBS* were amplified from *L. aestivum* and *N. papyraceus* bulbs cDNA, respectively. PCR amplifications were done using 200 μ M dNTPs, 0,50 μ L of Q5 High-fidelity DNA polymerase in 50 μ L reaction and 0,5 μ M forward and reverse gateway primers (**Table S2**). PCR cycling parameters were 30 s 98° 1 cycle, 10 s 98°, 20 s 61°, 30 s 72° for 34 cycles, 2 mins 72 °C 1 cycle. *AttB*-flanked PCR products of *LaNBS* and *NpNBS* were migrated on a 1,2% agarose gel and extracted using the Gel/PCR DNA fragments extraction kit (Geneaid, New Taipei City, Taiwan). Gateway cloning technology was used to clone *LaNBS* and *NpNBS* according to manufacturer's protocol (ThermoFisher Scientific, Canada). PCR-amplified *NBS* were inserted in the *attP*-flanked pDONR221 vector using BP enzyme, which catalyzes the BP recombination reaction. The pDONR221 plasmid bearing a kanamycin resistance gene generated *attL*-flanked entry clones using the *attB*-flanked *LaNBS* and *NpNBS* DNA fragments. These entry clones were transformed by electroporation into *E. coli* DH5 α competent cells and positive clones were selected on kanamycin (50 μ g/mL) Luria-Bertani (LB) agar plates. After sequencing verification, *attL*-flanked products *LaNBS*, *NpNBS1* and *NpNBS2* (2 transcript variants were detected by sequencing in the case of *NpNBS*) were used to perform LR recombination reaction with *attR*-flanked pB7FWG2 vector. The coding sequences were transferred in frame into pB7FWG2 plasmid containing a C-terminal green fluorescent protein (GFP) tag, p35S promoter, T35S terminator and a streptomycin/spectinomycin resistance gene using LR recombinase enzyme. The generated constructs 35S::*LaNBS*-GFP, 35S::*NpNBS1*-GFP and 35S::*NpNBS2*-GFP (**Fig. S3**) were transformed into *E. coli* DH5 α and positively selected on 50 μ g/mL streptomycin/spectinomycin LB plates incubated overnight at 37 °C. The positive colonies were confirmed by Sanger sequencing at Genome Quebec Innovation Centre (Montréal, QC, Canada).

As of positive control MLP124466, which is a fungus effector protein from *Melampsora larici-populina* that is expressed in infected plant host and contribute to pathogen virulence, has been cloned and transformed as described by Germain et al. (2018). Briefly, DNA sequence has been ordered from GenScript. The construct was

transferred by recombination into pDONR Zeo vector (Invitrogen, Carlsbad, CA, USA) and then inserted into pB7FWG2 (Karimi et al. 2002) vector to generate C-terminal GFP-tagged protein. This positive control was selected because of its availability and its well-known localization to both the nucleus and the cytoplasm. Thus, we did not feel the need to performed a nuclear staining such as DAPI.

Plant transformation and confocal microscopy

All GFP constructions (35S::*La*NBS-GFP, 35S::*Np*NBS1-GFP, 35S::*Np*NBS2-GFP and 35S::*124466*-GFP) were transformed into *Agrobacterium tumefaciens* strain C58C1 resistance to rifampicin (50 µg/µL) by electroporation. Cultures of *A. tumefaciens* containing pB7FWG2 vectors were grown overnight in Luria-Broth with 50 µg/mL streptomycin/spectinomycin and rifampicin, at 28 °C. The bacterial cultures were centrifuged at 5000 rpm for 5 minutes and the pellets were resuspended with infiltration buffer (MgCl₂ 10 mM and 5 mM acetosyringone). The resultant bacterial suspensions harboring different T-DNA were diluted to obtain a final OD at 600 nm (OD₆₀₀) of 0.5. The mixtures were kept 1 h at room temperature and then infiltrated into the young leaves of *N. benthamiana* plants with a 1 mL syringe. Four-week-old *N. benthamiana* grown at 22 °C in a 16/8 h light/dark cycle was used for the infiltration procedure.

Two days post infiltration, the abaxial epidermis of leaves was placed in a water drop, covered by a slip, and imaged immediately. Images were captured with a Leica TCS SP8 confocal laser scanning microscope (Leica Microsystems) with a 40X/1.30 oil immersion objective. The GFP excitation wavelength used was 488 nm and the emission of fluorescence signals was detected from 500 to 525 nm. Chlorophyll autofluorescence was observed with an excitation wavelength of 552 nm and the emission of fluorescence signals was detected from 630 to 670 nm. The images were first processed in the Las AF Lite software and the combined images were generated using the Las X program.

Protein extraction from *N. benthamiana* leaves and Western blotting

In addition to the validation of the DNA sequence using Sanger sequencing, the integrity of the fusion protein NBS-GFP was validated using Western blot analysis (**Fig. S4**). For protein extraction from *N. benthamiana* leaves, four leaf discs (1 cm diameter) were frozen in liquid nitrogen, homogenized in 250 μ L extraction buffer (100 mM Tris [pH 7.5], 1% [v/v] Triton X-100, 1 mM PMSF, and 0.1x protease inhibitor cocktail), and centrifuged at 17,000g for 30 min at 4 °C. The clear supernatant was collected, and protein (30 μ L) was fractionated by 10% (v/v) SDS-PAGE. Proteins from gels were transferred onto Polyvinylidene difluoride (PVDF) membrane using Trans-Blot Turbo transfer system (Bio-Rad). The membrane was equilibrated with Tris-buffered saline (TBS) buffer (20 mM Tris, 150 mM NaCl pH 7.6) for 15 min, followed by blocking of membrane for 3 h with TBS buffer containing 0.1% tween 20 (TBST), and 3% bovine serum albumin (BSA). The membrane was incubated overnight at 4 °C in TBST with 3% BSA containing 1:1000 dilution of mouse anti-GFP monoclonal antibody (Cedarlane labs). Following primary antibody incubation, the membrane was washed three times each for 5 min in TBST buffer and incubated for 30 min in TBST containing 5% skim milk and goat anti-mouse horseradish peroxidase (GAM)-HRP conjugate in 1:5000 dilutions. The immunoblot was washed three times for 5 min each in TBST buffer and developed using clarity Western ECL substrate (Bio-Rad). Finally, the PVDF membrane was washed with TBST and stained with Ponceau S stain and photographed using Gel Doc XR system (Bio-Rad).

Accession Numbers

The RNA-seq data discussed in this publication have been deposited in the NCBI Sequence Read Archive under the accession number PRJNA720900 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA720900>). The gene transcript sequences described herein were deposited in GenBank with the following accession numbers for nucleotide sequences: *L. aestivum* 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1 (*LaDAHPS1*; MW971931), (*LaDAHPS2*; MW971932), (*LaDAHPS3*; MW971933), *L. aestivum* 3-dehydroquinate synthase 1 (*LaDHQS1*; MW971934),

(*LaDHQS2*; MW971935), *L. aestivum* bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase 1 (*LaDHQD1*; MW971936), (*LaDHQD2*; MW971937), *L. aestivum* shikimate kinase 1 (*LaSK1*; MW971938), (*LaSK2*; MW971939), (*LaSK3*; MW971940), *L. aestivum* 5-enolpyruvylshikimate-3-phosphate synthase (*LaEPSPS*; MW971941), *L. aestivum* chorismate synthase (*LaCS*; MW971942), *L. aestivum* chorismate mutase (*LaCM*; MW971943), *L. aestivum* bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase 1 (*LaPAT1*; MW971944), (*LaPAT2*; MW971945), *L. aestivum* arogenate dehydratase/prephenate dehydratase 1 (*LaADT1*; MW971946), (*LaADT2*; MW971947), (*LaADT3*; MW971948), *L. aestivum* arogenate dehydrogenase 1 *LaADH1*; MW971949), (*LaADH2*; MW971950), (*LaADH3*; MW971951), *L. aestivum* tyrosine decarboxylase 1 (*LaTYDC1*; MW971952), (*LaTYDC2*; MW971953), *L. aestivum* phenylalanine ammonia lyase 1 (*LaPAL1*; MW971954), (*LaPAL2*; MW971955), (*LaPAL3*; MW971956), *L. aestivum* cinnamate 4-hydroxylase 1 (*LaC4H1*; MW971957), (*LaC4H2*; MW971958), *L. aestivum* coumarate 3-hydroxylase *LaC3H*; MW971959), *L. aestivum* 4-coumarate-CoA ligase 1 *La4CL1*; MW971960), (*La4CL2*; MW971961), (*La4CL3*; MW971962), (*La4CL4*; MW971963), (*La4CL5*; MW971964), (*La4CL6*; MW971965), (*La4CL7*; MW971966), *L. aestivum* hydroxycinnamoyltransferase 1 (*LaHCT1*; MW971967), (*LaHCT2*; MW971968), (*LaHCT3*; MW971969), (*LaHCT4*; MW971970), *L. aestivum* bifunctional p-coumarate 3-hydroxylase/ascorbate peroxidase 1 (*LaAPX/C3H1*; MW971971), (*LaAPX/C3H2*; MW971972), *L. aestivum* caffeoyl shikimate esterase 1 (*LaCES1*; MW971973), (*LaCES2*; MW971974), (*LaCES3*; MW971975), (*LaCES4*; MW971976), *L. aestivum* norbelladine synthase (*LaNBS*; MW971977), *L. aestivum* norbelladine 4'-O-methyltransferase (*LaN4OMT*; MW971978), *L. aestivum* noroxomaritidine synthase 1/CYP96T1 (*LaCYP96T1*; MW971979), (*LaCYP96T2*; MW971980), *L. aestivum* noroxomaritidine/norcraugsodine reductase 1 (*LaNorRed1*; MW971981), (*LaNorRed2*; MW971982), *L. aestivum* histone 3 *LaHIS3*; MW971983), *L. aestivum* glyceraldehyde-3-phosphate dehydrogenase (*LaGAPDH*; MW971984), and *L. aestivum* tubulin alpha chain (*LaTUBa*; MW971985). Also for *N. papyraceus* norbelladine synthase (*NpNBS1*; MZ054103 and *NpNBS2*; MZ054104).

Results

Alkaloids

To establish the presence of Amaryllidaceae alkaloids from *Leucojum aestivum*, an acid-base extraction was used to separate alkaloids from other metabolites. The alkaloid profiles from *L. aestivum* bulbs, stems, leaves and flowers were initially assessed. Thin-layer chromatography (TLC) showed qualitative differences among alkaloids from different *L. aestivum* plant parts (**Fig. S1**). We observed that all plant tissues contain various known, according to AA standards retention factor values, and unknown metabolites. Bulbs possess the largest diversity of alkaloids in comparison to other tissues, including large spots on TLC with Rf corresponding to galanthamine (Rf 0.02) and lycorine (Rf 0.13) standards. Thus, TLC analyses showed strong differences in the alkaloid profiles of the different tissues extracted from *L. aestivum* and bulbs displayed most AAs.

To quantify the alkaloids in bulbs extract, high-performance liquid chromatography (HPLC) with photodiode array (PDA) detector analyses were performed. HPLC-PDA chromatograms confirmed the presence of AAs (**Table 1**). Peaks were detected at similar retention times (Rt) as those of AA standards lycorine (Rt 4.4 min) and galanthamine (Rt 5.61 min), which correlates with our TLC data. Interestingly, six additional unknown compounds/alkaloids were detected (**Table 1**). Relative quantification of the alkaloid content was done after normalization with our papaverine internal standard, and the extraction yield was 70% (recovered 0.21 mg/g of 0.3 mg/g papaverine added; **Table 1**). The highest level of lycorine (3.89 mg/g) compared to galanthamine (1.17 mg/g) was found. In addition, alkaloids at Rt 7.6 min (3.89 mg/g) and Rt 13.1 min (5.78 mg/g) were detected at similar levels than lycorine (**Table 1**). Altogether, the results show the presence of different AA-types in *L. aestivum* bulbs.

RNA-Seq, de novo assembly and annotation of *L. aestivum* bulb's transcriptome

To better understand the AA metabolic pathway and to facilitate the discovery of genes involved in their biosynthesis, we performed RNA sequencing of *L. aestivum*. Bulbs were selected for transcriptomic analysis based on alkaloid accumulation profiles (**Table 1; Fig. S1**). RNA was extracted and used for cDNA library creation for deep sequenced using Illumina. A total of 63,351,063 raw paired reads were obtained, which were normalized using Trinity to give 57,887,920 surviving paired reads that corresponds to 91.4% (**Table 2**). For non-model Amaryllidaceae plants, whose genomic information is lacking, a *de novo* assembly was required. Normalized reads were used to assemble the transcriptome using the Trinity assembler generating 307,669 transcripts. Furthermore, BUSCO analysis revealed that *L. aestivum* bulb's transcriptome contained 85.3% of complete liliopsida odb10 orthologs. An overall alignment rate of 84.2% raw reads were mapped to *L. aestivum* transcriptome. A detailed overview of the procedure and the sequencing output, trimming, normalization, *de novo* assembly and annotation is provided (**Fig. S2; Table 2**).

The annotation was performed using the longest ORFs (207,578 total) which yielded a total of 158,155 transcripts (76.2%) homologous to known proteins or conserved hypothetical proteins were obtained from the BLAST annotation (**Table 2**). Strikingly, 49,423 (23.8%) transcripts showed no similarity with known sequences from the databases, which suggests that unknown transcripts of uncharacterized genes or sequences specific to *L. aestivum* have been uncovered. The distribution of transcript length is presented in **Fig. 2A**. The average length of the annotated transcripts was 1,510 bp with a N50 value of 1,799 bp. The GO classification of transcripts was performed to acquire functional information of transcripts at the macro level and classified into three main GO ontologies: biological process, cellular component and molecular function. The results indicated that the most highly annotated GO category was the cellular component (43.74%) with the nucleus being the largest, while only a few transcripts were associated with the plasmodesma (**Fig. 2B**). Under the category of molecular function (37.14%), ATP binding was the most abundant category. Regarding the biological process category (19.12%), the major subcategories included transcription and regulation of transcription

(**Fig. 2B**). In addition, some important categories of poor enrichment were also represented such as, transcription factor activity, protein binding and nucleic acid binding. Altogether, we concluded that a good quality transcriptome was developed with a high number of surviving pair reads and assembled transcripts.

Identification of transcripts involved in AA metabolism

Local BLASTx analyses were performed to identify specific gene transcripts encoding enzymes putatively involved in alkaloid biosynthesis (**Table 3**). From the precursor pathways leading to norbelladine (**Fig. 1**), several transcript variants of orthologous genes were identified. For example, the first steps in the formation of aromatic amino acids catalyzed by *LaDAHPS*, *LaDHQS*, *LaDHQD*, and *LaSK* are represented by 3, 2, 2, and 3 transcript variants, respectively (**Table 3**). PAL and TYDC genes have been cloned and characterized from various plants including the Amaryllidaceae species where two isoforms for each were identified (Jiang, Xia et al. 2011, Jiang, Xia et al. 2013, Singh and Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019). BLASTx searches for PAL gene transcripts in our transcriptome led to the identification of three (*LaPAL1*, *LaPAL2*, *LaPAL3*) full-length transcript variants. Each was most similar to their corresponding variants from *N. pseudonarcissus* with an E-value of zero (**Table 3**). Similarly, two full-length (*LaTYDC1*, *LaTYDC2*) gene transcripts were found for TYDC. For gene transcripts specific to AA biosynthesis, we were able to identify all four known genes (*NBS*, *N4OMT*, *NoroxSyn*, *NoroxRed*) with one full-length expressed variant for each (**Table 3**). Thus, we were able to find full-length transcripts and several isoforms with E-values corresponding to genes encoding enzymes involved in AA biosynthesis.

Interestingly, few transcripts were represented by only one sequence including *LaEPSPS*, *LaCS*, and *LaCM*, involved in the last steps for the formation of phenylalanine and tyrosine and *LaNBS*, *LaN4OMT*, *LaNoroxSyn*, and *LaNoroxRed* involved in the formation of AAs (**Fig. 1; Table 3**). These steps may be involved in important bottleneck reactions. Comparative TPM digital expression of the *L. aestivum* bulb transcriptome indicated that most transcripts were expressed at similar levels suggesting coordinated

regulation for the formation of AA precursors. However, in the central AA biosynthetic pathway, from norbelladine to oxomaritinamine, sequence reads corresponding to cytochrome P450s NoroxoSyn (CYP96T3) were most abundant among the “AA-specific” gene transcripts (**Fig. 1, Table 3**). Interestingly, *TYDC2*, was also expressed to similar level (**Table 3**). They may thus be coordinately regulated since both these genes encode enzymes catalyzing the first reactions in the AA biosynthetic pathway (**Fig. 1**). Furthermore, for the most part, the digital expression profile of the *L. aestivum* bulb is comparable to that reported for the *N. pseudonarcissus* and *N. papyraceus* bulbs (Singh and Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019). However, *LaNBS*, the first committed reaction in AA biosynthesis, appeared to be the lowest expressed transcripts among the AA specific ones. Thus, we selected to further study *LaNBS* in comparison with NBS from other Amaryllidaceae species.

PCR amplification of NBS

To further study NBS, we amplified *LaNBS* and *NpNBS* from the cDNA library generated from bulbs of *L. aestivum* and *N. papyraceus* using specific primers (**Table S2**). As expected, we obtained one PCR product for *LaNBS* but unexpectedly we amplified two products for *NpNBS* identified *NpNBS1* and *NpNBS2* with 98% nucleotide sequence identity and all of 480 bp. We pursued our investigation with these three NBS sequences: *LaNBS*, *NpNBS1*, and *NpNBS2*.

Phylogenetic analysis of norbelladine synthase

To investigate the evolutionary relationship of *LaNBS*, *NpNBS1*, and *NpNBS2* with other NBS, norcoclaurine synthase (NCS) and pathogen-related protein 10 (PR10), a phylogenetic tree was generated using NBS amino acid sequences. Even if NBS and NCS catalyze a similar reaction, which is a condensation of aldehyde and amine but with different substrates, the phylogenetic tree divided NBSs and NCSs into different clusters along with PR10 proteins forming a third cluster (**Fig. 3**). *N. papyraceus* *NpNBS1* and *NpNBS2* share higher homology with *N. pseudonarcissus* ‘King Alfred’ *NpKANBS* with

87% and 89% amino acid sequence identity, respectively. *Np*NBS1 shares 39% of sequence identity with *Papaver somniferum* *Ps*NCS1/2 and *Thalictrum flavum* *Tf*NCS, whereas *Np*NBS2 displayed 40% of identity with these proteins. *Np*NBS1/2 also has 40 and 41% of amino acid sequence identity with *Sinopodophyllum hexandrum* *Ph*NCS, respectively. The lower homology of *Np*NBS1/2 is with *Coptis japonica* *Cj*PR10A with 34 and 35 % of identity, respectively. Similarly, *L. aestivum* *La*NBS shares 86% sequence identity with *Np*KANBS, 44% with *P. somniferum* *Ps*NCS1/2, 44% with *S. hexandrum* *Ph*NCS, 43% with *T. flavum* *Tf*NCS, and displayed lower identity percentage with *C. japonica* *Cj*PR10A with 37%. Altogether, the results suggest that NBS and NCS evolved as distinctive genes over time. All NBS amino acid sequences share over 23% of identity with the PR10 protein family. Although NBS and PR10 display sequence identity between each other, NBSs are more related to the NCS group as it shows higher homology with those proteins. Altogether, the results suggest that NBS, NCS, and PR10 evolved as distinctive genes over time.

To study the evolutionary relationships among the different NBS, multiple sequences alignment was performed to identify shared regions. Thereby, *Np*KANBS, *Np*NBS1/2, and *La*NBS share over 39% of amino acid sequence identity with the ortholog norcochlorine synthase from *T. flavum* (*Tf*NCS) (**Fig. 4**). In addition, *Np*KANBS, *Np*NBS1/2, and *La*NBS share over 83% of identity between each other. The alignment showed that the catalytic residues Tyr, Glu, Lys, and Ile from NBS are well-conserved within each species (**Fig. 4**). Regarding *Tf*NCS, the alignment also showed that it shares common catalytic residues with NBS such as Tyr, Glu, and Lys (Ilari, Franceschini et al. 2009). The fourth catalytic residue Asp from *Tf*NCS was not found among NBS and was replaced by Ile as reported previously for *Np*KANBS (Singh, Massicotte et al. 2018).

Heterologous expression of *La*NBS in *E. coli*

The *La*NBS open reading frame (ORF) is 480 bp and encodes a 159-amino acid protein. Protein analysis of *La*NBS predicted a molecular weight (MW) of 17 kDa with a theoretical isoelectric point (pI) of 5.1. *La*NBS's complete ORF sequence was amplified

from cDNA of *L. aestivum* bulb to produce the recombinant *LaNBS* protein. The amplicon showed a band of 480 bp on the 1% agarose electrophoresis gel (**Fig. 5A**). The *LaNBS* PCR product was gel purified, cloned into the pMAL-c2X expression vector with a N-terminal MBP tag, transformed into *E. coli* DH5 α competent cells and the positive colonies were selected on LB supplemented with ampicillin. Colony PCR confirmed the presence of positive colonies (**Fig. 5B**). Two positive cloned plasmids were sequenced to ensure the correct sequence and transformed into *E. coli* Rosetta competent cells for protein expression. The positive colonies were selected on LB supplemented with ampicillin and chloramphenicol. Colony PCR confirmed the presence of transformed plasmid in all positive colonies (**Fig. 5C**), which were induced with IPTG at 18 °C and protein fractions were extracted. SDS-PAGE analysis was performed with crude and purified fractions obtained during *LaNBS* protein purification process (**Fig. 5D**). The crude fraction shows the expected induced band of 59 kDa (*LaNBS* 17 kDa + MBP-tag 42 kDa) *LaNBS*-MBP fusion protein and bands of different sizes, suggesting the presence of numerous bacterial cell proteins in this fraction. After purification using a high-affinity amylose resin bead, the *LaNBS*-MBP fusion protein showed a predicted molecular weight of 59 kDa (**Fig. 5D**; pure 1 and pure 2). The results showed successful cloning and heterologous production of recombinant *LaNBS* fusion protein in *E. coli*.

***LaNBS* is able to synthesize norbelladine in vitro**

To investigate *LaNBS* protein function, enzyme assays were performed using purified *LaNBS*. The resulting assay product was subjected to LC-MS/MS analysis using a Positive Electrospray Ionization mode (+ESI) for norbelladine, tyramine and papaverine and Negative (-ESI) for 3-4-DHBA. The QqQ dual +ESI source conditions were optimized using norbelladine standard (**Table S3**) to obtain high sensitivity results. The MS/MS parameters (*i.e.*, gas flow, gas temperature, nebulizer, sheath gas flow, sheath gas temperature, capillary voltage, and Nozzle voltage) were optimized to maximize the ionization in the source and the sensitivity to identify and characterize all possible fragmentation products of norbelladine and its precursors (**Table S5**). With standards, we observed predicted major mass spectral fragments for the norbelladine *m/z* 260

[M + H]⁺ at 3.435 min, for 3,4-DHBA *m/z* 137 [M - H]⁻ at 3.842 min, for tyramine *m/z* 138 [M + H]⁺ at 3.240 min, and papaverine *m/z* 340 [M + H]⁺ at 7.062 (Fig. 6A and Table S5). Fragmentation of norbelladine molecular ion *m/z* 260 [M + H]⁺ yielded ion fragments of *m/z* 138 and 121 (Table S5). The ion fragment *m/z* 138 was obtained by the elimination of the 4-ethylphenol moiety (122 Da) whereas *m/z* 121 was produced by loss of tyramine. Precisely for norbelladine, the transition [M + H]⁺ of 260→138 was selected for the qualifier ion fragment and the transition [M + H]⁺ of 260→121 was used as the quantifier ion (Table S5). Enzyme assays with *La*NBS yielded a peak at 3.435 min on LC-MS, which was the same retention time as those of the norbelladine standard (Fig. 6D). The negative control assays without substrates showed no norbelladine peak (Fig. 6B) whereas substrates without enzyme (Fig. 6C) showed very little peak close to background level suggesting a low level of spontaneous condensation. The +ESI-MS/MS analysis of the *La*NBS enzyme assay product showed the presence of the qualifier and quantifier ion fragments of norbelladine. In addition, the mass spectra obtained after the fragmentation of the product match those of the norbelladine standard (Fig. S5) showing spectral basis for the identification of the product. Our results confirm that *La*NBS catalyzes the condensation reaction between tyramine and 3,4 DHBA to produce norbelladine.

Predictions of subcellular localizations

Prior to our experiments of localization of NBS in *N. benthamiana* plants, a variety of online bioinformatics tools were used to predict signal peptides and subcellular localizations of all our NBS constructs using NBS amino acid sequences. From all the signal peptide prediction programs used, no targeting sequence such as signal peptide (SP), mitochondrial transit peptide (mTP) or chloroplast transit peptide (cTP) was predicted for *La*NBS, *Np*NBS1, and *Np*NBS2 (Table S2). Predictions obtained by both TargetP and SignalP indicated ‘other’, which means that there is no targeting sequence present in *La*NBS, *Np*NBS1, or *Np*NBS2. All three proteins were suggested to be localized in the cytosol according to Loctree3, PSLpred and WoLF PSORT programs (Table S2). The WoLF PSORT tool also predicted extracellular localization for *La*NBS, *Np*NBS1,

and *NpNBS2*. Furthermore, this program indicated a low probability of *LaNBS* localization in the cytoskeleton, and low probabilities for localizations of *NpNBS1* in the endoplasmic reticulum and *NpNBS2* in the Golgi apparatus. From all the above predictions, NBS most likely localized in the cytosol.

Cloning of NBS into pB7FWG2 vector

For localization studies in *N. benthamiana*, we prepared different constructs of NBS fused to GFP (**Fig. S3A**). In order to produce recombinant NBS, full-length cDNA of *LaNBS*, *NpNBS1* and *NpNBS2* were amplified by PCR. The amplified product of *NpNBS* shows a band of 548 bp on the agarose electrophoresis gel (**Fig. S3B**). The amplified *LaNBS* also shows a band of 548 bp on gel (**Fig. S3C**). The PCR products were gel purified, cloned into pB7FWG2 vector with a C-terminal GFP tag and transformed into competent cells. Colony PCR confirmed the presence of positive clones (**Fig. S3D**) for *LaNBS*, *NpNBS1* and *NpNBS2* with bands between 500 and 600 bp for every construct. The negative control empty vector pB7FWG2 shows no band on the gel. The results showed that NBS gene was successfully recombined in fusion with a GFP tag and confirmed by Sanger sequencing (data not shown).

Subcellular localization of GFP tagged at the C-terminus of NBS

To determine the subcellular localization of *LaNBS*, *NpNBS1* and *NpNBS2*, we used the GFP tag to the C-terminus of NBS genes. The fusion proteins were identically expressed under the control of the 35S promoter in *N. benthamiana* epidermal leaf cells, with the fungal effector *35S::124466-GFP* used as positive control for the experiment. Fungal effector 124466 is a small protein (8 kDa) which once fused to GFP (28 kDa) displayed a nucleo-cytosolic localization similar to GFP alone (Germain, Joly et al. 2018). As expected, 124466-GFP yielded its dual localization in the nuclear and the cytosol, with a stronger green fluorescence signal in the nucleus (**Fig. 7**). Similarly, the GFP signals for the NBS constructions *35S::LaNBS-GFP*, *35S::NpNBS1-GFP*, and *35S::NpNBS2-GFP* (**Fig. S3A**) occurred both in the cytosol and the nucleus of the

Agrobacterium-infiltrated *N. benthamiana* leaf cells (**Fig. 7**). The correct expression of the full-length NBS-GFP fusion proteins was further confirmed by Western blot analysis with GFP antibodies in the Agrobacterium-infiltrated *N. benthamiana* leaves. (**Fig. S4**). For all the fusion proteins, no GFP signal was recorded in the chloroplasts, specifically into the stomatal guard cell chloroplasts of the epidermal leaf cells, analyzing the autofluorescence of the chlorophyll. The cytosolic localization signals of *La*NBS, *Np*NBS1, and *Np*NBS2 appeared stronger than those of the positive control 124466 which is well-known for its dual localization. The above results demonstrated that NBSs from *L. aestivum* and *N. papyraceus*, when tagged with GFP at their C-terminal, were localized to the cytosol.

Discussion

Amaryllidaceae alkaloids are a vast and diversified group of specialized metabolites with over 650 reported structures and diverse biological activities. Additionally, these metabolites are of pharmaceutical significance (Hotchandani and Desgagne-Penix 2017, Ka, Koirala et al. 2020). Although the chemical steps in the AA biosynthesis have been established (**Fig. 1**), the genes and corresponding enzymes of the pathway remain predominantly unstudied. Recent transcriptome analysis correlating alkaloids accumulation and biosynthetic genes expression involved in AA metabolism have provided molecular insights on this pathway, including the identification and characterization of NBS, N4OMT, NoroxoSyn (CYP96T1) and NoroxoRed (Kilgore, Augustin et al. 2014, Kilgore, Augustin et al. 2016, Kilgore, Holland et al. 2016, Singh, Massicotte et al. 2018). The formation of norbelladine is a key step in the biosynthesis of AAs. The NBS enzyme, which catalyzes the condensation of tyramine with 3,4-DHBA to form norbelladine, has been partially characterized at a molecular level from wild daffodil (*Narcissus pseudonarcissus*) (Singh, Massicotte et al. 2018). The subcellular localizations of NBS are indicative of the sites of its catalytic function and therefore, in which intracellular compartments the AA are biosynthesized. In this study, we performed RNA sequencing and report *de novo* assembly of summer snowflake. We were able to identify various AA genes including NBS and we investigated the subcellular localizations of NBS

from *L. aestivum* and *N. papyraceus* using *N. benthamiana* as heterologous system for functional analysis.

L. aestivum displayed the most diverse AA profiles compared to *N. pseudonarcissus* ‘Carlton’, *Lycoris radiata* and *Ungernia victoris* (Berkov, Georgieva et al. 2009). Phytochemical studies have revealed the occurrence of about 36 alkaloids in *L. aestivum* (Berkov, Georgieva et al. 2009). We observed different alkaloid profiles within the different tissues of *L. aestivum* with higher concentration and diversity of AA in bulbs (**Table 1, Fig. S1**). These results relate to previously published data for on *N. papyraceus*, *N. pseudonarcissus* ‘king Alfred’, *N. pseudonarcissus* ‘Carlton’ and *N. confusus* where the highest amount of AAs occurred in bulbs (López, Bastida et al. 2003, Lubbe, Pomahacova et al. 2010, Berkov, Bastida et al. 2011, Lubbe, Verpoorte et al. 2012, Lubbe, Gude et al. 2013, Singh and Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019). Our alkaloid analysis showed predominantly lycorine and galanthamine which is consistent with published studies (Georgieva, Berkov et al. 2007, Berkov, Georgieva et al. 2009, El Tahchy 2010, Berkov, Georgieva et al. 2013, Ptak, Simlat et al. 2019, Ptak, Morańska et al. 2020). In addition, we also detected 6 alkaloids not identified in our study. Those could correspond to AA previously reported in *L. aestivum* such as crinine, chlidanthine, haemanthamine, demethylmaritidine, ismine, or tazettine (Berkov, Georgieva et al. 2009, Berkov, Georgieva et al. 2013). Altogether, the results show the presence of different AA-types in *L. aestivum* bulbs suggesting the presence of enzymes involved in their formation.

