Functional analysis of *Hypoxylon pulicicidum* genes required for heterologous biosynthesis of nodulisporic acids

This thesis is presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

in

Biochemistry

at

Victoria University of Wellington,

Wellington, New Zealand

Kyle Cornelius Van de Bittner

2018

ABSTRACT

Nature holds some of the greatest secrets in drug design and development and the ability to access these trade secrets has been revolutionised by modern bioengineering technologies. In order to exploit these technologies it is essential to understand what genes are involved in compound production and the enzymatic steps that limit flux to the desired product. This thesis describes the discovery of four secondary-metabolic enzymatic steps involved in the biosynthesis of a group of valuable natural products known as nodulisporic acids.

Nodulisporic acids are known for their potent insecticidal activities; however, biosynthesis of these compounds by the natural fungal producer, *Hypoxylon pulicicidum* (*Nodulisporium* sp.), is exceptionally difficult and has prevented the commercial development of novel nodulisporic acid-containing veterinary medicines and crop protects.

To discover how nodulisporic acids are biosynthesized:

1. the *H. pulicicidum* genome was sequenced
2. a gene cluster responsible for nodulisporic acid production was predicted
3. genes in the cluster were functionally characterised by pathway reconstitution in a common, fast growing mould, *Penicillium paxilli*

In turn, four genes involved in the biosynthesis of the nodulisporic acid core compound, nodulisporic acid F, have been functionally characterised. The four genes encode a geranylgeranyl transferase (NodC), a flavin adenine dinucleotide-dependent oxygenase (NodM), an indole diterpene cyclase (NodB) and a cytochrome P450 oxygenase (NodW). Two of the gene products (NodM and NodW) catalyse two previously unreported reactions that provide the enzymatic basis of the biosynthetic branch point unique to nodulisporic acid biosynthesis.

From here, future efforts will explore how these genes can be engineered to overcome flux bottlenecks and enable production of significantly increased, and even industrially relevant, product titres.

OUTPUTS FROM THIS THESIS

For all of the below outputs:

Emily Parker, Matt Nicholson, and Barry Scott conceived the fungal project. Emily Parker sourced funding, provided the lab space and materials, and acted as the biochemistry expert to link the predicted gene functionality to individual biosynthetic steps. Matt Nicholson assisted in sourcing funding, provided expertise around fungal work and acted as primary bioinformatician to identify the *NOD* cluster and annotate the genes. Barry Scott provided fungal strains and fungal expertise. Leyla Bustamante acted as lead science and lab manager, ensuring the projects ran smoothly at all times, provided molecular biology expertise and teaching, and performed some experimental work. I, **Kyle Van de Bittner**, established the gene characterisation pipeline in our lab including method development for plasmid construction, fungal transformations, protoplast preparations, chemical extraction and compound analysis by TLC, LC-MS, NMR, and preparation of compound standards. Craig van Dolleweerd conceived MIDAS, constructed the MIDAS vector suite, and provided expert advice on using the MIDAS system. Sarah Kessans prepared some of the plasmid assemblies and fungal transformations.

Additional information that is specific to individual outputs is listed below.

Publications:

- **Van de Bittner, K. C.**; Nicholson, M. J.; Bustamante, L. Y.; Kessans, S. A.; Ram, A.; van Dolleweerd, C. J.; Scott, B.; Parker, E. J., Heterologous biosynthesis of nodulisporic acid F. *J. Am. Chem. Soc.* **2018,** *140* (2), 582-585.[1]

I completed nearly all of the experimental work and wrote the manuscript (including generation of all figures and tables) with input from my supervisors, Emily Parker and Matt Nicholson. Emily Parker and Matt Nicholson edited the manuscript. Arvina Ram provided expertise around fungal transformation and protoplast preparation methods, prepared the *H. pulicicidum* DNA for genomic sequencing, and managed all of the material transfers from Barry Scott's lab.

- Nicholson, M. J.; **Van de Bittner, K. C.**; Ram, A.; Bustamante, L. Y.; Parker, E. J., Draft Genome Sequence of the Filamentous Fungus *Hypoxylon pulicicidum* ATCC 74245. *Genome A.* **2018,** *6* (2).[2]

Matt Nicholson wrote the manuscript, sourced the *H. pulicicidum* strain, identified the *NOD* cluster and annotated its genes. Arvina Ram, Leyla Bustamante, and I completed the experimental work.

- van Dolleweerd, C. J.; Kessans, S. A.; **Van de Bittner, K. C.**; Bustamante, L. Y.; Bundela, R.; Scott, B.; Nicholson, M. J.; Parker, E. J., MIDAS: A modular DNA assembly system for synthetic biology. *ACS Synth Biol.* **2018,** *7* (4), 1018-1029.[3]

Craig van Dolleweerd wrote the manuscript. Sarah Kessans, Leyla Bustamante, Rudy Bundela, and I performed the plasmid preparations and fungal transformations. I performed all of the LC-MS and NMR analysis and provided many figures and tables. Emily Parker, and I edited the manuscript.

Patents:

- Inventors: Parker, E. J.; Nicholson, M. J.; **Van de Bittner, K. C.**; Bustamante, L. Y.; Kessans, S. A.; van Dolleweerd, C. J., “Heterologous Biosynthesis of Nodulisporic Acids”, A.U. Patent AU2017903956 filed 29/09/2017.

Emily Parker, Matt Nicholson, and I wrote the patent with input from our patent attorney, Jon Ashen. I generated all of the figures and tables and performed nearly all of the experimental work required for this patent.

- Inventors: van Dolleweerd, C. J.; Parker, E. J.; Nicholson, M. J.; Beneficiaries: **Van de Bittner, K. C.**; Bustamante, L. Y.; Kessans, S. A.; “Modular DNA Assembly System”, A.U. Patent AU2017903955 filed 29/09/2017.

Craig van Dolleweerd conceived MIDAS, constructed the MIDAS vector suite, and wrote the patent with input from our patent attorney, Jon Ashen. Emily Parker, Matt Nicholson, and I edited the patent.

ACKNOWLEDGEMENTS

To reach the point of writing the acknowledgments for this piece of work was no small undertaking. By no means can I say that this work is mine alone, and so luckily, I have a place to acknowledge many of those who have contributed – not only to this work, but to granting me all the resources and support I have required to get to this point in my life.

To start I would like to thank those who were directly involved in the work that brought this thesis into fruition. I will start with the person who gave me the opportunity to get amongst this amazing science endeavour, my primary supervisor Professor Emily Parker. Although she was not physically present for the majority of the labwork - instead serving as what she deems the main “cheerleader”- this title she gives herself only highlights how humble she is; because without her, there would be no lab, no funding, and no past or future of this project. It comes with no surprise that when she announced she intended to relocate the lab from Christchurch to Wellington, nearly everyone followed. Secondly, comes my co-supervisor, Dr. Matthew Nicholson, who was ‘the man with a plan’ – the one who approached Emily with the initial idea to make a fungal factory for nodulisporic acids. In times of despair, Matt was my go-to father figure who consoled me. Thank you Emily and Matt for giving me the opportunity to work on a commercially relevant thesis project, which has inspired me to pursue future work in the biotech industry. I so look forward to what we can create together in the coming years. Next and most important to me, is my partner in crime, lab and science manager, super fancy scientist and extraordinary role-model, Dr. Leyla Bustamante. She has been with me nearly every step of the way, giving me support and love in the darkest of times and celebrating with me in the lightest. She is the main reason I decided to stay in New Zealand and pursue my PhD here rather than going back to the States. It is for her that I reserve my deepest gratitude for I cannot begin to explain all she has taught me over the years.

I would also like to thank all of our collaborators who have enabled this project to succeed. Namely, those in Professor Barry Scott’s lab at Massy University in Palmerston North – in particular, Prof. Barry Scott himself, for his insightful discussions and providing the project with all of the heterologous expression hosts, as well as Arvina Ram, who taught me how to transform fungi and ran quality control on the genomic DNA samples used to sequence the *Hypoxylon pulicicidum* genome.

Additionally, I would like to thank Dr. Craig van Dolleweerd from Callaghan Innovation who conceived the Modular Idempotent DNA Assembly System (MIDAS) used as our platform to functionally characterise genes and reconstitute metabolic pathways in *P. paxilli* and Dr. Sarah Kessans for assembling some of the MIDAS constructs used in this study. I am especially grateful to those who often go unnoticed; the technicians and mechanics who ensured that all of the labware was running properly, thereby preventing unneccesary disruptions in the flow of my research. In particular, I would like to thank Dr. Amelia Albrett from the University of Canterbury and Jan Vorster from Victoria University of Wellington for their technical service around liquid chromatography-mass spectrometry and nuclear magnetic resonance spectrometry, as well as Wayne Mackay who was the truly spectacular all-round handy man and go-to Mr. Fix-It at Univeristy of Canterbury.

Of course the project would not be possible without our funding bodies, the New Zealand Ministry of Business, Innovation, and Employment who supplied the Smart Ideas grant to get this project off the ground, Fulbright New Zealand for supplying funding for my first year of research, the University of Canterbury for supplying me with a PhD scholarship for my second year of research, and Victoria University of Wellington for supplying me with PhD scholarship for my third.

Finally, thank you Mom and Dad, Gen and Great Mom, for giving me your deepest love and support from when I was a little munchkin all the way through the present. I feel privileged to have been born into such a loving environment – full of inspiring people who taught me to never regard anything as impossible. Thank you for giving me more than what the world could ever offer.

Love forever,

Kyle

TABLE OF CONTENTS

1. INTRODUCTION 1

Fungal secondary metabolites 2

The indole diterpenes 2

A review of indole diterpene biosynthesis: From early labelled substrate feeding studies to recent pathway reconstitution studies 4

Primary metabolites involved in indole diterpene biosynthesis 7

Secondary metabolism step 1: Formation of geranylgeranyl pyrophosphate (3) . 8

Secondary metabolism step 2: Geranylgeranylation of indole-containing substrate 10

Secondary metabolism step 3: Epoxidation of 3-geranylgeranylindole 14

Secondary metabolism step 4: Formation of cyclic indole diterpene scaffolds .. 21

Decoration of the cyclic scaffolds: Gene identification and elucidation of the decoration steps 26

Identification of gene clusters enables rapid confirmation of individual gene function 27

Identification of the *PAX* gene clusters and functional analysis of *paxP*, *paxQ* and *paxD* 27

Identification of the *ATM* gene clusters and functional analysis of *atmP*, *atmQ* and *atmD* 39

Identification of the *LTM* gene clusters and functional analysis of *LTM* genes .. 44

Identification of the *TER* gene cluster and functional analysis of *terP* and *terQ* 52

Identification of the *PTM and PEN* gene clusters and functional analysis of the genes 57

Identification of the *JAN* gene cluster and functional analysis of the genes 63

Identification and functional analysis of unclustered indole diterpene genes, which led to the identification of another indole diterpene cluster, *ATS5* 71

Conclusions 75

Nodulisporic acids: the target for this research 77

The discovery of nodulisporic acid A's (1) insecticidal activity 79

Nodulisporic acid A (1) selectively targets glutamate gated chloride ion channels 79

Nodulisporic acid A (1) is an orally safe, systemically active, ectoparasiticide .. 80

Medicinal chemistry studies: elucidation of the nodulisporic acid A (1) pharmacophore and development of nodulisporic acid A (1) analogues with more desirable properties....81
Analysis of the ideal nodulisporic acid A (1) analogue drug target: *N*-tert-butyl nodulisporamide (78)....85
Difficulties with nodulisporic acid A (1) production prevent drug development...88

2. THE GENE CHARACTERISATION TOOLBOX....89
Bioinformatics: Identification of the *NOD* cluster....90
DNA assembly: The Modular Idempotent DNA Assembly System (MIDAS)...98
Heterologous expression....100
A library of expression hosts for confirming gene function....101
Compound analysis....103
Conclusion....104

3. FUNCTIONAL CHARACTERISATION OF NodC....105
Prediction of *nodC*....106
Functional analysis of NodC....107
Preparation of MIDAS Level-1 plasmids....109
Generation of PCR products....109
Assembly of Level-1 plasmids....110
Preparation of MIDAS Level-2 plasmids....113
Preparation of MIDAS Level-3 plasmids....117
Control plasmids....118
Fungal transformations....119
Resolution of MIDAS construct problem....120
Successful complementation of *nodC* into the Δ*paxC P. paxilli* deletion mutant....122
LCMS confirmation of paxilline (2) production....123
Conclusions....124

4. NodM, THE KEY EPOXIDASE....125
The indole diterpene epoxidases....126
Identifying the *NOD* epoxidase....127
Functional analysis of *nodM*....128
MIDAS assembly....129

Fungal transformations 130
A note about promoter choice 133
Novel compound identification 133
Conclusions 134
Explanation for emindole SB (18) production 134

5. NodB, A PROMISCUOUS INDOLE DITERPENE CYCLASE 136
Indole diterpene cyclases are notoriously promiscuous 137
Identifying the *NOD* cyclase 139
Functional analysis of *nodB* 140
Initial approach 140
Assembly of multigene constructs 141
Transformations of the multigene constructs 147
Analysis of LCMS results 149
Conclusions 150

6. NodW, THE FUNNEL TO NODULISPORIC ACIDS 151
Review of post cyclisation oxidations by cytochrome P450 oxygenases in indole diterpene biosynthesis 152
The *NOD* cluster lacks highly homologous cytochrome P450 oxygenases 152
P450 analysis: An approach to functionally confirm non-homologous genes 153
Transformation of *P. paxilli* with P450-encoded vectors 156
Identification of nodulisporic acid F (82) 158
Confirmation that five genes are required for nodulisporic acid biosynthesis 159
Enhancing nodulisporic acid F (82) biosynthesis 163
Conclusions 165

7. CONCLUSIONS 167
Key findings and results 168
Future work: Functional analysis of the remaining *NOD* genes 171
What is the benefit of this work and what will be done with all that has been learned? 172
The opportunity for a metabolic engineering business 173

8. MATERIALS AND METHODS 175

gDNA isolation for genome sequencing and transcription unit module amplification **176**
Genome sequencing, identification and annotation of the *H. pulicicidum* (ATCC® 74245™) *NOD* gene cluster and GenBank accession numbers **177**
Molecular biology reagents **178**
Bacterial strains **179**
PCR **179**
Construction of MIDAS vectors **179**
Construction of the pML1 source vector. **180**
Construction of pML2 shuttle vectors. **182**
Construction of the pML3 destination vector. **188**
Golden Gate assembly reactions **191**
Preparation of MIDAS constructs **191**
Primer design. **191**
Protocols for MIDAS Level-1 module cloning **193**
Protocols for MIDAS Level-2 transcription unit assembly **193**
Protocols for MIDAS Level-3 multigene assembly **194**
***Penicillium paxilli PAX* gene deletion mutants** **195**
Media and reagents used for fungal work **197**
Protoplast preparation **197**
Transformation of *P. paxilli* **198**
Fungal transformations and analysis of indole diterpene chemotypes **199**
Indole diterpene production and extraction **199**
Normal phase TLC **200**
Reverse phase LCMS **200**
Large scale indole diterpene purification for NMR spectroscopy analysis **201**
NMR spectroscopy **201**

9. APPENDIX **203**
Additional Notes: **204**
Introduction: Additional radiography feeding studies: 204
Chapter 3: Testing of the *paxC* promoter region (*paxC*$_{ProUTR-2}$) 204
Bioinformatics analyses to identify the potential genetic elements involved in the biosynthesis of nodulisporic acids **205**

Assembling of genetic elements (5'UTRs, CDS, 3'UTRs, etc.) into MIDAS vectors 213

LCMS and NMR metabolite characterisation 220

10. REFERENCES..... 249

1. INTRODUCTION

Fungal secondary metabolites

Secondary metabolites are natural products that are not required for the growth of the producing species. Although the systematic study of fungal secondary metabolites began in 1922,[4] it was not until the discovery and development of the most well known fungal secondary metabolite, penicillin – the first broad spectrum antibiotic – that pharmaceutical companies invested in screening programs directed toward fungal secondary metabolites.[5] Since then, thousands of fungal secondary metabolites have been isolated, characterised, and bioactively profiled. An impressive review of 1,500 fungal metabolites that were identified between 1993 and 2001 showed that over half of the compounds had antimicrobial or antitumor activities.[6]

One method used to classify the array of fungal secondary metabolites differentiates them based on the enzyme classes used in their biosynthesis.[5] According to this classification method, four main classes of fungal secondary metabolites exist: polyketides, non-ribosomal peptides, terpenes, and indole alkaloids. This introduction focuses on one subclass of the terpene secondary metabolites known as indole diterpenes.

The indole diterpenes

The indole diterpenes are of particular interest due to their wide range of chemical diversity and concomitant bioactivities, which include insecticidal,[7-11] anti-cancer,[12-13] anti-MRSA,[14] anti-H1N1,[15] and tremorgenic[16] activities. The differences in the bioactive properties of indole diterpenes arise from the variety of ways the indole diterpene core structure is cyclised and the functional groups that decorate the cyclised cores (Figure 1.1, Note: Table 9.10, located at the end of the **APPENDIX**, contains a list of numbered compounds and corresponding structures.). Genetic variation of the native indole diterpene producers gives rise to the different cyclisation mechanisms and functional groups of indole diterpenes. Understanding how these compounds are biosynthesised and subsequently identifying enzymes from biosynthetic pathways can enable enhanced production of compounds that are difficult or impossible to access from natural sources, and provides the opportunity to engineer efficiently novel compounds with enhanced bioactive properties.

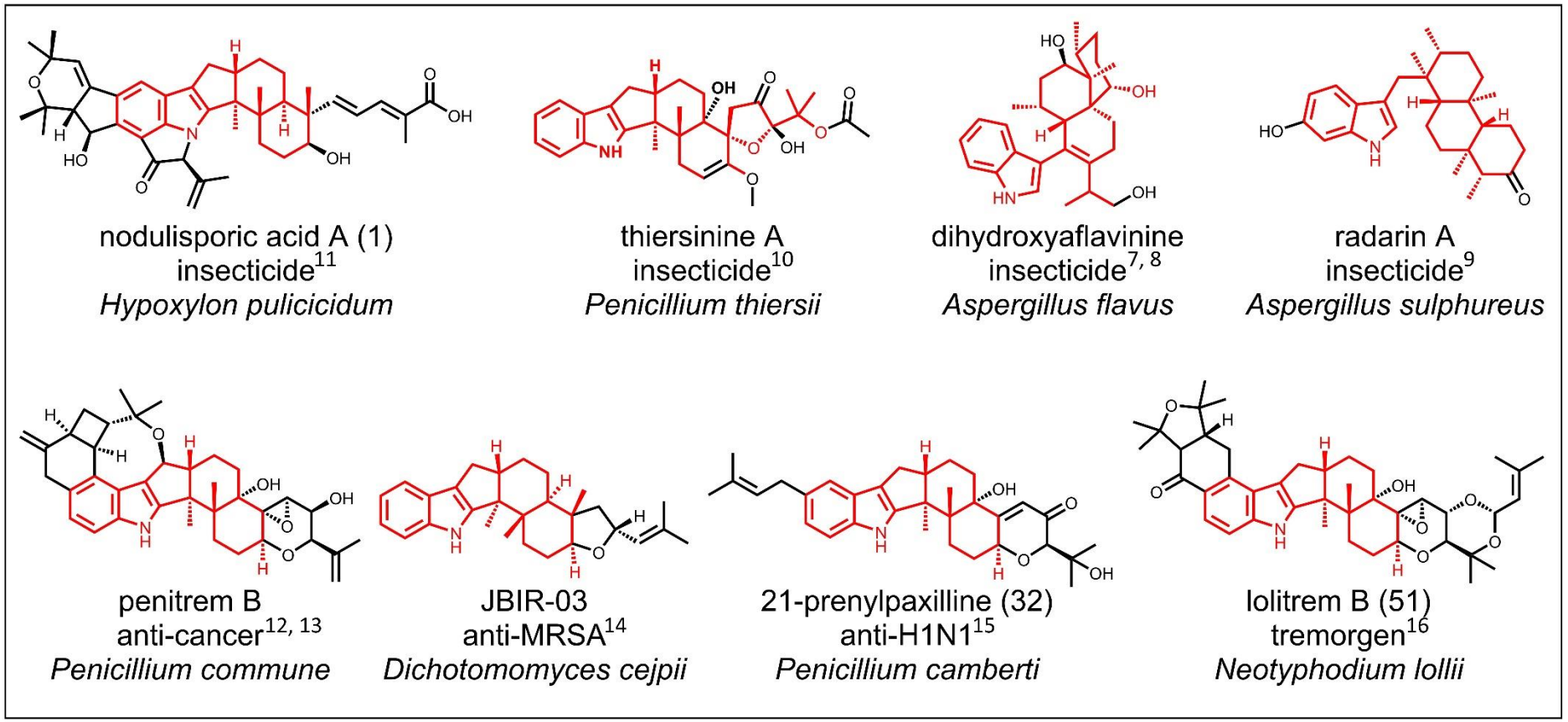


Figure 1.1. Examples of indole diterpenes with different bioactive properties. The different bioactive properties arise from the functional groups (shown in black) that decorate the shared indole diterpene core (shown in red).

The main goal of this thesis is to identify the genes involved in the biosynthesis of the indole diterpene nodulisporic acid A (1) – structure shown in Figure 1.1. Therefore, to start, what is currently known about the biosynthesis of indole diterpenes will be described to provide a foundation of knowledge that will be exploited throughout the rest of this thesis and used to examine how nodulisporic acids are produced by their native host, *Hypoxylon pulicicidum* (*Nodulisporium* sp.).[17]

A review of indole diterpene biosynthesis: From early labelled substrate feeding studies to recent pathway reconstitution studies

The biosynthesis of indole diterpenes was initially explored throughout the 1970s-90s using labelled substrate feeding studies;[18-22] where labelled precursors shed insight on how the indole diterpene core scaffolds were formed by their native fungal producers. In the late 1990s the first potential indole diterpene gene cluster, the *PAX* cluster, responsible for paxilline (2) production in the filamentous fungi *Penicillium paxilli*, was identified.[23-24] Throughout the 2000s to the early 2010s, the *PAX* cluster served as the pivotal reference for the identification of seven other indole diterpene clusters.[25-33] During this time, the individual functions of the secondary-metabolic genes involved in indole diterpene biosynthesis have been elucidated,[25-28, 31-43] which, when combined together, establish a database of information that can be used to probe for novel indole diterpene gene clusters. This section will review how the current understanding of indole diterpene biosynthesis has been established.