We generated the first *L. aestivum* transcriptome by using Illumina RNA sequencing. This method is useful to generate valuable resources for non-model plants such as Amaryllidaceae where there is lack of genomic information. The *de novo* assembly of paired reads resulted in a 273 Mbp transcriptome that yielded 307,669 assembled transcripts with a mean length of 887 bp and 158,155 of these transcripts were annotated (**Table 2, Fig. 2**). Similar numbers were obtained for *N. pseudonarcissus* transcriptome with 195 347 transcripts and a mean length of 761 bp (Singh and Desgagne-Penix 2017). In addition, *N. papyraceus* transcriptome also obtained similar results with

148 563 transcripts and a mean length of 854 bp (Hotchandani, de Villers et al. 2019) which support the quality and the depth of the transcriptome.

We were able to identify several transcripts variants for all known genes in the AA biosynthetic pathway (**Table 3**). The transcripts with the highest TPM values were tyrosine decarboxylase 2 (*LaTYDC2*), 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1 (*LaDHAPSI*), noroxomaritidine synthase 3 (*LaNoroxoSyn/CYP96T3*), and bifunctional coumarate 3-hydroxylase/ascorbate peroxidase (*LaAPX/C3'HI*). High expression of *LaNoroxoSyn/CYP96T3* in bulbs has been reported for *N. pseudonarcissus* 'King Alfred', *N. papyraceus* and *L. radiata* associated to a higher content of AAs (Singh and Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019, Park, Yeo et al. 2019). However, high expression of genes in the precursor pathway such as *TYDC* has not been shown either in *N. pseudonarcissus* 'King Alfred', nor in *N. papyraceus* (Singh and Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019). This could suggest that *TYDC* expression is not coordinately regulated with *PAL* in *L. aestivum* as it has been suggested for *Narcissus* spp. and instead it could be co-expressed with *APX/C3'HI*, although further studies are needed to confirm this assumption. About *APX/C3'HI*, it was recently reported that this non-membrane bound ascorbate peroxidase enzyme catalyzing the direct 3-hydroxylation of *p*-coumaric acid to caffeic acid (Barros et al. 2019). This path (**Fig. 1**) was proposed by (Zenk 1965) to be involved in vanillin biosynthesis, but not such genes or enzymes have been reported in Amaryllidaceae so far. Besides, the expression levels of *N4OMT* and *NBS* was lower than what has been reported for *N. pseudonarcissus* 'King Alfred'. Altogether the transcriptomic profile suggests that *L. aestivum* AAs biosynthesis is not regulated in the same manner as *Narcissus* species.

Sequences alignment of NBS proteins showed the absolute conservation of catalytic residues from the different plant species (**Fig. 4**). All NBS displayed over 39% of identity with *T. flavum* norcochlorine synthase *TfNCS* and share three common catalytic residues of the four with *TfNCS* (Ilari, Franceschini et al. 2009, Singh, Massicotte et al. 2018). Phylogenetic analysis also suggests that NBS and NCS have evolved from a common ancestor over time and share 37-44% of amino acid sequences identity between each other (**Fig. 3**). Whereas NBS lacks a signal peptide, it has been reported that *PsNCS2* for example, has an N-Terminus peptide signal and is initially localized into the endoplasmic reticulum before sorting to the vacuole (Lee and Facchini 2010). In addition, the homology between NBS and PR10 suggests that those genes have evolved differently over time, although they conserved sequence features to perform their function. The PR10 family is known to play a major role in defense mechanism against biotic and abiotic stresses and is classified into two distinct groups: intracellular pathogenesis-related proteins (IPR), which have homology with ribonucleases, and (s)-norcochlorine synthases (NCS) (Samanani, Liscombe et al. 2004, Agarwal and Agarwal 2014). Studies have shown that PR10 NCS group are distinctive of IPR proteins because they have putative N-terminal signal peptides, which indicated that they are localized in a subcellular compartment other than the cytosol, whereas IPR proteins display a cytosolic localization (Samanani, Liscombe et al. 2004, Liscombe, MacLeod et al. 2005, Liu and Ekramoddoullah 2006). The results suggest that NBS proteins have been recruited from a PR10 common ancestor as the primary evolution event that allowed Amaryllidaceae plants to produce AAs.

LaNBS enzyme catalyzes norbelladine synthesis by a condensation process between amine and aldehyde (**Fig. 6**). Among the proposed catalytic residues, the Glu⁷¹ was detected with +1 histidine shift in *LaNBS*, the same reported for *NpKANBS* (**Fig. 4**) (Singh, Massicotte et al. 2018), which might affect the interaction with catalytic and hinder the cyclization process. Moreover, unlike NCS, Amaryllidaceae NBS's ability to form norbelladine in the absence of Asp¹⁴¹ catalytic residues suggests that Asp residue is not responsible for basic catalysis and might be involved in imparting electrostatic stabilization (Lichman, Gershter et al. 2015, Singh, Massicotte et al. 2018).

We detected full-length *LaNBS* which is homologous with *NpKANBS* and *NpNBS* suggests similar action and localization site. NBS contains a conserved region, Bet v1 domain and belongs to Bet v1/PR10 protein family. Some studies have shown that Bet v1/PR10 members, such as NCS, have an important role in alkaloid biosynthesis (Samanani, Liscombe et al. 2004, Zulak, Cornish et al. 2007, Singh, Massicotte et al. 2018). We studied the subcellular localization predictions of NBS using online bioinformatics tools such as LocTree3, PSLpred and WoLF PSORT. All programs used predicted a cytoplasmic localization for *LaNBS* and *NpNBS* and no targeting sequences were predicted by SignalP, ChloroP or TargetP (**Table S2**). We then used GFP as a tag at the C-terminal (**Fig. S3**) of the NBS to not disturb any subcellular localization signals that NBS could have, since the software WoLF PSORT also predicted extracellular, Golgi and ER localizations for the studied protein. Our results showed that both predictors and microscopy imaging of the C-terminal tagging NBS indicated a cytosolic localization (**Table S2, Fig. 7**). Interestingly, as demonstrated by lower fluorescence signals, the GFP tag suggested a nuclear localization for *LaNBS*, *NpNBS1* and *NpNBS2* while none of the bioinformatic programs used predicted a subcellular localization to the nucleus (**Table S2, Fig. 7**). These results may indicate that the NBS does not contain a typical nuclear localization signal (NLS) (Fukasawa, Leung et al. 2014) but this is the result of GFP diffusion. Thereby, prediction program's databases search for similarities between the studied protein and those of typical NLS to predict a nuclear localization signal. As of today, the algorithms for NLS prediction usually relies on prior knowledge of NLS basic residues of specific species and may thus not recognize every NLS, since many of them have not been identified (Guo, Yang et al. 2020). Besides, the presence of a signal at the C-terminus of the NBS proteins could be masked by the fusion with GFP, however, we confirmed that the localization is not coming from GFP alone using Western-blot where we demonstrated the integrity of the fusion proteins NBS-GFP (**Fig. S4**). GFP may drive the localization of NBS since it has a lower molecular weight (17 kDa) compared to GFP (28 kDa). Thus, the fusion NBS-GFP can facilitate some diffusion of the GFP protein into the nucleus along with the NBS. However, various small fungal effectors (8-10 kDa) with different signal peptides, once fused to GFP in C-terminal displayed different localizations (nucleus, nucleolus, chloroplast, mitochondria, etc.) suggesting that protein

size may not influence localization (Germain, Joly et al. 2018). We cannot completely rule out the possibility of diffusion of the NBS-GFP protein inside the nucleus and cytosol. Alternatively, *in situ* immunolocalization could also help to clarify the localization of NBS. In addition, although no localization signal was predicted at the C-terminus, it remains a possibility. To further confirmed the localization of NBS, a GFP fusion at the N-terminus would need to be assayed. Indeed, the N-terminal GFP fusion of GFP-NBS is less likely to give the correct localization. It has been shown by multiple studies that signal peptides are mostly N-terminal targeting signals (*e.g.* the targeting signals for mitochondria, chloroplast and ER). In contrast, signals at the extreme C-terminal of the protein are mostly associated with peroxisomal targeting (Schatz and Dobberstein 1996, Kunze and Berger 2015). NBS is not expected to go to the peroxisome. NBS shares substantial identity with PR10 and Bet v1 proteins whose expected subcellular location is the cytosol.

A protein performs its functions once synthesized and transported to its specific subcellular localization (Huh, Falvo et al. 2003). Since *La*NBS, *Np*NBS1 and *Np*NBS2 are likely localized to the cytosol and are recognized as a major enzyme in AA biosynthesis, it suggests that NBS reaction should take place in the cytosol. Those results are correlated with the localization of the substrates used by NBS. Recently, the enzyme *La*TYDC1, which catalyzes the conversion of tyrosine to tyramine, one of the substrates for the condensation reaction catalyzed by NBS, was functionally characterized and showed to be localized at the cytosol (Wang, Han et al. 2019). Regarding 3,4-DHBA caffeic acid (its precursor) was found only in the cytosol of carrot roots (Nagahashi, Abney et al. 1996). Therefore, tyramine and 3,4-DHBA are likely to be co-localized to the cytosol where they can be condensed into norbelladine by NBS. In addition, the enzymes *O*-methyltransferase from *Lycoris aurea* (*La*OMT1) and the N4OMT from *Lycoris longituba* (*LIN*4OMT) which catalyze the *O*-Methylation of norbelladine to 4'-*O*-methylnorbelladine both localized in the cytosol whereas *La*OMT1 is additionally localized in the endosome (Sun, Wang et al. 2018, Awwad, Fantino et al. 2020). Thus, our results support that the AAs may be biosynthesized in the cytosol and/or the nucleus.

Here, we report that all three NBS proteins: *La*NBS, *Np*NBS1, and *Np*NBS2 likely localize to the cytosol, which suggests that the NBS exerts its function there. Considering that NBS is recognized as a crucial enzyme in AA metabolism (**Fig. 1**), precursors of AAs such as norbelladine, are most likely to be biosynthesized into the cytosol.

Conclusion

In this study, we generated the first summer snowflake transcriptome using RNAseq. We were able to identify 50 expressed gene transcripts involved in AA biosynthesis. Moreover, we identified and cloned NBS from the transcriptomic data of two AA producing plants *L. aestivum* and *N. papyraceus*. In addition, we characterized the activity and localization of these key enzymes necessary to produce all AAs. The identification of AA genes and the elucidation of the subcellular localization of the AA biosynthesis will facilitate the engineering of AA- and non-AA-containing plants, as well as engineering the AA biosynthetic pathway in microorganisms such as yeast, bacteria, or microalgae.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00425-021-03741-x>.

Author contributions statement LT and IDP conceived the study. LT, AMDG, BBM, SEG, AS and IDP performed the experiment, analyzed the data, help with data interpretation and manuscript preparation. LT drafted the manuscript. LT, AMDG, BBM, AS and IDP revised the manuscript. BBM and IDP supervised the research work. All authors read and approved the final manuscript.

Funding This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) award number RGPIN 05294-2014 (Discovery) to I.D-P. This work was also supported by the Canada Research Chair on plant specialized metabolism Award No 950-232164 to I.D-P. Thanks are extended to the Canadian

taxpayers and to the Canadian government for supporting the Canada Research Chairs Program.

Acknowledgements We gratefully thank Professor Hugo Germain for helpful discussion, generous comments and for providing the fungus effector 124466 used as positive control in the microscopy experiment and other lab materials. We thank Alicia Hericher for preliminary alkaloid analysis. We also thank Natacha Mérindol and Mélodie B. Plourde for the revision of the manuscript and helping on confocal imaging, respectively.

Availability of data and materials The sequencing datasets supporting the conclusions of this article are available in the NCBI Sequence Read Archive (SRA) repository, accession number PRJNA720900 under the following link (<https://www.ncbi.nlm.nih.gov/sra/PRJNA720900>).

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Agarwal P, Agarwal PK (2014) Pathogenesis related-10 proteins are small, structurally similar but with diverse role in stress signaling. *Mol Biol Rep* 41 (2):599-611. doi:10.1007/s11033-013-2897-4
- Akula R, Ravishankar GA (2011) Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signaling & Behavior* 6 (11):1720-1731. doi:10.4161/psb.6.11.17613
- Bastida J, Berkov S, Torras L, Pigni NB, de Andrade JP, Martínez V, Codina C, Viladomat F (2011) Chemical and biological aspects of Amaryllidaceae alkaloids. In: *Recent Advances in Pharmaceutical Sciences*. Transworld Research Network

- Berkov S, Bastida J, Viladomat F, Codina C (2011) Development and validation of a GC-MS method for rapid determination of galanthamine in *Leucojum aestivum* and *Narcissus* ssp.: a metabolomic approach. *Talanta* 83 (5):1455-1465. doi:10.1016/j.talanta.2010.11.029
- Berkov S, Georgieva L, Kondakova V, Atanassov A, Viladomat F, Bastida J, Codina C (2009) Plant Sources of Galanthamine: Phytochemical and Biotechnological Aspects. *Biotechnology & Biotechnological Equipment* 23 (2):1170-1176. doi:10.1080/13102818.2009.10817633
- Berkov S, Georgieva L, Kondakova V, Viladomat F, Bastida J, Atanassov A, Codina C (2013) The geographic isolation of *Leucojum aestivum* populations leads to divergence of alkaloid biosynthesis. *Biochemical Systematics and Ecology* 46 (0):152-161. doi:<http://dx.doi.org/10.1016/j.bse.2012.10.002>
- Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczeniński MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology* 17 (1):13. doi:10.1186/s13059-016-0881-8
- Croteau R, Kutchan TM, Lewis NG (2000) Natural Products (Secondary Metabolites). *Biochemistry and Molecular Biology of Plants* 24:1250-1318
- de Andrade JP, Pigni NB, Torras-Claveria L, Berkov S, Codina C, Viladomat F, Bastida J (2012) Bioactive alkaloid extracts from *Narcissus broussonetii*: mass spectral studies. *Journal of pharmaceutical and biomedical analysis* 70:13-25. doi:10.1016/j.jpba.2012.05.009
- Desgagne-Penix I (2021) Biosynthesis of alkaloids in Amaryllidaceae plants: a review. *Phytochemistry Reviews* 20:409-431. doi:10.1007/s11101-020-09678-5
- El Tahchy A (2010) Étude de la voie de biosynthèse de la galanthamine chez *Leucojum aestivum* L. – Criblage phytochimique de quelques Amaryllidaceae. Nancy Université
- Felsenstein J (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 39 (4):783-791. doi:10.1111/j.1558-5646.1985.tb00420.x
- Fukasawa Y, Leung RK, Tsui SK, Horton P (2014) Plus ça change—evolutionary sequence divergence predicts protein subcellular localization signals. *BMC Genomics* 15:46. doi:10.1186/1471-2164-15-46
- Furst R (2016) Narciclasine—an Amaryllidaceae Alkaloid with Potent Antitumor and Anti-Inflammatory Properties. *Planta Med* 82 (16):1389-1394. doi:10.1055/s-0042-115034

- Georgieva L, Berkov S, Kondakova V, Bastida J, Viladomat F, Atanassov A, Codina C (2007) Alkaloid variability in *Leucojum aestivum* from wild populations. *Z Naturforsch C J Biosci* 62 (9-10):627-635. doi:10.1515/znc-2007-9-1002
- Germain H, Joly DL, Mireault C, Plourde MB, Letanneur C, Stewart D, Morency MJ, Petre B, Duplessis S, Seguin A (2018) Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Mol Plant Pathol* 19 (1):191-200. doi:10.1111/mpp.12514
- Guo Y, Yang Y, Huang Y, Shen HB (2020) Discovering nuclear targeting signal sequence through protein language learning and multivariate analysis. *Anal Biochem* 591:113565. doi:10.1016/j.ab.2019.113565
- Heinrich M, Lee Teoh H (2004) Galanthamine from snowdrop—the development of a modern drug against Alzheimer's disease from local Caucasian knowledge. *J Ethnopharmacol* 92 (2-3):147-162. doi:10.1016/j.jep.2004.02.012
- Hotchandani T, de Villers J, Desgagne-Penix I (2019) Developmental Regulation of the Expression of Amaryllidaceae Alkaloid Biosynthetic Genes in *Narcissus papyraceus*. *Genes* 10 (8):594. doi:10.3390/genes10080594
- Hotchandani T, Desgagne-Penix I (2017) Heterocyclic Amaryllidaceae Alkaloids: Biosynthesis and Pharmacological Applications. *Current topics in medicinal chemistry* 17 (4):418-427
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* 425 (6959):686-691. doi:10.1038/nature02026
- Ilari A, Franceschini S, Bonamore A, Arengi F, Botta B, Macone A, Pasquo A, Bellucci L, Boffi A (2009) Structural basis of enzymatic (S)-norcoclaurine biosynthesis. *The Journal of biological chemistry* 284 (2):897-904. doi:10.1074/jbc.M803738200
- Jiang Y, Xia B, Liang L, Li X, Xu S, Peng F, Wang R (2013) Molecular and analysis of a phenylalanine ammonia-lyase gene (LrPAL2) from *Lycoris radiata*. *Molecular biology reports* 40 (3):2293-2300. doi:10.1007/s11033-012-2310-8
- Jiang Y, Xia N, Li X, Shen W, Liang L, Wang C, Wang R, Peng F, Xia B (2011) Molecular cloning and characterization of a phenylalanine ammonia-lyase gene (LrPAL) from *Lycoris radiata*. *Molecular biology reports* 38 (3):1935-1940. doi:10.1007/s11033-010-0314-9
- Jin Z, Yao G (2019) Amaryllidaceae and Scetium alkaloids. *Nat Prod Rep* 36 (10):1462-1488. doi:10.1039/c8np00055g

- Ka S, Koirala M, Mérindol N, Desgagne-Penix I (2020) Biosynthesis and biological activities of newly discovered Amaryllidaceae alkaloids. *Molecules* 25 (21):4901. doi:doi:10.3390/molecules25214901
- Karimi M, Inze D, Depicker A (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci* 7 (5):193-195. doi:10.1016/s1360-1385(02)02251-3
- Kilgore M, Augustin MM, May GD, Crow JA, Kutchan TM (2016a) CYP96T1 of *Narcissus* sp. aff. *pseudonarcissus* Catalyzes Formation of the Para-Para' C-C Phenol Couple in the Amaryllidaceae Alkaloids. *Frontiers in Plant Science* 7. doi:10.3389/fpls.2016.00225
- Kilgore M, Holland C, Jez JM, Crow JA, Kutchan TM (2016b) Identification of a Noroxomaritidine Reductase with Amaryllidaceae Alkaloid Biosynthesis Related Activities. *Journal of Biological Chemistry*
- Kilgore MB, Augustin MM, Starks CM, O'Neil-Johnson M, May GD, Crow JA, Kutchan TM (2014) Cloning and Characterization of a Norbelladine 4'-O-Methyltransferase Involved in the Biosynthesis of the Alzheimer's Drug Galanthamine in *Narcissus* sp. aff. *pseudonarcissus*. *PloS one* 9 (7):e103223. doi:10.1371/journal.pone.0103223
- Kilgore MB, Kutchan TM (2016) The Amaryllidaceae alkaloids: biosynthesis and methods for enzyme discovery. *Phytochemistry reviews: proceedings of the Phytochemical Society of Europe* 15 (3):317-337. doi:10.1007/s11101-015-9451-z
- Kornienko A, Evidente A (2008) Chemistry, biology, and medicinal potential of narciclasine and its congeners. *Chemical reviews* 108 (6):1982-2014. doi:10.1021/cr078198u
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35 (6):1547-1549. doi:10.1093/molbev/msy096
- Kunze M, Berger J (2015) The similarity between N-terminal targeting signals for protein import into different organelles and its evolutionary relevance. *Frontiers in Physiology* 6 (259). doi:10.3389/fphys.2015.00259
- Lamoral-Theys D, Andolfi A, Van Goietsenoven G, Cimmino A, Le Calve B, Wauthoz N, Megalizzi V, Gras T, Bruyere C, Dubois J, Mathieu V, Kornienko A, Kiss R, Evidente A (2009) Lycorine, the main phenanthridine Amaryllidaceae alkaloid, exhibits significant antitumor activity in cancer cells that display resistance to proapoptotic stimuli: an investigation of structure-activity relationship and mechanistic insight. *J Med Chem* 52 (20):6244-6256. doi:10.1021/jm901031h

- Lee EJ, Facchini P (2010) Norcochlorine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. *The Plant cell* 22 (10):3489-3503. doi:10.1105/tpc.110.077958
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi:10.1186/1471-2105-12-323
- Lichman BR, Gershater MC, Lamming ED, Pesnot T, Sula A, Keep NH, Hailes HC, Ward JM (2015) 'Dopamine-first' mechanism enables the rational engineering of the norcochlorine synthase aldehyde activity profile. *The FEBS journal* 282 (6):1137-1151
- Liscombe DK, MacLeod BP, Loukanina N, Nandi OI, Facchini PJ (2005) Evidence for the monophyletic evolution of benzyloisoquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry* 66 (11):1374-1393. doi:<http://dx.doi.org/10.1016/j.phytochem.2005.04.029>
- Liu J-J, Ekramoddoullah AKM (2006) The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses. *Physiological and Molecular Plant Pathology* 68 (1-3):3-13. doi:<http://dx.doi.org/10.1016/j.pmpp.2006.06.004>
- López S, Bastida J, Viladomat F, Codina C (2003) Galanthamine pattern in *Narcissus confusus* plants. *Planta medica* 69 (12):1166-1168.
- Lubbe A, Gude H, Verpoorte R, Choi YH (2013) Seasonal accumulation of major alkaloids in organs of pharmaceutical crop *Narcissus Carlton*. *Phytochemistry* 88:43-53. doi:10.1016/j.phytochem.2012.12.008
- Lubbe A, Pomahacova B, Choi YH, Verpoorte R (2010) Analysis of metabolic variation and galanthamine content in *Narcissus* bulbs by ¹H NMR. *Phytochemical analysis : PCA* 21 (1):66-72. doi:10.1002/pca.1157
- Lubbe A, Verpoorte R, Choi YH (2012) Effects of fungicides on galanthamine and metabolite profiles in *Narcissus* bulbs. *Plant physiology and biochemistry: Plant Physiology and Biochemistry* 58:116-123. doi:10.1016/j.plaphy.2012.06.022
- Mbandi S, Hesse U, Rees DJ, Christoffels A (2014) A glance at quality score: implication for de novo transcriptome reconstruction of Illumina reads. *Frontiers in Genetics* 5 (17). doi:10.3389/fgene.2014.00017
- Nagahashi G, Abney GD, Doner LW (1996) A comparative study of phenolic acids associated with cell walls and cytoplasmic extracts of host and non-host roots for AM fungi. *New Phytol* 133 (2):281-288. doi:10.1111/j.1469-8137.1996.tb01895.x

- Nair JJ, van Staden J (2013) Pharmacological and toxicological insights to the South African Amaryllidaceae. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* 62C:262-275. doi:10.1016/j.fct.2013.08.042
- Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York,
- Park CH, Yeo HJ, Park YE, Baek SA, Kim JK, Park SU (2019) Transcriptome Analysis and Metabolic Profiling of *Lycoris Radiata*. *Biology (Basel)* 8 (3). doi:10.3390/biology8030063
- Ptak A, Morańska E, Skrzypek E, Warchoń M, Spina R, Laurain-Mattar D, Simlat M (2020) Carbohydrates stimulated Amaryllidaceae alkaloids biosynthesis in *Leucojum aestivum* L. plants cultured in RITA® bioreactor. *PeerJ* 8 (e8688). doi:<https://doi.org/10.7717/peerj.8688>
- Ptak A, Simlat M, Morańska E, Skrzypek E, Warchoń M, Tarakemeh A, Laurain-Mattar D (2019) Exogenous melatonin stimulated Amaryllidaceae alkaloid biosynthesis in in vitro cultures of *Leucojum aestivum* L. *Industrial Crops and Products* 138:111458. doi:<https://doi.org/10.1016/j.indcrop.2019.06.021>
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4 (4):406-425. doi:10.1093/oxfordjournals.molbev.a040454
- Samanani N, Liscombe DK, Facchini PJ (2004) Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. *The Plant journal: for cell and molecular biology* 40 (2):302-313. doi:10.1111/j.1365-3113X.2004.02210.x
- Schatz G, Dobberstein B (1996) Common principles of protein translocation across membranes. *Science* 271 (5255):1519-1526. doi:10.1126/science.271.5255.1519
- Singh A, Desgagne-Penix I (2017) Transcriptome and metabolome profiling of *Narcissus pseudonarcissus* 'King Alfred' reveal components of Amaryllidaceae alkaloid metabolism. *Nature Scientific Reports* 7 (1):17356-17370. doi:10.1038/s41598-017-17724-0
- Singh A, Massicotte MA, Garand A, Tousignant L, Ouellette V, Berube G, Desgagne-Penix I (2018) Cloning and characterization of norbelladine synthase catalyzing the first committed reaction in Amaryllidaceae alkaloid biosynthesis. *BMC Plant Biol* 18 (1):338. doi:10.1186/s12870-018-1570-4

- Sun B, Wang P, Wang R, Li Y, Xu S (2018) Molecular Cloning and Characterization of a meta/para-O-Methyltransferase from *Lycoris aurea*. *International Journal of Molecular Sciences* 19 (7). doi:10.3390/ijms19071911
- TA. H (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT. *Nucleic Acids Symposium Series* 41.
- Wang R, Han X, Xu S, Xia B, Jiang Y, Xue Y, Wang R (2019) Cloning and characterization of a tyrosine decarboxylase involved in the biosynthesis of galanthamine in *Lycoris aurea*. *PeerJ* 7:e6729. doi:10.7717/peerj.6729
- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM (2017) BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* 35 (3):543-548. doi:10.1093/molbev/msx319
- Zenk M (1965) Biosynthese von vanillin in *Vanilla planifolia* Andr. *Z. Pflanzen Physiol* 53:404-414.
- Zulak KG, Cornish A, Daskalchuk TE, Deyholos MK, Goodenowe DB, Gordon PMK, Klassen D, Pelcher LE, Sensen CW, Facchini PJ (2007) Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism. *Planta* 225 (5):1085-1106. doi:10.1007/s00425-006-0419-5

Tables

Table 1 Alkaloids identified in *L. aestivum* bulbs using HPLC analysis. The extraction yield was of 70% i.e. we recovered 0.21 mg/g of fresh weight of 0.3 mg/g papaverine added. Data are presented as the mean \pm standard deviation (n = 3). Rt = Retention time.

Alkaloids	Concentration (mg equivalent papaverine/g tissues)
Lycorine (Rt 4.4)	3.89 \pm 1.04
Galanthamine (Rt 5.1)	1.17 \pm 0.25
Rt 5.4	0.75 \pm 0.14
Rt 5.9	0.16 \pm 0.05
Rt 6.3	0.03 \pm 0.01
Rt 7.6	3.04 \pm 0.76
Rt 9.4	0.75 \pm 0.12
Rt 13.1	5.78 \pm 1.56
Papaverine (Rt 17.8)	0.21 \pm 0.04

Table 2 Summary of the Illumina sequencing output and assembly overview of *L. aestivum* transcriptome

Normalization	
Raw Paired End Reads ^a	63 351 063
Total number of normalized paired end reads ^b	57 887 920
Surviving Paired Reads after normalization ^c (%)	91.4
Trinity de novo Assembly	
Nb. Transcripts ^d	307 669
Nb. Components ^d	131 408
Total Transcripts Length (bp)	273 047 423
Median Transcript Length (bp)	548
Mean Transcript Length (bp)	887.47
N50 (bp)	1 424
BLAST annotation	
Nb. Longest ORFs	207 578
Nb. Annotated Transcripts ^e	158 155

^a Number (Nb.) of Paired End Reads generated with Illumina HiSeq 2000

^b Number of remaining Paired End Reads after the normalization step

^c Percentage of Surviving Paired Reads after normalization/raw paired end reads

^d Trinity has created a list of transcripts (contigs) representing the transcriptome isoforms. The transcripts are grouped in components loosely representing genes. Transcript names are prefixed by the component/gene name e.g. transcripts c115_i1 and c115_i2 are derived from the same isolated de Bruijn graph and therefore share the same component/gene number c115

^e Each transcript has been aligned against the uniprot_sprot.trinotate_v2.0.pep protein database using blastx program from the NCBI BLAST family

Table 3 Summary of the biosynthetic genes identified from the *Leucojum aestivum* transcriptome proposed to be involved in Amaryllidaceae alkaloids metabolism with the normalized transcript per million (TPM), the size of the open reading frame (ORF) in nucleotide (nt) and the corresponding top annotation. Each transcript sequence was deposited in GenBank

Name	TPM	ORF (nt)	Top Annotation	Species of top annotation	E value	Accession number
<i>LaDHAPS1</i>	217.77	1593	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1	<i>Solanum tuberosum</i>	0	P21357.2
<i>LaDHAPS2</i>	9.63	1557	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 2	<i>Solanum lycopersicum</i>	0	P37216.1
<i>LaDHAPS3</i>	11.18	1464	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1	<i>Petunia × hybrida</i>	0	A0A067XH53.1
<i>LaDHQS1</i>	25.79	1302	3-dehydroquinate synthase	<i>Actinidia chinensis</i>	0	U3KRF2.3
<i>LaDHQS2</i>	39.06	1197	3-dehydroquinate synthase	<i>Methanospaera stadtmanae</i>	3e-78	Q2NI00.1
<i>LaDHQD1</i>	0.88	1554	Bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase	<i>Arabidopsis thaliana</i>	0	Q9SQT8.1
<i>LaDHQD2</i>	15.15	1554	Bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase	<i>Arabidopsis thaliana</i>	0	Q9SQT8.1
<i>LaSK1</i>	0.59	927	Shikimate kinase 2	<i>Oryza sativa</i>	1e-120	Q5NTH3.1
<i>LaSK2</i>	0.18	1047	Shikimate kinase 2	<i>Oryza sativa</i>	5e-119	Q5NTH3.1
<i>LaSK3</i>	1.14	1179	Probable inactive Shikimate kinase-like 2	<i>Arabidopsis thaliana</i>	6e-145	O82290.2
<i>LaEPSPS</i>	8.66	1554	5-enolpyruvylshikimate-3-phosphate synthase	<i>Arabidopsis thaliana</i>	0	P05466.3
<i>LaCS</i>	34.38	1314	Chorismate synthase 1	<i>Solanum lycopersicum</i>	0	Q42884.1
<i>LaCM</i>	3.46	927	Chorismate mutase 1	<i>Petunia × hybrida</i>	8e-136	D2CSU4.1
<i>LaPAT1</i>	11.22	1356	Bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase	<i>Petunia × hybrida</i>	0	E9L7A5.1