The biosynthesis of indole diterpenes has been established from the functional analysis of a number of indole diterpene genes through a combination of feeding studies with labelled[19-20, 22, 44-46] or unlabelled precursors,[29-30, 36, 39-40, 42] the analysis of indole diterpene products produced by knockout strains,[24, 26, 29-32, 38] and the reconstruction of biosynthetic pathways.[27-29, 31, 33, 38, 40-41, 43] Studies have found that the structural variability of indole diterpenes, which corresponds to their various bioactive properties, arises from genetic variation of specific indole diterpene genes within the genomes of the indole diterpene producing hosts (Figure 1.2). Typically, the complexity of the final indole diterpene product correlates with the number of genes required for its biosynthesis, where the more complex products like the lolitrems[26-27, 29] and penitrems[31-32] require more genes for their biosynthesis (10 genes and 17 genes

respectively), than the simpler indole diterpene products like paxillines[24, 34-40, 42] and aflatrems[25, 28, 39, 41] (7 genes each). To date, the functionality of genes in eight different indole diterpene biosynthetic pathways, which give rise to the paxillines,[24, 34-40, 42] aflatrems,[25, 28, 39, 41] lolitrems,[26-27, 29] shearinines,[32, 43] penitrems,[31-32] terpendoles,[30] aflavinines,[33] and emindoles,[33, 40] have been confirmed. Notably, the function of all the secondary-metabolic genes required for the biosynthesis of the major products in four of these indole diterpene classes, the paxillines,[35-36, 38, 40, 42] aflatrems,[39, 41] shearinines,[43] and penitrems,[31] have been confirmed and thus provide an excellent database to understand the complex indole diterpene biosynthetic mechanisms.

For example, the genes required to produce these four classes of indole diterpenes, six including the lolitrems and terpendoles, have all been found clustered together in the genomes of their native fungal producers (Figure 1.2), where each gene product in the cluster is linked to the catalysis of a specific biochemical reaction in the indole diterpene biosynthetic pathway. Naming of genes in indole diterpene clusters has followed the *Aspergillus nidulans* naming convention, where genes are given a name with a three-letter prefix in lower case that designates species, followed by a single letter suffix in upper case that designates gene function written in italic font (e.g. *paxG*). Naming of the corresponding protein product follows the same rules except that the initial letter of the prefix is upper case and the entire name is written in normal (non-italic) font (e.g. PaxG is the protein product of *paxG*). Therefore, homologous genes from different species (with the exception of the *ATM* clusters in *Aspergillus flavus* and *Aspergillus oryzae*)[28] contain a different three letter prefix but have been given the same single letter suffix (e.g. *paxG* is from *P. paxilli* and *ltmG* is from *Epicloë festucae*).

This introduction begins by highlighting the first four secondary-metabolic steps in indole diterpene biosynthesis that give rise to the library of indole diterpene cyclised cores that are often further decorated in distinct ways to install specific bioactive properties. From there, attention will turn to how these cyclised cores are decorated and aspects of the different indole diterpene gene clusters will be examined; including how the clusters were identified and how the genes have been functionally analysed. After elaborating on what is known about indole diterpene biosynthesis, the history of the main indole diterpenes of this thesis, the nodulisporic acids, will be presented.

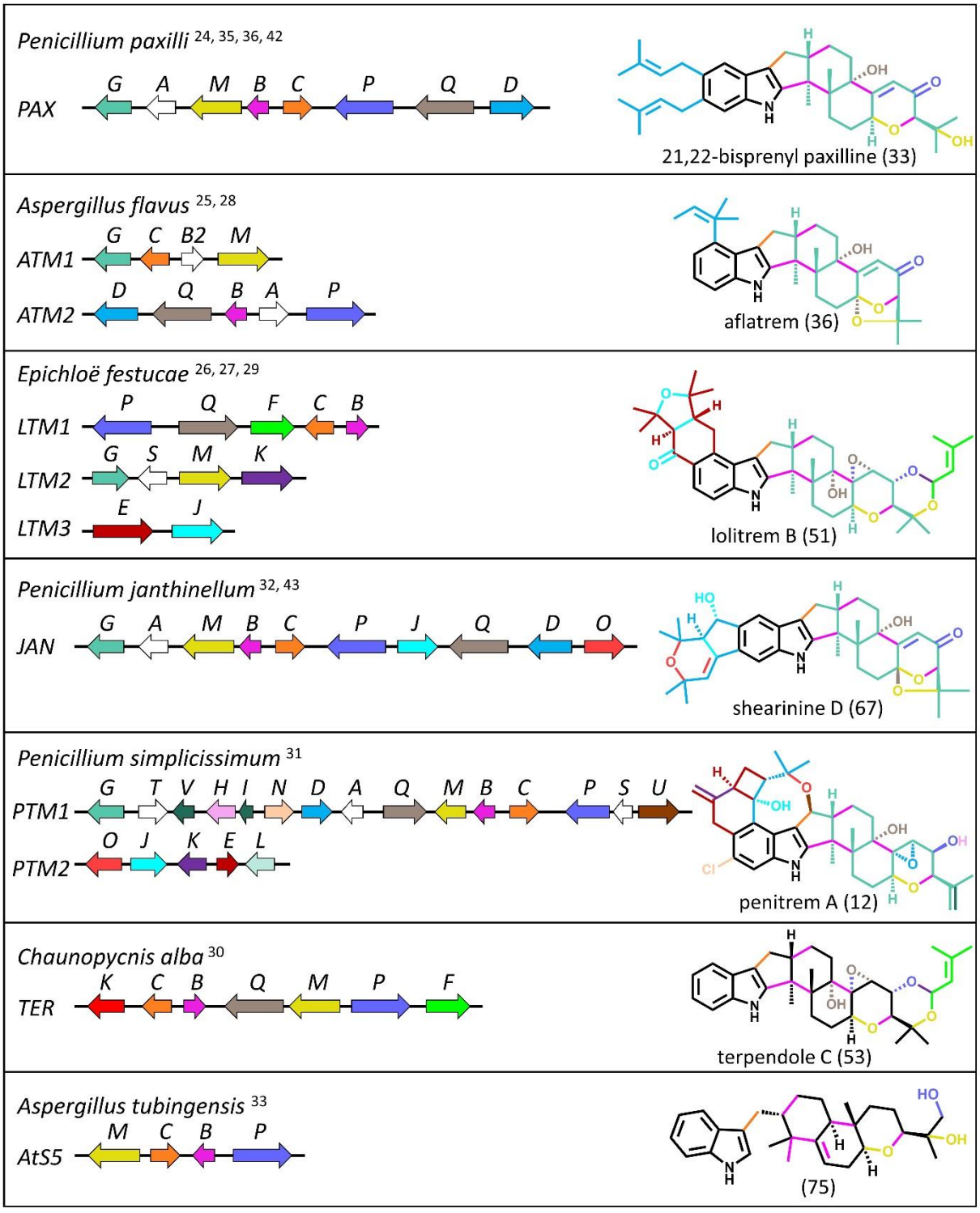


Figure 1.2. Compilation of known indole diterpene gene clusters. The name of the organism that each gene cluster was isolated from is listed above the cartoon of its corresponding gene cluster. Genes are represented by coloured arrows and named with a letter. The colour and letter of each gene corresponds to the function of the gene in indole diterpene biosynthesis. Genes that are homologous to one another have the same gene letter and colour. One of the indole diterpene products that each gene cluster codes for is listed to the right of the cluster, where the colour in the indole diterpene product indicates the specific moiety that is installed by the protein product of the corresponding coloured gene.

Primary metabolites involved in indole diterpene biosynthesis

The results from various indole diterpene biosynthesis studies have revealed that indole diterpenes are biosynthesised from the condensation of an indole-containing precursor and a twenty-carbon chain (diterpene) of geranylgeranyl pyrophosphate (3).[20-21, 40, 44] The indole-containing precursor is preferentially indole-3-glycerol phosphate (4)[40, 44] but alternate indole cores like indole (5)[40] and tryptophan (6)[19, 22] can also serve as a substrate for some geranylgeranyl transferase enzymes. These three indole-containing precursors are synthesised via the tryptophan (6) biosynthetic pathway, in which the key substrate, indole-3-glycerol phosphate (4), is specifically derived from anthranilate and phosphoribosyl pyrophosphate.[44]

Geranylgeranyl pyrophosphate (3) is synthesised from the two common terpene building blocks of activated isoprene units known as isopentyl pyrophosphate (7) and dimethylallyl pyrophosphate (8) shown in Figure 1.3. In fungi, animals, plants (cytosol), and archaea isopentyl pyrophosphate (7) is synthesised from 3 units of acetate (9) as an intermediate of the mevalonate pathway and dimethylallyl pyrophosphate (8) is synthesised from isopentyl pyrophosphate (7) by isopentyl pyrophosphate (7)-isomerase. In eubacteria, green algae, and higher plants isopentyl pyrophosphate and dimethylallyl pyrophosphate (8) are synthesised from D-glyceraldehyde 3-phosphate and pyruvate via the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway (syn. 1-deoxy-D-xylulose 5-phosphate or non-mevalonate pathway).[47] To form the fifteen-carbon chain of farnesyl pyrophosphate (10), one unit of dimethylallyl pyrophosphate (8) sequentially condenses with two units of isopentyl pyrophosphate (7). Farnesyl pyrophosphate (10) then condenses with isopentyl pyrophosphate (7) in a geranylgeranyl pyrophosphate (3) synthase-mediated condensation to form geranylgeranyl pyrophosphate (3) (Figure 1.3).

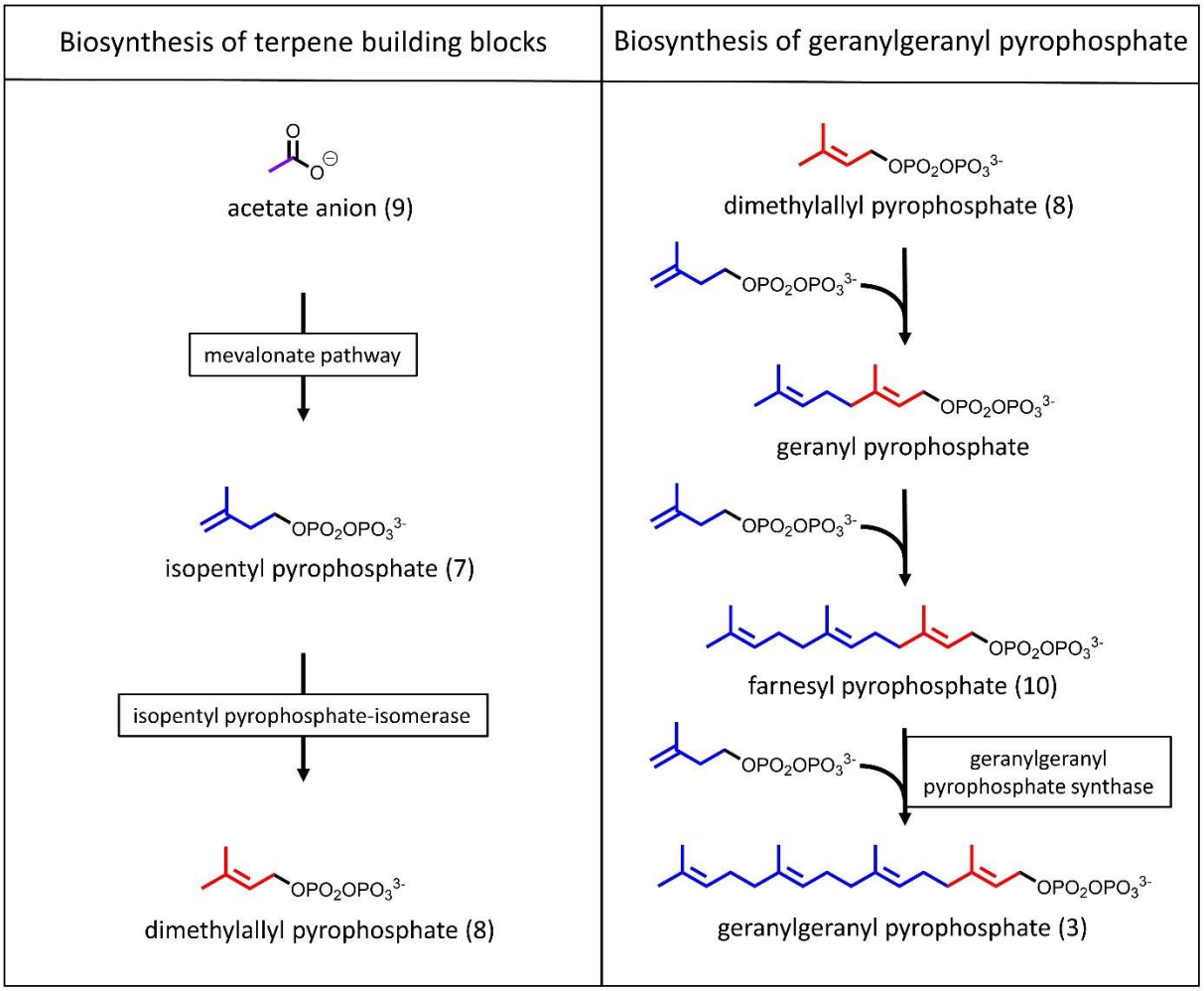


Figure 1.3. Depiction of the biosynthesis of the diterpene component [geranylgeranyl pyrophosphate (3)] of indole diterpenes.

Secondary metabolism step 1: Formation of geranylgeranyl pyrophosphate (3)

A combination of feeding studies with labelled substrates and gene functionality studies have established that the synthesis of is the first secondary-metabolic step in indole diterpene biosynthesis (Figure 1.4). Genomes of the known indole diterpene producers often contain two geranylgeranyl pyrophosphate (3) synthases – one dedicated to primary metabolism and one dedicated to secondary metabolism.[37] A key difference between the two is that the secondary- geranylgeranyl pyrophosphate (3) synthases contain a DDXDD motif in their *N*-terminus that is not present in any of the primary geranylgeranyl pyrophosphate (3) synthases.[24]

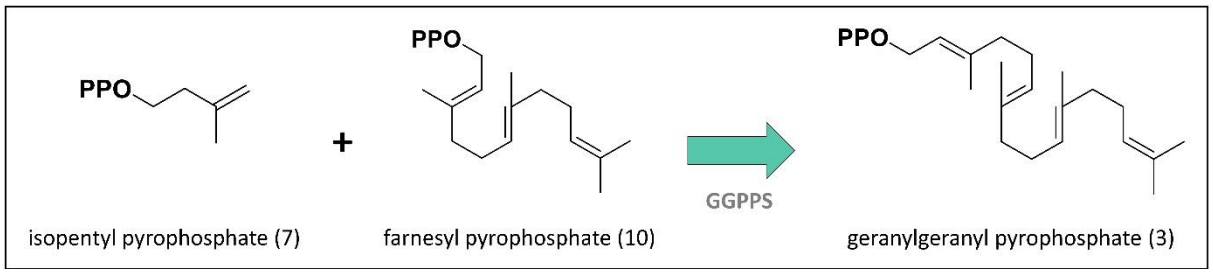


Figure 1.4. First secondary-metabolic step in indole diterpene biosynthesis catalysed by a geranylgeranyl pyrophosphate (3) synthase (GGPPS) enzyme (turquoise arrow). By convention, indole diterpene geranylgeranyl pyrophosphate (3) synthases are identified by the suffix G. For illustrative purposes block arrows have been used to link genes to chemistry. The name of the enzyme that the gene encodes is depicted below the arrow and the direction of the arrow indicates reaction direction. The likely cofactor for GGPPS enzymes is Mg^{2+} but not all enzymatic mechanisms have been thoroughly examined.

Feeding studies with labelled substrates have clearly shown that the diterpene moiety of the indole diterpene core is biosynthesised from twelve units of acetate (9), which make up the four isoprene units of geranylgeranyl pyrophosphate (3). Acklin *et al*. (1977) were the first to show that twelve units of [^{13}C]-acetate (9) were an essential component of the diterpene moiety in paspaline (11) biosynthesis in *Claviceps paspali*.[18] Further feeding studies on *Penicillium crustosum* by de Jesus *et al*. (1983) using ^{13}C and ^{2}H-labelled acetate (9), mevalonate, and mevalonolacetone concluded that six mevalonate units, four for the indole diterpene core and two for the northern region, were involved in the biosynthesis of penitrem A (12).[21] Similarly, feeding studies by Byrne *et al*. (2002) on *Hypoxylon pulicicidum* (formerly *Nodulisporium* sp.) using ^{13}C- labelled acetate (9) and mevalonolacetone showed that nodulisporic acids were biosynthesised from seven isoprene units, four for the indole diterpene core, two in the western region, and one in the southern region;[44] thereby confirming that twelve acetates (9) were involved in geranylgeranyl pyrophosphate (3) biosynthesis and nine involved in three isopentyl pyrophosphates (7) that form western and southern regions.

The identification of the first indole diterpene gene cluster, linked to paxilline (2) production in *P. paxilli*, further supported the involvement of a geranylgeranyl pyrophosphate (3) synthase due to the findings that a geranylgeranyl pyrophosphate (3) synthase (*paxG*), located within the *PAX* gene cluster, was essential for paxilline (2) biosynthesis.[24] Since then, five other secondary-metabolic geranylgeranyl pyrophosphate (3) synthase-encoding gene sequences [*atmG* (*A. flavus*),[28] *ltmG* (*E.*

festucae),[29] *janG* (*P. janthinellum*),[32] *ptmG* (*P. simplicissimum*),[31] and *penG* (*P. crustosum*)[32]] have been found clustered with their downstream indole diterpene genes (Figure 1.2).

Interestingly, there has been one reported indole diterpene producing organism, *Chaunopycnis alba*, that does not seem to have a secondary-metabolic geranylgeranyl pyrophosphate (3) synthase gene that is homologous to the known geranylgeranyl pyrophosphate (3) synthase genes (i.e. *paxG*, *atmG*, *ltmG*).[30] *C. alba* produces a group of indole diterpenes known as terpendoles, for which the *TER* gene cluster has been identified. Since the *TER* cluster clearly lacks a secondary-metabolic geranylgeranyl pyrophosphate (3) synthase gene, the *C. alba*, genome was screened with degenerate PCR primers designed to functionally-confirmed secondary-metabolic geranylgeranyl pyrophosphate (3) synthase genes but no homologues were identified. The high titres of indole diterpenes that *C. alba* produces suggest high flux of geranylgeranyl pyrophosphate into the pathway. Therefore, it has been speculated that *C. alba* does have a secondary-metabolic geranylgeranyl pyrophosphate (3) synthase gene but that no homolgoues were identified due to low sequence similarity.[30]

Secondary metabolism step 2: Geranylgeranylation of indole-containing substrate

The second biosynthetic step in indole diterpene biosynthesis is catalysed by a geranylgeranyl transferase (GGT) enzyme, which facilitates the indole condensation of geranylgeranyl pyrophosphate (3) and an indole-containing donor to make 3-geranylgeranylindole (13) as shown in Figure 1.5. Results from feeding experiments with labelled[19, 22, 44] and unlabelled[40] indole-core precursors have shown that geranylgeranyl transferases from different organisms have variability in their indole donor substrate preferences.

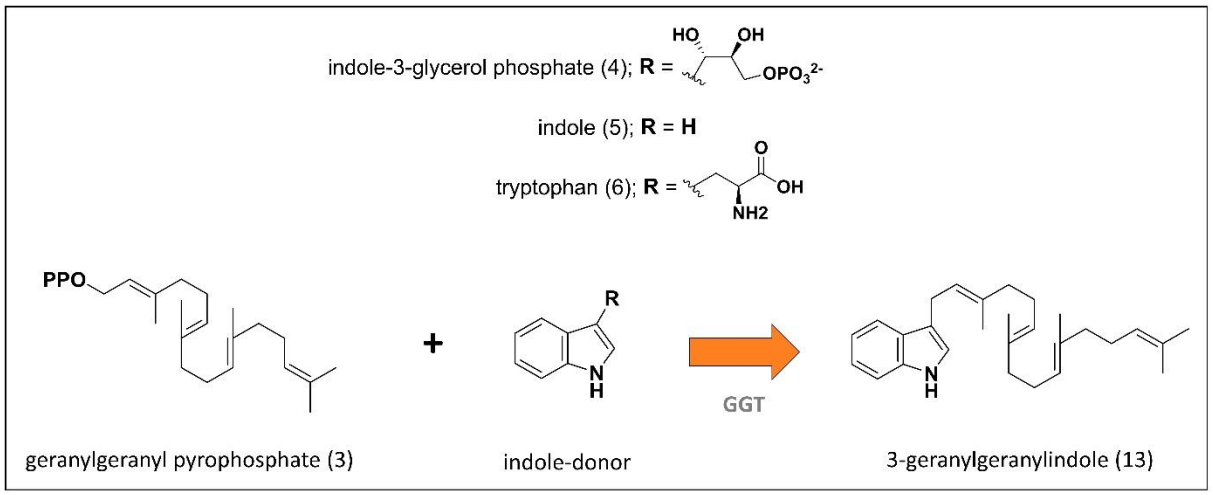


Figure 1.5. Second secondary-metabolic step in indole diterpene biosynthesis catalysed by a geranylgeranyl transferase (GGT) enzyme (orange arrow). By convention, indole diterpene geranylgeranyl transferases are identified by the suffix C. The likely cofactor for GGT enzymes is Mg^{2+} but not all enzymatic mechanisms have been thoroughly examined.

The first reported labelled indole-core precursor feeding studies on indole diterpene biosynthesis examined the effects of feeding two ^{14}C-labelled tryptophan (6) precursors, (2*S*)-[3-^{14}C]-tryptophan (6) and (2*R/S*)-[benzene-ring-U-^{14}C]-tryptophan (6), to *Penicillium crustosum*, a known producer of penitrems [see Figure 1.2 for srtructure of penitrem A (12)].[19] Results from this study revealed that the indole portion of tryptophan (6) was incorporated into penitrem A (12), but further examination into alternate indole-containing donors was never conducted.

For around 20 years it was generally assumed that tryptophan (6) was the preferred indole-donor in indole diterpene biosynthetic pathways - as the specific substrate for the geranylgeranyl transferase step of indole diterpene biosynthesis had not been identified. In further support of this, Mantle *et al*. published results in 1994, where they found that [^{14}C]-tryptophan (6) fed to *Penicillium paxilli* was incorporated into its indole diterpenes, the paxillines, but again, no other indole-containing precursors were tested.[22]

It was not until Byrne *et al*. (2002) published results of their indole-core precursor feeding studies with the nodulisporic acid indole diterpene producer *H. pulicicidum* (formerly *Nodulisporium* sp.[17]) that researchers in the field began to really question the substrate preference of the geranylgeranyl transferases. Byrne *et al*. (2002) were the first to propose the geranylgeranyl transferase substrate preference of indole-3-

glycerol phosphate (4), the direct precursor of tryptophan (6), on the basis that they were unable to incorporate indole-^{14}C-tryptophan (6), 2-^{13}C-indole-tryptophan (6), or 2-^{13}C-indole (5) into any of the nodulisporic acids but were successfully able to incorporate (ring-^{14}C6)-anthranilic acid and 1-^{13}C-D-ribose.[44]

To explain the conflicting results presented for the substrate specificity of the geranylgeranyl transferases, Tagami *et al*. (2013) carried out *in vitro* substrate conversion experiments using recombinant PaxC.[40] They individually subjected three indole-containing donors, indole-3-glycerol phosphate (4), indole (5), and tryptophan (6), to a PaxC-mediated reaction with geranylgeranyl pyrophosphate (3) and Mg^{2+} and conducted subsequent kinetic experiments. High-performance liquid chromatography (HPLC) analysis of their reaction products revealed high quantities of 3-geranylgeranylindole (13) in the reaction with indole-3-glycerol phosphate (4), moderate quantities of 3-geranylgeranylindole (13) in the reaction with indole (5), and trace quantities of 3-geranylgeranylindole (13) in the reaction with tryptophan (6). In accordance with the HPLC results, their subsequent kinetic experiments further exemplified that indole-3-glycerol phosphate (4) was the preferred indole-core substrate for PaxC (k_{cat}/K_M for indole-3-glycerol phosphate (4) was 28.2 mM^{-1} s^{-1} and for indole (5) was 3.6 mM^{-1} s^{-1}). Notably, the *in vitro* analysis of PaxC avoids interference from other enzymes in the fungi, which could convert the fed substrate [e.g. [^{14}C]-tryptophan (6)] into alternate compounds [e.g. [^{14}C]-indole-3-glycerol phosphate (4)] that in turn could be used as the geranylgeranyl transferase substrate. Therefore, it can be concluded that indole-3-glycerol phosphate (4) is the preferred indole-containing donor in the paxilline (2) biosynthetic pathway, but to ensure that this result can be extended to all indole diterpene biosynthetic pathways additional substrate conversion experiments using recombinant geranylgeranyl transferases and subsequent kinetic experiments are required.

In addition to PaxC's promiscuity around accepting various indole-containing donors, Liu *et al*. (2013) revealed PaxC's promiscuity around the prenyl donor.[39] PaxC kinetic experiments involving farnesyl pyrophosphate (10) as the prenyl donor revealed catalytic conversion to farnesylindole (14), albeit with a lower k_{cat}/K_M value (16.6 s^{-1} mM^{-1}) than Tagami *et al*. (2013) noted for geranylgeranyl pyrophosphate (3) (278.1 s^{-1} mM^{-1}).[40] These results indicate that geranylgeranyl pyrophosphate (3) is the preferred prenyl donor in comparison to farnesyl pyrophosphate (10). Unfortunately, no other

prenyl donors were tested, but this single observation of farnesyl pyrophosphate (10) conversion suggests that PaxC may readily accept a range of synthetic prenyl donor derivatives that contain a carbon tail composed of 15 to 25 carbons. One could presumably determine the extent of the substrate specificity through additional PaxC kinetic experiments using a wide range of prenyl donors.

Although PaxC is the only geranylgeranyl transferase that has been functionally analysed *in vitro*, functional analysis of two other geranylgeranyl transferases, AtmC from *A. flavus*[33] and LtmC from *E. festucae*,[27] have been completed *in vivo*. AtmC was functionally characterised using pathway reconstitution of *atmG* and *atmC* in *Saccharomyces cerevisiae*. Unexpectedly, the extract of the *atmG-atmC* transformant did not contain 3-geranylgeranylindole (13) [as seen in the *paxG-paxC* transformant from Tagami *et al*. (2013)[40]] but instead contained the 17,18-epoxy-3-geranylgeranylindole (15), which Tang *et al*. (2015) attributed to a putative endogenous yeast epoxidase.[33] Even though Tang *et al*. (2015) did not get 3-geranylgeranylindole (13) from the *atmG-atmC* transformant, they concluded that AtmC was a functional orthologue of PaxC. In contrast, LtmC was functionally characterised by Young *et al.* (2006) using complementation of *ltmC* with a *P. paxilli* Δ*paxC* deletion mutant. While the Δ*paxC* deletion mutant was unable to synthesise paxilline, the *ltmC* transformant restored paxilline (2) biosynthesis demonstrating that *ltmC* is a functional orthologue of *paxC*.

In addition to *paxC* and *atmC*, four other homologous geranylgeranyl transferase encoding genes, clustered with known indole diterpene genes, have been identified through sequence similarity [*ptmC* (*P. simplicissimum*),[31] *penC* (*P. crustosum*),[32] *janC* (*P. janthinellum*),[32] and *terC* (*Chaunopycnis alba*)[30]]. The high sequence similarity and their location within an indole diterpene cluster, suggests that these geranylgeranyl transferase encoding genes are functional orthologues of paxC. It may prove to be the case that the level of promiscuity of each gene product differs and thus further kinetic studies on recombinant geranylgeranyl transferases are required to identify subtle differences between these proteins, and to ascertain if sequence differences correspond to any differences in function.

Secondary metabolism step 3: Epoxidation of 3-geranylgeranylindole

At the third step in indole diterpene biosynthesis, the geranylgeranyl tail of 3-geranylgeranylindole (13) is activated for cyclisation by epoxidation. Fueki *et al*. (2004) first established that 3-geranylgeranylindole (13) was a common essential intermediate in indole diterpene biosynthesis through feeding studies with deuterium labelled [$^{2}H_{3}$]-3-geranylgeranylindole on *P. paxilli* and *Emericella desertorum*.[45] Results showed that [$^{2}H_{3}$]-3-geranylgeranylindole was readily incorporated into the respective indole diterpene products, paxilline (2) and emindole DA indicating that different epoxidised indole diterpene products utilised the same 3-geranylgeranylindole (13) substrate.

To date, three different types of epoxidations of the geranylgeranyl tail of 3-geranylgeranylindole (13) have been proposed (Figure 1.6) that result in three unique epoxidised products: (1) a single epoxide product, 13,14-epoxy-3-geranylgeranylindole (16), where the third alkene from the indole core is epoxidised, (2) a double epoxide product, bisepoxy-3-geranylgeranylindole (17)**,** where the third and fourth alkene from the indole core are epoxidised, and (3) a single epoxide product, 17,18-epoxy-3-geranylgeranylindole (15)**,** where the fourth alkene from the indole ring is epoxidised.

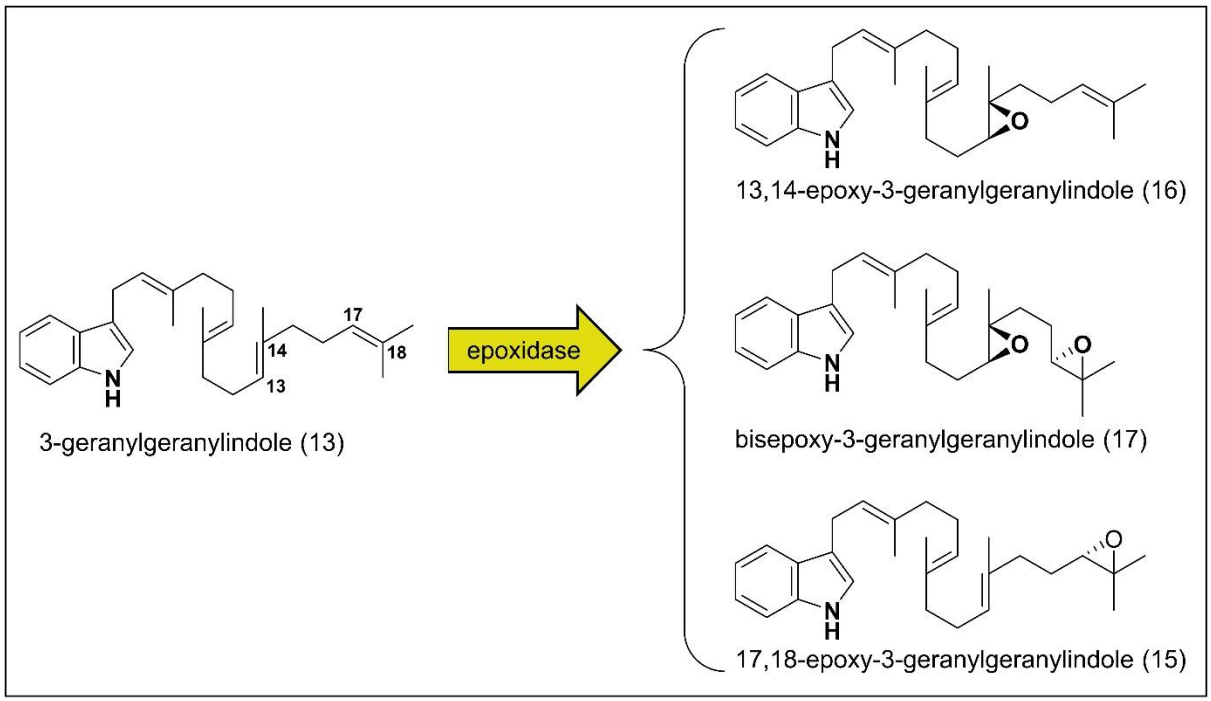


Figure 1.6. Third secondary-metabolic step in indole diterpene biosynthesis catalysed by an indole diterpene epoxidase enzyme (yellow arrow). Indole diterpene epoxidases from different species have the capacity to produce different epoxidised products [13,14-epoxy-3-geranylgeranylindole (16), bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15)]. By convention, indole diterpene 3-geranylgeranylindole (13)-epoxidases are identified by the suffix M. The likely cofactors for indole diterpene expoxidase enzymes are NADPH and O_2 but enzymatic mechanisms have not been thoroughly examined.

Full *in vitro* functional analysis of any of the indole diterpene epoxidases that give rise to 13,14-epoxy-3-geranylgeranylindole (16)**,** bisepoxy-3-geranylgeranylindole (17) and/or 17,18-epoxy-3-geranylgeranylindole (15) have not been completed, but putative functions have been deduced from the results of pathway reconstitution experiments.[33, 40] In 2013, Tagami *et al*. published results from their stepwise reconstruction of the *PAX* cluster from *P. paxilli* into in *Aspergillus oryzae*. Interestingly, HPLC analysis of the extract from the *paxG*, *paxC*, and *paxM* transformant produced two compounds not present in the wild type strain, 13,14-epoxy-3-geranylgeranylindole (16) and its respective diol, but no bisepoxy-3-geranylgeranylindole (17) was observed. Subsequent studies on the indole diterpene cyclase, *paxB*, revealed that *paxB* readily cyclised 13,14-epoxy-3-geranylgeranylindole (16) to emindole SB (18) and bisepoxy-3-geranylgeranylindole (17) to paspaline (11) leading Tagami *et al*. (2013) to propose the first two-path stepwise epoxidation scheme of 3-geranylgeranylindole (13) shown in Figure 1.7;[40] where in **path 1** the epoxidase (*paxM*) carries out a single epoxidation

to produce 13,14-epoxy-3-geranylgeranylindole (16) followed by a second epoxidation to produce bisepoxy-3-geranylgeranylindole (17) that gets passed onto *paxB* for cyclisation to paspaline (11), and in **path 2** 13,14-epoxy-3-geranylgeranylindole (16) is cyclised (by *paxB*) to form emindole SB (18) and then subjected to a second epoxidation (by *paxM*) and second cyclisation (by *paxB*) to form paspaline (11). Path 1 of their two path scheme seems the most reasonable due to likely substrate specificity problems of the *paxM* gene product, but the fact that no bisepoxy-3-geranylgeranylindole (17) was identified in the extract from the *paxG*, *paxC*, *paxM* transformant makes any conclusion on the preferred route from 3-geranylgeranylindole (13) to paspaline (11) unclear.

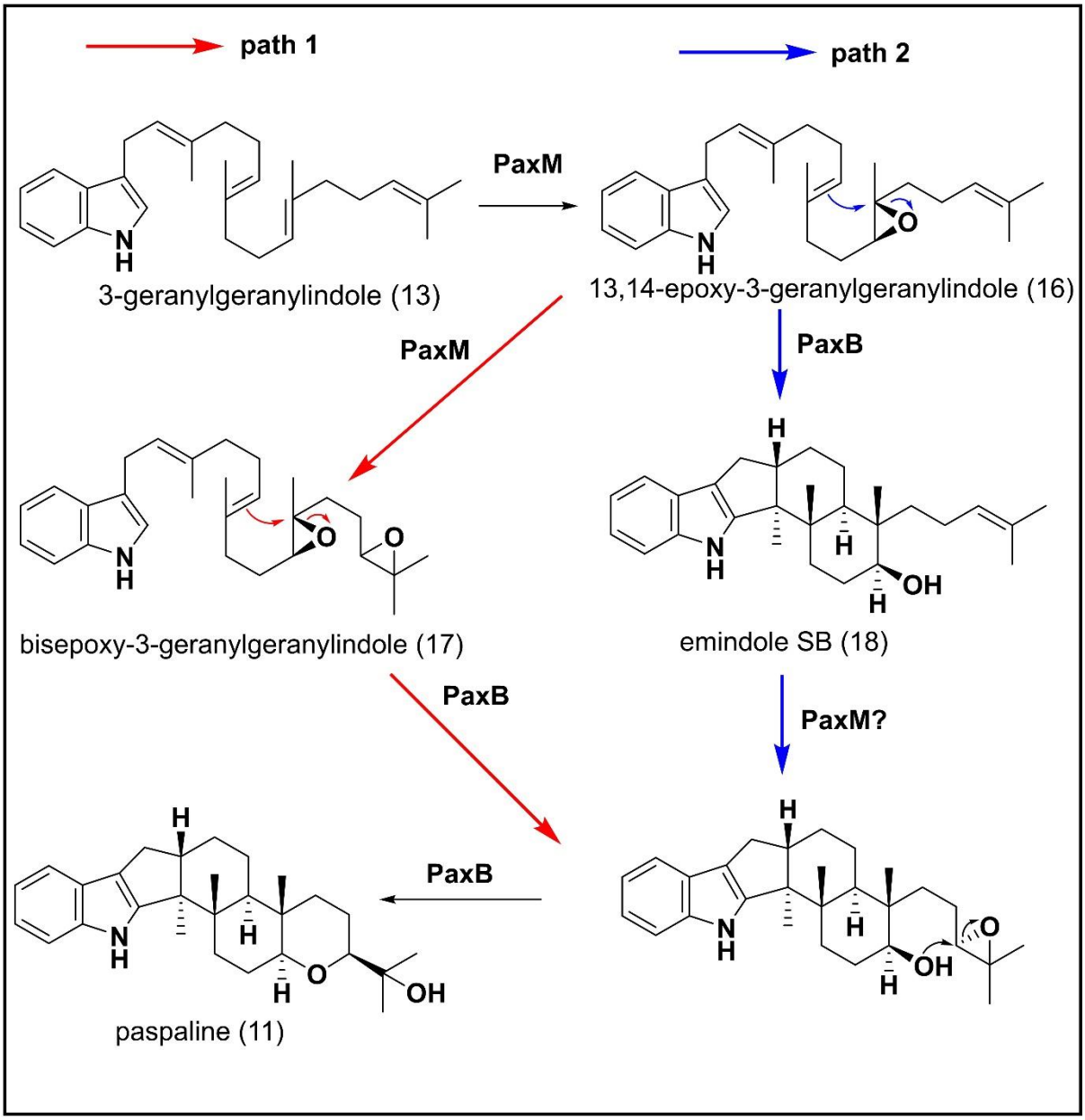


Figure 1.7. Two-path stepwise epoxidation scheme of 3-geranylgeranylindole (13) involving the epoxidase (PaxM) and the cyclase (PaxB) proposed by Tagami *et al*. (2013).[40]

A more recent study by Tang *et al*. (2015) published results from the reconstruction of part of the *ATM* gene cluster from *Aspergillus flavus* in *Saccharomyces cerevisiae*.[33] Notably, extract from the *atmG*, *atmC* transformant unexpectedly contained 17,18-epoxy-3-geranylgeranylindole (15), which they proposed was a product of an endogenous yeast epoxidase. Therefore, the bisepoxy-3-geranylgeranylindole (17) that they observed in the extract of the *atmG*, *atmC*, *atmM* transformant cannot conclusively be linked to activity of the *atmM* gene product alone. Overall, the

mechanism by which 3-geranylgeranylindole (13) gets epoxidised remains inconclusive and further *in vitro* substrate studies with recombinant epoxidases are required to reveal the preferred epoxidation mechanism.

In addition to the two epoxidases that were functionally analysed through pathway reconstitution [PaxM (*P. paxilli*)[40] and AtmM (*A. flavus*)[33]] and predicted to give rise to both the mono- and di-epoxide products 13,14-epoxy-3-geranylgeranylindole (16) and bisepoxy-3-geranylgeranylindole (17), one other epoxidase, LtmM (*E. festucae*),[29] has been functionally characterised through complementation of *ltmM* into a *P. paxilli* Δ*paxM* deletion mutant.[26] While the Δ*paxM* deletion mutant was unable to synthesise paxilline, the *ltmM* transformant restored paxilline (2) biosynthesis demonstrating that *ltmM* is a functional orthologue of *paxM* and thereby encodes the epoxidase capable of producing the necessary epoxide product for paxilline (2) biosynthesis; however discrepancies around whether that is the mono- or di-epoxide product [13,14-epoxy-3-geranylgeranylindole (16) or bisepoxy-3-geranylgeranylindole (17)] remains unclear.

In addition to these three functionally characterised epoxidases (PaxM, AtmM, and LtmM), four others [(PtmM (*P. simplicissimum*),[31] PenM (*P. crustosum*),[32] JanM (*P. janthinellum*),[32] TerM (*Chaunopycnis alba*)[30]], have been identified through sequence similarity and are proposed to be functional orthologues of PaxM, AtmM, and LtmM based on their hosts' abilities to produce paspaline (11)-derived indole diterpene products (i.e. penitrems, shearinines, and terpendoles; examples shown in Figure 1.2). Unlike the epoxidases that give rise to 13,14-epoxy-3-geranylgeranylindole (16) and bisepoxy-3-geranylgeranylindole (17), epoxidases that only give rise to the single epoxide product, 17,18-epoxy-3-geranylgeranylindole (15), have not been identified making the biosynthesis of compounds derived from 17,18-epoxy-3-geranylgeranylindole (15) relatively underexplored.

Together, the three epoxidised products [13,14-epoxy-3-geranylgeranylindole (16)**,** bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15)] establish the potential biosynthetic pathways by which the indole diterpenes can be cyclised (Figure 1.8). The first epoxidised product, 13,14-epoxy-3-geranylgeranylindole (16), where only the third alkene from the indole core is epoxidised, serves as a substrate in the biosynthesis of emindole SB (18) derived indole diterpenes (like the nodulisporic acids) and non-emindole SB derived compounds (like the aflavinines, anominines, tubingensins, eujindoles, emindole

SA/DA, and emeniveol-derivatives). The second epoxidised product, bisepoxy-3-geranylgeranylindole (17), where both the third and fourth alkenes are epoxidised, is used as a substrate in the biosynthesis of paspaline (11)-derived indole diterpenes (like the paxillines, lolitrems, aflatrems, penitrems, terpendoles, shearinines, janthitrems, and sulpinines) and some non-paspaline-derived indole diterpenes (like the thiersinines, anthcolorins, AtS5 compounds, emindole DB (19), and emindoles PC-PA). The third epoxidised product, 17,18-epoxy-3-geranylgeranylindole (15), where the fourth alkene from the indole core is epoxidised, acts as the substrate for the relatively unexplored biosynthesis of radarins and petromindole (20). The cyclisation mechanisms of the three epoxides [13,14-epoxy-3-geranylgeranylindole (16)**,** bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15)] will be explored in detail in the next section (**Indole diterpene biosynthesis: Step 4**).

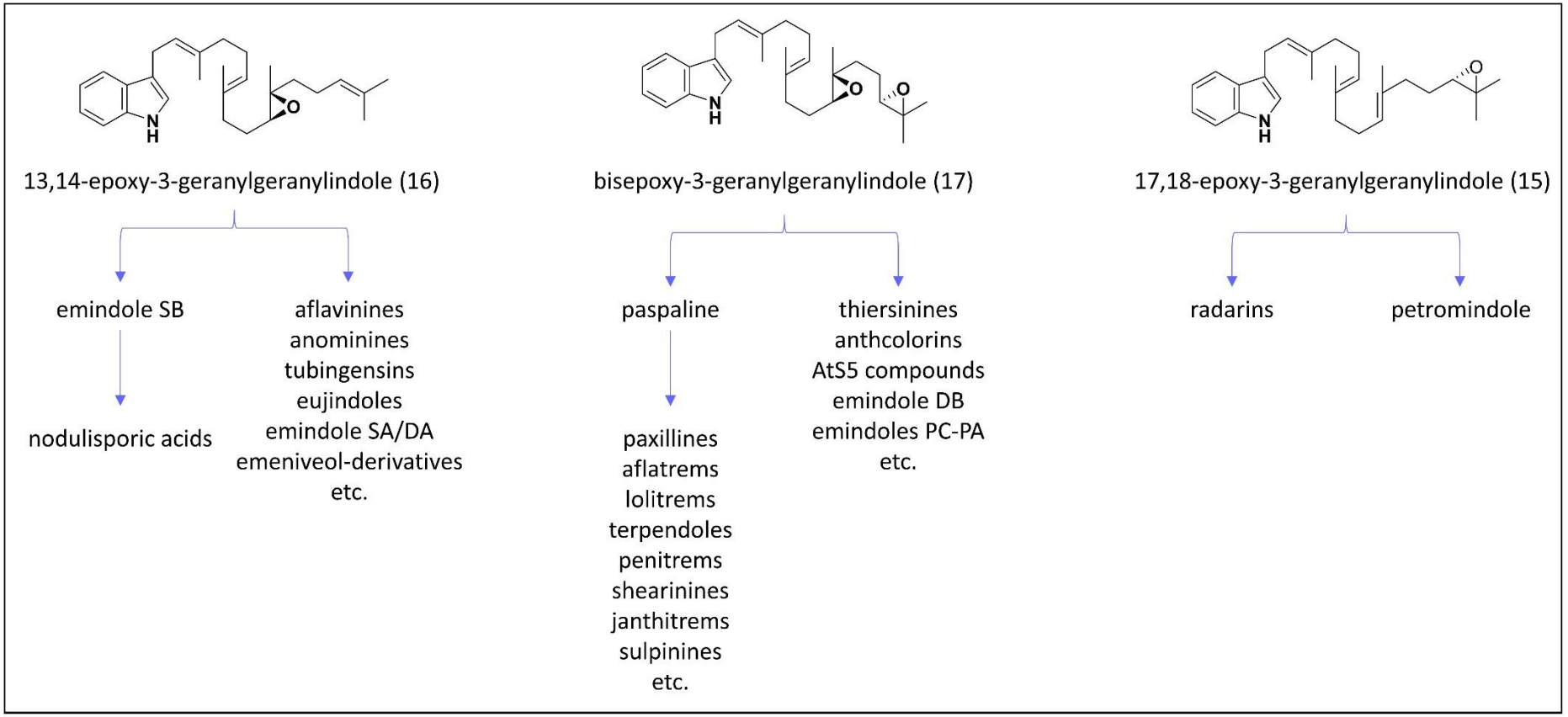


Figure 1.8. Depiction of the different biosynthetic pathways that arise from the three epoxidised products 13,14-epoxy-3-geranylgeranylindole (16), bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15).