Name	TPM	ORF (nt)	Top Annotation	Species of top annotation	E value	Accession number
<i>LaPAT2</i>	0.57	1272	Bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase	<i>Petunia × hybrida</i>	0	E9L7A5.1
<i>LaADT1</i>	0.82	1188	Arogenate dehydratase/prephenate dehydratase1	<i>Arabidopsis thaliana</i>	1e-163	Q9SA96.1
<i>LaADT2</i>	11.21	1152	Arogenate dehydratase 2	<i>Petunia × hybrida</i>	1e-169	D3U716.1
<i>LaADT3</i>	5.24	1254	Arogenate dehydratase 3	<i>Petunia × hybrida</i>	0	D3U717.1
<i>LaADH1</i>	14.36	1188	Arogenate dehydrogenase 2	<i>Arabidopsis thaliana</i>	6e-124	Q9LMR3.1
<i>LaADH2</i>	3.04	1182	Arogenate dehydrogenase 2	<i>Arabidopsis thaliana</i>	2e-122	Q9LMR3.1
<i>LaADH3</i>	8.76	990	Arogenate dehydrogenase 2	<i>Arabidopsis thaliana</i>	9e-130	Q9LMR3.1
<i>LaPAL1</i>	0.88	2127	Phenylalanine ammonia-lyase 1	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIW0.1
<i>LaPAL2</i>	19.62	2133	Phenylalanine ammonia-lyase 2	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIY6.1
<i>LaPAL3</i>	2.47	2133	Phenylalanine ammonia-lyase 1	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIW0.1
<i>LaC4H1</i>	32.21	1518	Cinnamic acid 4-hydroxylase	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIX7.1
<i>LaC4H2</i>	2.8	1518	Cinnamic acid 4-hydroxylase	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIX7.1
<i>LaC3H</i>	9.96	1719	<i>p</i> -coumarate 3-hydroxylase	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIX6.1
<i>La4CL1</i>	0.64	1650	4-coumarate:CoA ligase 1	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIX5.1
<i>La4CL2</i>	11.24	1689	4-coumarate:CoA ligase 2	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIY4.1
<i>La4CL3</i>	11.66	1626	4-coumarate:CoA ligase-like 7	<i>Arabidopsis thaliana</i>	0	Q9M0X9.1
<i>La4CL4</i>	4.38	1332	4-coumarate:CoA ligase-like 4	<i>Oryza sativa</i>	0	Q10S72.1
<i>La4CL5</i>	0.38	1365	4-coumarate:CoA ligase-like 5	<i>Arabidopsis thaliana</i>	6e-124	Q84P21.2
<i>La4CL6</i>	9.34	1680	4-coumarate:CoA ligase-like 7	<i>Oryza sativa</i>	0	Q69RG7.1
<i>La4CL7</i>	0.9	1641	4-coumarate:CoA ligase-like 9	<i>Arabidopsis thaliana</i>	0	Q84P23.2

Name	TPM	ORF (nt)	Top Annotation	Species of top annotation	E value	Accession number
<i>LaHCT1</i>	10.39	1311	Hydroxycinnamoyltransferase	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIZ1.1
<i>LaHCT2</i>	2.27	1287	Putrescine hydroxycinnamoyltransferase 1	<i>Oryza sativa</i>	1e-154	Q5SMM8.1
<i>LaHCT3</i>	3.65	1386	Putrescine hydroxycinnamoyltransferase 1	<i>Oryza sativa</i>	7e-103	Q5SMM8.1
<i>LaHCT4</i>	2.27	1329	Shikimate O-hydroxycinnamoyltransferase	<i>Nicotiana tabacum</i>	1e-102	Q8GSM7.1
<i>LaCES1</i>	15.69	978	Caffeoyl shikimate esterase	<i>Arabidopsis thaliana</i>	4e-165	Q9C942.1
<i>LaCES2</i>	1.96	1053	Caffeoyl shikimate esterase	<i>Arabidopsis thaliana</i>	2e-66	Q9C942.1
<i>LaCES3</i>	0.08	1035	Caffeoyl shikimate esterase	<i>Arabidopsis thaliana</i>	7e-61	Q9C942.1
<i>LaCES4</i>	2.61	1026	Caffeoyl shikimate esterase	<i>Arabidopsis thaliana</i>	8e-71	Q9C942.1
<i>LaAPX/C3'H1</i>	152.8	759	L- ascorbate peroxidase	<i>Pisum sativum</i>	4e-161	P48534.2
<i>LaAPX/C3'H2</i>	3.74	747	L- ascorbate peroxidase	<i>Pisum sativum</i>	1e-158	P48534.2
<i>LaTYDC1</i>	9.77	1536	Tyrosine decarboxylase 1	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIY0.1
<i>LaTYDC2</i>	227.28	1503	Tyrosine decarboxylase 2	<i>Narcissus pseudonarcissus</i>	0	A0A2H5AIY2.1
<i>LaNBS</i>	11.3	501	Norbelladine synthase	<i>Narcissus pseudonarcissus</i>	1e-95	A0A3G5BB24.1
<i>LaN4OMT</i>	38.14	720	Norbelladine 4'-O-methyltransferase	<i>Narcissus pseudonarcissus</i>	3e-165	A0A2H5AIZ6.1
<i>LaNoroxSyn1</i>	202.77	1539	Noroxomaritidine synthase 3 (CYP96T3)	<i>Narcissus pseudonarcissus</i>	0	A0A140IL92.1
<i>LaNoroxRed1</i>	48.31	816	Noroxomaritidine/norcroagsodine reductase	<i>Narcissus pseudonarcissus</i>	2e-138	A0A1A9TAK5.1

Figure legends

Fig. 1 Proposed Schematic of the shikimate and Amaryllidaceae alkaloids (AAs)

biosynthetic pathways. A solid arrow represents one enzymatic reaction whereas a dashed arrow represents multiple enzymatic steps. Enzymes encoded in the *Leucojum aestivum* transcriptome are in blue and question mark indicates unknown enzyme for the reaction. Abbreviations are: PEP, phosphoenolpyruvate; E4P, erythrose 4-phosphate; DAHPS, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase; DHQS, 3-dehydroquininate synthase; DHQD/SD, bifunctional 3-dehydroquininate dehydratase/shikimate dehydrogenase; SK, shikimate kinase; EPSPS, 5-enolpyruvylshikimate-3-phosphate synthase; CS, chorismate synthase; CM, chorismate mutase; PAT, bifunctional aspartate aminotransferase and glutamate/aspartate-prephenate aminotransferase; ADT, arogenate dehydratase; ADH, arogenate dehydrogenase; TYDC, tyrosine decarboxylase; PAL, phenylalanine ammonia lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; HCT, shikimate O-hydroxycinnamoyltransferase; C3H, *p*-coumarate 3-hydroxylase; CSE, caffeoyl shikimate esterase; C3'H/APX, bifunctional coumarate 3-hydroxylase/ascorbate peroxidase; NBS, norbelladine synthase; N4OMT, norbelladine 4'-O-methyltransferase; NoroxSyn, noroxomaritidine synthase/CYP96T1; and NoroxRed, noroxomaritidine/norcraugsodine reductase. *The formation of norbelladine can also be catalyzed by NoroxRed with 400X less fold efficacy than noroxomaritidine reduction.

Fig. 2 Size distribution and gene ontology analysis of the assembled transcripts.

(a) Assembled transcripts were evaluated for their size distribution and the number of transcripts in each size range is presented. **(b)** Gene ontology (GO) terms categorization of annotated *L. aestivum* transcripts on the basis of cellular components, molecular functions and biological processes. The numbers of transcripts in the ten most enriched GO terms for each of the categories are shown.

Fig. 3 Phylogenetic analysis among several NBS, NCS and PR10 proteins from a

variety of plants. Amino acid sequences for *Leucojum aestivum* LaNBS,

Narcissus papyraceus NpNBS1, NpNBS2, *Narcissus pseudonarcissus* “King Alfred” NpKANBS, *Papaver somniferum* PsNCS1, PsNCS2, *Thalictrum flavum* TfNCS, *Coptis japonica* CjPR10A, *Sinopodophyllum hexandrum* PhNCS, *Papaver bracteatum* PbNCS, *Pinus monticola* PmPR10-1, PmPR10-2, *Triticum turgidum* subsp. *Durum* TdPR10, *Triticum aestivum* TaPR10, *Zea mays* ZmPR10, ZmPR10.1, *Gossypium barbadense* GbPR10, *Arachis hypogaea* AhPR10, *Capsicum annuum* CaPR10, *Lupinus albus* LaPR10, *Hypericum perforatum* HYP-1 and *Solanum virginianum* SsPR10 were aligned and analyzed for phylogenetic relationships using the Neighbor-joining algorithm (Saitou and Nei 1987). Accession numbers are provided in subsection 2.11. The numbers refer to the bootstrap values for each branch point with 500 iterations. The optimal tree with the sum of branch length = 4.38852771 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of amino acid substitutions per site. This analysis involved 22 amino acid sequences. All ambiguous positions were removed. There were a total of 236 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar, Stecher et al. 2018).

Fig. 4 Alignment from the amino acid sequences of *LaNBS*, *NpNBS1/2*, *NpKANBS* and *TfNCS* using CLUSTAL W algorithm. Abbreviations are provided in the methods. Red circles represent the Tyr, Glu, Lys and Ile residues that form the catalytic site of NBS from different species.

Fig. 5 PCR amplification, cloning and heterologous expression of *LaNBS* in *E. coli*. (a) Agarose gel electrophoresis of PCR amplification of *LaNBS* ORF from two *L. aestivum* bulb cDNA library. The amplicon size is 480 bp. The sizes of the 1 kb ladder are positioned on the left. (b) Agarose gel electrophoresis of colony PCR amplification of *LaNBS* performed with 8 transformed *E. coli* DH5 α bacterial colonies (colony 1 to colony 8) to screen the positive clones. Colony number 1, 2, 5, and 7 shows a positive amplification of *LaNBS* ORF size of 480 bp. Positive control (C-positive) is the

PCR amplification from cDNA. The sizes of the 1 kb ladder are positioned on the left. **(c)** Agarose gel electrophoresis of colony PCR amplification of *LaNBS* performed with 4 positively transformed *E. coli* Rosetta (DE3) pLysS bacterial colonies (colony 1 to colony 4) shows a band of *LaNBS* ORF size of 480 bp. Four pMAL-c2X-empty plasmid transformed *E. coli* Rosetta bacterial colonies (colony 1 to colony 4) were used as negative control. The Positive control (C-positive) is the PCR amplification from cDNA. The sizes of the 1 kb ladder are positioned on the left. **(d)** SDS-PAGE analysis of *LaNBS*-MBP fusion protein. The lanes showed the crude extracted or purified proteins. Purified protein was extracted from 0.25 mM IPTG induced *E. coli* Rosetta (DE3) pLysS host strain at 18 °C for 20 h and purified proteins were eluted with 15 mM of maltose. The *LaNBS*-MBP fusion protein size is 59 kDa. Numbers on the left refer to the location of standard protein molecular weight markers in kDa.

Fig. 6 Extracted ion chromatograms showing the substrates and product of *LaNBS* enzyme assays. The tested substrates used were tyramine (10 µM) and 3,4 dihydroxybenzaldehyde (300 µM). The extracted ion chromatogram corresponds to the mixed standards **(a)**, or assays conducted without substrates **(b)**, without the recombinant *LaNBS* enzyme **(c)**, and the complete assay performed with recombinant *LaNBS* **(d)**. Parent ion mass-to-charge (m/z) of 260 for norbelladine, m/z 137 for 3,4-DHBA, m/z 138 for tyramine, and m/z 340 for papaverine were subjected to collision-induced dissociation analysis for identification and quantitation **(Supplementary Table 3)**.

Fig. 7 Subcellular localization of NBS. Confocal GFP fluorescence scanning of C-terminal tagged gene NBS in *Agrobacterium* infiltrated *N. benthamiana* leaf epidermal cells by agro-infiltration, and 2 days post infiltration, the cells were assessed for fluorescence under Leica TCS SP8 confocal laser scanning microscopy. For control, the upper panel shows fluorescence from GFP using a fungal effector construction, *35S::124466-GFP*, known to display a nucleocytoplasmic localization (Germain, Joly et al. 2018). All NBS-GFP fusion, GFP tagged at the C-terminus, displayed a nucleocytoplasmic localization. GFP-Gene; Bar = 30 µm.

Fig. 1 Proposed Schematic of the shikimate and Amaryllidaceae alkaloids (AAs) biosynthetic pathways.

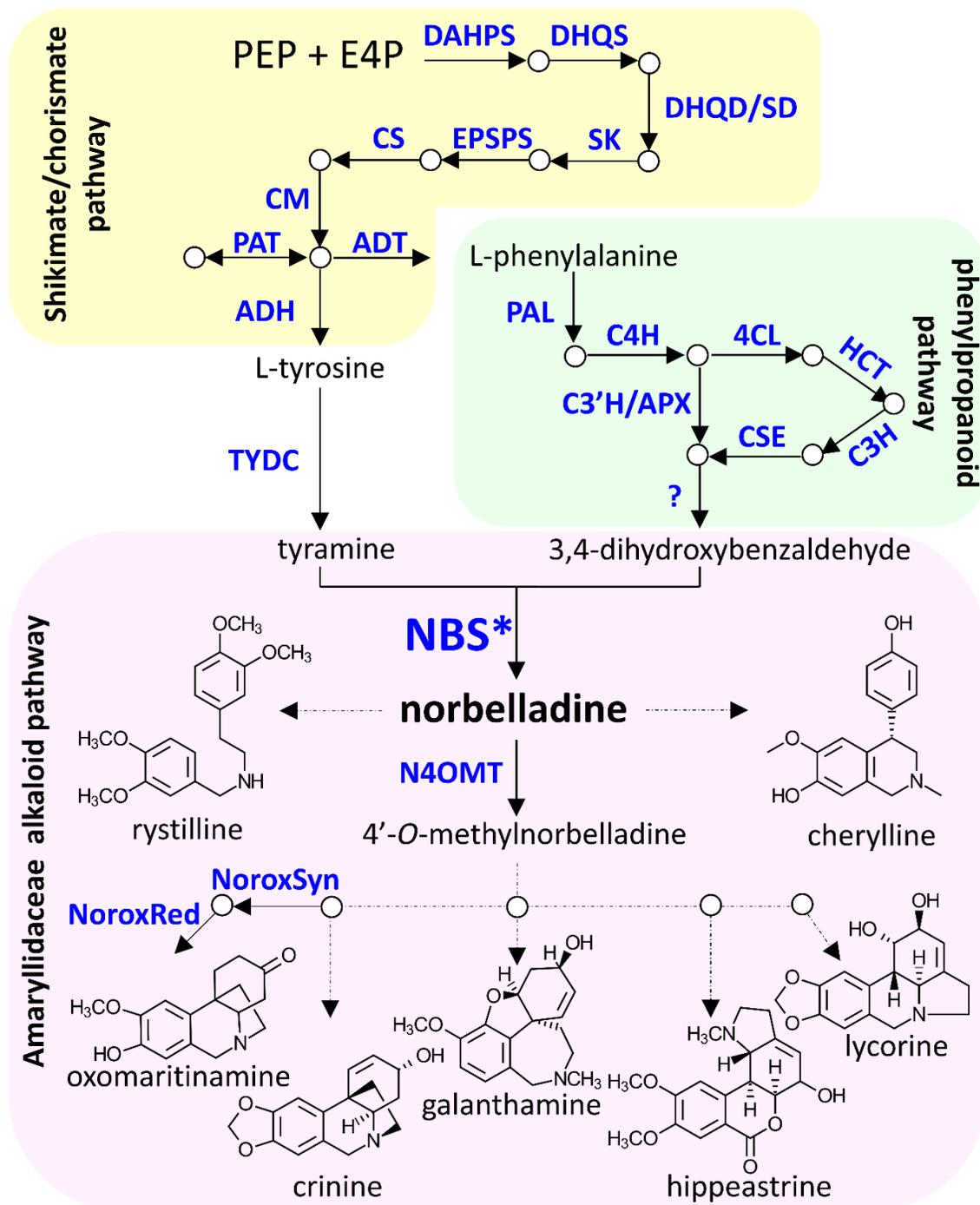


Fig. 2 Size distribution and gene ontology analysis of the assembled transcripts.

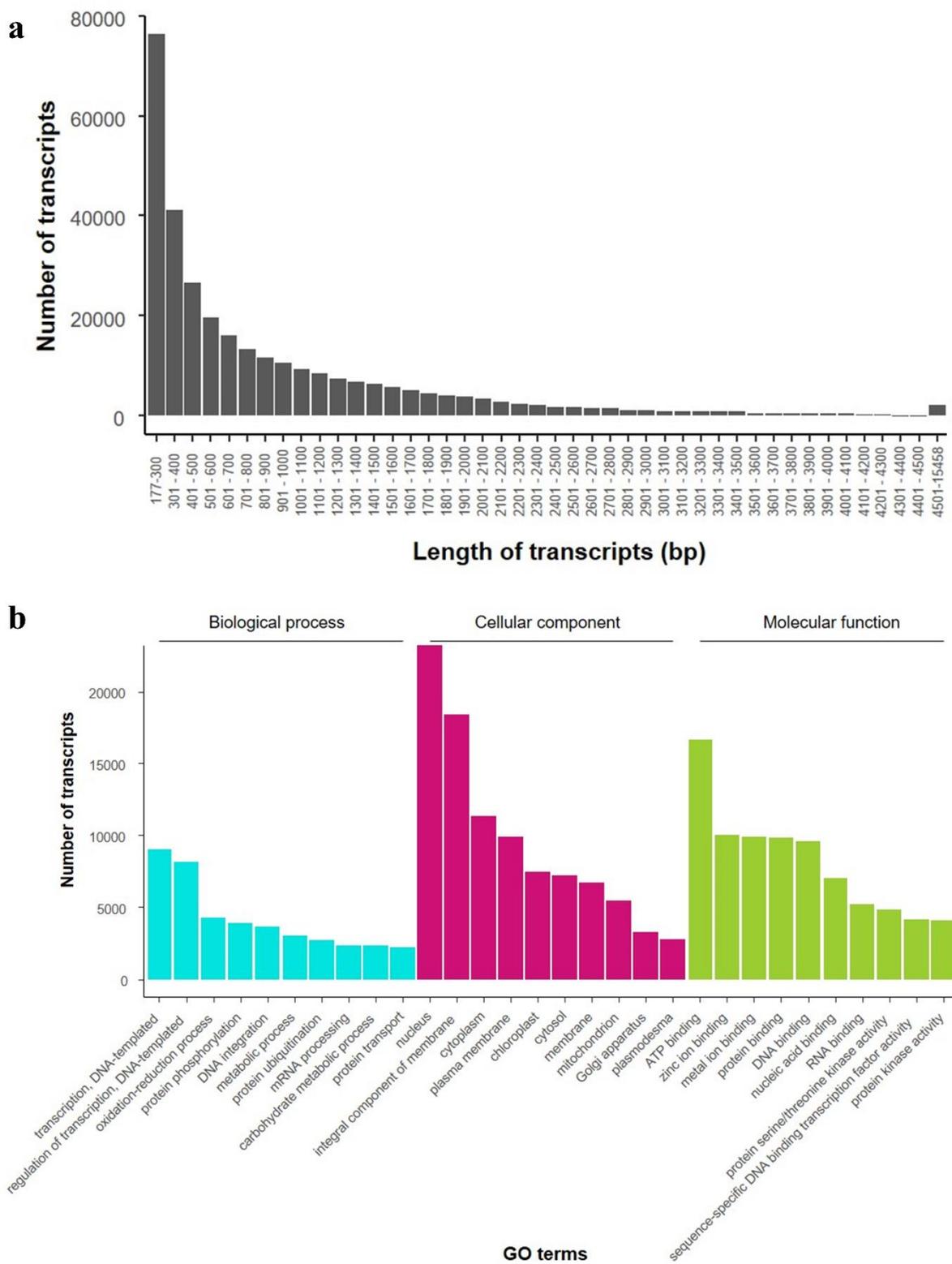


Fig. 3 Phylogenetic analysis among several NBS, NCS and PR10 proteins from a variety of plants.

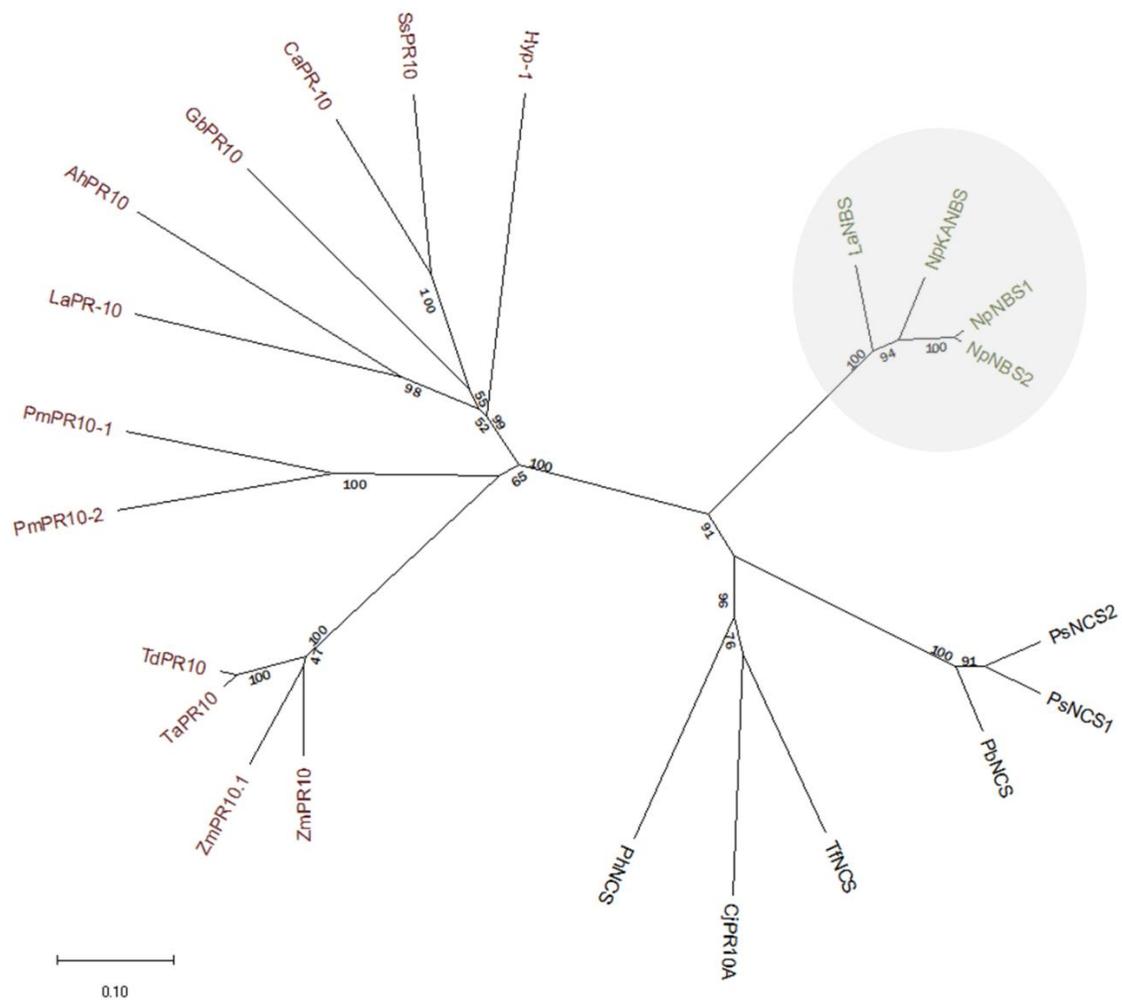


Fig. 4 Alignment from the amino acid sequences of *La*NBS, *Np*NBS1/2, *Np*KANBS and *Tf*NCS using CLUSTAL W algorithm.

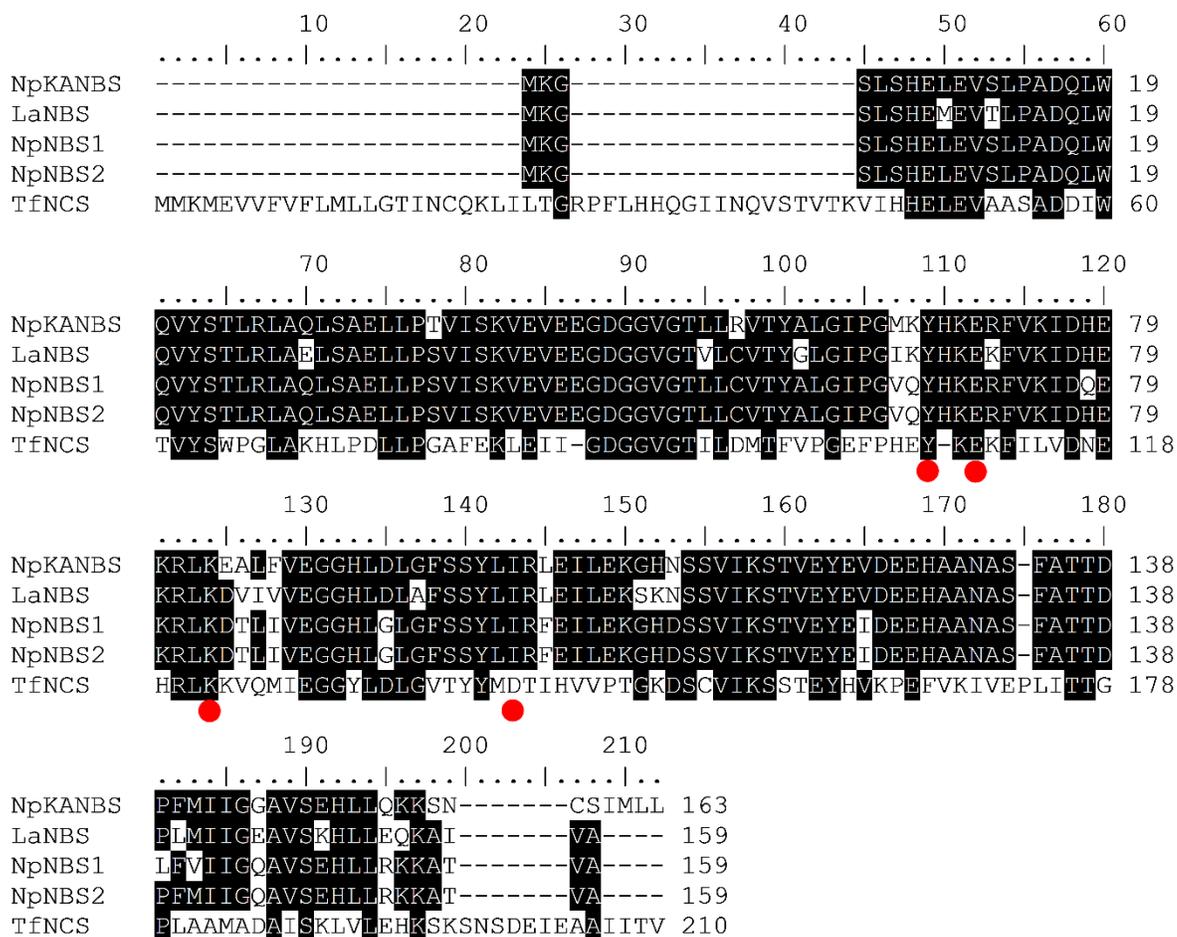


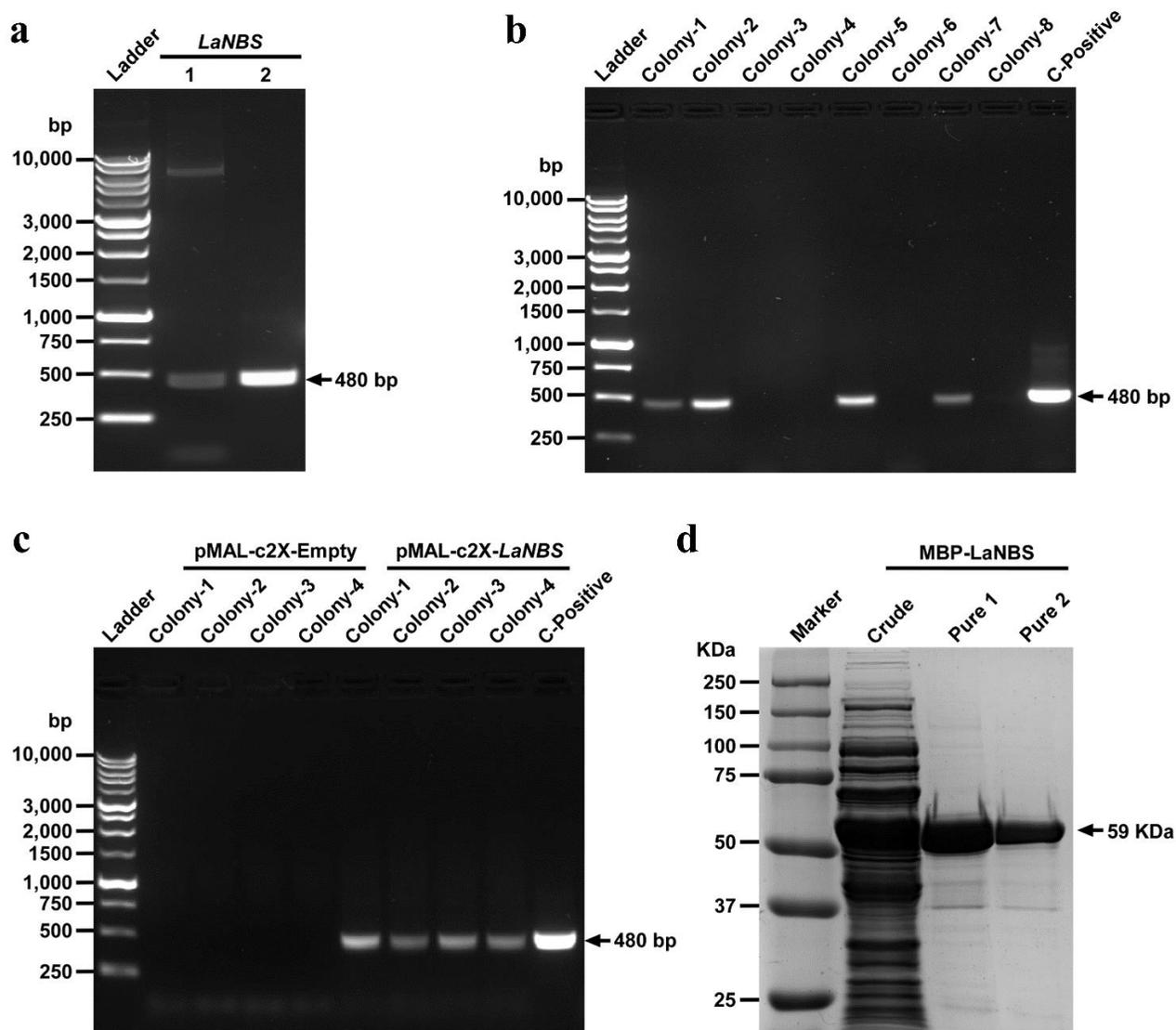
Fig. 5 PCR amplification, cloning and heterologous expression of *LaNBS* in *E. coli*.

Fig. 6 Extracted ion chromatograms showing the substrates and product of *La*NBS enzyme assays.