Secondary metabolism step 4: Formation of cyclic indole diterpene scaffolds

In the fourth indole diterpene biosynthesis step the activated 3-geranylgeranylindole (13) epoxides [13,14-epoxy-3-geranylgeranylindole (16)**,** bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15)] are cyclised to form the skeleton library of tetra-, penta-, and hexa-cyclic indole diterpene scaffolds as shown in Figure 1.9. Each skeleton is derived from a cyclisation mechanism that follows either a Markovnikov reaction mechanism or an anti-Markovnikov reaction mechanism often coupled with selected sigmatropic shifts (e.g. H shifts, CH_3 shifts, and bond shifts). The cyclisation mechanisms can be divided into four key paths (Figure 1.9, blue arrows) from which all known indole diterpenes are biosynthesised. Three of these paths involve Markovnikov cyclisation mechanisms (Figure 1.9, path 1, 2 and 4) and one of them involves an anti-Markovnikov cyclisation mechanism (Figure 1.9, path 3). Each cyclisation path has been proposed based on results from the metabolic profiles from native fungi,[48] or the metabolic profiles from precursor feeding studies[18, 21, 44] and/or pathway reconstruction studies.[33]

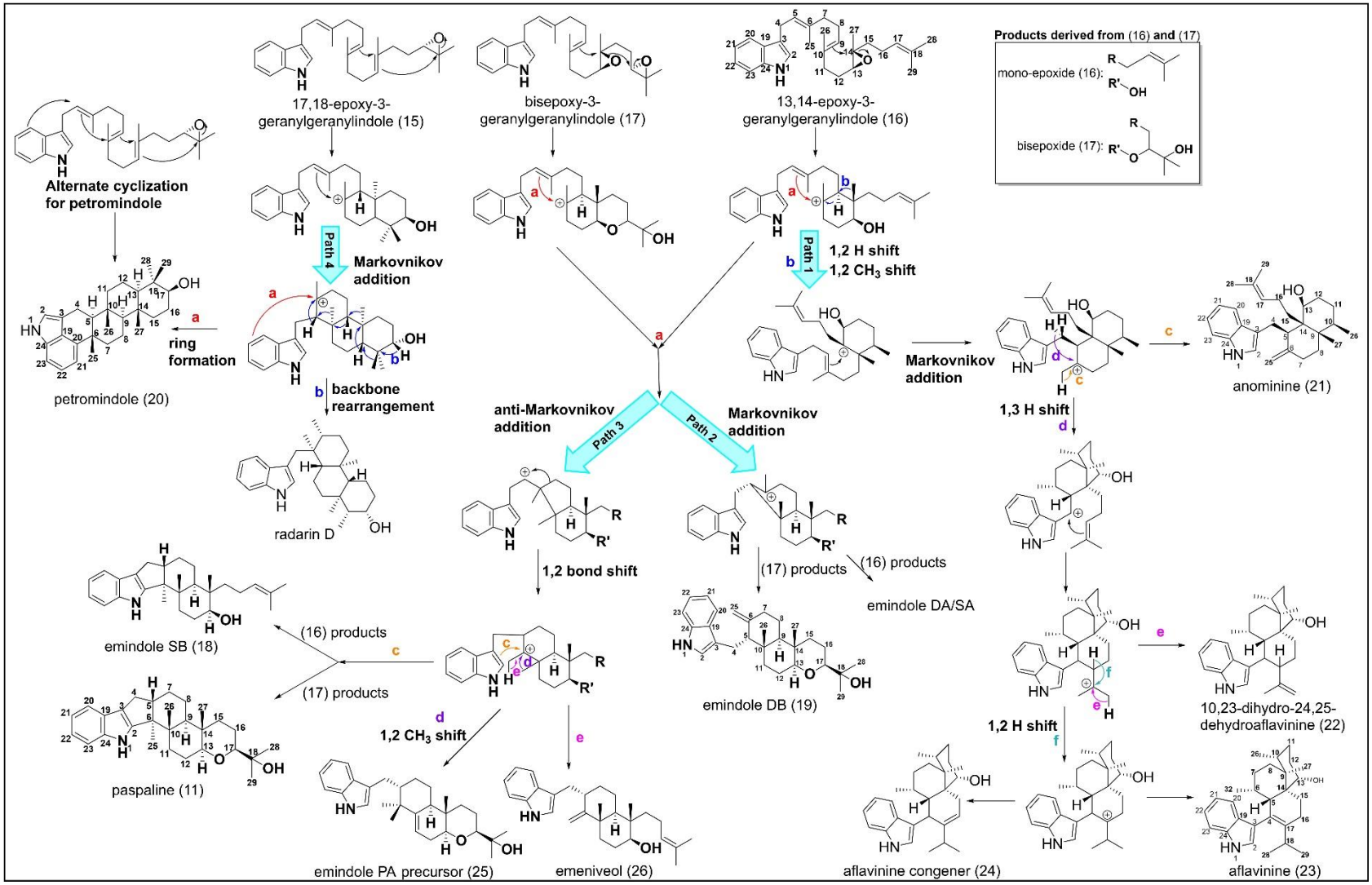


Figure 1.9. Fourth secondary-metabolic step in indole diterpene biosynthesis catalysed by an indole diterpene cyclase enzyme. Indole diterpene cyclases from different species have the capacity to cyclise the epoxidised products [13,14-epoxy-3-geranylgeranylindole (16), bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15)] in different ways.

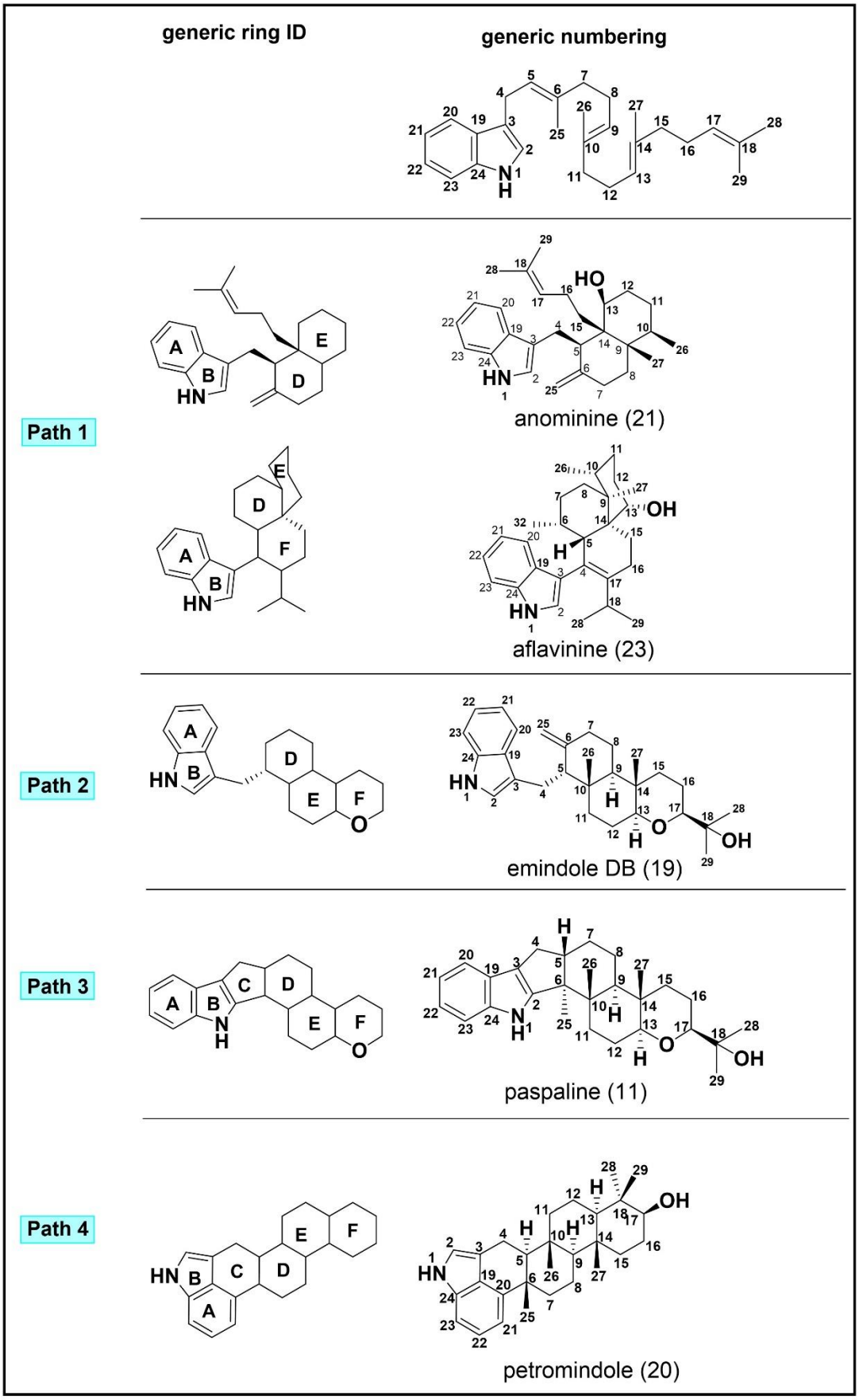


Figure 1.10. Summary of generic numbering scheme used in this thesis to classify the indole diterpene scaffolds.

Path 1 of Figure 1.9, first proposed by Tang *et al*. (2015), was established on the basis of phylogeny guided pathway reassembly of two unclustered indole diterpene cyclases (AtS2B, and AfB from *Aspergillus tubingensis* and *Aspergillus flavus* respectively) in *Saccharomyces cerevisiae*.[33] Results from their reconstitution studies revealed that AtS2B and AfB readily cyclised 13,14-epoxy-3-geranylgeranylindole (16) into anominine (21) and 10,23-dihydro-24,25-dehydroaflavinine (22) and that AfB could additionally cyclise 13,14-epoxy-3-geranylgeranylindole (16) into aflavinine (23) and its congener (24). They were then able to propose the most likely cyclisation mechanism that the two enzymes catalysed in which 13,14-epoxy-3-geranylgeranylindole (16) undergoes a 1,2-H shift and 1,2-CH_3 shift followed by subsequent Markovnikov driven cyclisation before splitting into two subpaths (c and d of Path 1) to produce anominine (21) as shown in path c and the aflavinines as shown in path d. Where subpath c of Path 1 involves a simple deprotonation, path d involves a 1,3-H shift and ring closure before branching into two additional subpaths e and f to deliver 10,23-dihydro-24,25-dehydroaflavinine (22) via deprotonation (path e) and aflavinine (23) and its congener (24) via a 1,2-H shift and subsequent deprotonation (path f). Since these cyclisation mechanisms are simply proposals, future work with labelled substrates is required to confirm the exact chemical mechanisms of Path 1 indole diterpene cyclases.

In contrast to Path1 of Figure 1.9, no sigmatropic shifts of the indole diterpene carbon skeleton are required to produce the various emindole indole diterpene products derived from Path 2 [e.g. emindole DA,SA, and DB (19)]. Instead, products from Path 2 are derived from the single or double epoxide products 13,14-epoxy-3-geranylgeranylindole (16) and bisepoxy-3-geranylgeranylindole (17) via a simple Markovnikov driven ring closure and subsequent deprotonation. Notably, Path 2 was first proposed by Nozawa *et al*. (1988) and the indole diterpene cyclases that catalyse this pathway have yet to be identified.

Similar to Path 2 of Figure 1.9, Path 3 involves cyclisation of both the single and double epoxide products 13,14-epoxy-3-geranylgeranylindole (16) and bisepoxy-3-geranylgeranylindole (17) but follows an anti-Markovnikov driven ring closure. The ring closure is the common precursor for a variety of subpaths (c,d, and e) to produce emindole SB (18), paspaline (11), the emindole PA precursor (25), and emeniveol (26). The involvement of a 1,2-bond shift has been proposed on the basis of coupling observed between C5 and C7 (See paspaline (11) in Figure 1.10) in the [2-^{13}C]-acetate

(9) derived indole diterpene products (paspalines, penitrems, and nodulisporic acids) from feeding studies on *C. paspali*,[18] *P. crustosum*,[21] and *H. pulicicidum*[44] respectively. Notably all of the indole diterpene products involved in the acetate (9)-labelled feeding studies are derived from subpath c of Path 3 and the functional examination of three indole diterpene cyclases [AtmB (*A. flavus*),[33] AtS5B1(*A. tubingensis*),[33] PaxB (*P. paxilli*)[40]], has revealed that all three cyclases have the ability to produce both subpath c products, emindole SB (18) and paspaline (11). In addition to the three cyclases that have been functionally confirmed, five others [PtmB (*P. simplicissimum*),[31] LtmB (*E. festucae*),[29] PenB (*P. crustosum*),[32] JanB (*P. janthinellum*),[32] TerB (*Chaunopycnis alba*)[30]], have been proposed to have the same functionality as AtmB, AtS5B1, and PaxB based on sequence identity and location within indole diterpene gene clusters. As for the other two subpaths (d and e) of Path 3, one indole diterpene cyclase [AtS5B1 (*A. tubingensis*)[33]] has been identified for subpath d involving a 1,2-CH_3 shift, and none have been identified for subpath e.

In comparison to Paths 1-3 of Figure 1.9, Path 4 is unique in that it does not involve a cyclisation scheme activated by an internal epoxide moiety. Rather, a cyclisation scheme involving the formation of two carbocyclic rings (E and F, See figure 1.10, path 4) from 17,18-epoxy-3-geranylgeranylindole (15) is proposed to give rise to petromindole (20) and the radarins via subpath a and b of Path 4 respectively. The lack of the internal epoxide is linked to *re*-face addition as opposed to the *si*-face addition seen in Paths 1-3 giving rise to the reversal in stereochemistry of the petromindole (20) and radarin stereocenters. To date, only one Path 4 cyclase has been identified and surprisingly it was not from a fungus known to produce petromindole (20) or radarins. Rather, the cyclase identified was a lupeole synthase (LUP1) from the plant species *Arabidopsis thaliana*, leading Xiong *et al*. (2003) to hypothesise that the cyclase in the petromindole (20) and radarin producing endophytes is a modified oxidosqualene cyclase distant from other indole diterpene cyclases.[49] They further proposed that the putative indole diterpene cyclase likely evolved from a lanosterol synthase or a pentacyclic triterpene synthase. Overall, more work is required to identify the petromindole (20) and radarin indole diterpene cyclases and confirm their mechanisms of action.

In conclusion, the indole diterpene cyclases encompass a class of enzymes that catalyse an extremely wide range of reactions to establish the core library of cyclised

indole diterpene compounds. These core compounds are often decorated further to install the potent bioactive properties seen across the known indole diterpenes. In the next section, genes encoding enzymes for these decorative steps will be reviewed.

Decoration of the cyclic scaffolds: Gene identification and elucidation of the decoration steps

Genes encoding decorative enzymes have been identified through sequence analysis followed by functional studies on gene knockout strains,[24, 29-30, 32, 34-35] pathway complementation[25-29] or reconstitution studies,[31, 33, 35, 40-41, 43] precursor feeding studies,[29, 31, 36] and *in vitro* recombinant enzyme analyses.[31, 39, 42-43] These decorative enzymes fall into five categories [cytochrome P450 oxygenases (P450s), FAD/FMN dependent oxygenases (FOs), $NAD(P)^+$ dependent oxidoreductases (NOs), prenyl transferases (PTs), acetyl transferases (ATs), and unknowns (UKs)] where each category represents a type of reaction. To date, indole diterpene decorative enzymes from seven fungal species within two fungal classes [Sordariomycetes (*Aspergillus flavus*, *Aspergillus terreus*, *Penicillium janthinellum*, *Penicillium paxilli*, *Penicillium simplicissimum*) and Eurotiomycetes (*Chaunopycnis alba*, *Epichloë festucae*)] have been functionally identified creating a library of 36 known decorative genes, which include 21 P450s [PaxPQ (*P. paxilli*),[36, 40] AtmPQ (*A. flavus*),[28, 41] LtmPQJK (*Epichloë festucae*),[29] TerPQK (*Chaunopycnis alba*),[30] PtmPQJKLU (*P. simplicissimum*),[31] JanPQJ (*Penicillium janthinellum*),[43] and AtS5-P450 (*A. terreus*)[33]], 4 FOs [PaxO (*P. paxilli*),[38] JanO (*Penicillium janthinellum*),[43] PtmNO (*P. simplicissimum*)[31]], 1 NO [PtmH (*P. simplicissimum*)[31]], 8 PTs [PaxD (*P. paxilli*),[42] AtmD (*A. flavus*),[39] LtmFE (*Epichloë festucae*),[29] TerF (*Chaunopycnis alba*),[30] PtmDE (*P. simplicissimum*),[31] JanD (*Penicillium janthinellum*)],[43] 1 AT [PtmV (*P. simplicissimum*)[31]], and 1 unknown [PtmI (*P. simplicissimum*)[31]]. This gene library establishes a suite of information that essentially provides a ‘recipe’ book to install selectively various functional groups onto the indole diterpene skeletons formed in Step 4 of indole diterpene biosynthesis. The various classes of indole diterpene decorative genes will be discussed in detail below.

Identification of gene clusters enables rapid confirmation of individual gene function

The various gene clusters shown in Figure 1.2 were discovered over the course of 2001-2015 and their discovery enabled subsequent gene functionality studies. During this time the technological advancements around genetic sequencing and molecular biology rapidly expedited the amount of time required to identify novel genes and functionally confirm their roles in indole diterpene biosynthesis. Although the feeding studies on native indole diterpene producers with labelled substrates gave insight into how the various indole diterpene cores were biosynthesised, it was the identification of the indole diterpene gene clusters themselves that was paramount in understanding how the various indole diterpene cores were decorated. Below, methods of how the known indole diterpene genes have been identified and functionally characterised will be examined.

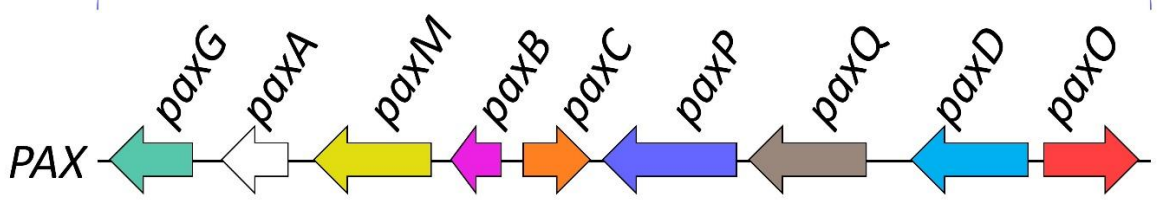
Identification of the *PAX* gene clusters and functional analysis of *paxP*, *paxQ* and *paxD*

The *PAX* cluster (Figure 1.11) was the first indole diterpene cluster identified. Using plasmid mutagenesis Young *et al*. (1998) isolated three paxilline-negative mutants, the first by random integration of the pAN7-1 plasmid into the *P. paxilli* genome and the other two by homologous integrations of a replacement construct at the locus identified by the first random integration.[23] Their results indicated that genes required for paxilline (2) biosynthesis were located between a 100-200 kb sequence on chromosome Va of the *P. paxilli* genome.

To determine where the *PAX* genes were located within the 100-200 kb sequence, Young *et al*. (2001) carried out a second round of plasmid mutagenesis with pAN7-1 using restriction enzyme-mediated integration.[24] From the thin-layer chromatography (TLC) analysis of the extracts of 750 transformants they isolated one transformant (LM-662) that did not produce paxilline. Molecular analysis of the plasmid insertion site revealed that the locus tagged with pAN7-1 was unlikely to be involved in paxilline (2) biosynthesis. Subsequent southern hybridisation of the LM-662 genome revealed a second site mutation corresponding to a 22.3 kb deletion that contained a portion of

the 17 genes (*paxGMCPQDNRSHTOUVWXY*) they hypothesised may be involved in paxilline (2) biosynthesis.

PAX cluster from *Penicillium paxilli* (strain PN2013)



Protein product	Proposed function	Secondary-metabolic biosynthetic step
PaxG	Geranylgeranyl pyrophosphate (3) synthase	1
PaxA	Unknown	-
PaxM	FAD-dependent oxygenase (epoxidase)	3
PaxB	Cyclase	4
PaxC	Geranylgeranyl transferase	2
PaxP	Cytochrome P450 oxygenase	5
PaxQ	Cytochrome P450 oxygenase	6
PaxD	Prenyl transferase	7
PaxO	FAD-dependent oxygenase	8

Figure 1.11. Depiction of the *PAX* cluster from *P. paxilli* strain PN2013.[38]

As this was the first indole diterpene cluster identified, and the whole *P. paxilli* genome was not known, Young *et al*. (2001) were surprised to find that *P. paxilli* paxilline-negative mutants were able to survive without a geranylgeranyl pyrophosphate (3) synthase (*paxG*). This led them to search for another geranylgeranyl pyrophosphate (3) synthase, which they found on chromosome IV, using degenerate primers designed to conserved domains of fungal geranylgeranyl pyrophosphate (3) synthases, and termed *ggs1*. Molecular analysis of each geranylgeranyl pyrophosphate (3) synthase revealed that *paxG* was unique from all other geranylgeranyl pyrophosphate (3) synthases characterised at the time as it

contained a DDXDD motif in the *N*-terminus that was not present in any other geranylgeranyl pyrophosphate (3) synthase. To confirm the involvement of *paxG* in paxilline (2) biosynthesis, they disrupted *paxG* in the wild-type strain and observed the loss of paxilline (2) production. Thus, they confirmed that *paxG* was a geranylgeranyl pyrophosphate (3) synthase dedicated to secondary metabolic synthesis and specifically the synthesis of paxilline.

Finally, Young *et al*. (2001) examined the expression levels of mRNA for *paxG*, *paxM*, and *paxP* to see if they were upregulated during paxilline (2) biosynthesis. Semi-quantitative RT-PCR monitoring revealed that all three genes were upregulated with the onset of paxilline (2) production (~60 h; Figure 1.12, A) further suggesting that they had identified the gene locus responsible for paxilline (2) biosynthesis and highlighting that there is a coregulation of the *PAX* genes with the onset of paxilline (2) production.