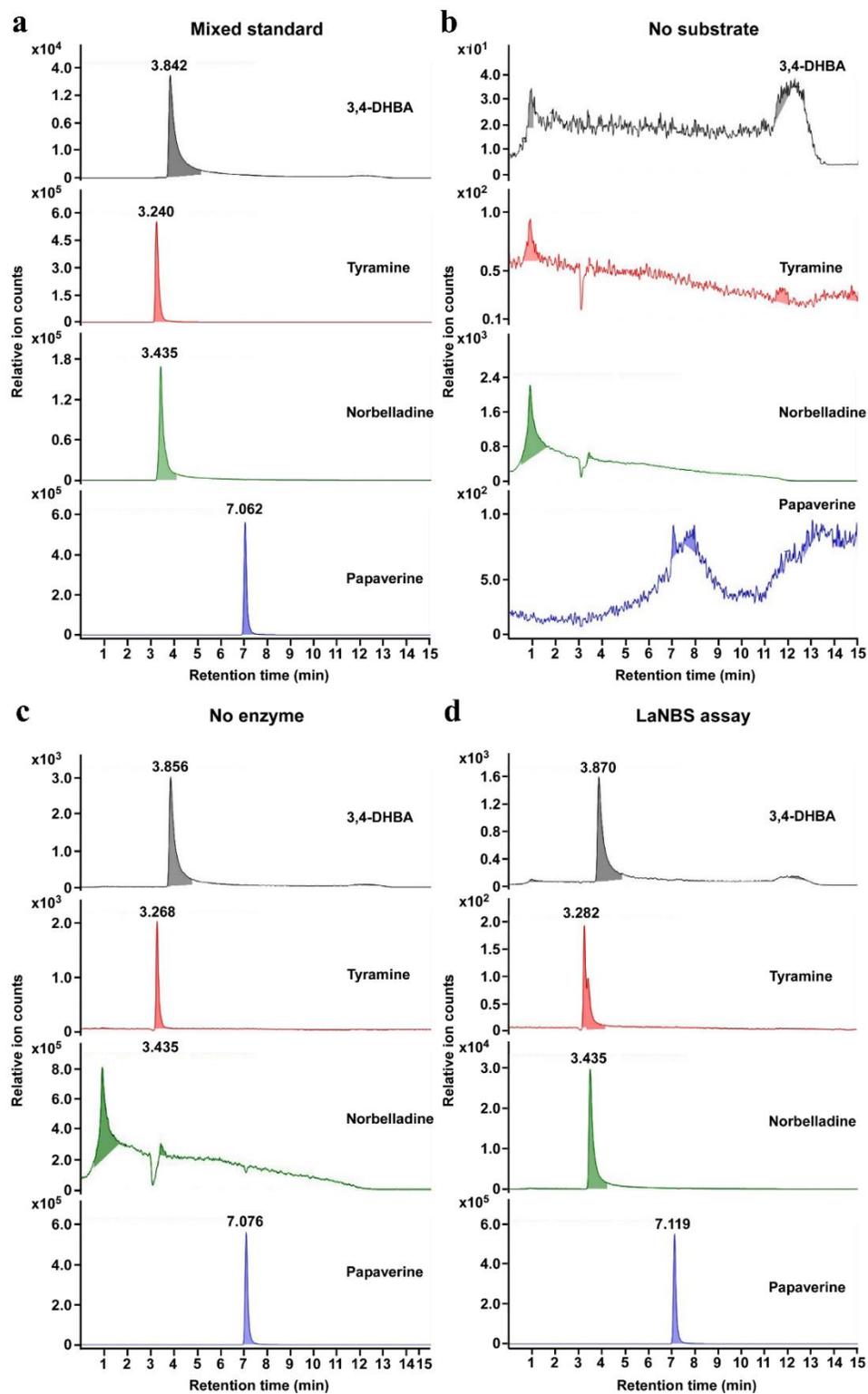
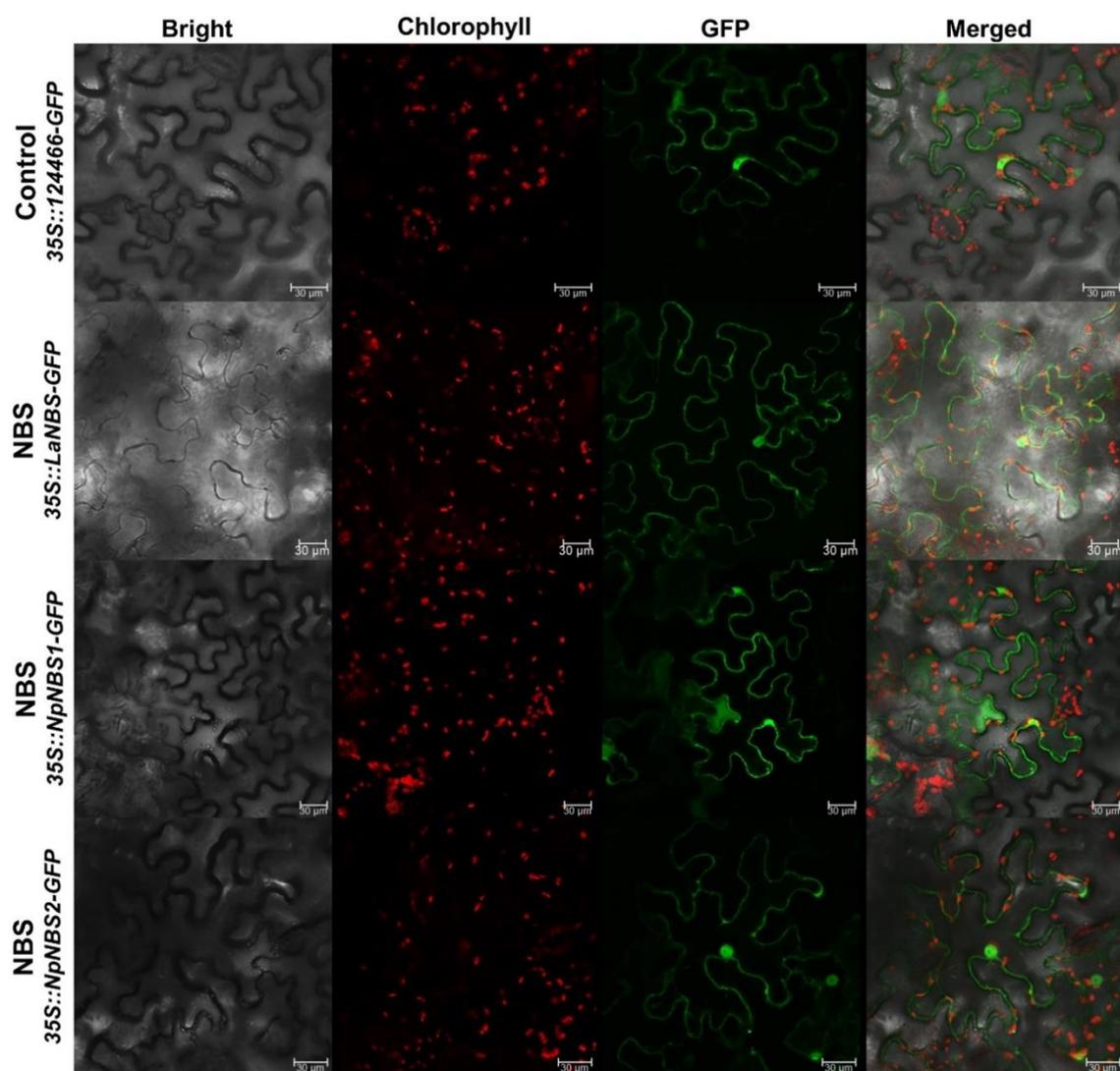


Fig. 7 Subcellular localization of NBS.

CHAPITRE III

DISCUSSION

Les Amaryllidacées sont une famille de plantes caractérisées par la présence d'un bulbe que l'on retrouve essentiellement dans les régions tropicales et tempérées. Cette famille est plus particulièrement connue en raison de sa capacité à produire des alcaloïdes de la classe des isoquinoléines. Les alcaloïdes isoquinoléiques produits par les plants d'Amaryllidacées se nomment donc les alcaloïdes des Amaryllidacées (AA). Jusqu'à aujourd'hui, plus de 650 structures des AA ont été rapportées avec une grande variété d'activités biologiques. De plus, ces métabolites ont une grande importance pharmaceutique (Hotchandani et Desgagne-Penix 2017, Ka, Koirala et al. 2020, Desgagne-Penix 2021). Bien que les étapes chimiques menant à la biosynthèse des AA aient été établies, les gènes et enzymes correspondantes de la voie se retrouvent majoritairement non étudiés. Récemment, les analyses transcriptomiques corrélant l'accumulation des AA et l'expression des gènes impliqués dans leur métabolisme ont fourni des connaissances importantes sur la voie de biosynthèse, ce qui a permis l'identification et la caractérisation de la NBS, N4OMT, CYP96T1 et NR (Kilgore, Augustin et al. 2014, Kilgore, Holland et al. 2016, Singh, Massicotte et al. 2018). La formation de la norbelladine est une étape cruciale de la voie, puisque c'est le précurseur commun de tous les AA. La NBS, catalysant la condensation de la tyramine avec le 3,4-DHBA pour former la norbelladine, a été identifiée et partiellement caractérisée à un niveau moléculaire chez le Narcisse jaune (*Narcissus pseudonarcissus*) (Singh, Massicotte et al. 2018). Cependant, aucun orthologue de la NBS n'a été identifié depuis, et la caractérisation fonctionnelle de la NBS n'a toujours pas été réalisée, ce qui contribue au manque de connaissances concernant les compartiments intracellulaires dans lesquels la NBS exerce ses fonctions catalytiques et par extension, les localisations où les AA sont biosynthétisés.

Ces travaux de recherche ont eu pour but principal d'identifier et de caractériser la NBS à partir des espèces de plantes *Leucojum aestivum* et *Narcissus papyraceus*. Dans cette étude, le séquençage de l'ARN de *L. aestivum* a été effectué et son assemblage *de novo* a été rapporté. Divers gènes des AA ont été identifiés, y compris la NBS. Les localisations subcellulaires de la NBS chez *L. aestivum* et *N. papyraceus* ont été étudiées à l'aide de l'expression transitoire de la NBS chez la plante *Nicotiana benthamiana*.

3.1 Retour sur les résultats de recherche

Maints résultats intéressants ont émergé du projet de recherche et ceux-ci sont bien identifiés et expliqués dans l'article présenté au chapitre II de ce mémoire. La présente section fera donc un récapitulatif de ces résultats.

Leucojum aestivum présente l'un des profils en AA les plus diversifiés. En effet, des études phytochimiques ont démontré la présence de plus de 36 alcaloïdes différents chez cette espèce (Berkov, Georgieva et al. 2009). Parmi les travaux de recherche effectués, le profil en alcaloïdes des différents tissus de la plante *L. aestivum* a été étudié. Les résultats obtenus ont démontré que la concentration la plus élevée et la plus grande diversité d'AA se retrouvent dans le bulbe. Ces résultats concordent très bien avec les données publiées sur le profil d'expression des AA chez *Narcissus pseudonarcissus* et *Narcissus papyraceus*, où la plus grande quantité des AA est produite dans le bulbe (López, Bastida et al. 2003, Lubbe, Gude et al. 2013, Singh et Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019). Les feuilles et les bulbes des Amaryllidacées possèdent une quantité élevée en alcaloïdes, car cela aiderait à la protection des réserves de carbohydrates contre les animaux herbivores et les pathogènes (Ruiz, Ward et al. 2002, García, Azorín et al. 2003). L'étude de l'expression des AA a également démontré que la galanthamine et la lycorine sont les plus fortement exprimées chez *L. aestivum*. Des résultats similaires ont déjà été publiés dans le passé chez d'autres plants des Amaryllidacées (Georgieva, Berkov et al. 2007, Berkov, Georgieva et al. 2009, El Tahchy 2010, Berkov, Георгиева et al. 2013, Ptak, Simlat et al. 2019, Ptak, Morańska et al. 2020).

Dans l'ensemble, les résultats montrent la présence de différents types d'AA dans le bulbe de *L. aestivum* suggérant la présence d'enzymes impliquées dans leur formation.

Par la suite, le premier transcriptome de *L. aestivum* a été généré en utilisant le séquençage d'ARN Illumina. Cette méthode est utile afin de générer des ressources et des connaissances précieuses sur les plantes non modèles comme *Leucojum aestivum* où le manque d'information génétique est proéminent. L'assemblage *de novo* des séquences de *L. aestivum* a abouti à un transcriptome de 273 Mpb, qui a donné 307 669 transcrits assemblés ayant une longueur moyenne de 887 pb et 158 155 de ces transcrits ont été annotés. Des nombres similaires avaient été obtenus pour le transcriptome de *N. pseudonarcissus* avec 195 347 transcrits annotés ayant une longueur moyenne de 761 pb (Singh et Desgagne-Penix 2017). De plus, le transcriptome de *N. papyraceus*, précédemment rapporté a également obtenu des résultats similaires avec 148 563 transcrits annotés ayant une longueur moyenne de 854 pb. (Hotchandani, de Villers et al. 2019). Ces similarités soutiennent donc la qualité du transcriptome généré pour *Leucojum aestivum*.

Lors de cette étude, nous avons identifié plusieurs variantes de transcrits pour tous les gènes connus à ce jour dans la voie de biosynthèse des AA. Les transcrits avec les valeurs de TPM (transcrits par million) les plus élevées sont la tyrosine décarboxylase 2 (LaTYDC2), la 3-désoxy-D-arabino-heptulosonate 7-phosphate synthase 1 (LaDHAPL1), la noroxomaritidine synthase 3 (LaNoroxoSyn/CYP96T3) et le coumarate 3-hydroxylase/ascorbate peroxydase (LaAPX/C3'H1). Une forte expression de LaNoroxoSyn/CYP96T3 a été rapportée dans les bulbes de *N. pseudonarcissus* King Alfred, *N. papyraceus* et *L. radiata* et est également associée à une concentration plus élevée en AA (Singh et Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019, Ptak, Simlat et al. 2019). Cependant, aucune étude ne supporte jusqu'à maintenant que l'expression des gènes précurseurs aux AA, tels que TYDC, serait plus élevée chez *N. pseudonarcissus* King Alfred ainsi que chez *N. papyraceus* (Singh et Desgagne-Penix 2017, Hotchandani, de Villers et al. 2019). Cela suggère que l'expression de TYDC n'est pas régulée de façon coordonnée avec PAL chez *L. aestivum*, alors qu'il a été suggéré que

ces deux enzymes travaillent de manière coordonnée chez *Narcissus* spp. Donc, il est possible que la TYDC soit coexprimée avec APX, bien que d'autres études soient nécessaires pour confirmer cette hypothèse. De plus, les niveaux d'expression de N4OMT et de la NBS se sont avérés inférieurs à ceux rapportés pour *N. pseudonarcissus* King Alfred (Singh et Desgagne-Penix 2017). En général, le profil transcriptomique des AA suggère que la voie de biosynthèse de *L. aestivum* n'est pas régulée de la même manière que les espèces de *Narcissus*.

L'alignement de séquences des différents orthologues de la NBS a montré la conservation absolue des résidus du site catalytique des différentes espèces de plantes à l'étude. Tous les orthologues de la NBS, soit LaNBS, NpNBS1 et NpNBS2 partagent plus de 39 % d'identité avec la norcochlorine synthase de *T. flavum* (TfNCS). L'alignement a également démontré que TfNCS et les trois NBS partagent trois résidus catalytiques communs sur quatre (Tyr, Glu et Lys) (Singh, Massicotte et al. 2018). L'analyse phylogénétique suggère que les NBS et les NCS ont évolué à partir d'un ancêtre commun, puisqu'ils partagent entre eux de 37 % à 44 % d'identité de séquences en acides aminés. Tandis que la NBS ne possède pas de signal peptidique, il a été rapporté dans la littérature que l'enzyme PsNCS2 aurait un signal peptidique en N-terminal et cette protéine serait initialement localisée dans le réticulum endoplasmique (RE) avant d'être entreposée dans la vacuole (Lee et Facchini 2010). Outre cela, l'homologie démontrée entre les NBS et les protéines des PR10 suggère le recrutement des NBS à partir d'un ancêtre commun des PR10, et que certains domaines de leurs séquences ont été conservés afin de leur permettre de réaliser leurs fonctions. La famille des PR10 est connue pour jouer un rôle majeur dans le mécanisme de défense de la plante contre les stress biotiques et abiotiques. Cette famille est classée en deux groupes distincts : les protéines intracellulaires liées à la pathogénèse (IPR), qui présente une homologie avec les ribonucléases, et les (S)-norcochlorine synthases (NCS) (Samanani, Liscombe et al. 2004, Agarwal et Agarwal 2014). Des études ont démontré que les NCS PR10 se distinguent des IPR par la présence d'un peptide signal en N-terminal, ce qui implique qu'elles sont localisées dans un compartiment intracellulaire autre que le cytosol alors que les protéines IPR affichent une localisation cytosolique (Samanani, Liscombe et al. 2004, Liscombe, MacLeod et al. 2005, Liu et

Ekramoddoullah 2006). Dans l'ensemble, les résultats suggèrent que l'évènement principal d'évolution qui a permis aux plantes des Amaryllidacées de produire des AA est le recrutement des NBS à partir d'un ancêtre commun des PR10.

Suite à l'expression hétérologue dans *Escherichia coli* et la purification de LaNBS, des essais enzymatiques par LC-MS/MS ont confirmé que l'enzyme catalyse effectivement la synthèse de la norbelladine par un processus de condensation entre un amine et un aldéhyde, plus particulièrement la tyramine et le 3,4-DHBA. Parmi les résidus catalytiques proposés pour la NBS, soit Tyr68, Glu71, Lys83 et Ile102, le Glu71 a été détecté avec un décalage d'histidine de +1 dans la séquence de LaNBS, ce qui a également été rapporté chez la NBS de *Narcissus pseudonarcissus* King Alfred (NpKANBS) (Singh, Massicotte et al. 2018). Ce décalage pourrait affecter les interactions avec le site catalytique et nuire au processus de cyclisation. De plus, contrairement aux NCS, la capacité de la NBS à former de la norbelladine en absence du résidu catalytique Aps141 suggère que celui-ci, qui est habituellement présent chez les NCS, n'est pas responsable directement de la réaction de condensation et pourrait plutôt être impliqué dans la stabilisation électrostatique (Lichman, Gershter et al. 2015, Singh, Massicotte et al. 2018).

Comme mentionné précédemment, ce projet a permis l'identification de séquences orthologues complètes de LaNBS, NpNBS1 et NpNBS2. L'homologie démontrée entre ces trois orthologues et NpKANBS suggère que ces enzymes ont toutes un site d'action et site de localisation similaire. En effet, la NBS contient une région bien conservée, un domaine Bet v1, qui appartient à la famille de protéines des Bet v1/PR10. Chez les NCS, plusieurs études ont démontré que les protéines membres de la famille des Bet v1/PR10, comme les NCS, jouent un rôle important dans la biosynthèse des alcaloïdes (Samanani, Liscombe et al. 2004, Zulak, Cornish et al. 2007, Singh, Massicotte et al. 2018). Pour cette raison, des outils bioinformatiques de prédiction de la localisation subcellulaire tels que LocTree3, PSLpred et WoLF PSORT ont été utilisés. Tous les programmes ont prédit une localisation cytoplasmique pour LaNBS, NpNBS1 et NpNBS2. De plus, aucune séquence signal ciblant un compartiment particulier n'a été

prédite par SignalP, ChloroP ou TargetP. Subséquemment, les résultats de prédiction de localisations subcellulaires obtenus nous ont mené à confirmer ces hypothèses à l'aide de la méthode de l'expression transitoire de la NBS dans la plante *Nicotiana benthamiana*. Pour ces expériences, la GFP a été utilisée comme protéine de fusion en C-terminale de la NBS, pour ne pas perturber les signaux de localisations subcellulaires que la NBS pourrait avoir en sa portion N-terminale. En effet, le logiciel de prédiction WoLF PSORT a également prédit une localisation extracellulaire, au RE et dans l'appareil de golgi. Au final, nos résultats de prédictions et de microscopie confocale de la NBS en fusion avec la GFP ont montré que la localisation de la NBS est cytosolique. En outre, un faible signal de fluorescence a été détecté au noyau pour LaNBS, NpNBS1 et NpNBS2, alors qu'aucun des outils de prédiction bioinformatiques suggérait ce compartiment. Ces résultats pourraient suggérer que la NBS ne contient pas un signal de localisation nucléaire (SLN) typique ou monopartite (Fukasawa, Leung et al. 2014). Par conséquent, puisque les bases de données des outils bioinformatiques de prédiction cherchent des similitudes entre les protéines à l'étude et les SLN typiques, il est possible que ceux-ci n'aient pas détecté un signal qui pourrait être présent dans la séquence de la NBS. Jusqu'à ce jour, les algorithmes pour la prédiction des SLN reposent sur des connaissances antérieures des résidus de base des SLN chez des espèces spécifiques. Pour cette raison, la reconnaissance de ces signaux peut être plus difficile puisque beaucoup de SLN n'ont toujours pas été identifiés (Guo, Yang et al. 2020). Aussi, la présence d'un signal à l'extrémité C-terminale de la NBS aurait pu être altérée par la fusion de la protéine GFP. Cependant, il a été confirmé que la localisation ne provenait pas de la GFP seulement puisque les résultats de western-blot ont démontré l'intégrité de la fusion NBS-GFP. Une autre possibilité est que la localisation de la NBS au noyau pourrait être le résultat de la diffusion de la GFP. En effet, la GFP peut entraîner la NBS au noyau puisque son poids moléculaire est beaucoup plus petit (17 kDa) comparativement à celui de la GFP (28 kDa). Alors, la fusion NBS-GFP pourrait faciliter une petite quantité de diffusion de la GFP dans le noyau, transportant avec elle la NBS. Toutefois, une variété de petits effecteurs fongiques (8-10 kDa) avec différents signaux peptidiques, en fusion avec la GFP en C-terminal, ont signalé des localisations intracellulaires différentes (noyau, nucléole, chloroplaste, mitochondrie, etc.) et suggèrent que la taille des protéines pourrait ne pas

influencer la localisation (Germain, Joly et al. 2018). Pour le moment, nous ne pouvons pas complètement exclure la possibilité de diffusion de la protéine NBS-GFP à l'intérieur du noyau. D'autres expérimentations pourraient être effectuées afin de clarifier la localisation de la NBS, comme par exemple l'immunolocalisation *in situ* ou bien analyser la localisation de la NBS en fusion avec la GFP en N-terminale. Bien que la fusion GFP-NBS est moins susceptible de donner la bonne localisation subcellulaire, cela permettrait de confirmer davantage les résultats obtenus. Plusieurs études ont montré que les signaux peptidiques se retrouvent principalement à l'extrémité N-terminale de la séquence (par exemple, les signaux de ciblage pour les mitochondries, le chloroplaste et le RE). En revanche, les signaux situés à l'extrémité C-terminale de la protéine sont plus associés au ciblage peroxysomal (Schatz et Dobberstein 1996, Kunze et Berger 2015). Dans le cas de la NBS, la localisation aux peroxysomes n'est pas attendue. La NBS partage une identité substantielle avec les protéines Bet v1 et PR10 dont l'emplacement subcellulaire attendu est au cytosol.

Une protéine remplit ses fonctions une fois synthétisée et transportée vers sa localisation subcellulaire spécifique (Huh, Falvo et al. 2003). Puisque les protéines LaNBS, NpNBS1 et NpNBS2 sont probablement localisées dans le cytosol et sont reconnues comme des enzymes clés dans la biosynthèse des AA, cela suggère que la réaction de condensation de la NBS devrait avoir lieu dans le cytosol. Les résultats obtenus sont en corrélation avec la localisation des substrats utilisés par la NBS. Récemment, l'enzyme TYDC1, qui catalyse la conversion de la tyrosine en tyramine pour l'obtention d'un précurseur des AA, a été caractérisée au niveau fonctionnel et il a été démontré que cette enzyme est localisée au niveau du cytosol (Wang, Han et al. 2019). En ce qui concerne le substrat 3,4-DHBA, son précurseur l'acide caféique n'a été trouvé que dans le cytosol dans les racines de carottes (Nagahashi, Abney et al. 1996). Par conséquent, la tyramine et le 3,4-DHBA sont susceptibles d'être colocalisés dans le cytosol où ils peuvent être condensés en norbelladine à l'aide de la NBS. De plus, l'enzyme *O*-méthyltransférase de la plante *Lycoris aurea* (LaOMT1) et N4OMT de *Lycoris longituba* (LIN4OMT) qui catalysent l'*O*-méthylation de la norbelladine en 4'-*O*-méthylnorbelladine sont toutes les deux localisées dans le cytosol. LaOMT1 a également montré une localisation aux

endosomes (Sun, Wang et al. 2018). Dans l'ensemble, les résultats de l'étude soutiennent que les AA peuvent être biosynthétisés dans le cytosol et/ou dans le noyau.

Les résultats obtenus quant à la localisation subcellulaire de la NBS au niveau du cytosol concordent avec les résultats préalablement obtenus sur le pH optimal de l'enzyme orthologue NpKANBS (Singh, Massicotte et al. 2018). En effet, la NBS semblerait avoir une plus grande activité enzymatique à pH 4. Il a été démontré que l'acidification du cytosol est un processus habituellement enclenché lorsque la plante est exposée à un stress, ce qui permettrait d'augmenter l'activité du métabolisme spécialisé comparativement au métabolisme primaire, qui est plus actif à pH neutre (Sakano 2001). La réponse des plantes à une attaque par les pathogènes entraînent un changement dans le flux des ions dans les différents compartiments, en plus de modifier le pH (Felle 2001). La NBS est une enzyme impliquée dans la biosynthèse de MS, qui est associée au mécanisme de défense des plantes. Sa localisation intracellulaire au cytosol et son pH optimal de 4 supportent son rôle de défense dans la plante et suggèrent que l'acidification du cytosol serait une précondition pour l'activation de la voie biosynthétique des AA.

En général, nous rapportons que les trois orthologues de la NBS : LaNBS, NpNBS1 et NpNBS2 se localisent dans le cytosol, ce qui suggère que la NBS exerce ses fonctions dans ce compartiment intracellulaire. Considérant que la NBS est reconnue comme une enzyme cruciale dans le métabolisme des AA, les précurseurs de AA tels que la norbelladine sont plus susceptibles d'être biosynthétisés dans le cytosol.

3.2 Perspectives du projet de recherche

Ce mémoire présente la première étude de localisation subcellulaire de la norbelladine synthase. Bien que la plupart des enzymes de la voie de biosynthèse des AA restent à identifier et caractériser, l'approfondissement des connaissances quant à la localisation de la voie métabolique apporte des informations importantes pouvant être utiles afin de développer des outils de biologie synthétique.

Parmi les résultats présentés dans ce mémoire, il est montré que la protéine NBS se retrouve majoritairement dans le cytosol des cellules et des indications expérimentales d'une faible présence au noyau ont aussi été obtenues. Toutefois, il est suspecté que la localisation de la NBS au noyau soit due à la diffusion de la GFP, entraînant donc avec elle la protéine à l'étude. Il serait intéressant de faire d'autres expériences confirmant ou non la compartimentation intracellulaire de la NBS au noyau. Par exemple, l'étude du protéome nucléaire pourrait être envisagée. En effet, l'extraction des protéines du noyau des plantes des Amaryllidacées permettrait de confirmer si celles-ci présentent des enzymes de la voie de biosynthèse des AA dans ce compartiment. Une autre technique n'impliquant pas une extraction des protéines du noyau est l'immunolocalisation *in situ*. Dans cette méthode, les protéines d'intérêts sont détectées par des anticorps marqués et celles-ci peuvent ensuite être caractérisées par spectrométrie de masse. Dans la situation hypothétique où la NBS se retrouverait réellement au noyau par elle-même, il serait également intéressant d'étudier son rôle à cette localisation, donc d'étudier les interactions de la NBS avec l'ADN, l'ARN ou les protéines nucléaires. Pour ce faire, plusieurs techniques sont possibles. Par exemple, la méthode du double hybride est une technique de biologie moléculaire permettant de détecter des interactions entre deux protéines. Cette méthode détecte les interactions physiques entre la protéine d'intérêt et une autre protéine, par l'entremise d'un gène rapporteur (Maple et Møller 2007). De cette façon, il serait possible de déterminer les possibles interactions de la NBS avec des protéines, l'ARN ou l'ADN. Puisque la NBS réalise la première étape de la voie des AA, il est possible que celle-ci exerce un rôle de régulation dans ce compartiment. Sachant que la NBS réalise ses fonctions dans le cytosol, il serait également intéressant d'effectuer des analyses protéomiques des protéines contenues dans le cytoplasme des Amaryllidacées. Cela permettrait peut-être d'identifier de nouveaux gènes impliqués dans la biosynthèse des AA, car seulement quatre enzymes ont été identifiées et beaucoup d'enzymes sont encore inconnues jusqu'à ce jour.

Récemment, les progrès des technologies « omiques » ont accéléré l'étude des composés phytochimiques des plantes non modèles comme les Amaryllidacées. Pour ce projet, nous avons généré un profil transcriptomique de *L. aestivum* tel que décrit au

chapitre II. Ce transcriptome pourrait être utilisé pour une étude approfondie afin d'identifier de nouveaux gènes, isoformes de gènes et facteurs de transcription. Une recherche d'homologie pour un gène cible et une analyse de co-expression d'une voie de biosynthèse correspondante dans le transcriptome serait une bonne stratégie pour la découverte de gènes inconnus. Par exemple, la noroxomaritidine réductase (NR) a été identifiée à l'aide d'une approche similaire, ce qui a permis de caractériser son activité enzymatique permettant la conversion de la noroxomaritidine en oxomaritidine (Kilgore, Holland et al. 2016).

3.3 Conclusions

Ce projet de recherche a permis de générer le premier transcriptome de *L. aestivum* à l'aide de séquence d'ARN. Nous avons également pu identifier 50 transcrits de gènes impliqués dans la biosynthèse des AA. À partir des données transcriptomiques des deux plantes productrices d'AA *L. aestivum* et *N. papyraceus*, nous avons identifié et cloné la NBS. De plus, nous avons caractérisé l'activité et la localisation de cette enzyme clé nécessaire à la production de tous les AA. L'identification des nouveaux gènes des AA et l'élucidation de la localisation subcellulaire de la biosynthèse des AA faciliteront l'ingénierie métabolique des plantes contenant ou non des AA, ainsi que le développement de microorganismes modifiés par ingénierie afin d'insérer la voie de biosynthèse des AA dans des levures, bactéries ou microalgues.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Agarwal, P. and P. K. Agarwal (2014). "Pathogenesis related-10 proteins are small, structurally similar but with diverse role in stress signaling." Molecular Biology Reports **41**(2): 599-611.
- Akula, R. and G. A. Ravishankar (2011). "Influence of abiotic stress signals on secondary metabolites in plants." Plant Signaling & Behavior **6**(11): 1720-1731.
- Alcantara, J., D. A. Bird, V. R. Franceschi and P. J. Facchini (2005). "Sanguinarine biosynthesis is associated with the endoplasmic reticulum in cultured opium poppy cells after elicitor treatment." Plant physiology **138**(1): 173-183.
- Ascherio, A., M. G. Weisskopf, E. J. O'Reilly, M. L. McCullough, E. E. Calle, C. Rodriguez and M. J. Thun (2004). "Coffee consumption, gender, and Parkinson's disease mortality in the cancer prevention study II cohort: the modifying effects of estrogen." American Journal of Epidemiology **160**(10): 977-984.
- Ashour, M., M. Wink and J. Gershenzon (2010). Biochemistry of Terpenoids: Monoterpenes, Sesquiterpenes and Diterpenes. Annual Plant Reviews Volume 40: Biochemistry of Plant Secondary Metabolism: 258-303.
- Awwad, F., E. I. Fantino, F. Meddeb-Mouelhi, J. Li, M. Héneault, C. Damour, J.-F. Lemay, B. Karas and I. Desgagne-Penix (2020). "Getting the hype elsewhere: on the heterologous production of cannabinoid in bioengineered diatom *Phaeodactylum tricorutum* with minimal gene insertion." Nature Biotechnology **Submitted**.
- Barik, J., F. Dajas-Bailador and S. Wonnacott (2005). "Cellular responses to nicotinic receptor activation are decreased after prolonged exposure to galantamine in human neuroblastoma cells." British Journal of Pharmacology **145**(8): 1084-1092.
- Barton, D. and T. Cohen (1957). "Festschrift A. Stoll." Birkhäuser, Basel **117**.
- Barton, D., G. Kirby, J. Taylor and G. Thomas (1963). "866. Phenol oxidation and biosynthesis. Part VI. The biogenesis of amaryllidaceae alkaloids." Journal of the Chemical Society (Resumed): 4545-4558.
- Bastida, J., S. Berkov, L. Torras, N. B. Pigni, J. P. de Andrade, V. Martínez, C. Codina and F. Viladomat (2011). Chemical and biological aspects of Amaryllidaceae alkaloids., Transworld Research Network.

- Battersby, A., H. Fales and W. Wildman (1961). "Biosynthesis in the Amaryllidaceae. Tyrosine and norbelladine as precursors of haemanthamine." Journal of the American Chemical Society **83**(19): 4098-4099.
- Berkov, S., J. Bastida, F. Viladomat and C. Codina (2011). "Development and validation of a GC-MS method for rapid determination of galanthamine in *Leucojum aestivum* and *Narcissus* ssp.: a metabolomic approach." Talanta **83**(5): 1455-1465.
- Berkov, S., L. Georgieva, V. Kondakova, A. Atanassov, F. Viladomat, J. Bastida and C. Codina (2009). "Plant Sources of Galanthamine: Phytochemical and Biotechnological Aspects." Biotechnology & Biotechnological Equipment **23**(2): 1170-1176.
- Berkov, S., L. Georgieva, V. Kondakova, F. Viladomat, J. Bastida, A. Atanassov and C. Codina (2013). "The geographic isolation of *Leucojum aestivum* populations leads to divergence of alkaloid biosynthesis." Biochemical Systematics and Ecology **46**(0): 152-161.
- Bhattacharya, A., P. Sood and V. Citovsky (2010). "The roles of plant phenolics in defence and communication during *Agrobacterium* and *Rhizobium* infection." Molecular Plant Pathology **11**(5): 705-719.
- Bidlack, W. R., S. T. Omaye, M. S. Meskin and D. K. W. Topham (2000). Phytochemicals as bioactive agents.
- Bird, D. A., V. R. Franceschi and P. J. Facchini (2003). "A Tale of Three Cell Types: Alkaloid Biosynthesis Is Localized to Sieve Elements in Opium Poppy." The Plant Cell **15**(11): 2626-2635.
- Böttger, A., U. Vothknecht, C. Bolle and A. Wolf (2018). Plant Secondary Metabolites and Their General Function in Plants. Lessons on Caffeine, Cannabis & Co: Plant-derived Drugs and their Interaction with Human Receptors. Cham, Springer International Publishing: 3-17.
- Bourgaud, F., A. Gravot, S. Milesi and E. Gontier (2001). "Production of plant secondary metabolites: a historical perspective." Plant Science **161**(5): 839-851.
- Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Analytical biochemistry **72**(1-2): 248-254.
- Brown, S., M. Clastre, V. Courdavault and S. E. O'Connor (2015). "De novo production of the plant-derived alkaloid strictosidine in yeast." Proceedings of the National Academy of Sciences **112**(11): 3205-3210.

- Bruneton, J. (2009). Pharmacognosie, phytochimie, plantes médicinales (4e ed.), Lavoisier.
- Carocho, M. and I. C. Ferreira (2013). "The role of phenolic compounds in the fight against cancer--a review." Anticancer Agents Med Chem **13**(8): 1236-1258.
- Chan, W.-T., C. S. Verma, D. P. Lane and S. K.-E. Gan (2013). "A comparison and optimization of methods and factors affecting the transformation of *Escherichia coli*." Bioscience reports **33**(6): e00086.
- Cho, K. S., Y.-R. Lim, K. Lee, J. Lee, J. H. Lee and I.-S. Lee (2017). "Terpenes from Forests and Human Health." Toxicological research **33**(2): 97-106.
- Christenhusz, M. and J. Byng (2016). "The number of known plant species in the world and its annual increase." Phytotaxa **261**: 201-217.
- Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szczesniak, D. J. Gaffney, L. L. Elo, X. Zhang and A. Mortazavi (2016). "A survey of best practices for RNA-seq data analysis." Genome Biology **17**(1): 13.
- Cosme, P., A. B. Rodríguez, J. Espino and M. Garrido (2020). "Plant Phenolics: Bioavailability as a Key Determinant of Their Potential Health-Promoting Applications." Antioxidants (Basel) **9**(12).
- Cox-Georgian, D., N. Ramadoss, C. Dona and C. Basu (2019). "Therapeutic and Medicinal Uses of Terpenes." Medicinal Plants: From Farm to Pharmacy: 333-359.
- Cragg, G. M. and D. J. Newman (2005). "Plants as a source of anti-cancer agents." Journal of Ethnopharmacology **100**(1): 72-79.
- Croteau, R., T. M. Kutchan and N. G. Lewis (2000). "Natural Products (Secondary Metabolites)." Biochemistry and Molecular Biology of Plants **24**: 1250-1318.
- Dai, J. and R. J. Mumper (2010). "Plant phenolics: extraction, analysis and their antioxidant and anticancer properties." Molecules (Basel, Switzerland) **15**(10): 7313-7352.
- De Andrade, J. P., N. B. Pigni, L. Torras-Claveria, S. Berkov, C. Codina, F. Viladomat and J. Bastida (2012). "Bioactive alkaloid extracts from *Narcissus broussonetii*: mass spectral studies." Journal of Pharmaceutical and Biomedical Analysis **70**: 13-25.

- De Luca, V., V. Salim, A. Thamm, S. A. Masada and F. Yu (2014). "Making iridoids/secoiridoids and monoterpene indole alkaloids: progress on pathway elucidation." Current Opinion in Plant Biology **19**: 35-42.
- De Luca, V. and B. St Pierre (2000). "The cell and developmental biology of alkaloid biosynthesis." Trends in Plant Science **5**(4): 168-173.
- Desgagne-Penix, I. (2021). "Biosynthesis of alkaloids in Amaryllidaceae plants: a review." Phytochemistry Reviews **20**: 409-431.
- Desgagne-Penix, I. and P. Facchini (2011). Benzylisoquinoline Alkaloid Biosynthesis: 241-261.
- Dewick, P. M. (2009). Alkaloids. Medicinal Natural Products: 311-420.
- Dewick, P. M. (2011). Medicinal Natural Products: A Biosynthetic Approach, Wiley.
- Dey, P., A. Kundu, A. Kumar, M. Gupta, B. M. Lee, T. Bhakta, S. Dash and H. S. Kim (2020). "Analysis of alkaloids (indole alkaloids, isoquinoline alkaloids, tropane alkaloids)." Recent Advances in Natural Products Analysis: 505-567.
- Diamond, A. and I. Desgagné-Penix (2016). "Metabolic engineering for the production of plant isoquinoline alkaloids." Plant Biotechnology Journal **14**(6): 1319-1328.
- Dias, D. A., S. Urban and U. Roessner (2012). "A historical overview of natural products in drug discovery." Metabolites **2**(2): 303-336.
- Dräger, B. (1995). "Identification and quantification of calystegines, polyhydroxyl nortropane alkaloids." Phytochemical Analysis **6**(1): 31-37.
- Efferth, T., M. R. Romero, D. G. Wolf, T. Stamminger, J. J. Marin and M. Marschall (2008). "The antiviral activities of artemisinin and artesunate." Clinical Infectious Diseases **47**(6): 804-811.
- Eichhorn, J., T. Takada, Y. Kita and M. H. Zenk (1998). "Biosynthesis of the amaryllidaceae alkaloid Galanthaminefn2." Phytochemistry **49**(4): 1037-1047.
- El Tahchy, A. (2010). Étude de la voie de biosynthèse de la galanthamine chez *Leucojum aestivum* L. – Criblage phytochimique de quelques Amaryllidaceae. Ph. D., Nancy Université.

- El Tahchy, A., A. Ptak, M. Boisbrun, E. Barre, C. Guillou, F. Dupire, F. Chrétien, M. Henry, Y. Chapleur and D. Laurain-Mattar (2011). "Kinetic study of the rearrangement of deuterium-labeled 4'-O-methylnorbelladine in *Leucojum aestivum* shoot cultures by mass spectrometry. Influence of precursor feeding on Amaryllidaceae alkaloid accumulation." Journal of natural products **74**(11): 2356-2361.
- Facchini, P. J. and B. St-Pierre (2005). "Synthesis and trafficking of alkaloid biosynthetic enzymes." Current Opinion in Plant Biology **8**(6): 657-666.
- Falodun, A. (2010). "Herbal medicine in Africa-distribution, standardization and prospects." Research Journal of Phytochemistry **4**(3): 154-161.
- Felle, H. H. (2001). "pH: Signal and Messenger in Plant Cells." Plant Biology **3**(6): 577-591.
- Felsenstein, J. (1985). "Confidence Limits on Phylogenies: An Approach Using the Bootstrap." Evolution **39**(4): 783-791.
- Fernandes, H., K. Michalska, M. Sikorski and M. Jaskolski (2013). "Structural and functional aspects of PR-10 proteins." FEBS Journal **280**(5): 1169-1199.
- Ferreira, M.-J. U. and A. Paterna (2019). "Monoterpene indole alkaloids as leads for targeting multidrug resistant cancer cells from the African medicinal plant *Tabernaemontana elegans*." Phytochemistry Reviews **18**(4): 971-987.
- Frick, K. M., L. G. Kamphuis, K. H. M. Siddique, K. B. Singh and R. C. Foley (2017). "Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement." Frontiers in Plant Science **8**.
- Fukasawa, Y., R. K. Leung, S. K. Tsui and P. Horton (2014). "Plus ca change—evolutionary sequence divergence predicts protein subcellular localization signals." BMC Genomics **15**: 46.
- Furst, R. (2016). "Narciclasine—an Amaryllidaceae Alkaloid with Potent Antitumor and Anti-Inflammatory Properties." Planta Med **82**(16): 1389-1394.
- García, D., J. Azorín, J. Bastida, V. Francesc and C. Codina (2003). "Seasonal and Spatial Variations of Alkaloids in *Merendera montana* in Relation to Chemical Defense and Phenology." Journal of chemical ecology **29**: 1117-1126.

- Georgieva, L., S. Berkov, V. Kondakova, J. Bastida, F. Viladomat, A. Atanassov and C. Codina (2007). "Alkaloid variability in *Leucojum aestivum* from wild populations." Z Naturforsch C J Biosci **62**(9-10): 627-635.
- Germain, H., D. L. Joly, C. Mireault, M. B. Plourde, C. Letanneur, D. Stewart, M. J. Morency, B. Petre, S. Duplessis and A. Seguin (2018). "Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens." Molecular Plant Pathology **19**(1): 191-200.
- Gerrard, A. (1877). "The proximate principles of the *Narcissus pseudonarcissus*." Pharmaceutical Journal **8**: 214.
- González-Burgos, E. and M. P. Gómez-Serranillos (2012). "Terpene compounds in nature: a review of their potential antioxidant activity." Current Medicinal Chemistry **19**(31): 5319-5341.
- Grisebach, H. (1973). "Comparative biosynthetic pathways in higher plants." Pure and Applied Chemistry **34**(3-4): 487-514.
- Grycová, L., J. Dostál and R. Marek (2007). "Quaternary protoberberine alkaloids." Phytochemistry **68**(2): 150-175.
- Guimarães, P., J. Jose, M. Galetti and J. Trigo (2003). "Quinolizidine Alkaloids in *Ormosia arborea* Seeds Inhibit Predation But Not Hoarding by Agoutis (*Dasyprocta leporina*)." Journal of chemical ecology **29**: 1065-1072.
- Guo, Y., Y. Yang, Y. Huang and H. B. Shen (2020). "Discovering nuclear targeting signal sequence through protein language learning and multivariate analysis." Analytical Biochemistry **591**: 113565.
- Hagel, J. M. and P. J. Facchini (2013). "Benzyloquinoline Alkaloid Metabolism: A Century of Discovery and a Brave New World." Plant and Cell Physiology **54**(5): 647-672.
- Hartmann, T. (1999). "Chemical ecology of pyrrolizidine alkaloids." Planta **207**(4): 483-495.
- He, M., C. Qu, O. Gao, X. Hu and X. Hong (2015). "Biological and pharmacological activities of amaryllidaceae alkaloids." RSC Advances **5**(21): 16562-16574.

- Heinrich, M. and H. Lee Teoh (2004). "Galanthamine from snowdrop--the development of a modern drug against Alzheimer's disease from local Caucasian knowledge." Journal of Ethnopharmacology **92**(2-3): 147-162.
- Heinrich, M., J. Mah and V. Amirikia (2021). "Alkaloids Used as Medicines: Structural Phytochemistry Meets Biodiversity-An Update and Forward Look." Molecules (Basel, Switzerland) **26**(7): 1836.
- Hollingsworth, R. G., J. W. Armstrong and E. Campbell (2002). "Caffeine as a repellent for slugs and snails." Nature **417**(6892): 915-916.
- Hotchandani, T., J. de Villers and I. Desgagné-Penix (2019). "Developmental Regulation of the Expression of Amaryllidaceae Alkaloid Biosynthetic Genes in *Narcissus papyraceus*." Genes **10**(8): 594.
- Hotchandani, T. and I. Desgagne-Penix (2017). "Heterocyclic Amaryllidaceae Alkaloids: Biosynthesis and Pharmacological Applications." Current topics in medicinal chemistry **17**(4): 418-427.
- Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman and E. K. O'Shea (2003). "Global analysis of protein localization in budding yeast." Nature **425**(6959): 686-691.
- Ilari, A., S. Franceschini, A. Bonamore, F. Arengi, B. Botta, A. Macone, A. Pasquo, L. Bellucci and A. Boffi (2009). "Structural basis of enzymatic (S)-norcoclaurine biosynthesis." Journal of Biological Chemistry **284**(2): 897-904.
- Isah, T. (2019). "Stress and defense responses in plant secondary metabolites production." Biological Research **52**(1): 39.
- Jiang, Y., B. Xia, L. Liang, X. Li, S. Xu, F. Peng and R. Wang (2013). "Molecular and analysis of a phenylalanine ammonia-lyase gene (LrPAL2) from *Lycoris radiata*." Molecular Biology Reports **40**(3): 2293-2300.
- Jiang, Y., N. Xia, X. Li, W. Shen, L. Liang, C. Wang, R. Wang, F. Peng and B. Xia (2011). "Molecular cloning and characterization of a phenylalanine ammonia-lyase gene (LrPAL) from *Lycoris radiata*." Molecular Biology Reports **38**(3): 1935-1940.
- Jin, Z. and X.-H. Xu (2013). Amaryllidaceae Alkaloids. Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes. K. G. Ramawat and J.-M. Mérillon. Berlin, Heidelberg, Springer Berlin Heidelberg: 479-522.

- Jin, Z. and G. Yao (2019). "Amaryllidaceae and Sceletium alkaloids." Natural Product Reports **36**(10): 1462-1488.
- Ka, S., M. Koirala, N. Mérindol and I. Desgagné-Penix (2020). "Biosynthesis and Biological Activities of Newly Discovered Amaryllidaceae Alkaloids." Molecules **25**(21): 4901.
- Karban, R., L. H. Yang and K. F. Edwards (2014). "Volatile communication between plants that affects herbivory: a meta-analysis." Ecology Letters **17**(1): 44-52.
- Karimi, M., D. Inze and A. Depicker (2002). "GATEWAY vectors for Agrobacterium-mediated plant transformation." Trends in Plant Science **7**(5): 193-195.
- Kaur, R. (2015). "Alkaloids – Important Therapeutic Secondary Metabolites of Plants origin." Journal of Critical Reviews **2**: 1-8.
- Khan, A. Y. and G. Suresh Kumar (2015). "Natural isoquinoline alkaloids: binding aspects to functional proteins, serum albumins, hemoglobin, and lysozyme." Biophysical reviews **7**(4): 407-420.
- Kilgore, M., M. M. Augustin, G. D. May, J. A. Crow and T. M. Kutchan (2016). "CYP96T1 of *Narcissus* sp. aff. *pseudonarcissus* Catalyzes Formation of the Para-Para' C-C Phenol Couple in the Amaryllidaceae Alkaloids." Frontiers in Plant Science **7**.
- Kilgore, M. B., M. M. Augustin, C. M. Starks, M. O'Neil-Johnson, G. D. May, J. A. Crow and T. M. Kutchan (2014). "Cloning and characterization of a norbelladine 4'-O-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in *Narcissus* sp. aff. *pseudonarcissus*." PLoS One **9**(7): e103223.
- Kilgore, M. B., C. K. Holland, J. M. Jez and T. M. Kutchan (2016). "Identification of a noroxomaritidine reductase with Amaryllidaceae alkaloid biosynthesis related activities." Journal of Biological Chemistry **291**(32): 16740-16752.
- Kilgore, M. B. and T. M. Kutchan (2016). "The Amaryllidaceae alkaloids: biosynthesis and methods for enzyme discovery." Phytochemistry reviews : proceedings of the Phytochemical Society of Europe **15**(3): 317-337.
- Kohnen-Johannsen, K. L. and O. Kayser (2019). "Tropane Alkaloids: Chemistry, Pharmacology, Biosynthesis and Production." Molecules (Basel, Switzerland) **24**(4): 796.

- Kornienko, A. and A. Evidente (2008). "Chemistry, biology, and medicinal potential of narciclasine and its congeners." Chemical Reviews **108**(6): 1982-2014.
- Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura (2018). "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." Mol Biol Evol **35**(6): 1547-1549.
- Kunze, M. and J. Berger (2015). "The similarity between N-terminal targeting signals for protein import into different organelles and its evolutionary relevance." Frontiers in Physiology **6**(259).
- Kutchan, T. M. (1993). "Strictosidine: From alkaloid to enzyme to gene." Phytochemistry **32**(3): 493-506.
- Lamoral-Theys, D., A. Andolfi, G. Van Goietsenoven, A. Cimmino, B. Le Calve, N. Wauthoz, V. Megalizzi, T. Gras, C. Bruyere, J. Dubois, V. Mathieu, A. Kornienko, R. Kiss and A. Evidente (2009). "Lycorine, the main phenanthridine Amaryllidaceae alkaloid, exhibits significant antitumor activity in cancer cells that display resistance to proapoptotic stimuli: an investigation of structure-activity relationship and mechanistic insight." Journal of Medicinal Chemistry **52**(20): 6244-6256.
- Lee, E.-J. and P. Facchini (2010). "Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family." The Plant cell **22**(10): 3489-3503.
- Lee, M. J., J. S. Pate, D. J. Harris and C. A. Atkins (2007). "Synthesis, transport and accumulation of quinolizidine alkaloids in *Lupinus albus* L. and *L. angustifolius* L." Journal of Experimental Botany **58**(5): 935-946.
- Li, B. and C. N. Dewey (2011). "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome." BMC Bioinformatics **12**: 323.
- Li, J., E. J. Lee, L. Chang and P. J. Facchini (2016). "Genes encoding norcoclaurine synthase occur as tandem fusions in the Papaveraceae." Scientific Reports **6**: 39256.
- Li, L., H.-J. Dai, M. Ye, S.-L. Wang, X.-J. Xiao, J. Zheng, H.-Y. Chen, Y.-h. Luo and J. Liu (2012). "Lycorine induces cell-cycle arrest in the G0/G1 phase in K562 cells via HDAC inhibition." Cancer cell international **12**(1): 49.
- Lichman, B. R., M. C. Gershater, E. D. Lamming, T. Pesnot, A. Sula, N. H. Keep, H. C. Hailes and J. M. Ward (2015). "'Dopamine-first' mechanism enables the rational engineering of the norcoclaurine synthase aldehyde activity profile." The FEBS Journal **282**(6): 1137-1151.

- Lin, D., M. Xiao, J. Zhao, Z. Li, B. Xing, X. Li, M. Kong, L. Li, Q. Zhang, Y. Liu, H. Chen, W. Qin, H. Wu and S. Chen (2016). "An Overview of Plant Phenolic Compounds and Their Importance in Human Nutrition and Management of Type 2 Diabetes." Molecules (Basel, Switzerland) **21**(10): 1374.
- Liscombe, D. K., B. P. MacLeod, N. Loukanina, O. I. Nandi and P. J. Facchini (2005). "Evidence for the monophyletic evolution of benzyloquinoline alkaloid biosynthesis in angiosperms." Phytochemistry **66**(11): 1374-1393.
- Liu, J.-J. and A. K. M. Ekramoddoullah (2006). "The family 10 of plant pathogenesis-related proteins: Their structure, regulation, and function in response to biotic and abiotic stresses." Physiological and Molecular Plant Pathology **68**(1-3): 3-13.
- Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.
- López, S., J. Bastida, F. Viladomat and C. Codina (2003). "Galanthamine pattern in *Narcissus confusus* plants." Planta medica **69**(12): 1166-1168.
- Lubbe, A., H. Gude, R. Verpoorte and Y. H. Choi (2013). "Seasonal accumulation of major alkaloids in organs of pharmaceutical crop *Narcissus Carlton*." Phytochemistry **88**: 43-53.
- Lubbe, A., B. Pomahacova, Y. H. Choi and R. Verpoorte (2010). "Analysis of metabolic variation and galanthamine content in *Narcissus* bulbs by ¹H NMR." Phytochemical Analysis **21**(1): 66-72.
- Lubbe, A., R. Verpoorte and Y. H. Choi (2012). "Effects of fungicides on galanthamine and metabolite profiles in *Narcissus* bulbs." Plant Physiology and Biochemistry **58**: 116-123.
- Maeda, H. A. (2019). "Evolutionary Diversification of Primary Metabolism and Its Contribution to Plant Chemical Diversity." Frontiers in Plant Science **10**: 881.
- Maple, J. and S. G. Møller (2007). "Yeast two-hybrid screening." Methods in Molecular Biology **362**: 207-223.
- Matsuura, H. N., S. Malik, F. de Costa, M. Yousefzadi, M. H. Mirjalili, R. Arroo, A. S. Bhambra, M. Strnad, M. Bonfill and A. G. Fett-Neto (2018). "Specialized Plant Metabolism Characteristics and Impact on Target Molecule Biotechnological Production." Molecular Biotechnology **60**(2): 169-183.

- Mazid, M., T. A. Khan and F. Mohammad (2011). "Role of secondary metabolites in defense mechanisms of plants." Biology and Medicine **3**: 232-249.
- Mbandi, S., U. Hesse, D. J. Rees and A. Christoffels (2014). "A glance at quality score: implication for de novo transcriptome reconstruction of Illumina reads." Frontiers in Genetics **5**(17).
- Mohammed, A. E., Z. H. Abdul-Hameed, M. O. Alotaibi, N. O. Bawakid, T. R. Sobahi, A. Abdel-Lateff and W. M. Alarif (2021). "Chemical Diversity and Bioactivities of Monoterpene Indole Alkaloids (MIAs) from Six Apocynaceae Genera." Molecules (Basel, Switzerland) **26**(2): 488.
- Nagahashi, G., G. D. Abney and L. W. Doner (1996). "A comparative study of phenolic acids associated with cell walls and cytoplasmic extracts of host and non-host roots for AM fungi." New Phytologist **133**(2): 281-288.
- Nair, J. J. and J. van Staden (2013). "Pharmacological and toxicological insights to the South African Amaryllidaceae." Food and Chemical Toxicology **62**: 262-275.
- Ncube, B. and J. Van Staden (2015). "Tilting Plant Metabolism for Improved Metabolite Biosynthesis and Enhanced Human Benefit." Molecules **20**(7): 12698-12731.
- Nei, M. and S. Kumar (2000). Molecular evolution and phylogenetics, Oxford university press.
- Nishihachijo, M., Y. Hirai, S. Kawano, A. Nishiyama, H. Minami, T. Katayama, Y. Yasohara, F. Sato and H. Kumagai (2014). "Asymmetric synthesis of tetrahydroisoquinolines by enzymatic Pictet–Spengler reaction." Bioscience, biotechnology, and biochemistry **78**(4): 701-707.
- Nour-Eldin, H. H. and B. A. Halkier (2013). "The emerging field of transport engineering of plant specialized metabolites." Current Opinion in Biotechnology **24**(2): 263-270.
- O'Connor, S. E. and J. J. Maresh (2006). "Chemistry and biology of monoterpene indole alkaloid biosynthesis." Natural Product Reports **23**(4): 532-547.
- Osborn, A. and V. Lanzotti (2009). Plant-derived natural products: Synthesis, function, and application.
- Pallardy, S. G. (2008). CHAPTER 9 - Nitrogen Metabolism. Physiology of Woody Plants (Third Edition). S. G. Pallardy. San Diego, Academic Press: 233-254.

- Park, C. H., H. J. Yeo, Y. E. Park, S. A. Baek, J. K. Kim and S. U. Park (2019). "Transcriptome Analysis and Metabolic Profiling of *Lycoris Radiata*." Biology (Basel) **8**(3).
- Park, J. B. (2014). "Synthesis and characterization of norbelladine, a precursor of Amaryllidaceae alkaloid, as an anti-inflammatory/anti-COX compound." Bioorganic & medicinal chemistry letters **24**(23): 5381-5384.
- Pereira, D. M., P. Valentão, J. A. Pereira and P. B. Andrade (2009). "Phenolics: From Chemistry to Biology." Molecules **14**(6): 2202-2211.
- Petrovska, B. B. (2012). "Historical review of medicinal plants' usage." Pharmacognosy Reviews **6**(11): 1-5.
- Piasecka, A., N. Jedrzejczak-Rey and P. Bednarek (2015). "Secondary metabolites in plant innate immunity: conserved function of divergent chemicals." New Phytologist **206**(3): 948-964.
- Pichersky, E. and R. A. Raguso (2018). "Why do plants produce so many terpenoid compounds?" New Phytologist **220**(3): 692-702.
- Prvulovic, D., H. Hampel and J. Pantel (2010). "Galantamine for Alzheimer's disease." Expert opinion on drug metabolism & toxicology **6**(3): 345-354.
- Ptak, A., E. Morańska, E. Skrzypek, M. Warchoń, R. Spina, D. Laurain-Mattar and M. Simlat (2020). "Carbohydrates stimulated Amaryllidaceae alkaloids biosynthesis in *Leucojum aestivum* L. plants cultured in RITA® bioreactor." Peer Journal **8**(e8688).
- Ptak, A., M. Simlat, E. Morańska, E. Skrzypek, M. Warchoń, A. Tarakemeh and D. Laurain-Mattar (2019). "Exogenous melatonin stimulated Amaryllidaceae alkaloid biosynthesis in in vitro cultures of *Leucojum aestivum* L." Industrial Crops and Products **138**: 111458.
- Pusztahelyi, T., I. J. Holb and I. Pócsi (2015). "Secondary metabolites in fungus-plant interactions." Frontiers in Plant Science **6**: 573.
- Radauer, C. and H. Breiteneder (2007). "Evolutionary biology of plant food allergens." Journal of Allergy and Clinical Immunology **120**(3): 518-525.
- Rambaran, T. F. (2020). "Nanopolyphenols: a review of their encapsulation and anti-diabetic effects." SN Applied Sciences **2**(8): 1335.

- Rosemeyer, H. (2004). "The chemodiversity of purine as a constituent of natural products." Chemistry and Biodiversity **1**(3): 361-401.
- Rueffer, M. and M. H. Zenk (1987). "Distant Precursors of Benzyloisoquinoline Alkaloids and their Enzymatic Formation." Zeitschrift für Naturforschung C **42**(4): 319-332.
- Ruiz, N., D. Ward and D. Saltz (2002). "Calcium oxalate crystals in leaves of *Pancreaticum sickenbergeri*: constitutive or induced defence?" Functional Ecology **16**(1): 99-105.
- Saito, K. and I. Murakoshi (1995). Chemistry, biochemistry and chemotaxonomy of lupine alkaloids in the leguminosae. Studies in Natural Products Chemistry. R. Atta ur, Elsevier. **15**: 519-549.
- Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Molecular Biology and Evolution **4**(4): 406-425.
- Sakano, K. (2001). Metabolic regulation of pH in plant cells: Role of cytoplasmic pH in defense reaction and secondary metabolism. International Review of Cytology, Academic Press. **206**: 1-44.
- Saliba, S., A. Ptak and D. Laurain-Mattar (2015). "4'-O-Methylnorbelladine feeding enhances galanthamine and lycorine production by *Leucojum aestivum* L. shoot cultures." Engineering in Life Sciences **15**(6): 640-645.
- Samanani, N., J. Alcantara, R. Bourgault, K. G. Zulak and P. J. Facchini (2006). "The role of phloem sieve elements and laticifers in the biosynthesis and accumulation of alkaloids in opium poppy." The Plant Journal **47**(4): 547-563.
- Samanani, N. and P. J. Facchini (2001). "Isolation and partial characterization of norcoclaurine synthase, the first committed step in benzyloisoquinoline alkaloid biosynthesis, from opium poppy." Planta **213**(6): 898-906.
- Samanani, N., D. K. Liscombe and P. J. Facchini (2004). "Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzyloisoquinoline alkaloid biosynthesis." The Plant Journal **40**(2): 302-313.
- Samanani, N., S.-U. Park and P. J. Facchini (2005). "Cell type-specific localization of transcripts encoding nine consecutive enzymes involved in protoberberine alkaloid biosynthesis." The Plant Cell **17**(3): 915-926.

- Schatz, G. and B. Dobberstein (1996). "Common principles of protein translocation across membranes." Science **271**(5255): 1519-1526.
- Seremet, O. C., O. T. Olaru, C. M. Gutu, G. M. Nitulescu, M. Ilie, S. Negres, C. E. Zbarcea, C. N. Purdel, D. A. Spandidos, A. M. Tsatsakis, M. D. Coleman and D. M. Margina (2018). "Toxicity of plant extracts containing pyrrolizidine alkaloids using alternative invertebrate models." Molecular Medicine Reports **17**(6): 7757-7763.
- ingh, A. and I. Desgagne-Penix (2017). "Transcriptome and metabolome profiling of *Narcissus pseudonarcissus* 'King Alfred' reveal components of Amaryllidaceae alkaloid metabolism." Nature Scientific Reports **7**(1): 17356-17370.
- Singh, A. and I. Desgagné-Penix (2014). "Biosynthesis of the Amaryllidaceae alkaloids." Plant Science Today **1**(3): 114-120.
- Singh, A., M. A. Massicotte, A. Garand, L. Tousignant, V. Ouellette, G. Berube and I. Desgagne-Penix (2018). "Cloning and characterization of norbelladine synthase catalyzing the first committed reaction in Amaryllidaceae alkaloid biosynthesis." BMC Plant Biology **18**(1): 338.
- Smith, P. M. C. and C. A. Atkins (2002). "Purine Biosynthesis. Big in Cell Division, Even Bigger in Nitrogen Assimilation." Plant Physiology **128**(3): 793-802.
- Springob, K. and T. Kutchan (2009). Introduction to the Different Classes of Natural Products: 3-50.
- Stefano, G. B., N. Pilonis, R. Ptacek and R. M. Kream (2017). "Reciprocal Evolution of Opiate Science from Medical and Cultural Perspectives." Medical Science Monitor **23**: 2890-2896.
- Sun, B., P. Wang, R. Wang, Y. Li and S. Xu (2018). "Molecular Cloning and Characterization of a meta/para-O-Methyltransferase from *Lycoris aurea*." Int J Mol Sci **19**(7).
- TA., H. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT.
- Tahchy, A. E., M. Boisbrun, A. Ptak, F. Dupire, F. Chrétien, M. Henry, Y. Chapleur and D. Laurain-Mattar (2010). "New method for the study of Amaryllidaceae alkaloid biosynthesis using biotransformation of deuterium-labeled precursor in tissue cultures." Acta Biochimica Polonica **57**(1): 75.