From there, attention was shifted to two of the P450 oxygenases in the *PAX* cluster, PaxP and PaxQ. McMillan *et al*. (2003) showed that PaxP and PaxQ deletion mutants were unable to produce paxilline (2) and instead accumulated paspaline (11) and 9-desoxypaxilline (27), respectively as shown in Figure 1.12, C[34] (notably, the '9' in 9-desoxypaxilline (27) is numbered based on the numbering scheme shown in Figure 1.10 and will be so numbered throughout this thesis). In reports published by Barry Scott's lab they have used a different numbering scheme where this compound was named 13-desoxypaxilline).[34, 36, 38] These results provided the first evidence that PaxP was responsible for the C14 demethylation and subsequent C16 oxidation to form 9-desoxypaxilline (27) and that PaxQ catalysed the C9 oxidation of 9-desoxypaxilline (27) to form paxilline.

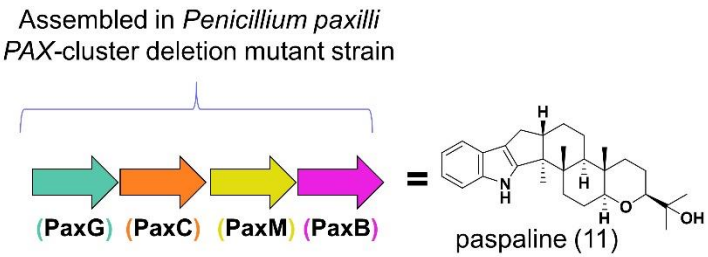
Working from the insights of Young *et al*. (2001)[24] and McMillan *et al*. (2003),[34] Saikia *et al*. (2006)[35] planned to define the minimum number of genes required for paspaline (11) biosynthesis. Results from *PAX* gene deletion experiments revealed that *paxG* [Young *et al*. (2001)], *paxC* (unpublished results from Young *et al*.),[35] *paxM* (unpublished McMillan *et al*.),[35] *paxA*, and *paxB* [Saikia *et al*. (2006)] deletion mutants lacked the ability to produce identifiable indole diterpenes. Thus, they predicted that a maximum of five genes (*paxGCMBA*) may be involved in paspaline (11) biosynthesis. They prepared plasmids containing various combinations of these five genes and transformed them into a *P. paxilli* knockout mutant that lacked the entire *PAX* cluster (strain CY2). Their results indicated that only four genes, *paxGCMB*, were necessary

for paspaline (11) production (Figure 1.12, B) thereby leaving discrepancies around the absence of indole diterpenes produced by the previously generated *paxA* deletion mutant. One possibility is that a portion of the *paxA*-flanking genes, *paxG* or *paxM*, may have been deleted as well in the construction of the mutant, which prevented paspaline (11) biosynthesis; however, re-introduction of wild-type *paxA* restored paxilline (2) production in the *paxA* knockout strain indicating that *paxA* may have some significant role. Notably, *paxA* homologues are conserved throughout many other known indole diterpene clusters (i.e. *atmA*, *janA*, *ptmA*) indicating they may have a role in indole diterpene biosynthesis; however, they are not required for heterologous biosynthesis of indole diterpenes. For example, Tagami *et al.* (2013) initiated paxilline (2) production in *A. oryzae* by heterologously expressing six *PAX* genes (*paxGCMBPQ*)[40] and van Dolleweerd *et al.* (2018) restored paxilline (2) production in a full *PAX* cluster *P. paxilli* knockout strain (CY2) by reconstituting the same six *PAX* genes (*paxGCMBPQ*),[3] clearly demonstrating that *paxA* is not required for paxilline (2) biosynthesis.

A. Gene expression and deletion mutants [Young *et al*. (2001), McMillan *et al*. (2003), Saikia *et al*. (2006)]

PAX	Gene expression	Products of gene deletion mutants
G	Upregulated with onset of paxilline production (~60 h) [Young *et al*. (2001)]	no identifiable IDTs [Young *et al*. (2001)]
M	Upregulated with onset of paxilline production (~60 h) [Young *et al*. (2001)]	no identifiable IDTs (McMillan *et al*., unpublished)
C		no identifiable IDTs (Young *et al*., unpublished)
P	Upregulated with onset of paxilline production (~60 h) [Young *et al*. (2001)]	paspaline [McMillan *et al*. (2003)]
Q		9-desoxypaxilline [McMillan *et al*. (2003)]
A		no identifiable IDTs [Saikia *et al*. (2006)]
B		no identifiable IDTs [Saikia *et al*. (2006)]

B. Pathway reconstitution [Saikia *et al*. (2006)]



C. Gene deletion analysis [McMillan *et al*. (2003)]

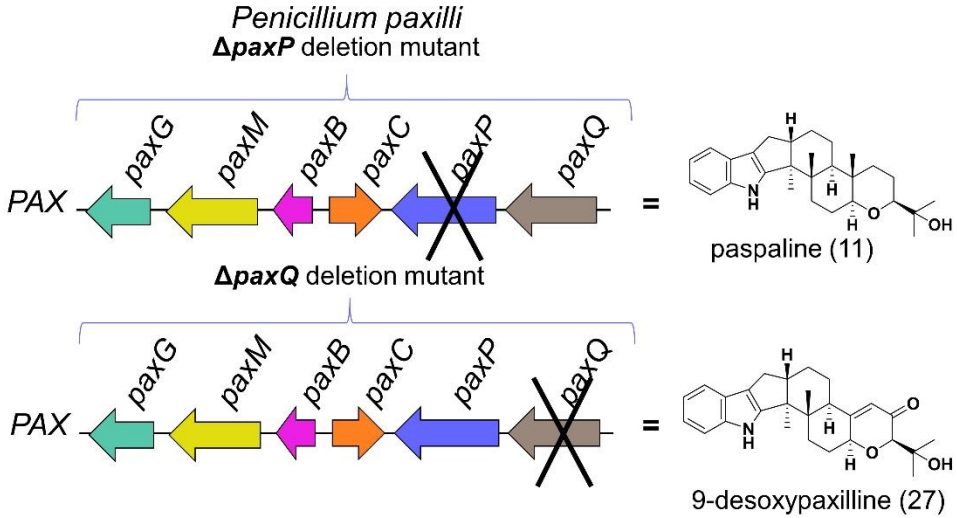


Figure 1.12. Summary of *PAX* gene expression, deletion, and pathway reconstitution results from Young *et al*. (2001),[24] McMillan *et al*. (2003),[34] and Saikia *et al*. (2006).[35]

Identification of the minimum number of genes required for paspaline (11) biosynthesis enabled Tagami *et al*. (2013) to evaluate the individual biosynthetic steps involved in paspaline (11) biosynthesis by reconstituting paspaline (11) biosynthesis in the heterologous host *Aspergillus oryzae*.[40] Stepwise introduction of *paxGCMB* combined with *in vitro* reconstituted protein studies revealed that PaxC is a geranylgeranyl transferase responsible for condensation of the indole-containing core and geranylgeranyl pyrophosphate (3) to form 3-geranylgeranylindole (13) and that PaxM and PaxB catalyse the stepwise epoxidation and cyclisation mechanism to form emindole SB (18) and paspaline (11) as shown in Figure 1.13.

Following the functional analysis of *paxC*, *paxM* and *paxB*, Tagami *et al*. (2013) tandemly transformed *paxP* and *paxPQ* into their four gene (*paxGCMB*) paspaline (11)-producing *A. oryzae* mutant, which initiated 9-desoxypaxilline (27) and paxilline (2) production from the respective *paxGCMBP* and *paxGCMBPQ* strains. Taken together, all of these results indicate that paxilline (2) production requires six genes (*paxGCMBPQ*; Figure 1.13).

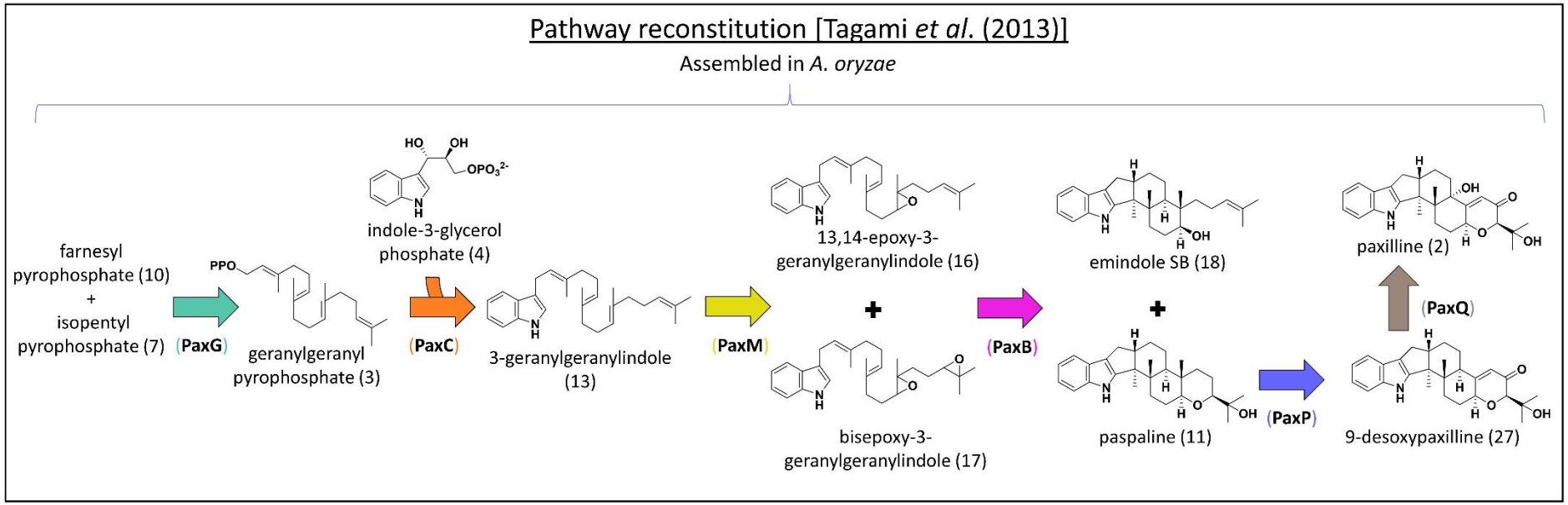


Figure 1.13. Summary of results pathway reconstitution experiments by Tagami *et al*. (2013).[40] Arrows represent enzymatic steps encoded by the genes in the *PAX* cluster.

Prior to the reconstitution study by Tagami *et al*. (2013), Saikia *et al*. (2007) completed the first functional analysis of the two P450 oxygenase-encoding genes in the *PAX* cluster, *paxP* and *paxQ*.[36] As suggested by the *paxPQ* gene knockout results from McMillan *et al*. (2003), the conversion of paspaline (11) to paxilline (2) is catalysed by *paxP* and *paxQ*. Thus, Saikia *et al*. (2007) conducted substrate feeding experiments on *P. paxilli* mutants containing either *paxP* or *paxQ* to confirm the catalytic abilities of each gene product. Upon feeding paspaline (11) to the *paxP* mutant strain, they observed conversion to two new metabolites, one major and one minor, identified as 9-desoxypaxilline (27) and β-PC-M6 (28) respectively (Figure 1.14, i.). They predicted that PaxP was capable of stepwise oxidation of paspaline (11) to first form β-PC-M6 (28) and then 9-desoxypaxilline (27). Subsequent feeding studies of β-PC-M6 (28) and α-PC-M6 (29) on the *paxP* mutant showed that PaxP readily converted β-PC-M6 (28) to 9-desoxypaxilline (27) as shown in Figure 1.14, ii. Notably, they saw no conversion of α-PC-M6 (29) to 9-desoxypaxilline (27) as shown in Figure 1.14, iv., demonstrating that the stereochemistry around the C16 hydroxyl group is a determining factor in the substrate specificity of PaxP. In their final PaxP feeding experiment, they fed α-paxitriol (30) and β-paxitriol (31) to the *paxP* mutant strain and found that β-paxitriol (31)-fed strains produced paxilline (2) whereas no conversion of α-paxitriol (30) was observed (Figure 1.14, iii. and iv.). Together these results showed that PaxP catalyses the demethylation of C14 and the complete oxidation of C16.

Saikia *et al*. (2007) completed similar feeding experiments on a *paxQ* mutant strain and found that PaxQ readily metabolises 9-desoxypaxilline (27) and α-PC-M6 (29) to form respectively paxilline (2) and α-paxitriol (30) as shown in Figure 1.14 v. and vi. Intriguingly, they did not see any PaxQ-mediated metabolism of β-PC-M6 (28) to form β-paxitriol (31) as shown in Figure 1.14 vii. Together these results indicate that PaxQ catalyses the hydroxylation at the C9 position and that the stereochemistry around the oxidised C16 must be in the α-conformation (in contrast to the β-conformation for PaxP) or have the carbonyl moiety in order for PaxQ to metabolise the substrate. This functional analysis of *paxP* and *paxQ* by Saikia *et al*. (2007) was the first functional analysis of indole diterpene “decoration” genes and established a foundation for further analysis of other decoration genes like the prenyl transferase, *paxD,* which flanks *paxQ* (Figure 1.11) and is potentially poorly expressed.[43]

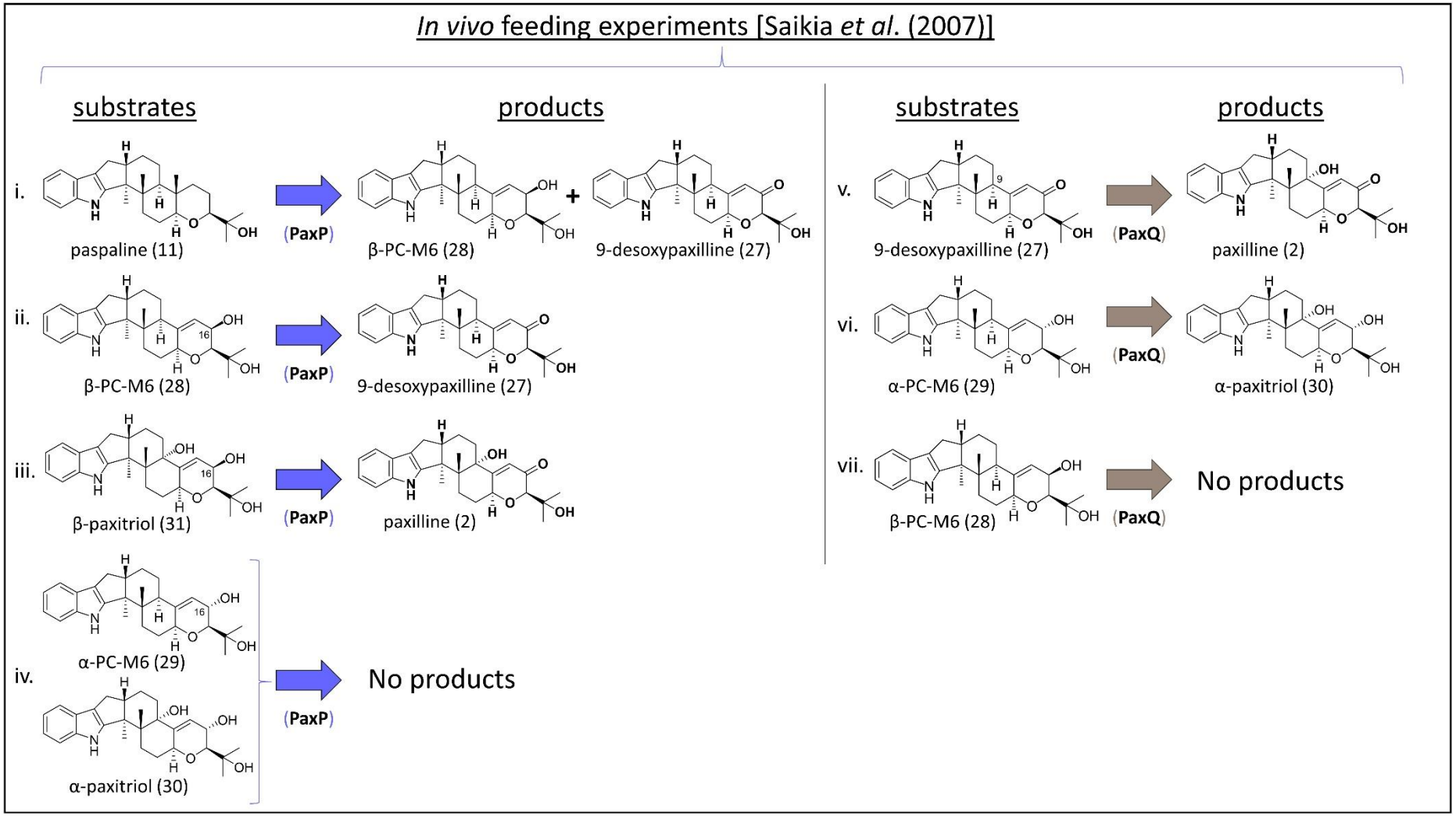


Figure 1.14. Summary of results from the feeding experiments by Saikia *et al*. (2007).[36]

To functionally analyse *paxD*, Liu *et al*. (2014) expressed *paxD* in *E. coli* and purified PaxD for *in vitro* analysis.[42] Recombinant PaxD was incubated with dimethylallyl pyrophosphate (8) and paxilline (2) and yielded two products, determined to be 21-prenylpaxilline (32) and 21,22-bisprenylpaxilline (33) as the minor and major products respectively (Figure 1.15, i.). Since these prenylated paxillines had not been observed in *P. paxilli* extracts, they explored the natural production of prenylated paxillines and found that they were present in production medium in quantities ~0.5% of paxilline. Therefore, they concluded that PaxD is naturally expressed at low levels or not partitioned correctly within the cell. Further analysis of PaxD by Liu *et al*. (2013) showed that it catalysed the monoprenylation of paspaline (11) at C21 to form 21-prenylpaspaline (34) and at C22 to form 22-prenylpaspaline (35) as shown in Figure 1.15, ii., indicating that indole diterpene D-enzymes can prenylate a variety of indole diterpene substrates and that the type of prenylation is substrate specific.[39] Intriguingly, Liu *et al*. (2013) also experimented with the substrate specificity of PaxC and found that it readily converted farnesyl pyrophosphate (10) and indole-3-glycerol phosphate (4) to farnesylindole (14) as shown in Figure 1.15, iii., demonstrating that many of these indole diterpene enzymes are rather promiscuous in their choice of substrate.

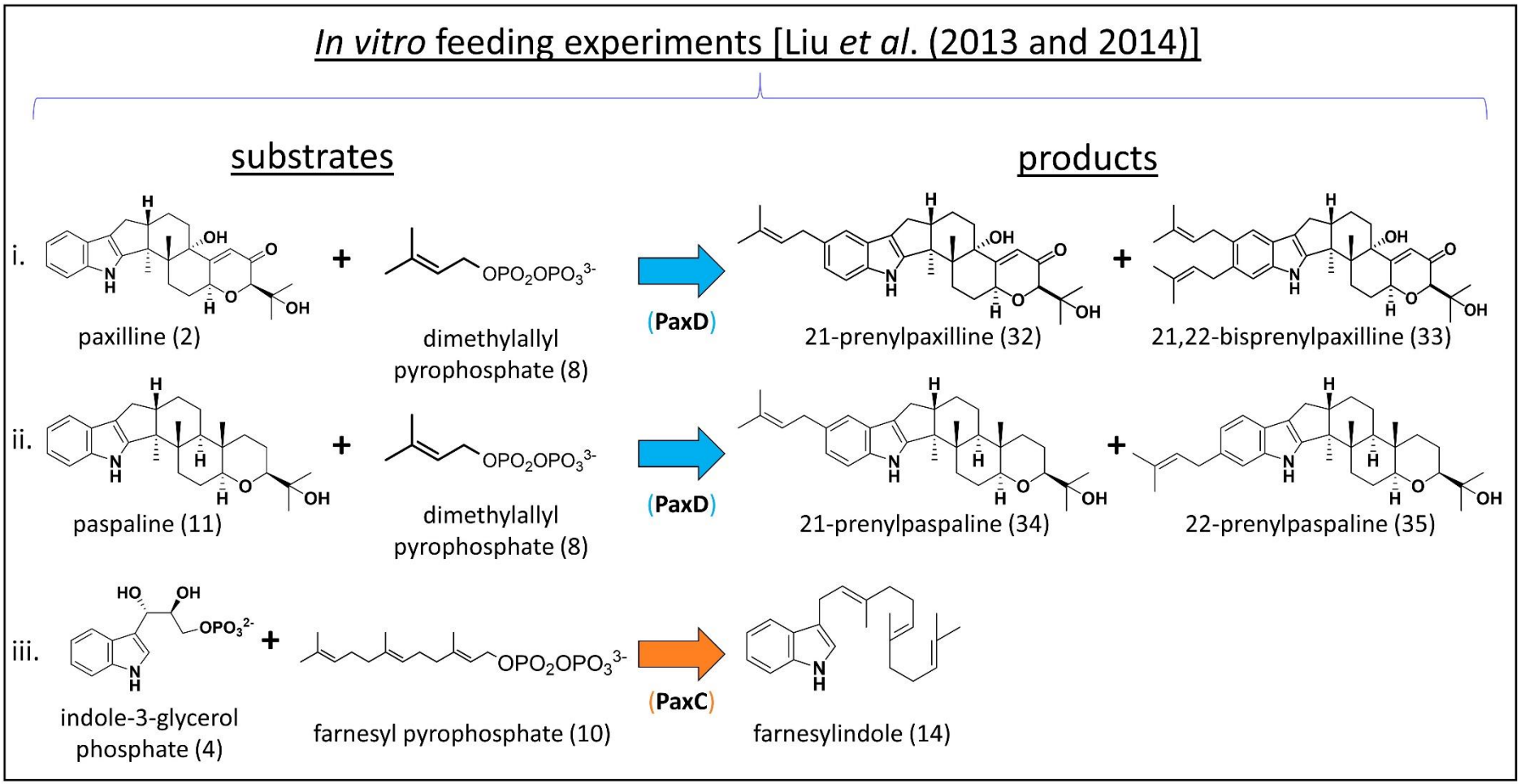
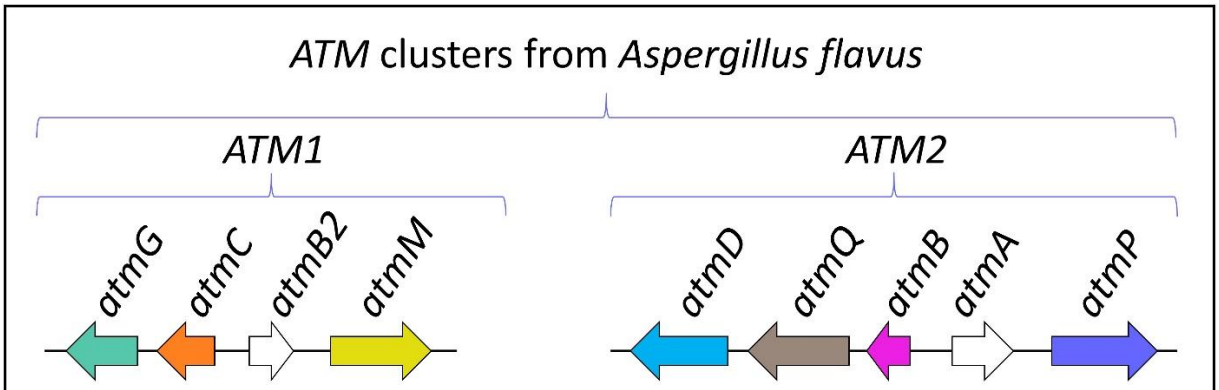


Figure 1.15. Summary of results from feeding experiments by Liu *et al*. (2013/2014).[39, 42]

Flanking *paxD* is a “decorative” FAD-dependent oxygenase-encoding gene, *paxO*. To date, *paxO* has not been functionally analysed but its close homologue *janO* has been shown to encode a protein product that catalyses the oxidative ring closure in janthitremane biosynthesis.[43] Additional evidence that PaxD and PaxO may have at one point induced janthitremane synthesis in *P. paxilli* is supported by the fact that they are upregulated during paxilline (2) biosynthesis.[38] It seems that alterations in gene expression/regulation or mutations of *paxD* and *paxO* over time may have somehow altered the expression or effectiveness of their protein products and thus the main indole diterpene produced by *P. paxilli* is paxilline.