- Taiz, L. and E. Zeiger (2002). "Secondary metabolites and plant defense." Plant Physiology **4**: 283-308.
- Takos, A. M. and F. Rook (2013). "Towards a molecular understanding of the biosynthesis of amaryllidaceae alkaloids in support of their expanding medical use." International Journal of Molecular Sciences **14**(6): 11713-11741.
- Tamura, K., F. U. Battistuzzi, P. Billing-Ross, O. Murillo, A. Filipski and S. Kumar (2012). "Estimating divergence times in large molecular phylogenies." Proceedings of the National Academy of Sciences **109**(47): 19333-19338.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: molecular evolutionary genetics analysis version 6.0." Molecular biology and evolution **30**(12): 2725-2729.
- Velderrain-Rodríguez, G. R., H. Palafox-Carlos, A. Wall-Medrano, J. F. Ayala-Zavala, C. Y. Chen, M. Robles-Sánchez, H. Astiazaran-García, E. Alvarez-Parrilla and G. A. González-Aguilar (2014). "Phenolic compounds: their journey after intake." Food and Function **5**(2): 189-197.
- Vuolo, M. M., V. S. Lima and M. R. Maróstica Junior (2019). Chapter 2 – Phenolic Compounds: Structure, Classification, and Antioxidant Power. Bioactive Compounds. M. R. S. Campos, Woodhead Publishing: 33-50.
- Wang, C.-S., J.-C. Huang and J.-H. Hu (1999). "Characterization of two subclasses of PR-10 transcripts in lily anthers and induction of their genes through separate signal transduction pathways." Plant Molecular Biology **40**(5): 807-814.
- Wang, R., X. Han, S. Xu, B. Xia, Y. Jiang, Y. Xue and R. Wang (2019). "Cloning and characterization of a tyrosine decarboxylase involved in the biosynthesis of galanthamine in *Lycoris aurea*." PeerJ **7**: e6729.
- Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva and E. M. Zdobnov (2017). "BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics." Molecular Biology and Evolution **35**(3): 543-548.
- Wen, J., M. Vanek-Krebitz, K. Hoffmann-Sommergruber, O. Scheiner and H. Breiteneder (1997). "The Potential of *Betv1* Homologues, a nuclear multigene family, as phylogenetic markers in flowering plants." Molecular phylogenetics and evolution **8**(3): 317-333.
- Wiedenfeld, H. (2011). "Plants containing pyrrolizidine alkaloids: toxicity and problems." Food Addit Contam Part A Chem Anal Control Expo Risk Assess **28**(3): 282-292.

- Wink, M. (2003). "Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective." Phytochemistry **64**(1): 3-19.
- Wink, M. (2007). Ecological Roles of Alkaloids. Modern Alkaloids: 1-24.
- Wink, M. (2010). Introduction. Annual Plant Reviews Volume 39: Functions and Biotechnology of Plant Secondary Metabolites: 1-20.
- Wink, M. (2010). Introduction: Biochemistry, Physiology and Ecological Functions of Secondary Metabolites. Annual Plant Reviews Volume 40: Biochemistry of Plant Secondary Metabolism: 1-19.
- Wink, M. (2019). "Quinolizidine and Pyrrolizidine Alkaloid Chemical Ecology – a Mini-Review on Their Similarities and Differences." Journal of Chemical Ecology **45**(2): 109-115.
- Yamada, M., K. Ono, T. Hamaguchi and M. Noguchi-Shinohara (2015). "Natural Phenolic Compounds as Therapeutic and Preventive Agents for Cerebral Amyloidosis." Advances in Experimental Medicine and Biology **863**: 79-94.
- Yang, L., K. S. Wen, X. Ruan, Y. X. Zhao, F. Wei and Q. Wang (2018). "Response of Plant Secondary Metabolites to Environmental Factors." Molecules **23**(4).
- Yu, X., A. K. Ekramoddoullah and S. Misra (2000). "Characterization of Pin m III cDNA in western white pine." Tree physiology **20**(10): 663-671.
- Zenk, M. (1965). "Biosynthese von vanillin in *Vanilla planifolia* Andr. Z." Pflanzen Physiol. **53**: 404-414.
- Zhao, H.-x., H.-s. Zhang and S.-f. Yang (2014). "Phenolic compounds and its antioxidant activities in ethanolic extracts from seven cultivars of Chinese jujube." Food Science and Human Wellness **3**(3): 183-190.
- Ziegler, J. and P. J. Facchini (2008). "Alkaloid Biosynthesis: Metabolism and Trafficking." Annual Review of Plant Biology **59**(1): 735-769.
- Zulak, K. G., A. Cornish, T. E. Daskalchuk, M. K. Deyholos, D. B. Goodenowe, P. M. Gordon, D. Klassen, L. E. Pelcher, C. W. Sensen and P. J. Facchini (2007). "Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism." Planta **225**(5): 1085-1106.

ANNEXE A

SUPPLEMENTARY DATA

Transcriptome analysis of *Leucojum aestivum* and identification of genes involved in norbelladine biosynthesis

Laurence Tousignant¹, Aracely Maribel Diaz-Garza¹, Bharat Bhusan Majhi¹,
Sarah-Eve Gélinas¹, Aparna Singh¹, Isabel Desgagne-Penix^{1,2*}

¹Department of Chemistry, Biochemistry and Physics, Université du Québec à
Trois-Rivières, Trois-Rivières, Québec, Canada

²Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières,
Québec, Canada

*Correspondence: Isabel.Desgagne-Penix@uqtr.ca, Telephone: 1-819-376-5011

Table S1 Specific primers for *NBS* used in this study

Primer name	Sequence (5'-3')
<i>LaNBS</i> -Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCGAAGGAGATAATGAAGGGAAGTCTCTCC
<i>LaNBS</i> -Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTGCGCTACAATAGCTTTTTGCTC
<i>NpNBS</i> -Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCGAAGGAGATAATGAAGGGAAGTCTGTCC
<i>NpNBS</i> -Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGCTACAGTAGCTTTTTTCC

Table S2 Comparison of signal peptides, localization predictions and experimental cell localizations of *LaNBS*, *NpNBS1* and *NpNBS2*

Gene name	Length (AA)	Signal peptide				Localization prediction			Cell localization (this study)
		SignalP		TargetP	ChloroP	Loctree3	PSLpred	WoLF PSORT	
		SP	Other	LOC	cTP				
<i>LaNBS</i>	161	0.002	0.998	O	N	CYTO	CYTO	CYTO EXTR CYSK	CYTO/NUC
<i>NpNBS1</i>	161	0.003	0.997	O	N	CYTO	CYTO	CYTO EXTR ER	CYTO/NUC
<i>NpNBS2</i>	161	0.003	0.997	O	N	CYTO	CYTO	CYTO EXTR GOLG	CYTO/NUC

AA, amino acids; SP, signal peptide; LOC, localization; cTP, chloroplast transit peptide; mTP, mitochondrial transit peptide; O, other location (except SP, cTP and mTP); N, no; CYTO, cytoplasm; EXTR, extracellular; CYSK, cytoskeleton; ER, endoplasmic reticulum; GOLG, golgi; NUC, nucleus

Table S3 Method used in high-performance liquid chromatograph (HPLC)

Time (min)	10 mM Ammonium acetate (pH 5.0) (%)	Acetonitrile (%)	Rate of flow (mL/min)
0	60	40	0.40
7	0	100	
9	0	100	
10	60	40	
15	60	40	

Table S4 Parameters used for the tandem mass spectrometer source

Parameter	Value used
Gas flow (L/min)	5
Gas temperature (°C)	300
Nebulizer (psi)	45
Sheat gas flow (L/min)	11
Sheat gas temperature (°C)	300
Capillary voltage (V) (+ / -)	3000/3500
Nozzle voltage (V) (+ / -)	500/500

Table S5 MRM transitions and optimized instrumental parameters used for MS/MS analysis in ESI+ and ESI-. Quantitation fragments are bold while those of confirmation are not

Compound	Transition MRM (m/z)	Retention time (min)	Fragmentor (V)	Collision energy (eV)	Polarity
3,4-DHBA	137 → 108	3.865	114	26	Negative
	137 → 92	3.867	114	26	Negative
Norbelladine	260 → 138	3.444	86	5	Positive
	260 → 121	3.446	86	17	Positive
Tyramine	138 → 77	3.235	81	20	Positive
	138 → 121	3.234	81	5	Positive
Papaverine	340 → 202	7.040	165	27	Positive
	340 → 324	7.038	165	31	Positive

Fig. S1 Chromatographic analysis of compounds obtained from an acid-base extraction of different tissues of *L. aestivum*. Thin-layer chromatography of extracts developed at 254 nm (left panel) and at 365 nm (right panel).

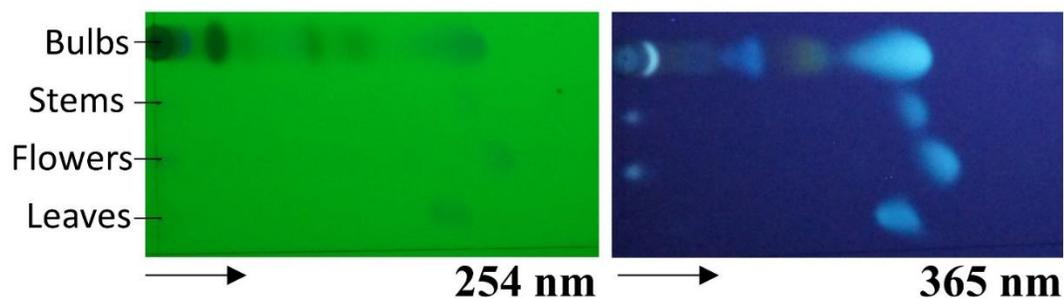


Fig. S2 Transcriptomic pipeline.

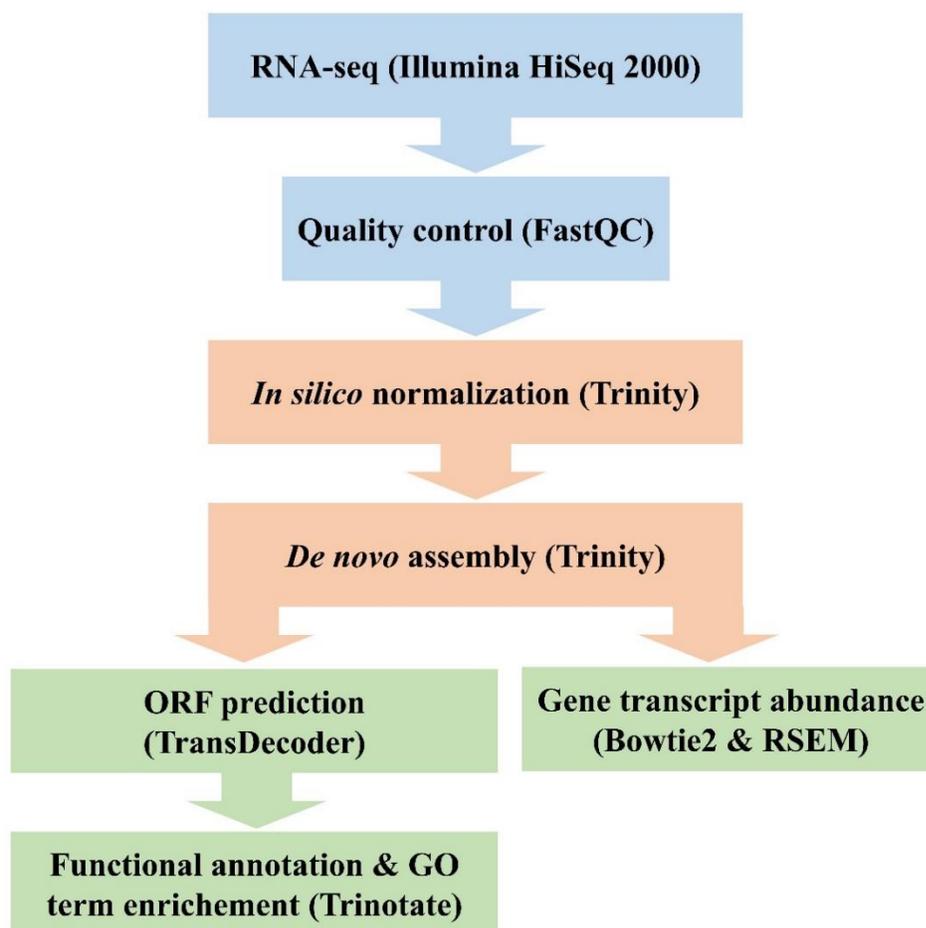


Fig. S3 Plasmid constructs and cloning of NBS from *Leucojum aestivum* and *Narcissus papyraceus* used in this study for localization analysis. (a) Representation of *La*NBS, *Np*NBS1, *Np*NBS2 and 124466 plasmid constructs. **(b)** Agarose gel of PCR amplified candidate *Np*NBS from *N. papyraceus*. Product size is 548 bp (ORF 477 bp + Gateway segments 71 bp). Numbers on the left refer to the location of standard DNA molecular markers in bp. **(c)** Agarose gel of PCR amplified candidate *La*NBS from *L. aestivum*. Product size is 548 bp including the Gateway adapters, L is for ladder. **(d)** Agarose gel of PCR amplified *La*NBS, *Np*NBS1 and *Np*NBS2 performed with positive colonies after transformation into *E. coli* DH5 α . Lanes 1 and 2 is *La*NBS, lane 3 is *Np*NBS2, lanes 4 and 5 is *Np*NBS1 and lane 6 is empty pB7FWG2 vector as negative control. All product sizes are 548 bp.

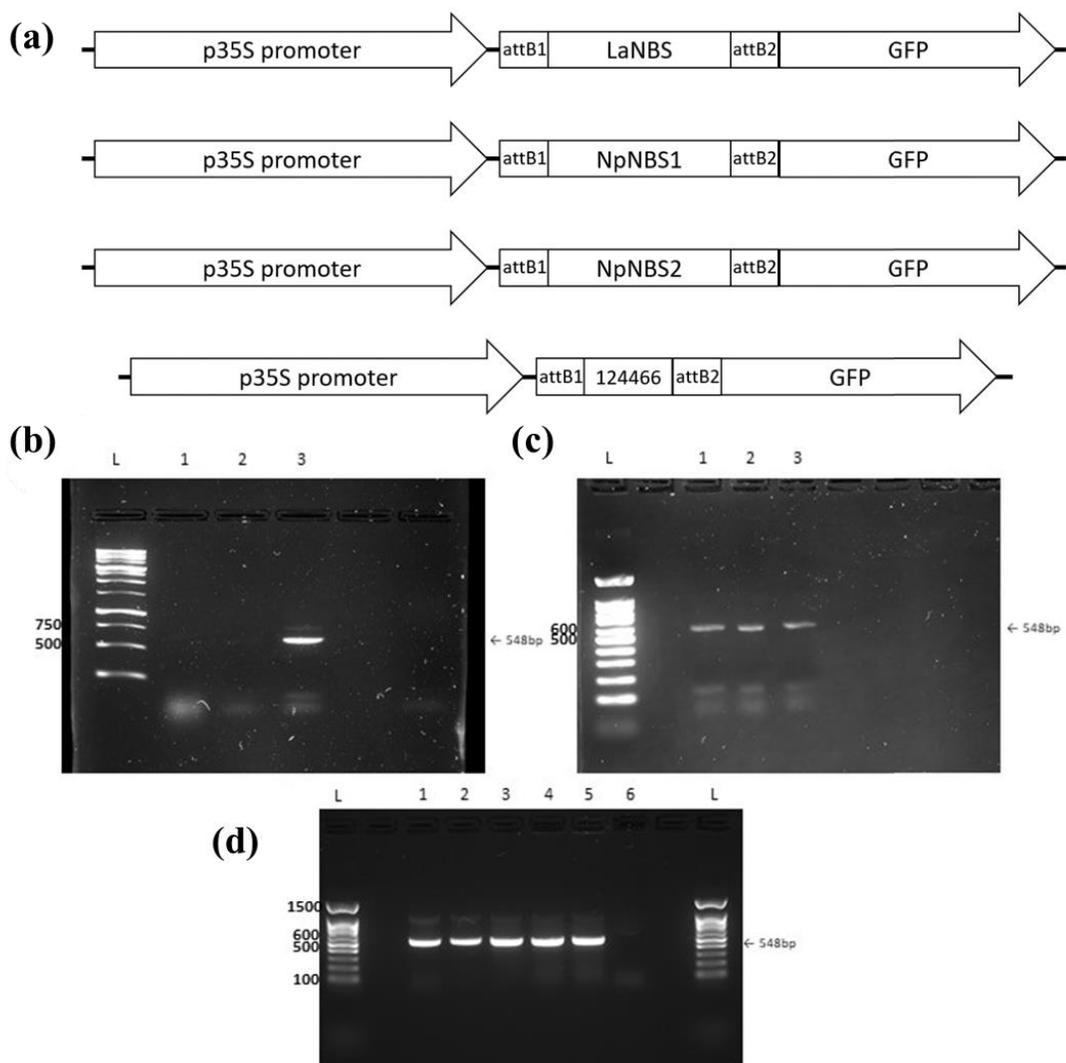


Fig. S4 Protein expression of GFP-fusion *La*NBS, *Np*NBS1 and *Np*NBS2 protein *in planta*. Expression in *N. benthamiana* leaves of the *La*NBS, *Np*NBS1 and *Np*NBS2 proteins fused to the C-terminal GFP tag was detected by Western blot analysis with anti-GFP antibody (α :GFP). The fusion proteins showed the expected size of 44 kDa (NBS-17 kDa + GFP-27 kDa). The Ponceau S staining of the membrane is shown as a loading control. Numbers on the right refer to the location of standard protein molecular weight markers in kDa.

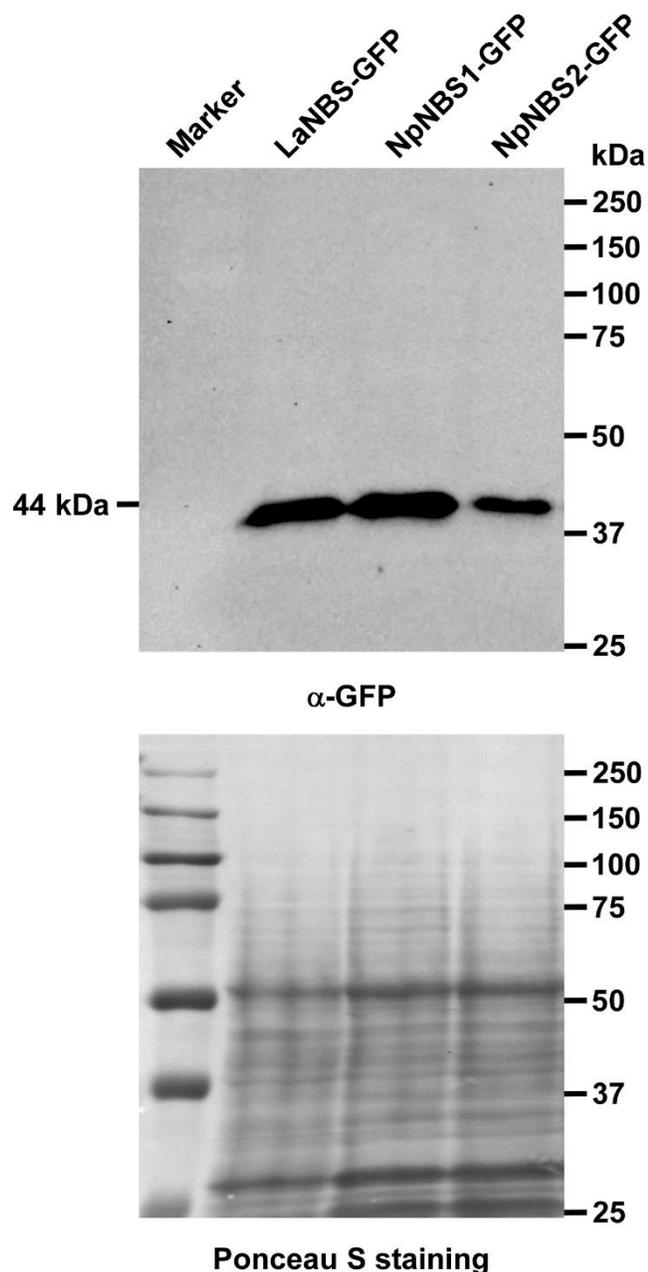
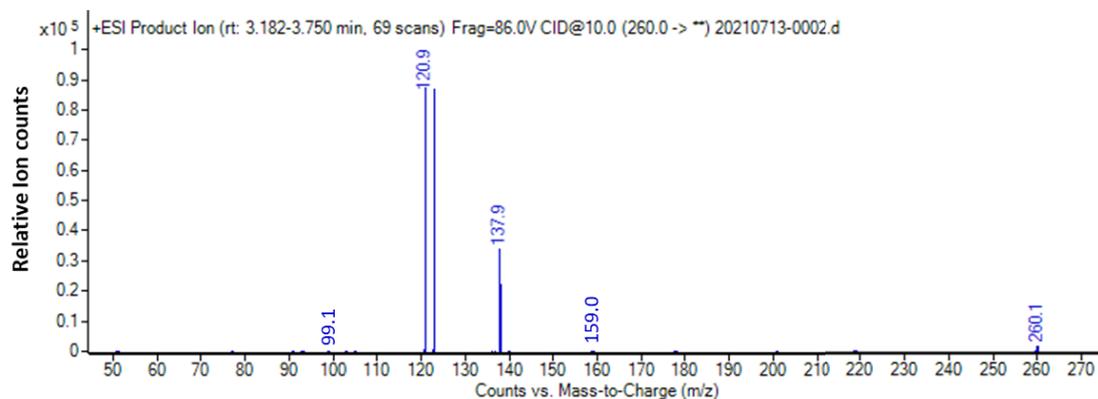
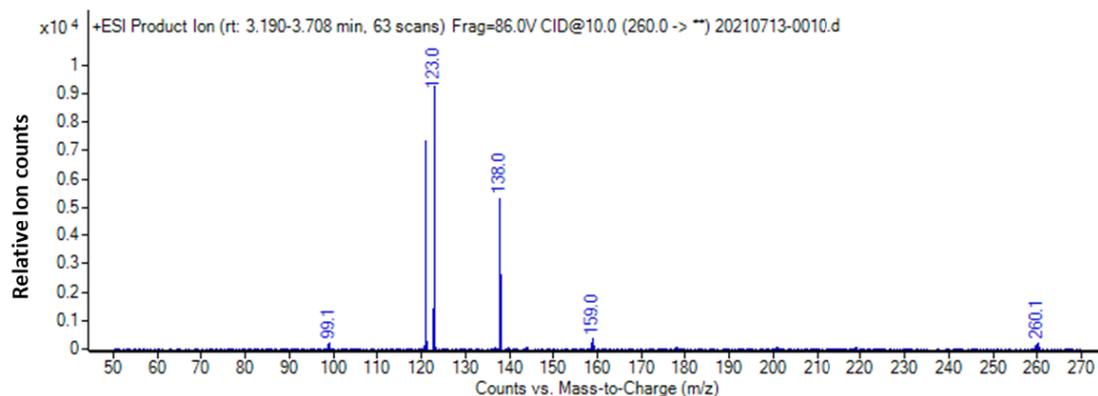


Fig. S5 Mass spectra of total ion chromatogram (TIC) obtained after the fragmentation using collision-induced dissociation (CID) of 10 electron Volts (eV) of the norbelladine standard compared to the norbelladine produced by *LaNBS in vitro* enzyme assay

a) Norbelladine (standard)



b) Norbelladine (produced in enzyme assay)



ANNEXE B

CLONING AND CHARACTERIZATION OF NORBELLADINE SYNTHASE CATALYZING THE FIRST COMMITTED REACTION IN AMARYLLIDACEAE ALKALOID BIOSYNTHESIS

Aparna Singh¹, Marie-Ange Massicotte¹, Ariane Garand¹, Laurence Tousignant¹,
Vincent Ouellette¹, Gervais Bérubé¹ and Isabel Desgagné-Penix^{1,2}

¹Department of Chemistry, Biochemistry and Physics, Université du Québec à
Trois-Rivières, 3351 boul. des Forges, Trois-Rivières, QC, G9A 5H7, Canada

²Plant Biology Research Group, Université du Québec à Trois-Rivières, 3351 boul. des
Forges, Trois-Rivières, QC, G9A 5H7, Canada

Author for correspondence: Isabel.Desgagne-Penix@uqtr.ca

Tel: 819-376-5011

Fax: 819-376-5014

Email addresses:

AS: Aparna.Singh@uqtr.ca

MAM: Marie-Ange.Massicotte@uqtr.ca

LT: Laurence.Tousignant@uqtr.ca

AG: Ariane.Garand@uqtr.ca

VO: Vincent.Ouellette@uqtr.ca

GB: Gervais.Berube@uqtr.ca

IDP: Isabel.Desgagne-Penix@uqtr.ca

Abstract

Background: Amaryllidaceae alkaloids (AAs) are a large group of plant-specialized metabolites displaying an array of biological and pharmacological properties. Previous investigations on AA biosynthesis have revealed that all AAs share a common precursor, norbelladine, presumably synthesized by an enzyme catalyzing a Mannich reaction involving the condensation of tyramine and 3,4-dihydroxybenzaldehyde. Similar reactions have been reported. Specifically, norcoclaurine synthase (NCS) which catalyzes the condensation of dopamine and 4-hydroxyphenylacetaldehyde as the first step in benzyloisoquinoline alkaloid biosynthesis.

Results: With the availability of wild daffodil (*Narcissus pseudonarcissus*) database, a transcriptome-mining search was performed for *NCS* orthologs. A candidate gene sequence was identified and named *norbelladine synthase (NBS)*. *NpNBS* encodes for a small protein of 19 kDa with an anticipated pI of 5.5. Phylogenetic analysis showed that *NpNBS* belongs to a unique clade of PR10/Bet v1 proteins and shared 41% amino acid identity to opium poppy NCS1. Expression of *NpNBS* cDNA in *Escherichia coli* produced a recombinant enzyme able to condense tyramine and 3,4-DHBA into norbelladine as determined by high-resolution tandem mass spectrometry.

Conclusions: Here, we describe a novel enzyme catalyzing the first committed step of AA biosynthesis, which will facilitate the establishment of metabolic engineering and synthetic biology platforms for the production of AAs.

Keywords: Amaryllidaceae alkaloid, pathogenesis related protein 10, alkaloid biosynthesis, *Narcissus pseudonarcissus*, norcoclaurine synthase, norbelladine.

Background

The Amaryllidaceae alkaloids (AAs) are a group of naturally synthesized molecules with more than 600 renowned complex structures (Jin and Xu 2013). They are

pharmacologically active compounds that are classified under three different groups of C-C phenol coupling namely *para-para'*, *ortho-para'* and *para-ortho'* (Singh and Desgagné-Penix 2014). An outsized variety of pharmacologically active AAs have been identified with the bioactive properties including the acetylcholine esterase inhibitor galanthamine, anti-tumor activity of lycorine and the cytotoxic haemanthamine (Li, Dai et al. 2012, He, Qu et al. 2015, Hotchandani and Desgagne-Penix 2017). AAs are obtained chiefly from the extracts of plants from *Galanthus*, *Leucojum* and *Narcissus* species, as their complicated structures do not enable cost-effective high-yield organic synthesis (Saliba, Ptak et al. 2015). Though AAs display a large range of pharmaceutical applications, only galanthamine is accessible in markets as an Alzheimer's treatment drug because of its ability to stabilize behavioral symptoms in the course of six months treatment in comparison to chemically synthesized acetylcholinesterase inhibiting drugs, donepezil and rivastigmine (Prvulovic, Hampel et al. 2010).

Previous investigations on the biosynthesis of AAs *in planta* have revealed that all AAs are made from the common metabolic intermediate, norbelladine (Fig. 1) (Barton and Cohen 1957, Battersby, Fales et al. 1961, Barton, Kirby et al. 1963, Grisebach 1973, Eichhorn, Takada et al. 1998, Tahchy, Boisbrun et al. 2010, El Tahchy, Ptak et al. 2011). For example, radiolabeled studies showed that deuterium labelled 4'-*O*-methylnorbelladine was incorporated in all three different groups of AAs (Tahchy, Boisbrun et al. 2010). To date, only three AAs biosynthetic genes have been identified including the *norbelladine 4'-O-methyltransferases (N4OMT)*, encoding N4OMT enzyme catalyzing norbelladine methylation at 4' position to form 4'-*O*-methylnorbelladine (Kilgore, Augustin et al. 2014), *CYP96T1*, encoding a cytochrome P450 enzyme which catalyzes the synthesis of (*S,R*)-noroxomaritidine (Kilgore, Augustin et al. 2016) and *noroxomaritidine reductase (NR)* encoding the enzyme that catalyzes the formation of oxomaritinamine (Kilgore, Holland et al. 2016). However, to this date, the enzyme catalyzing the first committed step leading to norbelladine synthesis has not been identified.

It has been proposed that norbelladine is formed by the condensation of tyramine and 3,4-dihydroxybenzaldehyde (3,4-DHBA) via a Mannich reaction (Sobaro-Sanchez, 2015). The involvement of the Mannich reaction has been proposed in many biosynthetic pathways, especially for alkaloids. For example, norcoclaurine synthase (NCS), catalyzing the first committed step in benzyloquinoline alkaloids formation, condenses the two precursors molecules; dopamine and 4-hydroxyphenylacetaldehyde to form norcoclaurine (Samanani, Liscombe et al. 2004). Since the AA biosynthetic pathway shares analogy with the benzyloquinoline alkaloid biosynthesis, we proposed that a *NCS* ortholog in Amaryllidaceae may be a candidate that is capable of catalyzing the formation of norbelladine.

Hence for this study, we applied BLAST searches for candidate sequences orthologous to *NCS* in our previously generated *de novo* transcriptome from wild daffodil (*Narcissus pseudonarcissus*) (Singh and Desgagné-Penix 2017). The *NCS* orthologous sequence identified named *norbelladine synthase (NBS)* was isolated, cloned in an expression vector and transformed into *E. coli*. Following heterologous expression and purification, NBS recombinant protein was used for enzyme assay and LC-MS/MS analysis of the product showed that NBS has the ability to convert 3,4-DHBA and tyramine to norbelladine. Thus, we report on the first characterization of norbelladine synthase, the enzyme which catalyzes the first committed step in AAs biosynthesis.

Results

Candidate gene identification

RNA from *N. pseudonarcissus* 'King Alfred' bulb was isolated and corresponding cDNA was used for Illumina sequencing, which provided a deep transcriptome database to search for expressed genes encoding enzymes involved in AA biosynthesis (Singh and Desgagné-Penix 2017). To identify orthologous sequence to *NCS* in *N. pseudonarcissus* transcriptome, we performed BLASTx sequence similarity searches with *NCS* from *Thalictrum flavum*, *Papaver somniferum*, *Argemone Mexicana*, *Eschscholzia californica*,

Chelidonium majus and *Coptis japonica*). Three resulting orthologous full-length transcripts were obtained (TR17354|c0_g1_i1, TR17354|c0_g1_i2 and TR17354|c0_g1_i3) consisting of 593, 575 and 583 bp with corresponding open reading frames of 492, 480 and 483 bp, respectively. Their E-value to NCS ranges between $1e^{-20}$ to $4e^{-20}$. The three transcripts were 97% identical, though small differences were observed towards the 3' end (Additional file 1). The candidate displaying the longest predicted ORF of 492 bp was named *N. pseudonarcissus Norbelladine Synthase (NpNBS)* and was used for subsequent functional analysis. The ORF encoded protein consisted of 163 amino acids with an anticipated molecular weight of 19 kDa and a theoretical pI of 5.5. BLAST and motif scan analyses revealed the presence of conserved Bet v1 and Pathogenesis-Related (PR-10) protein domain. The PR-10/Bet v1 allergen proteins do not possess NCS activity although they share similarity to NCS protein. Thus, *NpNBS*, similarly to NCS, belongs to the PR-10/Bet v1 protein family.