Overall, the sequence identities of genes within the *PAX* cluster established a foundation for further sequence analysis guided identification of homologous indole diterpene genes in other indole diterpene producing fungi; thereby enabling more rapid identification of additional indole diterpene producing gene clusters and especially facilitating the localization of clusters of indole diterpene genes found in more than one specific locus as seen with the *ATM*, *LTM and PTM* clusters from *A. flavus*, *E. festucae* and *P. simplicissimum* respectively (Figure 1.2).

Identification of the *ATM* gene clusters and functional analysis of *atmP*, *atmQ* and *atmD*



Protein product	Proposed function	Secondary-metabolic biosynthetic step
AtmG	Geranylgeranyl pyrophosphate (3) synthase	1
AtmC	Geranylgeranyl transferase	2
AtmB2	Non-functional	-
AtmM	FAD-dependent oxygenase (epoxidase)	3
AtmD	Prenyl transferase	7
AtmQ	Cytochrome P450 oxygenase	6
AtmB	Cyclase	4
AtmA	Unknown	-
AtmP	Cytochrome P450 oxygenase	5

Figure 1.16. Depiction of *ATM* gene clusters from *Aspergillus flavus*.[25, 28]

Zhang *et al*. (2004) provided the first genetic evidence for the biosynthetic pathway of aflatrem (36) in *A. flavus* by locating the first cluster of *ATM* genes (*atmG*, *atmC*, and *atmM*; Figure 1.16) that were highly homologous to respective gene products in the *PAX* cluster (49% to PaxG, 55% to PaxC, and 52% to PaxM respectively).[25] Degenerate primers for conserved domains of geranylgeranyl pyrophosphate (3) synthases were used to identify *atmG* and subsequent chromosome walking revealed the closely clustered *atmC* and *atmM*. They showed that gene expression for these three *ATM* genes increased during the onset of aflatrem (36) biosynthesis (~108 hr; Figure 1.17, A) and that *atmM* was orthologous to *paxM* through complementation of *atmM* into the *P. paxilli* Δ*paxM* knockout strain (Figure 1.17, B).

With the availability of complete genome sequences, Nicholson *et al*. (2009) were able to tBLASTn the genomes of *A. flavus* and *A. oryzae* to locate the second *ATM* cluster containing five homologous indole diterpene genes (*atmD*, *atmQ*, *atmB*, *atmA*, and *atmP*; Figure 1.16) to those found in the *PAX* cluster of *P. paxilli* (29% to PaxD, 55% to PaxQ, 61% to PaxB, 29% to PaxA, and 61% to PaxP).[28] In contrast to the *ATM* genes in cluster 1, they found that these five decorative genes were expressed prior to the onset of aflatrem (36) biosynthesis at 108 h (48 h for *atmQ*, *atmB*, and *atmA* and 60 h for *atmD* and *atmP*; Figure 1.17, A). Complementation experiments of *atmP* and *atmQ* into respective Δ*paxP* and Δ*paxQ P. paxilli* mutant strains revealed that *atmP* is a functional orthologue of *paxP* [i.e. *atmP* readily demethylates C14 and oxidises C16 of paspaline (11)] and that *atmQ* is not a functional orthologue of *paxQ* (Figure 1.17, B). Instead, the *atmQ* transformant was shown to produce two new indole diterpene products, paspalicine (37) and paspalinine (38); thereby indicating that AtmQ is able to catalyse the same oxidation as PaxQ (i.e. the hydroxylation at C9) in addition to the acetal ring formation at C13.

A. Gene expression [Zhang *et al*. (2004), Nicholson *et al*. (2009)]

ATM	Gene Expression
G	upregulated at onset of aflatrem biosynthesis [Zhang *et al*. (2004)]
C	upregulated at onset of aflatrem biosynthesis [Zhang *et al*. (2004)]
M	upregulated at onset of aflatrem biosynthesis [Zhang *et al*. (2004)]
D	upregulated at 60 hr, prior to onset of aflatrem biosynthesis [Nicholson *et al*. (2009)]
Q	upregulated at 60 hr, prior to onset of aflatrem biosynthesis [Nicholson *et al*. (2009)]
B	upregulated at 48 hr, prior to onset of aflatrem biosynthesis [Nicholson *et al*. (2009)]
A	upregulated at 48 hr, prior to onset of aflatrem biosynthesis [Nicholson *et al*. (2009)]
P	upregulated at 48 hr, prior to onset of aflatrem biosynthesis [Nicholson *et al*. (2009)]

B. Complementation studies (Zhang et al. 2004, Nicholson et al. 2009)

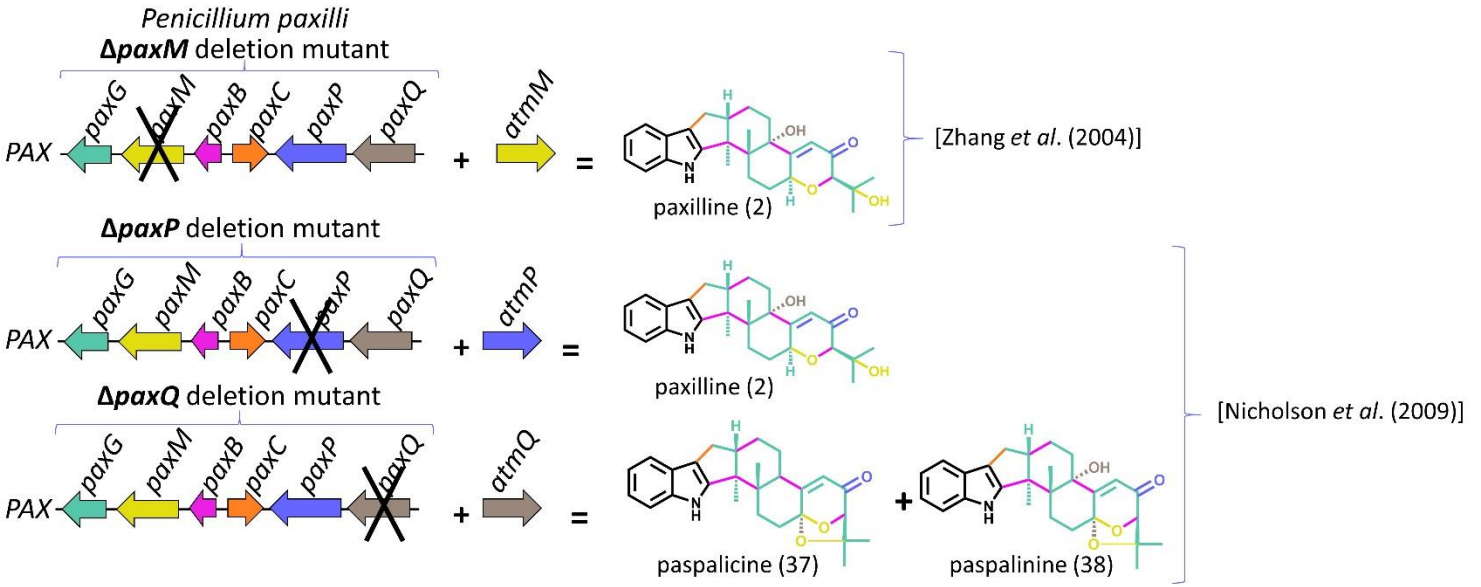


Figure 1.17. Summary of results from gene expression and complementation experiments done by Zhang *et al*. (2004) and Nicholson *et al*. (2009).[25, 28]

Subsequent *in vitro* analysis of recombinant AtmD by Liu *et al*. (2013) showed that AtmD was highly promiscuous.[39] AtmD was incubated with dimethylallyl pyrophosphate (8) and one of four different substrates: farnesylindole (14), 3-geranylgeranylindole (13), paspaline (11), or paxilline. Firstly, the incubation with paxilline (2) led to the formation of three monoprenylated products, two major and one minor (Figure 1.18). The two major products were determined to be reversely 20-prenylated paxilline (39) and reversely 21-prenylated paxilline (40) (Liu *et al*. (2013) used 'regularly' and 'reversely' to describe the prenylations they observed, which correspond respectively to a 'head' or 'tail' isoprenoid condensation). As for the minor product, the most likely candidate is reversely 22-prenylated paxilline (41) but not enough compound was produced to complete the structural identification. These results suggested that AtmD catalyses the same reverse prenylations of paspalinine (38) to form aflatrem (36) and β-aflatrem (42).

Liu *et al*. (2013) then investigated incubation of AtmD with farnesylindole (14), 3-geranylgeranylindole (13), and paspaline (11). Liquid chromatography-mass spectrometry (LCMS) analysis of the completed reactions revealed that all three substrates were monoprenylated and that the yield of the 3-geranylgeranylindole (13)-reaction was too low for clear product identification (Figure 1.18). They identified the products of the farnesylindole (14)-reaction to be regularly 5-prenylated farnesylindole (43) and regularly 6-prenylated farnesylindole (44). The products from the paspaline (11) reaction were determined to be regularly 21-prenylated paspaline (34) and regularly 22-prenylated paspaline (35). Finally, the nuclear magnetic resonance (NMR) spectroscopy results of the product mixture from the 3-geranylgeranylindole (13)-reaction revealed typical signals for regular dimethylallyl moiety indicating that 3-geranylgeranylindole (13) was monoprenylated in the same regular prenylation fashion as farnesylindole (14) and paspaline (11) to produce regularly 5-prenylated 3-geranylgeranylindole (45) and regularly 6-prenylated 3-geranylgeranylindole (46).

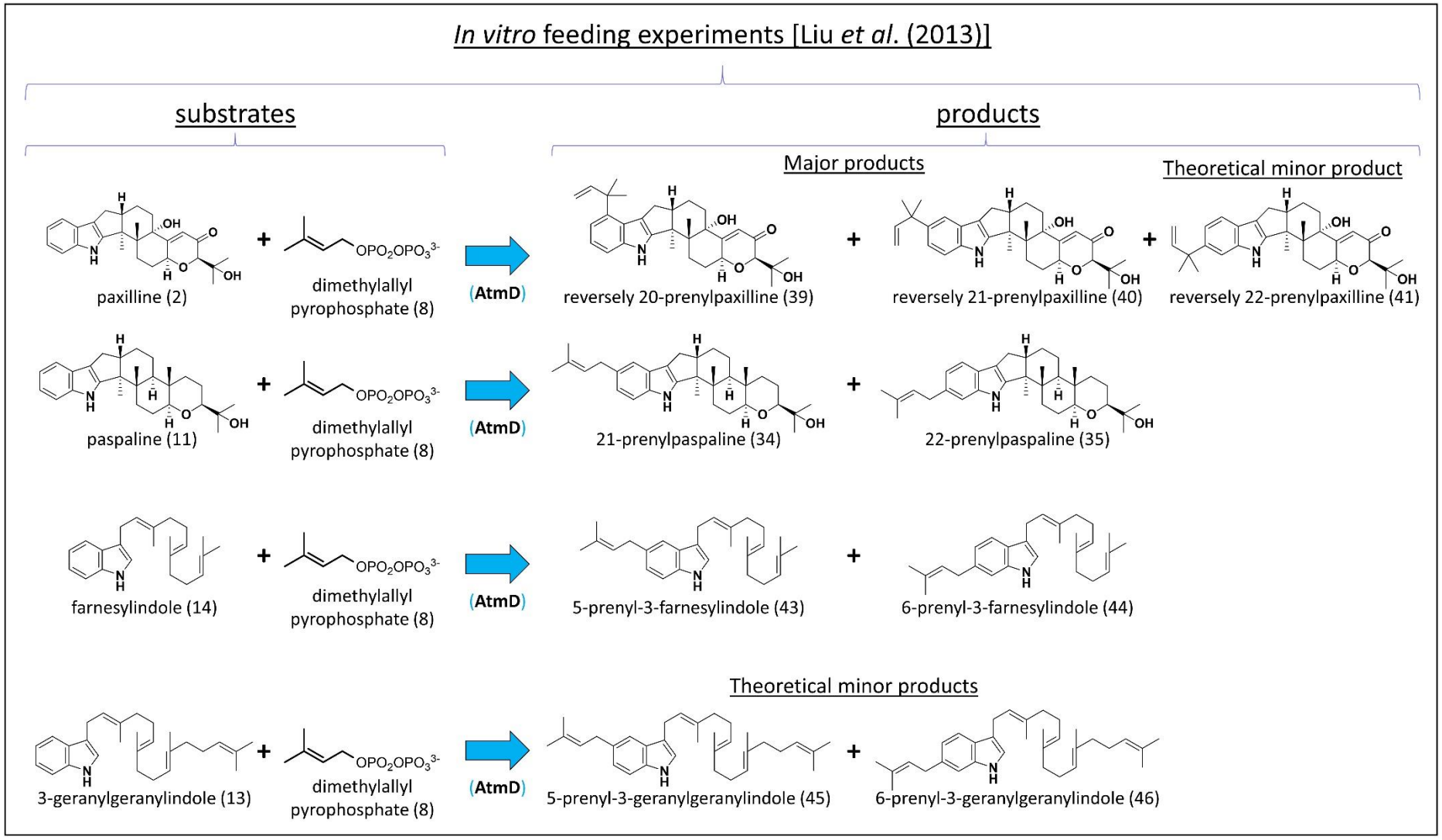


Figure 1.18. Summary of results from *in vitro* feeding experiments of AtmD by Liu *et al*. (2013).[39]

Tagami *et al*. (2014) concluded that seven genes (*atmGCMBPQD*) are involved in aflatrem (36) biosynthesis, and confirmed that AtmD reversely prenylated paspalinine (38) at the C20 and C21 positions to form aflatrem (36) and β-aflatrem (42) respectively, by rapidly reconstituting four *PAX* genes (*paxGCMB*) and three ATM genes (*atmPQD*) in a heterologous expression host, *A. oryzae* (Figure 1.19).[41] Amazingly, only two rounds of tandem transformations, involving four different plasmids (containing *paxGC*, *paxMB*, *atmPQ*, or *atmD*), were used to fully reconstruct the aflatrem (36) biosynthetic machinery in *A. oryzae*.

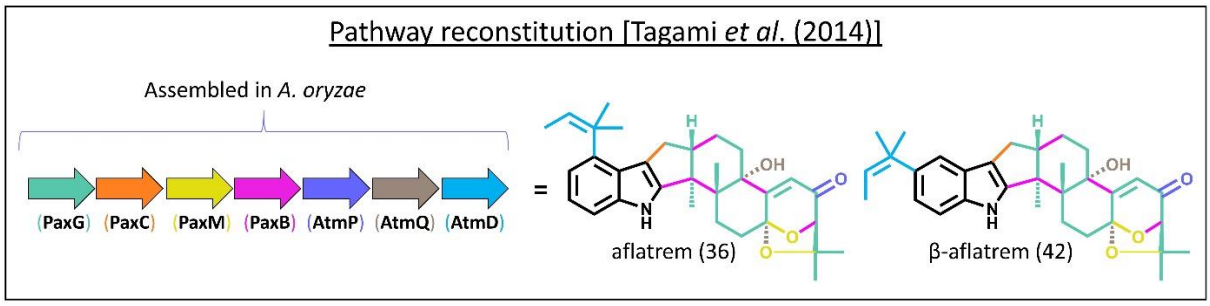
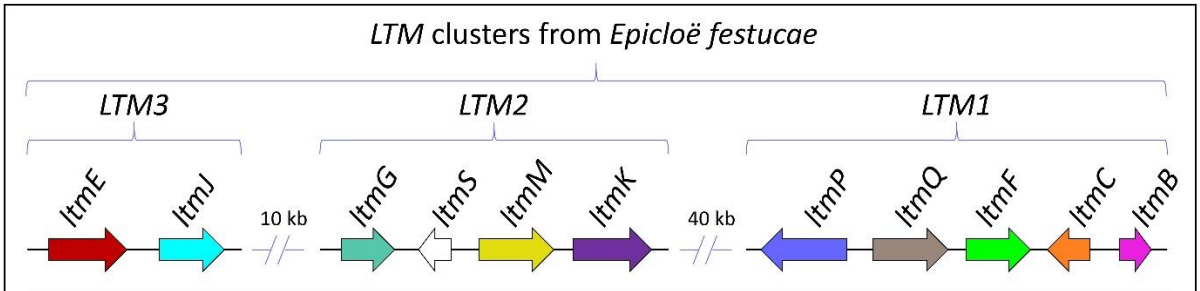


Figure 1.19. Summary of results from pathway reconstitution studies by Tagami *et al*. (2014).[41]

Identification of the *LTM* gene clusters and functional analysis of *LTM* genes

During the discovery of the *ATM* biosynthetic genes from *Aspergillus* spp., similar studies were being conducted on *Neotyphodium lolii* and *E. festucae* to discover the *LTM* gene clusters responsible for producing the potent rye-grass stagger-inducing lolitrems. The first genes that were proposed to be essential for lolitrem biosynthesis in *Neotyphodium lolii* and *E. festucae* were discovered by Young *et al*. (2005).[26] Just as with the discovery of the first *ATM* gene cluster, degenerate primers designed to conserved domains of fungal geranylgeranyl pyrophosphate (3) synthases were used to discover *ltmG* and subsequent chromosome walking was used to locate the closely clustered *ltmM* and *ltmK* (Figure 1.20).



Protein product	Proposed function	Secondary-metabolic biosynthetic step
LtmE	Prenyl transferase	10 or 8
LtmJ	Cytochrome P450 oxygenase	11 or 9
LtmG	Geranylgeranyl pyrophosphate (3) synthase	1
LtmS	Unknown	-
LtmM	FAD-dependent oxygenase (epoxidase)	3
LtmK	Cytochrome P450 oxygenase	9 or 11
LtmP	Cytochrome P450 oxygenase	6
LtmQ	Cytochrome P450 oxygenase	5 and 7
LtmF	Prenyl transferase	8 or 10
LtmC	Geranylgeranyl transferase	2
LtmB	Cyclase	4

Figure 1.20. Schematic of *LTM* clusters identified in *E. festucae*.[26-27, 29]

Young *et al*. (2005) showed that *ltmM* was a functional orthologue of *paxM* through complementation of *paxM* into the *E. festucae* Δ*ltmM* mutant strain (Figure 1.21). Intriguingly, this three gene *LTM* cluster was flanked by large blocks of retrotransposon sequence preventing them from using chromosome walking to identify any other potential *LTM* genes. Therefore, they changed tactics to suppression subtractive hybridisation, and generated subtracted Expressed Sequence Tag libraries. Young *et al*. (2006) compared these Expressed Sequence Tag libraries to GenBank databases, SwissProt databases, and *P. paxilli* biosynthesis genes using BLAST algorithms.[27] Results from these BLAST sequence comparisons revealed three new *LTM* genes: *ltmC*, *ltmP*, and *ltmJ*. Subsequent contig generation revealed that these three genes were part of two additional separate clusters, named *LTM* cluster 2 and *LTM* cluster 3. Five genes were found in *LTM* cluster 2 (*ltmP*, *ltmQ*, *ltmF*, *ltmC*, and *ltmB*) and two genes were found in *LTM* cluster 3 (*ltmE* and *ltmJ*) (Figure 1.20). Of these seven new *LTM* genes, Young *et al*. 2006 showed that *ltmC* is a functional orthologue of *paxC* through complementation of *ltmC* into the *P. paxilli paxC* mutant strain (Figure 1.21).

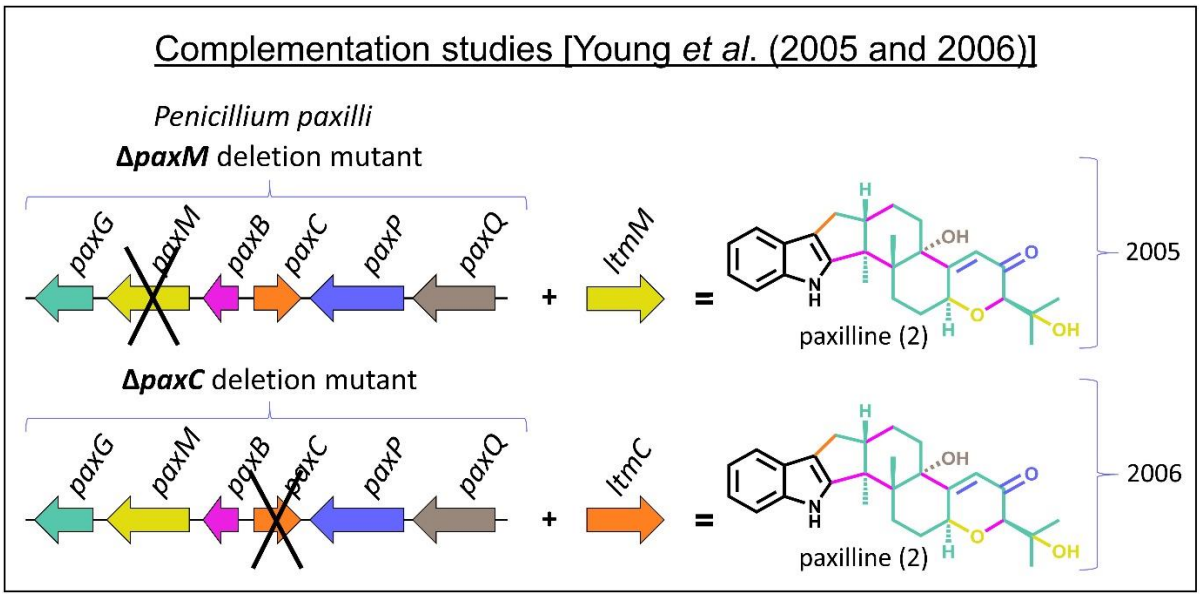


Figure 1.21. Summary of results from complementation experiments by Young *et al*. (2005 and 2006).[26-27]

For the next six years, the functionality of the rest of the *LTM* genes remained a mystery; then Saikia *et al*. (2012) published results from their studies on the exploration of the functionality of *E. festucae* genes in these three *LTM* clusters.[29] They used a mixture of gene knockout studies of *LTM* genes in *E. festucae*, complementation experiments of *LTM* genes into *P. paxilli* knockout strains, and feeding studies on *LTM* gene-containing *P. paxilli* mutants to propose a potential metabolic biosynthetic grid for lolitrem biosynthesis. Unfortunately, results from their studies did not resolve the functionality of the first two cytochrome P450 oxygenases, LtmP and LtmQ, because the novel compounds produced from feeding experiments were never structurally confirmed by NMR spectroscopy analysis. That said, they were still able to provide a great deal of valuable information about these two cytochrome P450 oxygenases.

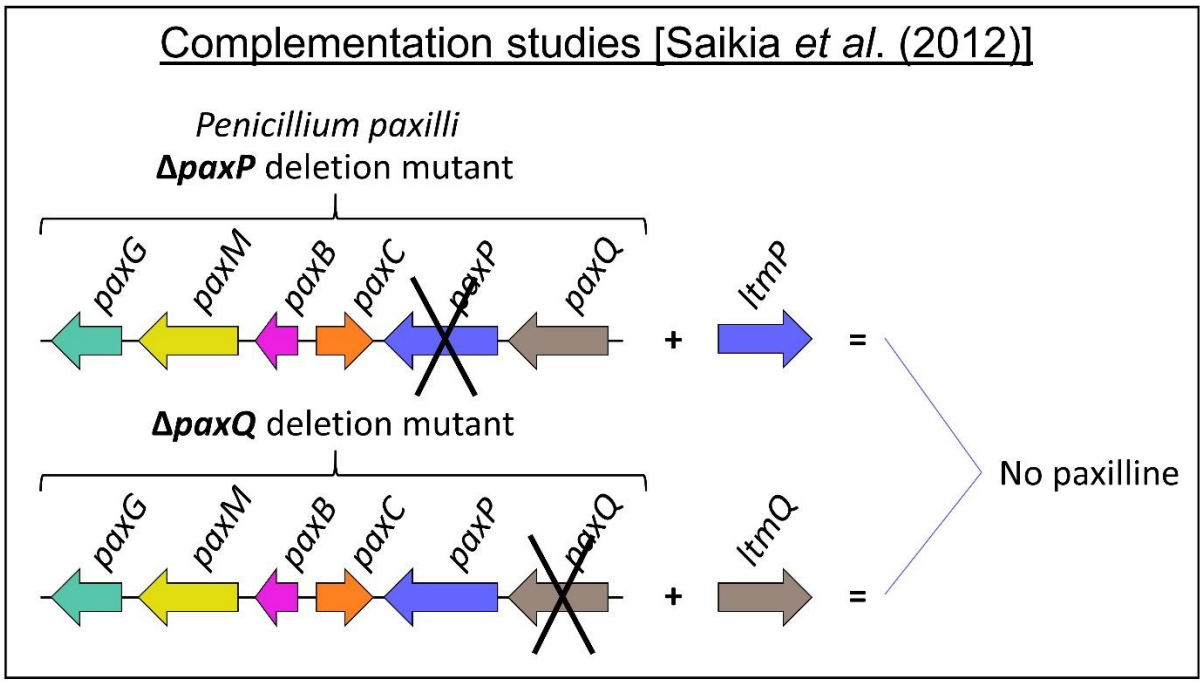


Figure 1.22. Summary of *LTM* gene complementation attempts by Saikia *et al*. (2012).[29]

To start, *ltmP* and *ltmQ* were unable to complement respective Δ*paxP* and Δ*paxQ* *P. paxilli* mutant strains (Figure 1.22) indicating that *ltmP* and *ltmQ* were not orthologous to *paxP* and *paxQ*. Therefore, three *P. paxilli* mutant strains containing only *ltmP*, *ltmQ*, or both *ltmP* and *ltmQ* were prepared for substrate feeding experiments to analyse the LtmP and LtmQ functionality *in vivo*. Paspaline (11) fed to the *ltmP* containing mutant was readily converted to PC-M6 [they identified this as β-PC-M6 (28), but based on the lack of NMR spectroscopy analysis and the α-stereochemistry of the C16 hydroxyl group of lolitrems, this product was most likely α-PC-M6 (29)] and two unknown indole diterpenes with $[M+H]^+$ 438 *m/z* that were paspaline (11) derivatives based on tandem mass spectrometry fragmentation patterns (Figure 1.23, i). Intriguingly, no 9-desoxypaxilline was produced from this experiment highlighting why *ltmP* was unable to complement the *P. paxilli* Δ*paxP* mutant strain. In the LtmQ analysis, they found that extract from the Δ*ltmQ* gene knockout *E. festucae* strain revealed wild-type levels of paspaline (11) production and a large accumulation of 9-desoxypaxilline (27) - see Figure 1.23, iv - supporting the potential involvement of 9-desoxypaxilline (27) as an intermediate in lolitrem biosynthesis; however, feeding of 9-desoxypaxilline (27) to the *ltmQ*-containing and *ltmP*/*ltmQ*-containing *P. paxilli* mutant strains showed only minor conversion to paxilline (2) and did not yield any other

novel indole diterpene products (Figure 1.23, ii) indicating that 9-desoxypaxilline (27) is unlikely a direct intermediate in lolitrem biosynthesis. Thus, they fed paspaline (11) to the *ltmPQ* double mutant and found conversion to seven indole diterpene products (Figure 1.23, iii); the same three metabolites observed when paspaline (11) was fed to the single *ltmP*-mutant (PC-M6, and the two unknown indole diterpenes), and four additional indole diterpene products: terpendole E (47), terpendole F (48), and two additional unknown compounds with $[M+H]^+$ 422 *m/z* and 452 *m/z*. Notably, none of these compounds, including terpendole E (47) and F (48), were structurally confirmed by NMR spectroscopy nor compared to standards and thus, further structural confirmation is required to make definitive conclusions.

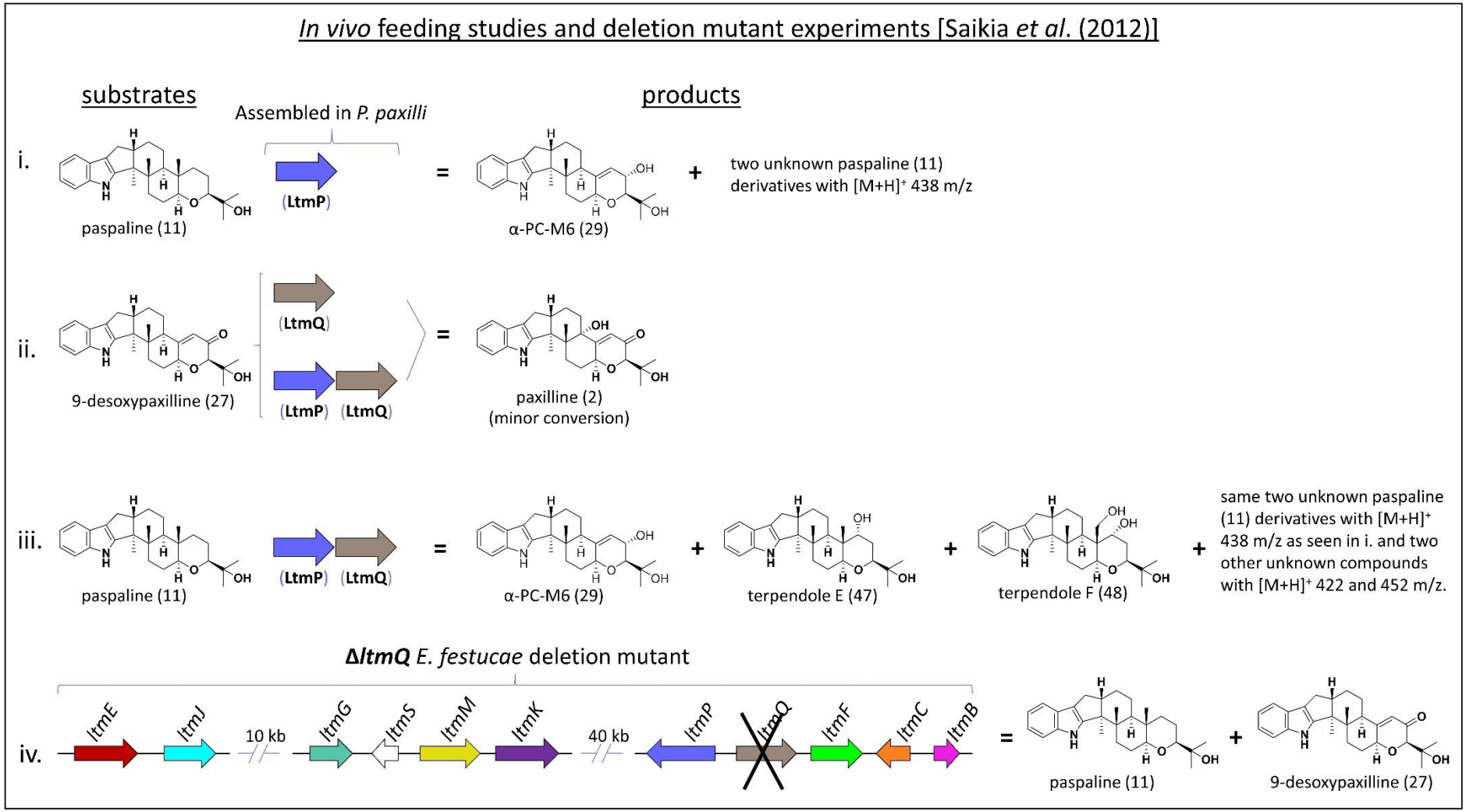


Figure 1.23. Summary of results from experiments used to elucidate the function of LtmP and LtmQ (Saikia *et al*. 2012).[29] Saikia *et al*. 2012 identified the PC-M6 as β-PC-M6 (28), but based on the lack of NMR spectroscopy analysis and the α-stereochemistry of the C16 hydroxyl group of lolitrems, this product was most likely α-PC-M6 (29). Therefore, α-PC-M6 (29) is shown in the figure.

Overall, Saikia *et al*. (2012) showed that LtmP was able to oxidise C16 (with unconfirmed stereochemistry) and LtmQ was able to oxidise C9 but more studies must be completed to clearly functionally characterise LtmP and LtmQ.[29] A good starting point would be to analyse the novel compounds produced in feeding studies by NMR spectroscopy and feed the LtmQ strain paspaline (11) to see if terpendole E (47) is produced.

In addition to LtmP and LtmQ, Saikia *et al*. (2012) also analysed two more cytochrome P450 oxygenases (LtmK and LtmJ) and two prenyl transferases (LtmF and LtmE) by studying the chemical compounds of the respective gene knockout strains. To start, the *ltmF* knockout mutant produced lolitriol (49) and lolitrem J (50) but no lolitrem B (51) – structure shown in Figure 1.2, lolitrem E (52), or *O*-prenylated paspalanes (i.e. terpendoles C (8) and J) (Figure 1.24, i). Thus they concluded that LtmF catalysed the prenylation of the C18-hydroxyl group. In contrast, the *ltmK* knockout strains lacked lolitrem B (51) and terpendole C (53) but contained lolitrem E (52) as shown in Figure 1.24, ii, showing that LtmK is responsible for oxidative acetyl ring formation. Finally, based on absence of lolitremanes in the *ltmE* and *ltmJ* knockout strains but the presence of simple and *O*-prenylated paspalanes like terpendole C (53) and terpendole M (54) shown in Figure 1.24, iii-iv, they proposed that LtmE catalyses the prenylations at position C20 and C21 of the lolitremane core and that LtmJ catalyses the oxidative cyclisation of the two prenyl groups to form the upper left portion of lolitremanes. With these results, they were able to propose a biosynthetic metabolic grid of lolitremane biosynthesis, but as noted above, there is still a great deal of work to be done to elucidate the precise biosynthetic pathway of the lolitremanes.

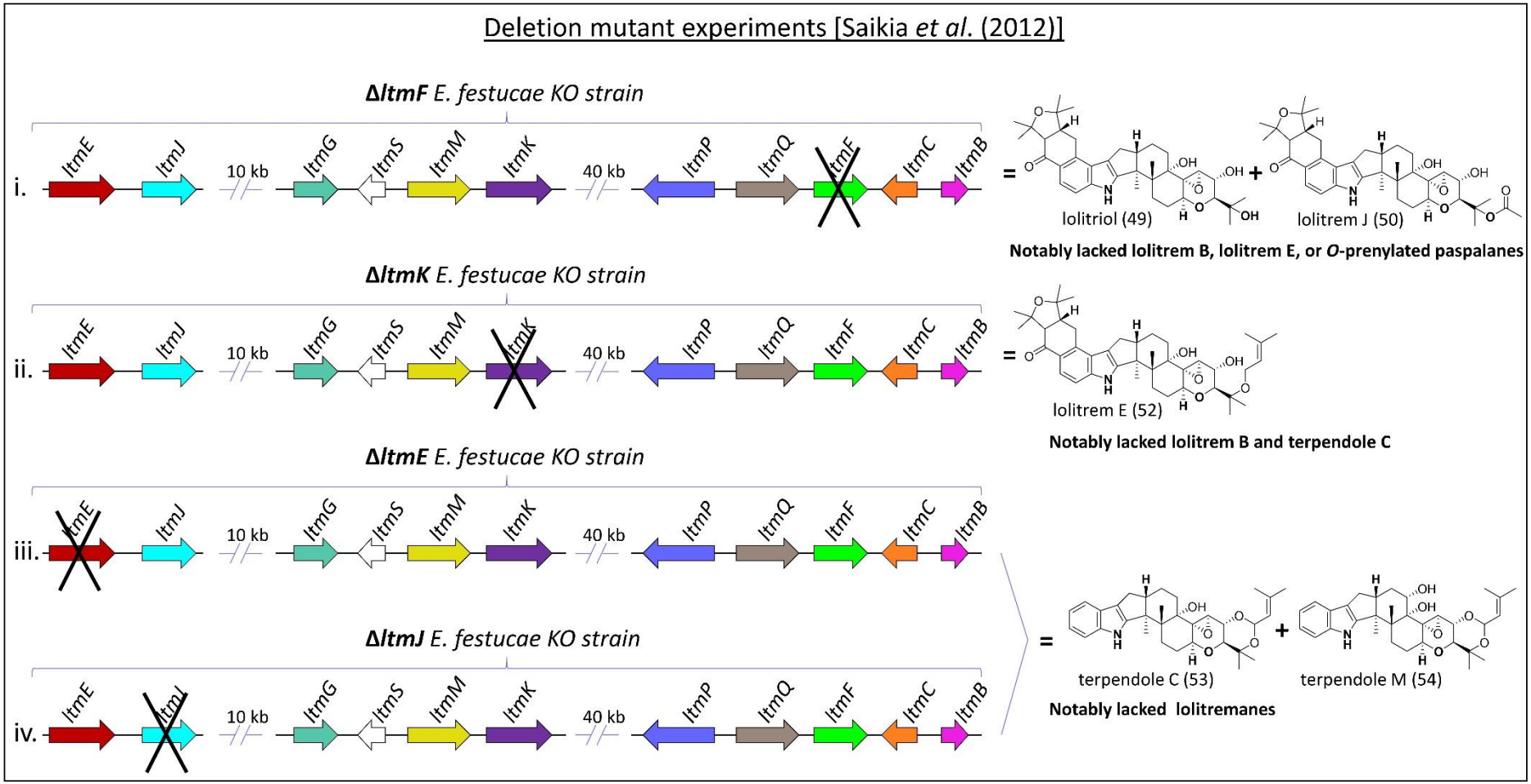
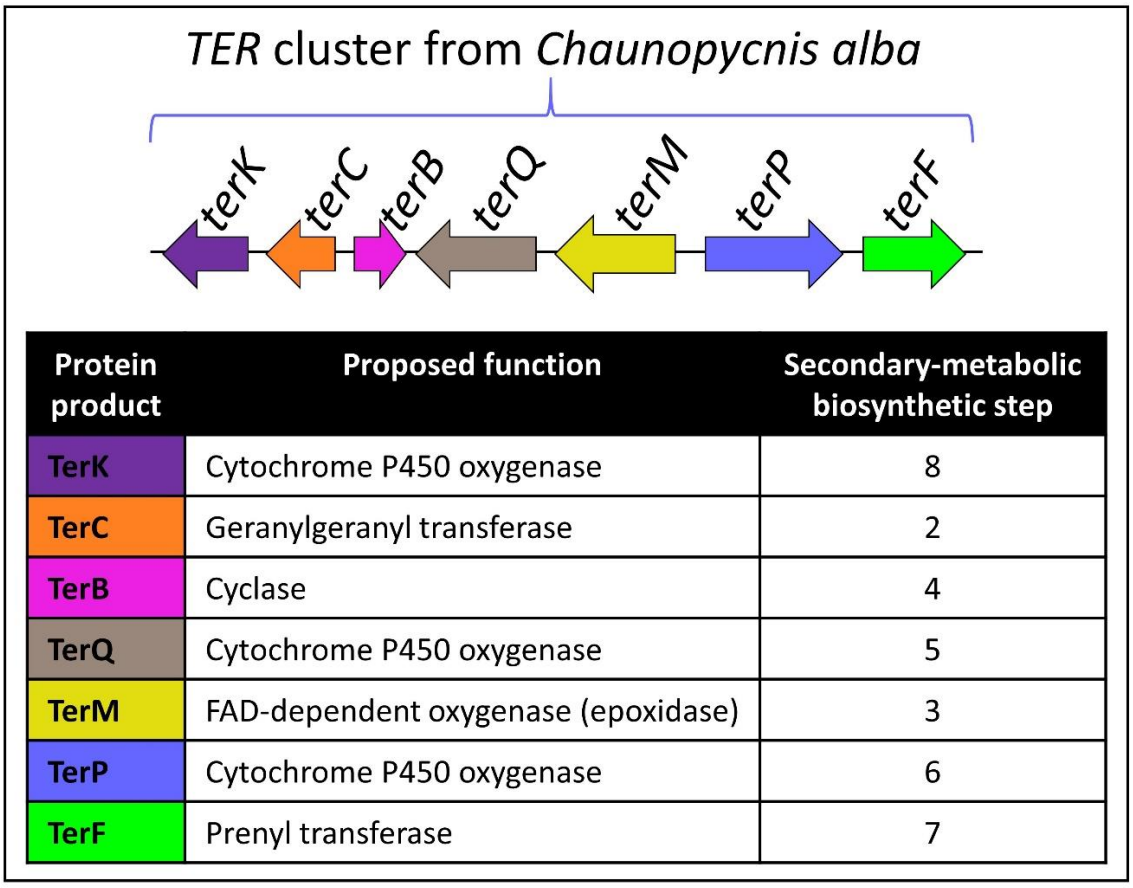


Figure 1.24. Summary of results from *E. festucae* gene knockout studies (Saikia *et al*. 2012).[29]

Identification of the *TER* gene cluster and functional analysis of *terP* and *terQ*

The *TER* gene cluster is responsible for terpendole production in *Chaunopycnis alba* and was identified in hope of generating a way to produce efficiently terpendole E (47), a known kinesin Eg5 inhibitor for use as an anti-cancer agent. Using a similar degenerate primer method that Zhang *et al*. (2004) and Young *et al*. (2005) used to identify the *ATM* and *LTM* clusters, Motoyama *et al*. (2012) designed degenerate primers to locate a *paxP* homologue within the *C. alba* genome.[30] Upon location of the *paxP* homologue, named *terP*, a 25 kb region of DNA around *terP* was cloned by genome walking and sequenced. This 25 kb region (Figure 1.25) contained seven genes (*terK*, *terC*, *terB*, *terQ*, *terM*, *terP*, and *terF*) that were highly similar to known indole diterpene genes and thus predicted to encode the gene products required for terpendole biosynthesis. Notably, the amino acid sequences of these seven gene products were most similar to those found in the lolitrem B (51) producer *N. lolii* (LtmK, 54%; LtmC, 55%; LtmB, 70%; LtmQ, 61%; LtmM, 49%; LtmP, 61%; and LtmF, 54%).



Protein product	Proposed function	Secondary-metabolic biosynthetic step
TerK	Cytochrome P450 oxygenase	8
TerC	Geranylgeranyl transferase	2
TerB	Cyclase	4
TerQ	Cytochrome P450 oxygenase	5
TerM	FAD-dependent oxygenase (epoxidase)	3
TerP	Cytochrome P450 oxygenase	6
TerF	Prenyl transferase	7

Figure 1.25. Depiction of *TER* gene cluster identified in *C. alba* by Motoyama *et al*. (2012).[30]

As the main goal of the study conducted by Motoyama *et al*. (2012) was to create more terpendole E (47), their experiments focused on confirming the function of two key terpendole genes, *terP* and *terQ*, which they predicted could be involved in terpendole E (47) biosynthesis. Generation of three *C. alba* gene knockout strains (*terP*-knockout, *terQ*-knockout, and *terP*/*terQ*-knockout) showed that the *terP* knockout strain accumulated terpendole E (47) in reasonably high quantities (36 mg/L after 2 days growth), the *terQ* knockout strain initially accumulated paspaline (11) that was readily converted to 9-desoxypaxilline (27), α-PC-M6 (29), and a putative 9-desoxypaxilline (27) analogue with + 16 mass of 9-desoxypaxilline (27), and that the *terQ*/*terP*-double knockout strain only accumulated paspaline (11) (Figure 1.26).