Phylogenetic analysis

To investigate the evolutionary history of *NpNBS* and its relationship with PR-10/Bet v1 proteins, a phylogenetic tree was constructed with the predicted amino acid sequence of *NpNBS*. Although all proteins are members of the PR-10 protein family, the phylogenetic tree was clearly divided into two main clusters: the PR-10 and the Norcochlorine synthases (Fig. 2; Additional file 2). This type of clustering was reported previously (Lee and Facchini 2010). Notably, *NpNBS* shares higher homology with proteins from the NBS cluster including 38% amino acid sequence identity with *Thalictrum flavum TfNCS1/2* and 41% with *Papaver somniferum PsNCS1/2*. *NpNBS* displayed comparatively lower identity to proteins of the PR-10 cluster including 19% identity to *Betula verrucosa BvPR10* and only 7% to *Hyacinthus orientalis HoPR10*. Although *NpNBS* sequence is located the edge of the PR-10/NCS cluster, it is more closely related with the NCS group since it showed higher homology with *TfNCS* (Fig. 2). Altogether the results suggest that NBS, NCS and PR-10 evolved as distinctive genes over time.

To identify regions of similarity that may be important for defining functional and structural relationships, multiple sequences alignment were performed using *NpNBS* and the NCS amino acid sequences from various plant species (Fig. 3). Thus, *EcNCS1*, *PsNCS1*, *CjNCS2* (formely *CjPR10A*) and *TfNCS1* genes shared over 38% identity in amino acid sequences with each other, while they showed the same identity (38%) with *NpNBS*. The alignment also showed the presence of NCS conserved catalytic residues Tyr, Glu, Lys and Asp (Fig. 3). The residues were renamed as Tyr, Glu, Lys, according to their position in *NpNBS* protein sequence. However, the NCS Asp was not found in *NpNBS* sequence and was replaced by an isoleucine residue. All amino acid sequences contained a phosphate-binding loop (P-loop) glycine-rich region (Fig. 3). The glycine-rich loop is the conserved ligand-binding domain of Bet v1 protein family (Fernandes, Michalska et al. 2013). Subcellular localization of the *NpNBS* amino acid sequence was predicted using several programs, including SignalP4.1 and WoLF PSORT, and the result indicated the absence of signal peptides. In addition, localization predictor program (PSLPred) identified *NpNBS* as a cytosolic protein. Altogether, the data suggest that *NpNBS* proteins are localized in cytoplasm.

RT-qPCR analysis

The study of *NpNBS* expression profile in different tissues of *N. pseudonarcissus* ‘King Alfred’ by reverse transcription quantitative PCR (RT-qPCR) revealed high expression of *NpNBS* in bulbs compared to roots, stems, leaves and flowers (Fig. 4). The normalized ddCT expression of *NpNBS* in other tissues was detected below 60, whereas its expression in bulbs was approximately 1500 folds higher. High expression of *NpNBS* relates to the higher AAs content in bulbs of *N. pseudonarcissus* ‘King Alfred’ (Singh and Desgagné-Penix 2017).

***NpNBS* gene cloning and heterologous expression**

To produce recombinant *NpNBS*, a full-length *NpNBS* cDNA was PCR amplified from *N. pseudonarcissus*. The amplified product of *NpNBS* shows a band of 566 bp on the

1% agarose electrophoresis gel (Fig. 5A). *NpNBS* was gel purified, cloned into the pET301 expression vector with a C-terminal 6x-His tag, transformed into *E. coli* competent cells and the positive colonies were selected on LB supplemented with ampicillin and chloramphenicol. Colony PCR confirmed the presence of positive colonies (Fig. 5B) which were induced with IPTG at 37 °C and protein fractions were extracted. SDS-PAGE analysis was performed with different fractions obtained during *NpNBS* protein purification process which show bands of different sizes, suggesting the presence of numerous bacterial cell proteins in fractions from the crude (Cr), lysate (L1 and L2) and wash buffer (W1 and W2). An apparent molecular mass of 19 kDa in elution buffer 1 (E1) was observed, which was absent in non-induced protein (NI) (Fig. 5C). Western blot analysis was performed using 6x His-tag monoclonal antibody. A His-tag *NpNBS* crude (Cr) and purified extracts (E1) isolated from transformed bacterial cell culture grown at 37 °C showed the expected protein at molecular weight of 19 kDa. No protein expression was observed in cultures without IPTG (NI) (Fig. 5D). The results showed successful production of recombinant *NpNBS* from IPTG-induced bacterial cell culture.

***NpNBS* enzyme assay**

For the biosynthesis of all AAs, the precursor norbelladine is made from a the condensation of tyramine and 3,4-DHBA which forms the imine norcraugsodine followed by its reduction to norbelladine [17]. It is not clear if norbelladine formation is one stepwise reaction catalyzed by a single enzyme or two separate reactions (condensation and reduction) catalyzed by different enzymes. To confirm *NpNBS* protein function, enzyme assays were carried out with purified *NpNBS* recombinant protein. The resulting assay product was subjected to LC-MS analysis using a Positive Electrospray Ionization mode (+ESI). The QqQ dual +ESI source conditions were optimized using freshly synthesized alkaloid standards, norbelladine and norcraugsodine (Additional file 3) to obtain a high sensitivity. The MS/MS parameters such as capillary voltage, spray voltage and skimmer voltage were enhanced to maximize the ionization in the source and sensitivity to identify and characterize all possible fragmentation products. With standards, we observed predicted major mass spectral fragments for the nobelladine

m/z 260 $[M+H]^+$ and norcraugsodine m/z 258 $[M+H]^+$, both at 5.5 min (Additional file 3). Fragmentation of norbelladine molecular ion m/z 260 $[M+H]^+$ yielded ion fragments of m/z 159, 138, 123 and 121 (Additional file 3). The ion fragment m/z 138 was obtained by the elimination of the 4-ethylphenol moiety (122 Da) whereas m/z 121 was produced by loss of tyramine. Precisely for norbelladine, the transition $[M+H]^+$ of 260→138 was selected for the qualifier ion fragment and the transition $[M+H]^+$ of 260→121 was used as the quantifier ion (Additional file 3). Similar MS/MS ion fragments spectra were reported for norbelladine suggesting a good fragmentation of our standard (Kilgore, Holland et al. 2016).

Enzyme assays with *Np*NBS yielded a peak at 5.5 minute on LC-MS which was the same retention time as those of the norbelladine and norcraugsodine standards. The +ESI-MS/MS analysis of the *Np*NBS enzyme assay product showed the presence of the qualifier and quantifier ion fragments of norbelladine (Fig. 6). The level of product was detected in the *Np*NBS assay pH 4 and at lower levels at pH 7. Norcraugsodine molecular ion and fragment ions were not detected in the assay. It should be noted that a very low background reaction was observed in control assay/non-induced protein assay (*i.e.* when the two substrates were in contact without *Np*NBS) suggesting a low level of non-enzymatic condensation. However, these levels were 1000 times lower than those measured in presence of *Np*NBS. *Np*NBS enzyme assay product shows that *Np*NBS catalyzes the condensation reaction between tyramine and 3,4 DHBA to produce norbelladine. Thus, the *in vitro* enzyme assay demonstrated ability of norbelladine synthase activity of *Np*NBS with a preference at pH 4.

Discussion

The formation of norbelladine is a crucial step in the formation of Amaryllidaceae alkaloids. In this study, we identified the enzyme norbelladine synthase in wild daffodil (*N. pseudonarcissus*) responsible for the condensation of tyramine with 3,4-dihydroxybenzaldehyde (3,4-DHBA) to form norbelladine. *Np*NBS contains conserved Bet v1 domain and thus is a member of the Bet v1/PR10 protein family.

The PR-10 family universally exist in plants of monocots and dicots group (Wen, Vanek-Krebitz et al. 1997, Wang, Huang et al. 1999, Yu, Ekramoddoullah et al. 2000). Apart from the PR-10 family, recently Bet v1 have shown an extended low level similarity with three more protein families; major latex and ripening-related proteins, norcoclaurine synthase proteins, and cytokinin-binding proteins from legumes (Radauer and Breiteneder 2007). Several studies have also demonstrated that members of the Bet v1/PR-10, specifically of the NCS group, have a crucial role in alkaloid production (Samanani, Park et al. 2005, Zulak, Cornish et al. 2007). For example, *PsNCS* from opium poppy shares significant amino acid identity with PR-10 and Bet v1 protein and NCS has a major role in benzyloquinoline alkaloid biosynthesis. Indeed, Virus-Induced Gene Silencing (VIGS) of *PsNCS1* and *PsNCS2* transcripts resulted in a 75% to 82% decrease of main alkaloids in opium poppy latex compared to control (Lee and Facchini 2010).

NCS catalyzes the condensation of dopamine with 4-hydroxyphenylacetaldehyde to yield (*S*)-norcoclaurine, the central precursor of benzyloquinoline alkaloids. *NpNBS* was identified from transcriptome data search using *NCS* sequence homology. The high level of *NpNBS* expression in bulbs correlates with elevated content (and diversity) of AAs compared to other plant parts suggesting the importance of *NpNBS* in AA metabolism. Interestingly, UPLC-QTOF-MS detected 4-hydroxybenzaldehyde, the pathway precursor to 3,4-DHBA the co-substrate for *NpNBS*, only in underground tissues such as bulb and roots (Singh and Desgagné-Penix 2017). *NpNBS* shared 41% amino acid sequence identity with *P. somniferum* NCS1 and NCS2 but lacked the signal peptide that could be involved in subcellular localization. Furthermore, motif searches predicted that *NpNBS* is a cytosolic protein and is not associated with a subcellular compartment such as endoplasmic reticulum, mitochondria or peroxisomes. The observed catalytic activity of *NpNBS* confirmed its role in AA biosynthesis. The shorter N-terminal region of *NpNBS*, 41 amino acid less compared to *TfNCS* (Fig. 3), and the absence of the signal peptide in the sequence did not affect its activity suggesting that it is not required. These results are consistent with studies conducted on *TfNCS* which reported that a signal peptide had no role in the catalytic activity of an enzyme. Samanani et al. (2004) (Samanani, Liscombe et al. 2004) studied the activity of *TfNCS* protein encoding truncation of first 10 and

29 N-terminal amino acids which resulted in formation of (*S*)-norcoclaurine. Both recombinant proteins displayed enzymatic activity (Samanani, Liscombe et al. 2004). In addition, a study conducted on four N-terminally truncated variants of *CjNCS* of 10, 19, 29 and 42 amino acids residues revealed an increasing fold of enzymatic activity. For example, *CjNCS*- Δ 19 displayed 10-fold (0.30 U/mL) and *CjNCS*- Δ 29 had shown 40-fold (1.20 U/mL) higher activity compared to full length *CjNCS* (0.03 U/mL) (Nishihachijo, Hirai et al. 2014). Recently, Li et al. (2016) reported on the isolation and characterization of NCS variants from several alkaloid-producing plant species. They found that in some case, NCS orthologs possess two, three or four repeated catalytic domains and their presence was found to be proportional with increase in catalytic efficiency (Li, Lee et al. 2016).

NpNBS enzyme catalyzes norbelladine synthesis by a condensation process between amine and aldehyde. Among the proposed catalytic residues, the Glu was detected with +1 histidine shift in *NpNBS* which might affect the interaction with catalytic and hinder the cyclization process. Moreover, *NpNBS* ability to form norbelladine in the absence of Asp¹⁴¹ catalytic residues suggests that Asp¹⁴¹ residue is not responsible for basic catalysis and might be involved in imparting electrostatic stabilization (Lichman, Gershater et al. 2015). The *NpNBS* enzyme displays higher activity at pH 4 indicating that the reaction may prefer an acidic environment. Similarly, pH of 6.2 has been reported for optimum activity of *PsNCS* (Samanani and Facchini 2001).

The high expression of *NpNBS* in bulbs is supported by its high FPKM value obtained in the transcriptome (Singh and Desgagné-Penix 2017). *NpNBS* was reported among the top expressed genes, comprising 16.12% of the *N. pseudonarcissus* 'King Alfred' transcriptome. *NpNBS* high expression was consistent with the expression of *NpN4OMT* in bulbs, which supported the abundant occurrence of AAs in bulb indicating that bulb is an important site for enzyme localization and AAs biosynthesis (Singh and Desgagné-Penix 2017). Moreover, Kilgore et al. 2014, also presented a similar correlation between high expression and read count of *N4OMT* with abundant galanthamine accumulation in bulb tissues (Kilgore, Augustin et al. 2014). Therefore, based on the

expression data, we conclude that *Np*NBS has a crucial role in AAs metabolism. A similar High *Tf*NCS expression and enzymatic activity was reported in underground rhizome and root tissues compared to other parts, suggesting tissues a specific localization of NCS enzyme (Samanani, Liscombe et al. 2004).

Conclusion

The rapid progress in the discovery of AAs biosynthesis pathway genes along with newly characterized *Np*NBS enzyme is useful for reconstituting a short synthetic AAs biosynthesis pathway in yeasts or microbes that will allow feasible and high-scale production of targeted medicinally important Amaryllidaceae alkaloids which are gaining interest for their new pharmaceutical applications, including an Alzheimer's treatment drug, galanthamine (Heinrich and Lee Teoh 2004, Barik, Dajas-Bailador et al. 2005). The applied strategy could be implemented in characterization of genes from any class of specialized metabolites as our discovery shows the utility of the designed workflow.

Although the origin of alkaloid biosynthesis is not always clear, the key genes involved in these pathway are gradually being isolated. For example, norcoclaurine synthase couples dopamine and 4-hydroxyphenylacetaldehyde as the first committed step in benzyloquinoline alkaloid biosynthesis (Samanani, Liscombe et al. 2004, Lee and Facchini 2010). Another example is strictosidine synthase, which catalyzes the first specific step in monoterpene indole alkaloid biosynthesis (Kutchan 1993). *Np*NBS could offer a new perspective for studies on the diversity and evolution of alkaloid biosynthesis. The condensation of tyramine and 3,4-DHBA is central to the biosynthesis of myriad AAs in plants. We demonstrated that the enzyme responsible for this reaction is a member of the PR10/Bet v1 family of proteins. Further work on *Np*NBS will include attempts to determine the optimum pH and temperature and to probe the scope of alternate substrates in the reaction. The use of this enzyme in biotechnological microbial platforms will provide an attractive alternative to the production of AAs, including several pharmacologically important compounds such as galanthamine and lycorine.

Methods

Plant tissue and chemicals

Wild daffodil (*Narcissus pseudonarcissus* ‘King Alfred’) bulbs were purchased from Fraser’s Thimble farms (BC, Canada). Bulbs were planted in the field in Trois-Rivieres (Québec, Canada) during the month of October and were not harvested until flowering stage in early May. Different tissues such as bulbs, roots, stems, leaves and flowers were collected separately, flash frozen in liquid nitrogen and stored at -80°C . Ampicillin, HPLC grade acetonitrile, agarose, methanol, 6x His-tag epitope tag antibody and monosodium phosphate were purchased from Fisher Scientific (Janssen Pharmaceuticaan, Geel, Belgium). 3, 4-dihydroxybenzaldehyde (3,4-DHBA) were obtained from Acros organic (New Jersey, USA). Tyramine and chloramphenicol were bought from Sigma-Aldrich (MO, USA). *Taq* DNA polymerase was purchased from gene direx. Sodium chloride (NaCl), isopropyl β -D-1-thiogalactopyranoside (IPTG), and imidazole were bought from Fisher Bioreagents/scientific (New Jersey, USA). The SensiFAST SYBER Lo-Rox kit for reverse transcription-quantitative PCR was obtained from Bioline (London, U.K). The Gateway cloning kit was purchased from Invitrogen (CA, USA). Ni-NTA his-tag affinity columns were purchased from Qiagen (Germany), plasmid miniprep kit acquired from Geneaid (Taipei, Taiwan). Mini-protean TGX stain-free precast gels, 4x-Laemmli buffer, 10-250 kD precision plus kaleidoscope prestained protein standard ladder and clarity western ECL substrate were obtained from Bio-rad (USA) and kanamycin was purchased from Bioshop (Burlington, ON, Canada).

RNA extraction, Illumina sequencing and transcriptome assembly

The transcriptome was generated in a previously published study by Singh and Desgagné-Penix, 2017 (Singh and Desgagné-Penix 2017). Briefly, total RNA from bulbs of *N. pseudonarcissus* ‘King Alfred’ was extracted using the Cetyltrimethylammonium bromide method, converted to cDNA and sequenced using Illumina HiSeq 2000 PE. The raw pair reads were trimmed, cleaned, normalized, and assembled into a

transcriptome (Singh and Desgagné-Penix 2017). Of 73,081,603 raw paired reads, a total of 10,523,999 surviving paired reads after normalization were assembled into 195,347 transcripts. The sequences were deposited in the National Center for Biotechnology Information Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/?term=SRR5788585>) under the accession number SRR5788585.

Candidate gene identification

The sequence for norbelladine synthase was deposited to GenBank and the GenBank accession number for the nucleotide sequence is MG948545.

Phylogenetic tree and protein alignment

Sequences listed in Additional file 2 were aligned using CLUSTAL W in MEGA 6 software with default parameters. The evolutionary history was inferred using the Neighbor-Joining method using the branch lengths contained in the inferred tree (Saitou and Nei 1987). Divergence times for all branching points in the topology were calculated with the RelTime method (Tamura, Battistuzzi et al. 2012). Phylogenetic analysis was conducted on MEGA 6 (Tamura, Stecher et al. 2013). Protein sequence of *NpNBS* was aligned with norcoclaurine synthase sequences from *Coptis japonica* (*CjNCS2*), *Papaver somniferum* (*PsNCS1*), *Eschscholzia californica* (*EcNCS1*) and *Thalictrum flavum* (*TfNCS1*) using Clustal omega. *NpNBS* sequence was analyzed for signal peptide using Signal-BLAST (<http://sigpep.services.came.sbg.ac.at/signalblast>) and PSLpred (<http://crdd.osdd.net/raghava/pslpred>). Motif scan search for *NpNBS*, *PsNCS1*, *BtPR10* and *CjNCS* sequences was performed using myhits (http://myhits.isb-sib.ch/cgi-bin/motif_scan).

Reverse transcription-quantitative PCR (RT-qPCR)

N. pseudonarcissus cDNA for bulbs, roots, stems, leaves, and flowers was generated from 1-2 µg of RNA using the Qiagen omniscrypt RT kit according to manufacturer's protocol (QIAGEN, Germany). The experiment was performed in triplicate. A total reaction volume of 20 µL containing 1x SensiFAST SYBR Lo-ROX mix, 200 µM of each forward and reverse primers (Additional file 4) and cDNA sample was used for RT-qPCR analysis. Histone was used as internal reference gene. Real-time quantitative PCR was performed on CFX Connect RT-qPCR System from Bio-rad (USA). PCR conditions for amplification were 95 °C for 3 mins, 95 °C for 10 seconds, annealing temperature 52 °C for 30 seconds for 40 cycles. This was followed by dissociation step (as provided by software) – 95 °C for 10 seconds, 65 °C for 5 seconds and 95 °C for 5 seconds. The amplification efficiency was determined at 92% and a melting curve analysis confirmed *NpNBS* PCR product specificity. Norbelladine synthase relative expression values were determined by comparative $2^{-\Delta\Delta C_t}$ method and were scaled to lowest ddCq value by dividing ddCq of a sample with a minimum ddCq value identified among the samples (Livak and Schmittgen 2001).

PCR and Cloning

The open reading frame (ORF) of full length *NpNBS* was amplified from *N. pseudonarcissus* bulbs cDNA with 200 µM dNTPs, 1.25 unit *Taq* DNA polymerase in 50 µL reaction and 0.2 µM forward and reverse gateway primers (Additional file 4). PCR program parameters: 3 mins 94° 1 cycle, 30 s 94°, 45 s 52°, 1 min 72° for 30 cycles, 5 min 72 °C 1 cycle. A Gateway cloning technology was used to clone *NpNBS* according to manufacturer's protocol. The gateway-adapted *attP*-flanked pDONR 221 vector with kanamycin resistance gene was used for BP recombinase reaction (catalyzed by BP clonase enzyme) to generate an *attL*-flanked entry clones with *attB*-flanked *NpNBS* DNA fragment. These entry clones were transformed into *E. coli* DH10β competent cells and positive clones were obtained on a kanamycin selection plate. Chloramphenicol was used for counterselection of positive clones. These clones were further used to perform a LR recombination reaction between an *attL*-containing entry clone and *attR*-containing

pET301/CT-DEST destination vector (150 ng/ μ L) with a histidine tags and T7 promoter, lac operator and *attB* recombination site while C-terminal contain 6x His-tag, T7 reverse priming site and T7 terminator. Positively transformed *E. coli* DH10 β competent cells were selected from ampicillin-Luria-Bertani (LB) media plates incubated overnight at 37 °C. The positive colonies were obtained using 100 μ g/mL ampicillin and confirmed by sequencing.

Protein expression

The above extracted plasmid were transformed into *E. coli* RosettaTM (DE3) pLysS host strain for protein expression using the heat shock transformation protocol (Chan, Verma et al. 2013). Transformed cells were placed onto Luria-Bertani (LB) media with ampicillin and chloramphenicol selection plates overnight at 37 °C. A single colony was picked and grown overnight at 37 °C at 200 rpm in 7 mL LB broth containing ampicillin (100 μ g/mL) and chloramphenicol (34 μ g/mL). The overnight grown pre-culture was diluted 1:100 in fresh LB broth containing ampicillin (100 μ g/mL) and chloramphenicol (34 μ g/mL) and grown at 200 rpm at 37 °C to an A_{600nm} between 0.5-0.8. Isopropyl- β -D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.9 mM to induce protein expression. The culture was incubated for 8 hrs at 37 °C at 200 rpm. The supernatant and pellet of the IPTG-induced bacterial culture were separated by centrifugation at 9032 x *g* for 15 minutes and stored at -80 °C. *E. coli* crude cell pellet were used to purify *Np*NBS protein. Non-induced protein (not treated with IPTG) was also collected from bacterial pellet.

Protein purification and Western blotting

Protein purification was performed by resuspending cell pellet, obtained from IPTG induced *E. coli* Rosetta (DE3)pLys cell cultures, in 10 mg/mL lysozyme and lysis buffer containing 50 mM NaH₂PO₄, 300 mM NaCl and 10 mM imidazole pH 8 and incubated on ice for 30 minutes. After sonication and centrifugation at 10,000 x *g* for 20 minutes at 4 °C, supernatant containing the 6x His-tagged protein was collected (Lysate 1),

which was loaded on pre-equilibrated nickel affinity spin columns from Qiagen (Germany) and centrifuged at 900 x g for 5 minutes. Clear lysate (Lysate 2) was collected and saved for SDS-PAGE analysis. Nickel affinity columns were washed twice with 600 μ L wash buffer containing 50 mM sodium phosphate monobasic (NaH_2PO_4), 300 mM sodium chloride (NaCl) and 20 mM imidazole pH 8 by centrifugation at 900 x g for 2 minutes and flow through were saved as wash 1 and wash 2 for SDS-PAGE analysis. At last, protein was eluted twice (elution 1 and 2) each time in 300 μ L elution buffer with 50 mM NaH_2PO_4 , 300 mM NaCl and 300 mM imidazole pH 8 by centrifugation at 900 x g for 2 minutes and eluate was collected. Protein quantification was done according to Bradford assay (Bradford 1976). Five different BSA concentrations (5,10,15, 20, 25 and 50 μ g) were prepared and absorbance (mean of three) was determined at 595 nm. A graph was plotted with above BSA concentrations at x-axis and their absorbances at y-axis respectively, which gave a linear equation of standard curve ($y = 0.112x + 0.0357$) and extinction coefficient ($R^2 = 0.999$). The NBS protein concentration (x) was calculated using above equation by replacing y with NBS absorbance recorded at 595 nm. Protein was resolved on 15% Mini-protean TGX stain-free precast gels. Gels were transferred on nitrocellulose membrane, equilibrated with TBS buffer (20 mM Tris, 150 mM NaCl pH 7.5) for 15 minutes on a rotatory shaker, followed by blocking of membrane for 1 h. with Tris-buffered saline (TBS) containing tween 20 (TBST) and 1% bovine serum albumin (BSA). Nitrocellulose membrane was incubated overnight at 4 °C in TBST with 1% BSA containing 6x-His epitope tag antibody in 1:1000 dilution. After primary antibody incubation, membrane washed five times each for five minutes in TBST buffer and incubated for 1 h. in TBST containing 2.5% dry milk and goat anti-mouse horse radish peroxidase (GAM)-HRP conjugate in 1: 20,000 dilution. Immunoblot was washed six times for 5 minutes each in TBST buffer and developed using clarity Western ECL substrate from Bio-rad (USA).

Nuclear magnetic resonance spectroscopy and mass spectrometry

Proton and carbon NMR spectra were recorded on a Varian 200 MHz NMR apparatus. Chemicals shifts (δ) are recorded in parts per million (ppm). Coupling constant are

expressed in Hz. Samples were dissolved in DMSO- d_6 for data acquisition (δ 2.49 ppm for ^1H NMR and 39.95 ppm for ^{13}C NMR) using TMS as internal standard (δ 0.00 ppm). Multiplicities were described by the following abbreviations: s for singlet, d for doublet, dd for doublet of doublets, t for triplet and m for multiplet. Mass spectral data was obtained from NanoQAM (Université du Québec à Montréal) using a Time-of-Flight LC/MS (LC/MS-TOF), Agilent Technologie, LC 1200 Series/6210 TOF-LCMS with electrospray ionization and positive mode (ESI+).

Norbelladine synthesis

Norbelladine was synthesized using a previously published protocol with modifications (Park 2014). A summary of the method is described below along with the result of the NMR spectral data (Additional file 5).

Step A: Synthesis of the imine norcraugsodine

An equimolar quantity of 3,4-dihydroxybenzaldehyde (251.6 mg, 1.82 mmol) and tyramine (249.8 mg, 1.82 mmol) were added as powders to a flask containing dichloromethane (7 mL). The solution was stirred gently for 6 hours at room temperature to give the imine intermediate. The solvent was evaporated under reduced pressure with a rotatory evaporator and then, with a mechanical pump to remove solvent residue and water. The product obtained was sufficiently pure to be used as such in the next step. Crude yield, 99%; yellow solid.

^1H NMR (200 MHz, DMSO- d_6) δ : 7.98 (1H, s, CH imine), 7.15 (1H, d, $J = 2.0$ Hz, CH-Ar), 6.99 (2 H, d, $J = 8.6$ Hz, 2 x CH-Ar), 6.90 (1 H, dd, $J_1 = 2$ Hz and $J_2 = 8.2$ Hz, CH-Ar), 6.65 (1 H, d, $J = 8.6$ Hz, CH-Ar), 6.63 (2 H, d, $J = 8.6$ Hz, 2 x CH-Ar), 3.63 (2H, t, $J = 7.2$ Hz, CH = NCH $_2$ CH $_2$), 2.73 (2H, t, $J = 7.4$ Hz, CH = NCH $_2$ CH $_2$); ^{13}C NMR (200 MHz, DMSO- d_6) δ : 160.85, 155.90, 149.49, 146.08, 130.46, 130.12, 127.76, 121.88, 115.77, 115.42, 113.85, 62.41, 36.76; ESI+ HRMS: (M+H) $^+$ calculated for $\text{C}_{15}\text{H}_{16}\text{NO}_3 = 258.1125$; found = 258.1078.

Step B: Synthesis of norbelladine

The imine norcraugsodine (50.6 mg, 0.98 mmol) was dissolved in methanol (5 mL) and was hydrogenated to the amine norbelladine using 30 mol% palladium on carbon (Pd/C 10%) under a H₂ atmosphere. The hydrogen was bubbled three times (t = 0, 30 and 60 minutes) during the hydrogenation process. The mixture is agitated for a total of 2 hours and then filtered on a silica gel to remove the Pd/C and impurities. Methanol was evaporated under reduced pressure with a rotatory evaporator and then, with a mechanical pump to give norbelladine. The product is a brownish solid. Yield: 98%. ¹H NMR (200 MHz, DMSO-d₆) δ: 6.94 (2 H, d, J = 8.6 Hz, CH-Ar), 6.63 (4 H, m, 4 x CH-Ar), 6.52 (1H, dd, J₁ = 1.7 Hz and J₂ = 7.7 Hz, CH-Ar), 3.49 (2H, s, Ar-CH₂-NH), 2.58 (4 H, m, NH-CH₂CH₂-Ar); ¹³C NMR (200 MHz, DMSO-d₆) δ: 155.8, 145.4, 144.3, 132.1, 130.9, 129.8, 119.2, 116.0, 115.6, 115.5, 53.1, 51.1, 35.4; ESI+ HRMS: (M+H)⁺ calculated for C₁₅H₁₈NO₃ = 260.1281; found = 260.1182.

Enzyme assays

The screening assays contained 10 µg of purified protein, 10 µM tyramine, 300 µM 3,4-DHBA, 100 mM Tris buffer in a total volume of 90 µL. The assay was incubated at 37 °C for 2 hrs followed by termination using 3 µL of 20% trichloroacetic acid (TCA). Negative control includes non-induced proteins from *E. coli* (0.0 mM IPTG) whereas negative control assays include purified *Np*NBS protein boiled at 95 °C for 15 minutes.

The analysis of the enzymatic product of NBS was performed on a Waters 2690 high performance liquid chromatograph coupled to a Micromass Quattro LC mass spectrometer using a Kinetex C18 column (150 mm long x 4,6 mm inside diameter, 5 µM particle size). Samples were subjected to positive-mode electrospray ionization (ESI[+]) liquid chromatography (Zulak, Cornish et al.)-tandem mass spectrometry (MS/MS) for reaction product characterization, including collision-induced dissociation (CID) fragmentation analysis. Ten microlitres of each sample was injected onto the column and

compounds were eluted at a flow rate of 0.25 mL/min using ammonium acetate 10 mM, pH 5.0 (solvent A) and acetonitrile 100% (solvent B). The LC program started with 40% solvent B, a gradient began at 0 min to 98% at 7 min, 98% at 9 min, 40% at 10 min, and 40% at 11 min. The total run time was 12 minutes per sample. Analytes were detected using a triple-quadrupole mass analyzer operating in positive ion mode (ESI⁺). For MS/MS analyses, norbelladine and norcraugsodine standards were characterized by the isolation of the parent mass in Q1, the specific fractionation of the parent molecules in a collision cell at a selected energy in q, and finally the scan of the characteristic ions fragments in Q2. The conditions of the MS/MS section were set to acquire in positive ion mode as follows: desolvation gas flow rate 708 L/hr, desolvation gas temperature 400 °C, source temperature 120 °C, capillary voltage 1000 V, cone voltage 15 V, scan mass range from 100 to 265 +ESI and collision energy of 0-30 V. MassLynx software from Waters was used for data acquisition and processing.

Synthesized standard norbelladine (m/z 260) was analyzed using MS mode where the first two quadrupoles were set to radio frequency (RF) only and the third quadrupole scanned the mass range of 253-265 m/z . The obtained mass-to-charge (m/z) value and retention time were used to develop subsequent collision-induced dissociation (CID) experiments. Fragmentation spectrum was obtained using daughter mode at optimized collision energy (15 V) and the third quadrupole scanned the mass range of 50-275 m/z . Norbelladine eluted at 5.5 minute with five significant $[M+H]^+$ in daughter ion mode (260, 159, 138, 123, 121). Abundant fragment 138 m/z (using cone voltage 20 V and collision energy 10eV), 123 m/z (cone voltage 15 V and collision energy 20 eV) and 121 m/z (cone voltage 15V and collision energy 20eV) were optimized for qualifier and quantifier analysis in multi reaction monitoring mode (MRM) based on their signal intensity at an applied voltage (collision energy). MRM was used to measure the intensity of selected fragments to mark them as qualifier and quantifier ion. Norcraugsodine was also eluted at the similar retention time as norbelladine (5.5 min) with two significant $[M+H]^+$ in daughter ion mode (258, 257 and 121). The fragment ion m/z 121 was detected in both standards and considered as the quantitative ion for analysis. CID fragmentation spectra for standards are available in Additional file 3. Qualitative ions for norbelladine and

norcraugsodine were m/z 138 and m/z 157 respectively for validation. The MRM method thus developed was subsequently used to analyze the enzymatic assay samples. Quantification of standards and enzyme assay were performed using integration function of m/z 121 (quantifier daughter ion) on MassLynx software to obtain area under the chromatogram peak. Quantification of norbelladine in the enzyme assay sample was achieved by using the standard curve of norbelladine which has a linear equation of $y = 31264x + 2879$ and a correlation coefficient (R^2) of 0.999. Solutions of concentration between 0 and 10 ppm of norbelladine were prepared and each of them have been injected in triplicate in the LC-MS using the optimized MRM method for the quantitative daughter ion of norbelladine. The area under the chromatographic peaks for each solution was obtained by using the integration function of MassLynx software. The enzyme assay product was confirmed by comparison of LC-ESI-MS/MS data to standards.

List of abbreviations

AA	Amaryllidaceae Alkaloid
3,4-DHBA	3,4-dihydroxybenzaldehyde
ESI	ElectroSpray Ionization
LC-MS	Liquid chromatography-Mass-spectrometry
NBS	Norbelladine synthase
NCS	Norcloclaurine synthase
N4OMT	Norbelladine 4-O-methyltransferase
NR	Noroxomaritidine reductase
MRM	Multiple reaction monitoring
m/z	Mass to charge ratio
PR-10	Pathogenesis Related protein-10

Declarations

Acknowledgements

We thank Professors Hugo Germain, Céline Van Themsche and Charles Goulet for helpful discussion and revision of the manuscript. We also thank Professor André Lajeunesse for his helpful comments and suggestions to optimize the LC-MS/MS detection method.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) award number RGPIN 05294-2014 (Discovery) to IDP. This work was also supported by the NSERC award number EQPEQ 472990-2015 (Research tools and instruments) for the acquisition of the HPLC-PDA and the qPCR. Authors declare that none of the funding bodies have any role in the design of the study and collection, analysis, and interpretation of data as well as in writing the manuscript.

Availability of data and materials

All the data pertaining to the present study has been included in table and/or figure form in the present manuscript and authors are pleased to share analyzed/raw data and plant materials upon reasonable request. All raw sequence reads have been deposited in NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/?term=SRR5788585>). The sequence for Norbelladine Synthase GenBank accession number for the nucleotide sequence is MG948545.

Authors' contributions

IDP and AS conceived and designed the experiments. AS performed the experiments. MAM, AG and LT performed LC-MS/MS analysis. VO and GB chemically synthesized

the norbelladine and norcraugsodine standards and validate them using NMR analysis. IDP and AS interpreted results, prepared the figures, wrote and reviewed the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

The present manuscript does not contain potentially-identifying information of patients/participants needed of consent for publication.

Competing interests

The authors declare that they have no competing interests.

ADDITIONAL FILES

Additional file 1: Three orthologs of NBS obtained from the *N. pseudonarcissus* transcriptome.

Additional file 2: Sequences for NCS and PR-10 used in phylogenetic analysis.

Additional file 3: Fragmentation spectra obtained from LC-MS/MS analysis of standards norbelladine and norcraugsodine followed by a table listing the parameters used for LC-MS/MS analysis.

Additional file 4: List of primer sequences used in this study.

Additional file 5: Proton and carbon NMR spectral data of newly synthesized norcraugsodine and norbelladine.

REFERENCES

1. Jin Z, Xu X-H: **Amaryllidaceae Alkaloids**. In: *Natural Products*. Edited by Ramawat KG, Mérillon J-M: Springer Berlin Heidelberg; 2013: 479-522.
2. Singh A, Desgagné-Penix I: **Biosynthesis of the Amaryllidaceae alkaloids**. *Plant Science Today* 2014, **1**(3):114-120.
3. Hotchandani T, Desgagne-Penix I: **Heterocyclic Amaryllidaceae Alkaloids: Biosynthesis and Pharmacological Applications**. *Current topics in medicinal chemistry* 2017, **17**(4):418-427.
4. He M, Qu C, Gao O, Hu X, Hong X: **Biological and pharmacological activities of amaryllidaceae alkaloids**. *RSC Advances* 2015, **5**(21):16562-16574.
5. Li L, Dai H-J, Ye M, Wang S-L, Xiao X-J, Zheng J, Chen H-Y, Luo Y-h, Liu J: **Lycorine induces cell-cycle arrest in the G0/G1 phase in K562 cells via HDAC inhibition**. *Cancer cell international* 2012, **12**(1):49.
6. Saliba S, Ptak A, Laurain-Mattar D: **4'-O-Methylnorbelladine feeding enhances galanthamine and lycorine production by *Leucojum aestivum* L. shoot cultures**. *Engineering in Life Sciences* 2015, **15**(6):640-645.
7. Prvulovic D, Hampel H, Pantel J: **Galantamine for Alzheimer's disease**. *Expert opinion on drug metabolism & toxicology* 2010, **6**(3):345-354.
8. Barton D, Cohen T: **Festschrift A. Stoll**. *Birkhiiuser, Basel* 1957, **117**.
9. Battersby A, Fales H, Wildman W: **Biosynthesis in the Amaryllidaceae. Tyrosine and norbelladine as precursors of haemanthamine**. *Journal of the American Chemical Society* 1961, **83**(19):4098-4099.
10. Eichhorn J, Takada T, Kita Y, Zenk MH: **Biosynthesis of the amaryllidaceae alkaloid Galanthaminefn2**. *Phytochemistry* 1998, **49**(4):1037-1047.
11. Tahchy AE, Boisbrun M, Ptak A, Dupire F, Chrétien F, Henry M, Chapleur Y, Laurain-Mattar D: **New method for the study of Amaryllidaceae alkaloid biosynthesis using biotransformation of deuterium-labeled precursor in tissue cultures**. *Acta Biochimica Polonica* 2010, **57**(1):75.
12. El Tahchy A, Ptak A, Boisbrun M, Barre E, Guillou C, Dupire F, Chrétien F, Henry M, Chapleur Y, Laurain-Mattar D: **Kinetic study of the rearrangement of deuterium-labeled 4'-O-methylnorbelladine in *Leucojum aestivum* shoot cultures by mass spectrometry. Influence of precursor feeding on Amaryllidaceae alkaloid accumulation**. *Journal of natural products* 2011, **74**(11):2356-2361.

13. Grisebach H: **Comparative biosynthetic pathways in higher plants.** *Pure and Applied Chemistry* 1973, **34**(3-4):487-514.
14. Barton D, Kirby G, Taylor J, Thomas G: **866. Phenol oxidation and biosynthesis. Part VI. The biogenesis of amaryllidaceae alkaloids.** *Journal of the Chemical Society (Resumed)* 1963:4545-4558.
15. Kilgore MB, Augustin MM, Starks CM, O'Neil-Johnson M, May GD, Crow JA, Kutchan TM: **Cloning and characterization of a norbelladine 4'-O-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in *Narcissus* sp. aff. *pseudonarcissus*.** *PloS one* 2014, **9**(7):e103223.
16. Kilgore MB, Augustin MM, May GD, Crow JA, Kutchan TM: **CYP96T1 of *Narcissus* sp. aff. *pseudonarcissus* Catalyzes Formation of the Para-Para'CC Phenol Couple in the Amaryllidaceae Alkaloids.** *Frontiers in plant science* 2016, **7**:225.
17. Kilgore MB, Holland CK, Jez JM, Kutchan TM: **Identification of a noroxomaritidine reductase with Amaryllidaceae alkaloid biosynthesis related activities.** *Journal of Biological Chemistry* 2016, **291**(32):16740-16752.
18. Samanani N, Liscombe DK, Facchini PJ: **Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis.** *The Plant journal : for cell and molecular biology* 2004, **40**(2):302-313.
19. Singh A, Desgagne-Penix I: **Transcriptome and metabolome profiling of *Narcissus pseudonarcissus* 'King Alfred' reveal components of Amaryllidaceae alkaloid metabolism.** *Scientific reports* 2017, **7**(1):17356.
20. Lee EJ, Facchini P: **Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family.** *The Plant cell* 2010, **22**(10):3489-3503.
21. Fernandes H, Michalska K, Sikorski M, Jaskolski M: **Structural and functional aspects of PR-10 proteins.** *The FEBS journal* 2013, **280**(5):1169-1199.
22. Wen J, Vanek-Krebitz M, Hoffmann-Sommergruber K, Scheiner O, Breiteneder H: **The Potential of Betv1 Homologues, a nuclear multigene family, as phylogenetic markers in flowering plants.** *Molecular phylogenetics and evolution* 1997, **8**(3):317-333.
23. Wang C-S, Huang J-C, Hu J-H: **Characterization of two subclasses of PR-10 transcripts in lily anthers and induction of their genes through separate signal transduction pathways.** *Plant Mol Biol* 1999, **40**(5):807-814.

24. Yu X, Ekramoddoullah AK, Misra S: **Characterization of Pin m III cDNA in western white pine.** *Tree physiology* 2000, **20**(10):663-671.
25. Radauer C, Breiteneder H: **Evolutionary biology of plant food allergens.** *Journal of Allergy and Clinical Immunology* 2007, **120**(3):518-525.
26. Samanani N, Park S-U, Facchini PJ: **Cell type-specific localization of transcripts encoding nine consecutive enzymes involved in protoberberine alkaloid biosynthesis.** *The Plant Cell* 2005, **17**(3):915-926.
27. Zulak KG, Cornish A, Daskalchuk TE, Deyholos MK, Goodenowe DB, Gordon PM, Klassen D, Pelcher LE, Sensen CW, Facchini PJ: **Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism.** *Planta* 2007, **225**(5):1085-1106.
28. Lee E-J, Facchini P: **Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family.** *The Plant Cell* 2010, **22**(10):3489-3503.
29. Nishihachijo M, Hirai Y, Kawano S, Nishiyama A, Minami H, Katayama T, Yasohara Y, Sato F, Kumagai H: **Asymmetric synthesis of tetrahydroisoquinolines by enzymatic Pictet-Spengler reaction.** *Bioscience, biotechnology, and biochemistry* 2014, **78**(4):701-707.
30. Li J, Lee EJ, Chang L, Facchini PJ: **Genes encoding norcoclaurine synthase occur as tandem fusions in the Papaveraceae.** *Sci Rep* 2016, **6**:39256.
31. Lichman BR, Gershater MC, Lamming ED, Pesnot T, Sula A, Keep NH, Hailes HC, Ward JM: **'Dopamine-first' mechanism enables the rational engineering of the norcoclaurine synthase aldehyde activity profile.** *The FEBS journal* 2015, **282**(6):1137-1151.
32. Samanani N, Facchini PJ: **Isolation and partial characterization of norcoclaurine synthase, the first committed step in benzyloquinoline alkaloid biosynthesis, from opium poppy.** *Planta* 2001, **213**(6):898-906.
33. Samanani N, Liscombe DK, Facchini PJ: **Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzyloquinoline alkaloid biosynthesis.** *The Plant Journal* 2004, **40**(2):302-313.
34. Heinrich M, Lee Teoh H: **Galanthamine from snowdrop--the development of a modern drug against Alzheimer's disease from local Caucasian knowledge.** *J Ethnopharmacol* 2004, **92**(2-3):147-162.

35. Barik J, Dajas-Bailador F, Wonnacott S: **Cellular responses to nicotinic receptor activation are decreased after prolonged exposure to galantamine in human neuroblastoma cells.** *British journal of pharmacology* 2005, **145**(8):1084-1092.
36. Kutchan TM: **Strictosidine: From alkaloid to enzyme to gene.** *Phytochemistry* 1993, **32**(3):493-506.
37. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Molecular biology and evolution* 1987, **4**(4):406-425.
38. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S: **Estimating divergence times in large molecular phylogenies.** *Proceedings of the National Academy of Sciences* 2012, **109**(47):19333-19338.
39. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: molecular evolutionary genetics analysis version 6.0.** *Molecular biology and evolution* 2013, **30**(12):2725-2729.
40. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**(4):402-408.
41. Chan W-T, Verma CS, Lane DP, Gan SK-E: **A comparison and optimization of methods and factors affecting the transformation of Escherichia coli.** *Bioscience reports* 2013, **33**(6):e00086.
42. Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.** *Analytical biochemistry* 1976, **72**(1-2):248-254.
43. Park JB: **Synthesis and characterization of norbelladine, a precursor of Amaryllidaceae alkaloid, as an anti-inflammatory/anti-COX compound.** *Bioorganic & medicinal chemistry letters* 2014, **24**(23):5381-5384.
44. Nei M, Kumar S: **Molecular evolution and phylogenetics:** Oxford university press; 2000.

Figure Legends

Fig. 1 Norbelladine synthase (NBS) catalyzes the condensation of tyramine and 3,4-dihydroxybenzaldehyde (3,4-DHBA) to form norbelladine, the common precursor to all Amaryllidaceae alkaloids produced in plants including galanthamine, lycorine and haemanthamine.

Fig. 2 Phylogenetic relationships among several PR10/Bet v 1 proteins from a variety of plants. Phylogeny tree was based on the amino acid sequences for *Betula verrucosa* BvPR10, *Betula platyphylla* BpPR10, *Papaver somniferum* PsNCS, PsNCS2, *Thalictrum flavum* TfNCS, TfNCS2, TfNCS3, TfNCS4, TfNCS5, *Chelidonium majus* CmNCS1, *Argemone mexicana* AmNCS, *Eschscholzia californica* EcNCS1, EcNCS2, *Coptis japonica* CjNCS2 (formerly CjPR10A), *Daucus carota* DcPR10, *Solanum tuberosum* StPR10, *Hyacinthus orientalis* HoPR10, *Pinus monticola* PmPR10, *Hordeum vulgare* HvPR10, *Oryza sativa* OsPR10, *Lily regale* LrPR10, *Sorghum bicolor* SbPR10, *Zea mays* ZmPR10 and *Hypericum perforatum* Hyp-1. Accession numbers provided in Additional file 2. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987) with bootstrap value 500. The optimal tree with the sum of branch length = 5.78745187 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method (Nei and Kumar 2000) and are in the units of the number of amino acid differences per site. The analysis involved 25 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 139 positions in the final dataset. Bootstrap values were added to the figure. Evolutionary analyses were conducted in MEGA 6 (Tamura, Stecher et al. 2013).

Fig. 3 Clustal omega alignment of deduced amino acid sequence of *Narcissus pseudonarcissus* norbelladine synthase (NpNBS) with *norcoclaurine synthase* (NCS) amino acid sequences obtained from different species: *Eschscholzia californica* EcNCS1, *Thalictrum flavum* TfNCS1, *Papaver somniferum* PsNCS1, and *Coptis japonica* CjNCS2

(formely *CjPR10A*). Grey boxes corresponding to Tyr, Lys, Asp and Glu residues forms the *NpNBS* catalytic residue. The black bar depicts a glycine-rich P-loop domain conserved in *NCS* and *NpNBS* protein. The fully conserved residues are marked with an asterisks (*). The positions with conservation between amino acid residues of similar properties are marked with a colon (:). The positions with conservations between amino acids residues of weakly similar properties are marked with a period (.). Numbers on the right indicate the residues position in the sequence.

Fig. 4 Reverse transcription quantitative PCR analysis of *NpNBS* expression in different tissues of *N. pseudonarcissus* ‘King Alfred’. Plant tissues were harvest after flowering in spring. Graph is plotted using normalized ddCT values scaled to lowest ddCq value by dividing ddCq of a sample with a minimum ddCq value identified among the samples (Livak and Schmittgen 2001). Histone was used for internal reference. Expression fold change were calculated using the comparative $2^{-\Delta\Delta C_t}$ method from three independent replicates. Bars represent the mean standard deviation of three independent replicates. Abbreviations: tissues are BB, bulb; RT, root; ST, stem; LF, leaf; FL, flower.

Fig. 5 Heterologous expression of *NpNBS* in *Escherichia coli*. (A) Agarose gel electrophoresis of PCR amplified candidate *NpNBS* from a *N. pseudonarcissus* cDNA library. Product size is 560 bp (ORF 492bp + Gateway adapters 68bp). Numbers on the left refer to the location of standard DNA molecular markers in bp. (B) Agarose gel electrophoresis of PCR amplified *NpNBS* performed with 5 positively transformed bacterial colonies (col.1 to col.5) shows a band of *NpNBS* ORF transcript size of 492 bp, L is for-Ladder; (C) SDS-PAGE analysis of *NpNBS* protein. Purified protein was extracted from 0.9 mM IPTG induced *E. coli* RosettaTM (DE3) pLysS host strain at 37 °C for 8 h in elution 1 with 300 mM of imidazole. L-Ladder; +C- positive control with a 30 kDa protein; NI-non-induced protein; L1- lysate 1 (10 mM imidazole); L2- lysate 2 (10 mM imidazole); W1-wash 1 (20 mM imidazole); W2-wash 2 (20 mM imidazole); E1-elution 1 (300 mM imidazole); E2-elution 2 (300 mM imidazole). (D) Western blot analysis performed using 6x HIS-tag epitope antibody shows expression of recombinant protein at 37 °C in 0.9 mM IPTG induced crude and pure protein extracts. No detectable expression

of recombinant protein at 25 °C in 0.9 mM IPTG induced crude and pure protein extracts. Lane L-ladder; NI- protein non-induced ; +C-positive control protein with C-terminal His tags 30 kDa; Cr-crude protein 37 °C/8 h/0.9 mM; E1- elution 1, 37 °C/8 h/0.9 mM; E2- elution 2, 37 °C/8 h/0.9 mM; Cr-crude 25 °C/8 h/0.9 mM; E1- elution 1, 25 °C/8 h/0.9 mM; E2-elution 2, 25 °C/8 h/0.9 mM. Numbers on the left refer to the location of standard protein molecular weight markers in kDa.

Fig. 6 Extracted ion chromatograms showing the product of *Np*NBS enzyme assays. The tested substrates used were 3,4-dihydroxybenzaldehyde (300 µM) and tyramine (10 µM), the extracted ion chromatogram corresponds to standard NB or assays conducted with the non-induced protein fraction, the heat-denatured recombinant *Np*NBS enzyme; and the complete assay performed with recombinant *Np*NBS at pH 4 and pH 7. Parent ion mass-to-charge (m/z) of 260 for norbelladine and 258 for norcraugsodine were subjected to collision-induced dissociation analysis for identification and quantitation (Additional file 3).

Fig. 1

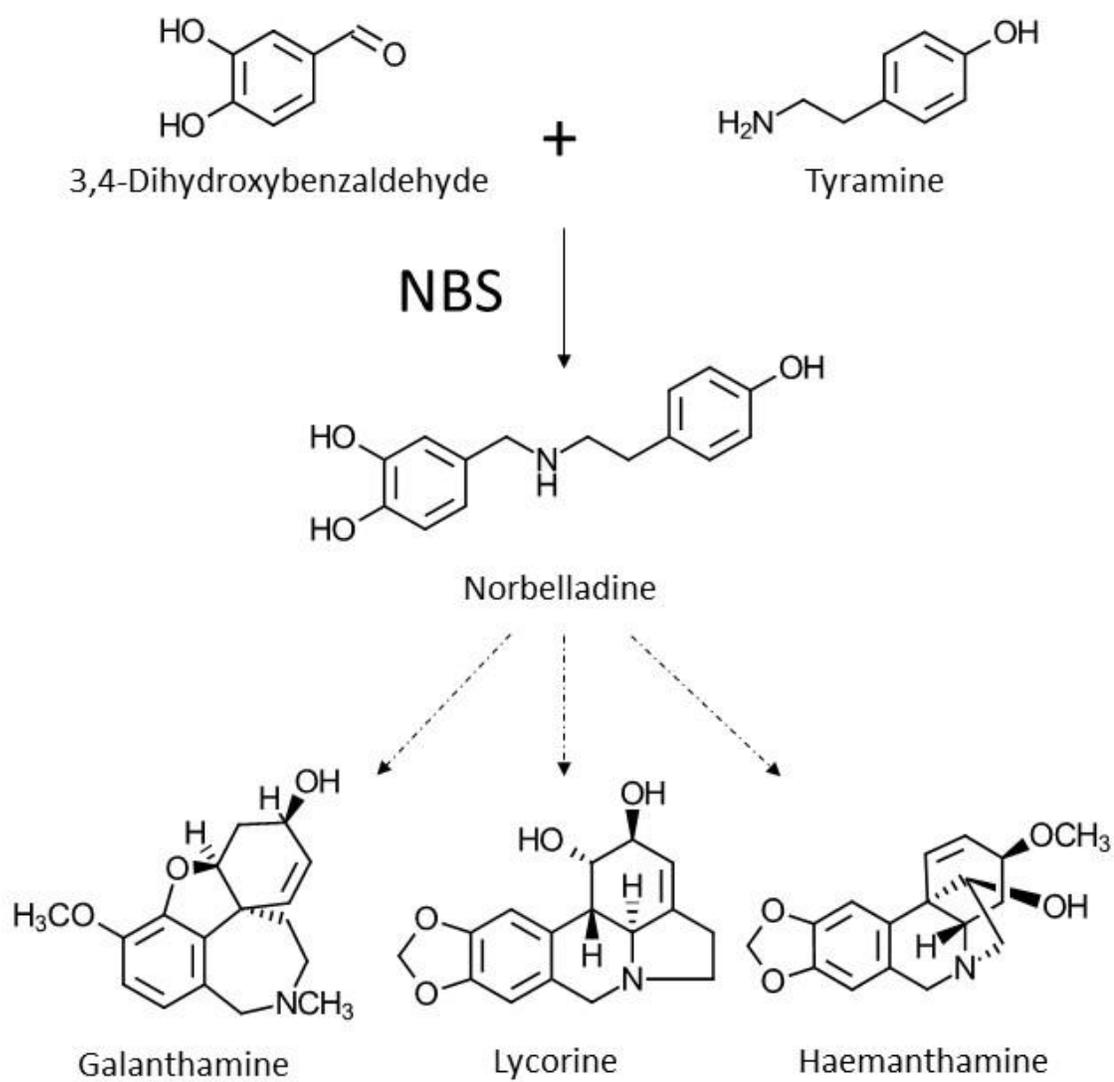


Fig. 2

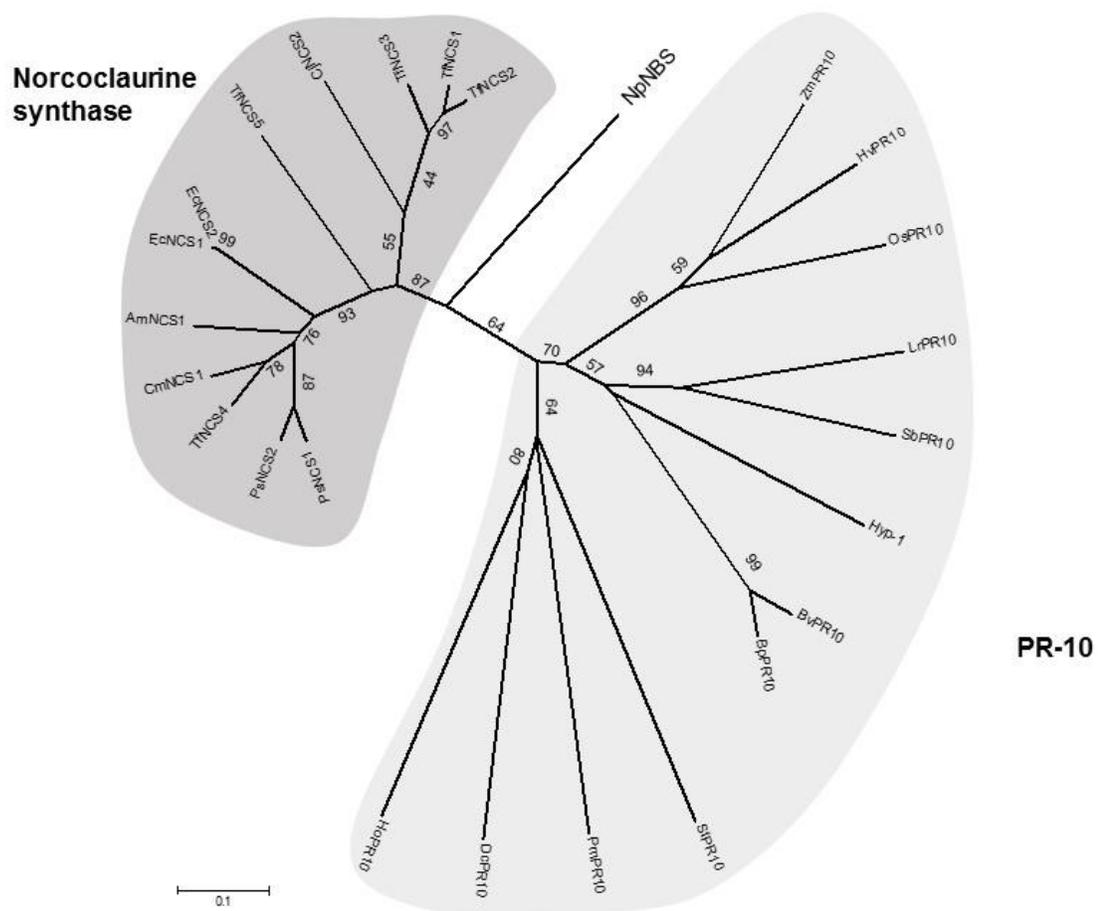


Fig. 4

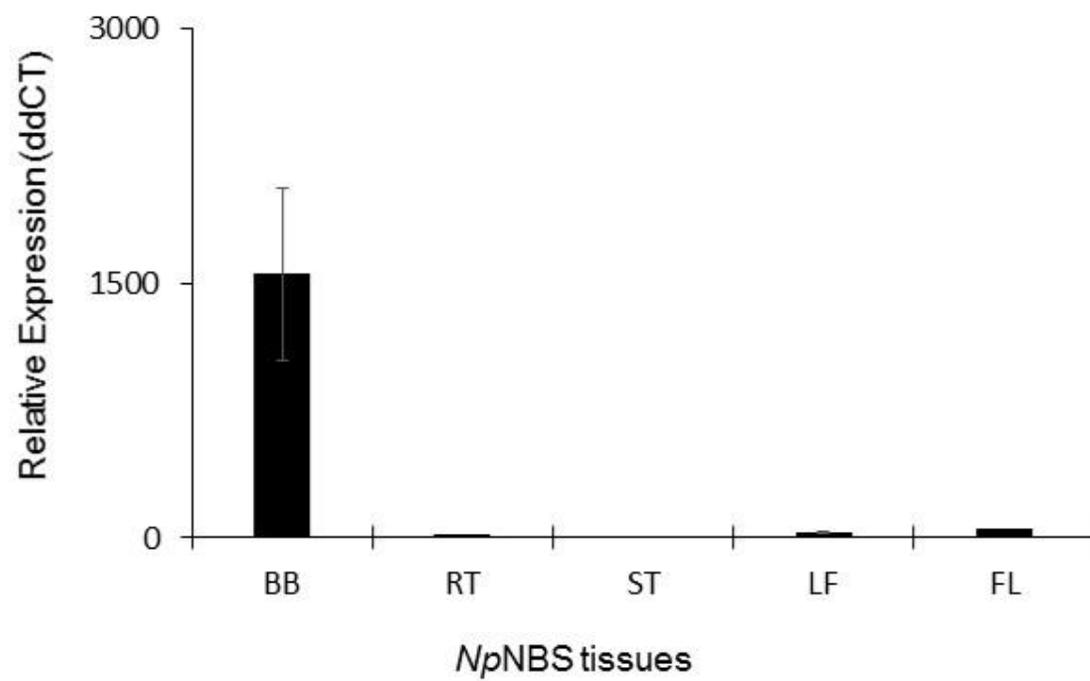


Fig. 5

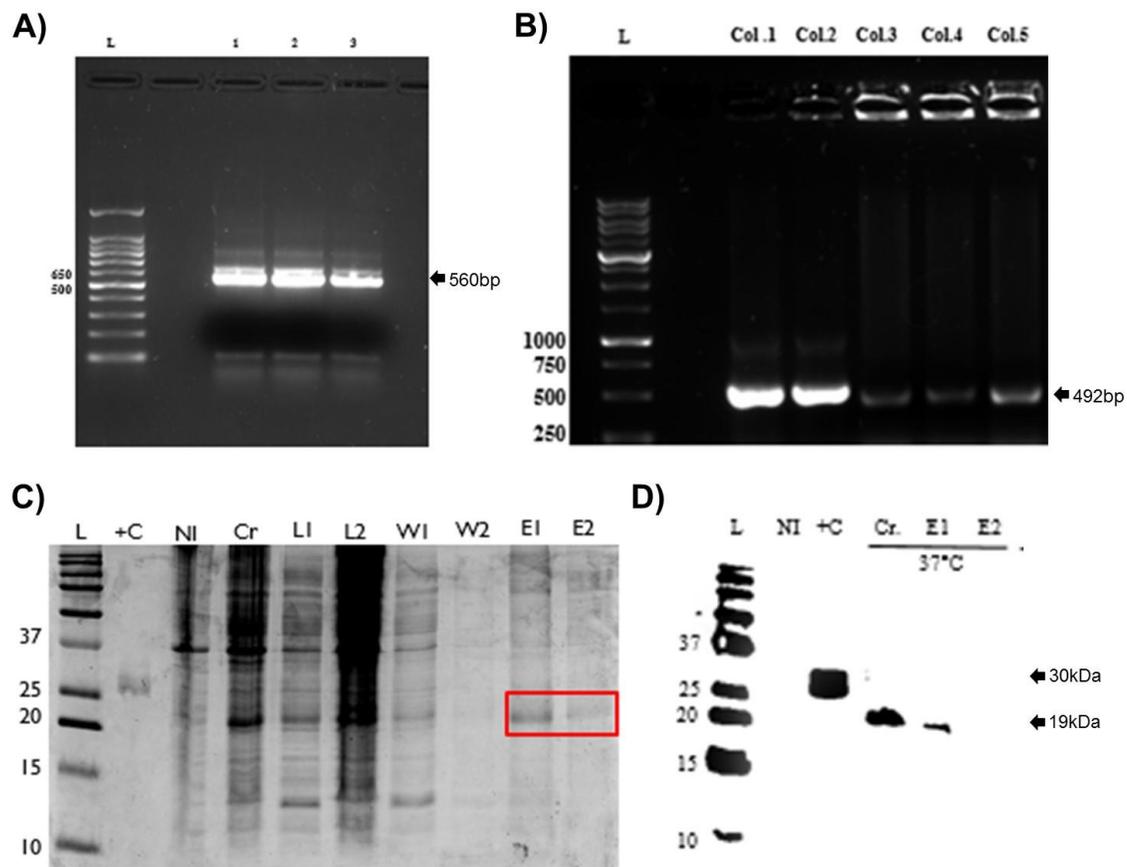


Fig. 6