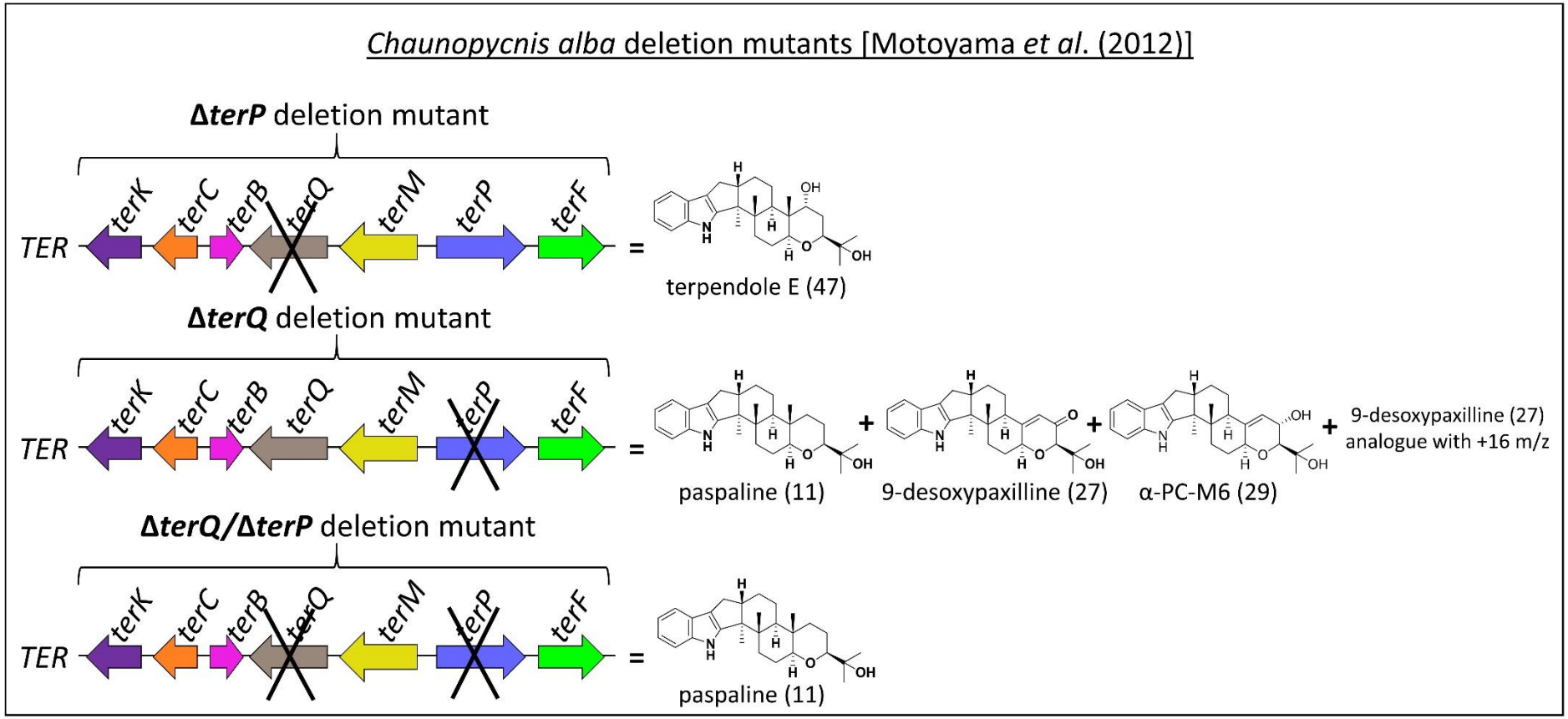


Figure 1.26. Summary of results from gene knockout studies completed by Motoyama *et al*. (2012).[30]

Subsequent feeding experiments on *C. alba* strains containing only *terP* or only *terQ* revealed the specific oxidative properties of TerP and TerQ. The *terP*-only strain converted terpendole E (47) to 9-desoxyterpendole I (55), paspaline (11) to 9-desoxypaxlline, α-PC-M6 (29), and the putative 9-desoxypaxilline (27) analogue, and α-PC-M6 (29) to 9-desoxypaxilline (27) as shown in Figure 1.27; and the *terQ*-only strain metabolised α-PC-M6 (29) to some unidentified compounds, 9-desoxypaxilline (27) to paxilline (2) and an unidentified paxilline (2) derivative with +14 mass of 9-desoxy paxilline, paspaline (11) to terpendole E (47), and 9-desoxyterpendole I (55) to terpendole I (56) as shown in Figure 1.27. Together these results indicate that TerP catalyses the C14 demethylation and C16 oxidation of terpendole E (47) and paspaline (11) and that TerQ catalyses the C9 hydroxylation of 9-desoxypaxilline (27) and 9-desoxyterpendole I (55) and the C15 hydroxylation of paspaline (11). Notably, the TerP catalysed C14 demethylation of terpendole E (47) triggers the C14-C15 epoxide formation to form 9-desoxyterpendole I (55). Additionally, since α-PC-M6 (29) and 9-desoxypaxilline (27) were not converted to any terpendoles by the *terP* or *terQ*-only strains, this suggests that 9-desoxypaxilline (27) is a shunt metabolite in terpendole biosynthesis. As a final note, they demonstrated that TerQ and PaxQ have different substrate specificities by feeding paspaline (11) and 9-deoxypaxilline to a *paxQ*-only strain. As expected, the *paxQ*-only strain readily converted 9-desoxypaxilline (27) to paxilline (2) but did not convert paspaline (11) to terpendole E (47).

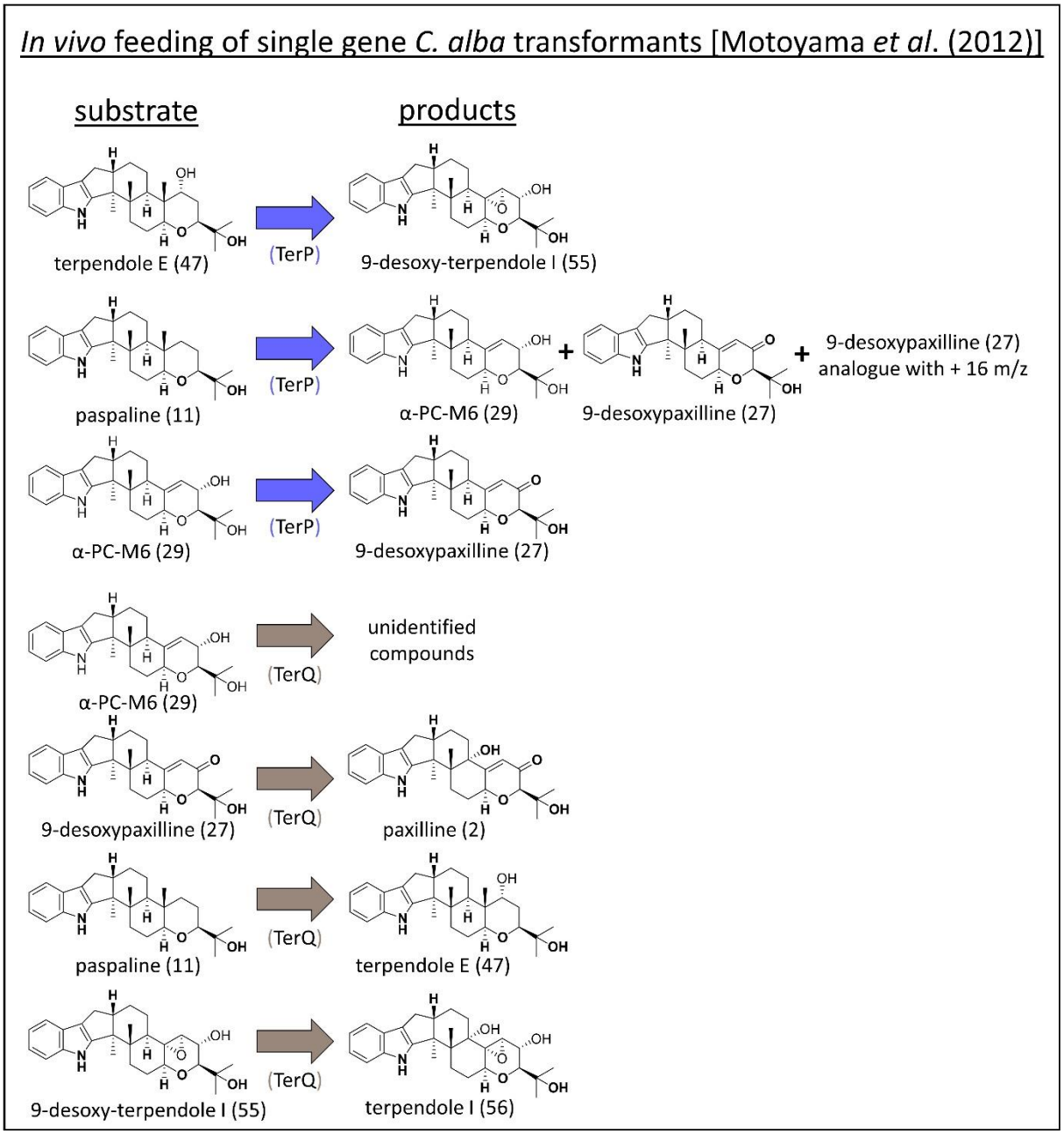


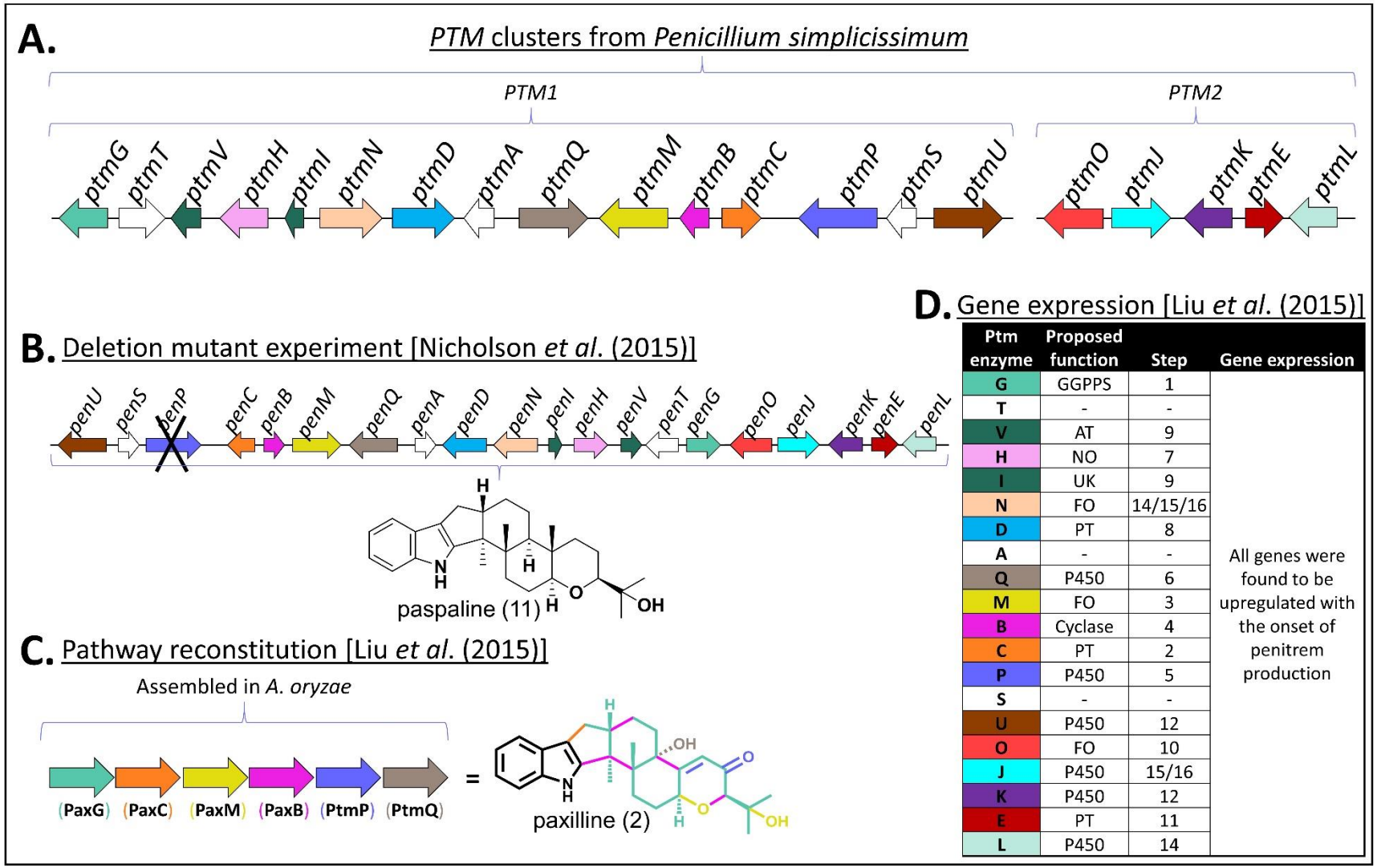
Figure 1.273. Summary of results from feeding experiments completed by Motoyama *et al*. (2012).[30]

To date, the *terP* and *terQ* genes are the only *TER* genes that have been functionally analysed, but based on sequence similarity to *LTM* gene products, Motoyama *et al*. 2012 proposed that the encoded products of *terCMB* deliver paspaline (11) and *terFK* convert terpendole I (56) to terpendole C (53).

Identification of the *PTM and PEN* gene clusters and functional analysis of the genes

The *PTM* and *PEN* gene clusters encode the genes responsible for allowing penitrem production in *Penicillium simplicissimum* and *P. crustosum* respectively. Liu *et al*. (2015) were the first to publish their results on the identification of the *PTM* cluster from *P. simplicissimum* (Figure 1.28, A) and functional characterisation of 13 of the 17 *PTM* genes through heterologous reconstitution in *A. oryzae*.[31] Shortly after, Nicholson *et al*. 2015 published their findings of the *PEN* cluster from *P. crustosum* and confirmation that *penP* was essential for penitrem production through generation of a Δ*penP P. crustosum* knockout strain (Figure 1.28, B).[32] Below, the methods each group used to identify the *PTM* and *PEN* clusters and how Liu *et al*. (2015) functionally confirmed the genes responsible for penitrem biosynthesis will be explored.

To start, Liu *et al*. (2015) employed gene cluster searching methods (i.e. Blast searching and 2ndFind program) to identify the first penitrem gene cluster spanning 44 kb and containing 15 putative genes (*ptmGTVHINDAQMBCPSU*) in *Penicillium simplicissimum*. Subsequent RNA sequencing analysis on another penitrem producer, *P. crustosum*, revealed a second 11 kb cluster containing 5 additional genes (*ptmOJKEL*) that were expressed with the onset of penitrem production (Figure 1.28, A and D). In contrast, Nicholson *et al*. (2015) used degenerate primers designed to conserved regions of indole diterpene biosynthetic genes homologous to *paxC* and *paxP* to create a fragment of the putative *PEN* gene cluster that was used to generate radioactively labelled hybridisation probes for cosmid library screening. Results from the cosmid screening were used to generate a contiguous sequence of 66 kb that contained 23 putative genes (PC-01 to PC-23, which included *penPCBMQADGO*) that were potentially involved in penitrem biosynthesis.



Ptm enzyme	Proposed function	Step	Gene expression
G	GGPPS	1	All genes were found to be upregulated with the onset of penitrem production
T	-	-	
V	AT	9	
H	NO	7	
I	UK	9	
N	FO	14/15/16	
D	PT	8	
A	-	-	
Q	P450	6	
M	FO	3	
B	Cyclase	4	
C	PT	2	
P	P450	5	
S	-	-	
U	P450	12	
O	FO	10	
J	P450	15/16	
K	P450	12	
E	PT	11	
L	P450	14	

Figure 1.28. Summary of initial results from Liu *et al*. (2015) (**A**, **C**, **D**)[31] and Nicholson *et al*. (2015) (**B**).[32] Abbreviations: GGPPS (geranylgeranyl pyrophosphate synthase), AT (acetyl transferase), NO [NAD(P)$^+$ dependent oxygenase], UK (unknown), FO (FAD/FMN dependent oxygenase), PT (prenyl transferase) and P450 (cytochrome P450 oxygenase).

After identifying the two *PTM* gene clusters, Liu *et al*. (2015) first examined the functionality of *ptmP* and *ptmQ*. They introduced *ptmP* and *ptmQ* into an *A. oryzae* paspaline (11)-producing strain containing *paxGCMB* and found that *ptmPQ* initiated production of paxilline, but not β-paxitriol (31) as shown in Figure 1.28, C. From there, they introduced *ptmH* and *ptmD* into an *A. oryzae* strain without any other indole diterpene genes. Paxilline (2) fed to the *ptmHD* strain was metabolised to β-paxitriol (31) and 20-prenyl-β-paxitriol (57) indicating that *ptmH* reduced the C16 ketone of paxilline (2) and *ptmD* prenylated C20 of B-paxitriol (Figure 1.29, i). Ambiguities over the genes required for the next step in penitrem biosynthesis led them to design four additional strains containing *ptmHDV*, *ptmHDIE*, *ptmHDVI*, *ptmHDVIE*, or *ptmHDVIOE* (Figure1.29, ii-vi). Whereas incubation of paxilline (2) with the *ptmHDV*, and *ptmHDIE* strains did not yield any new products, the *ptmHDVI* and *ptmHDVIE* strain showed metabolism of paxilline (2) to 20-prenylpenijanthine A (58), the dehydration product of 20-prenyl-β-paxitriol (57), showing that *ptmV* and *ptmI* are likely both essential for the dehydration. Incubation of paxilline (2) with the *ptmHDVIOE* strain showed that this strain readily converted paxilline (2) to PC-M4 (59) indicating that PtmE was capable of catalysing the construction of the highly unusual bicyclo[4.2.0]heptane skeleton of PC-M4 (59).

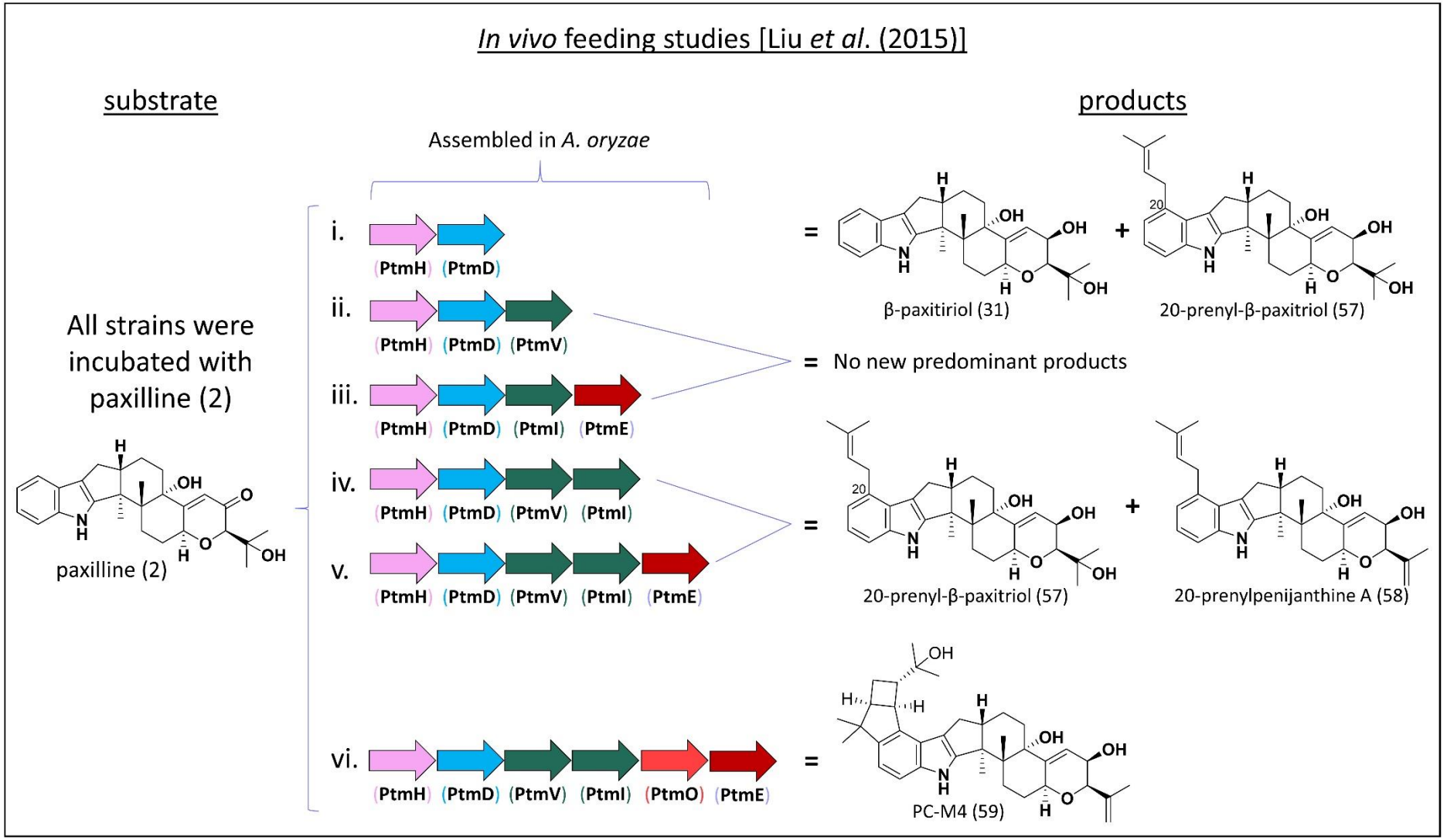


Figure 1.294. Summary of Liu *et al*. (2015) feeding studies used to delineate the biosynthetic steps from paxilline (2) to PC-M4 (59).[31]

Liu *et al*. (2015) then analysed whether PC-M5 (60) and PC-M4 (59) were shunt metabolites (i.e. metabolites not directly involved in the production of penitrems) by conducting feeding experiments on two additional strains that contained *ptmKULNJ* or *ptmKULNO*.[31] PC-M4 (59) was metabolised by the *ptmKULNJ*-strain to penitrem A (12) (Figure 1.30, vi) and the *ptmKULNO*-strain to penitrem F (61) as shown in Figure 1.30, iii, indicating that five genes (*ptmKULNJ*) are involved in the transformation of PC-M4 (59) to penitrem A (12) and that *ptmJ* is essential for the final benzylic hydroxylation. To elucidate the biosynthetic pathway from PC-M4 (59) to penitrem F (61), four more strains were assembled containing only *ptmKUN*, *ptmK*, *ptmN*, or *ptmJ*. Incubation of PC-M4 (59) with the *ptmKUN* strain resulted in penitrem C (62) as shown in Figure 1.30, ii, production showing that *ptmL* was essential for the C14-C15 epoxide formation. Incubation of PC-M4 (59) with the *ptmK*-strain showed conversion to secopenitrem D (63) as shown in Figure 1.30, i, indicating that PtmK catalyses the ring expansion and olefin formation of PC-M4 (59) to form secopenitrem D (63). Subsequent feeding of penitrem B (64) to the *ptmN*-strain showed conversion to penitrem F (61) as shown in Figure 1.30, v, indicating that PtmN catalyses the C22-chlorination in penitrem biosynthesis and that *ptmU* is essential for the 8-membered ether ring formation of penitrem D from secopenitrem D (63). Finally, penitrem F (61) fed to the *ptmJ*-strain was converted to penitrem A (12) (Figure 1.30, vi), confirming that *ptmJ* is essential for the final benzylic hydroxylation.

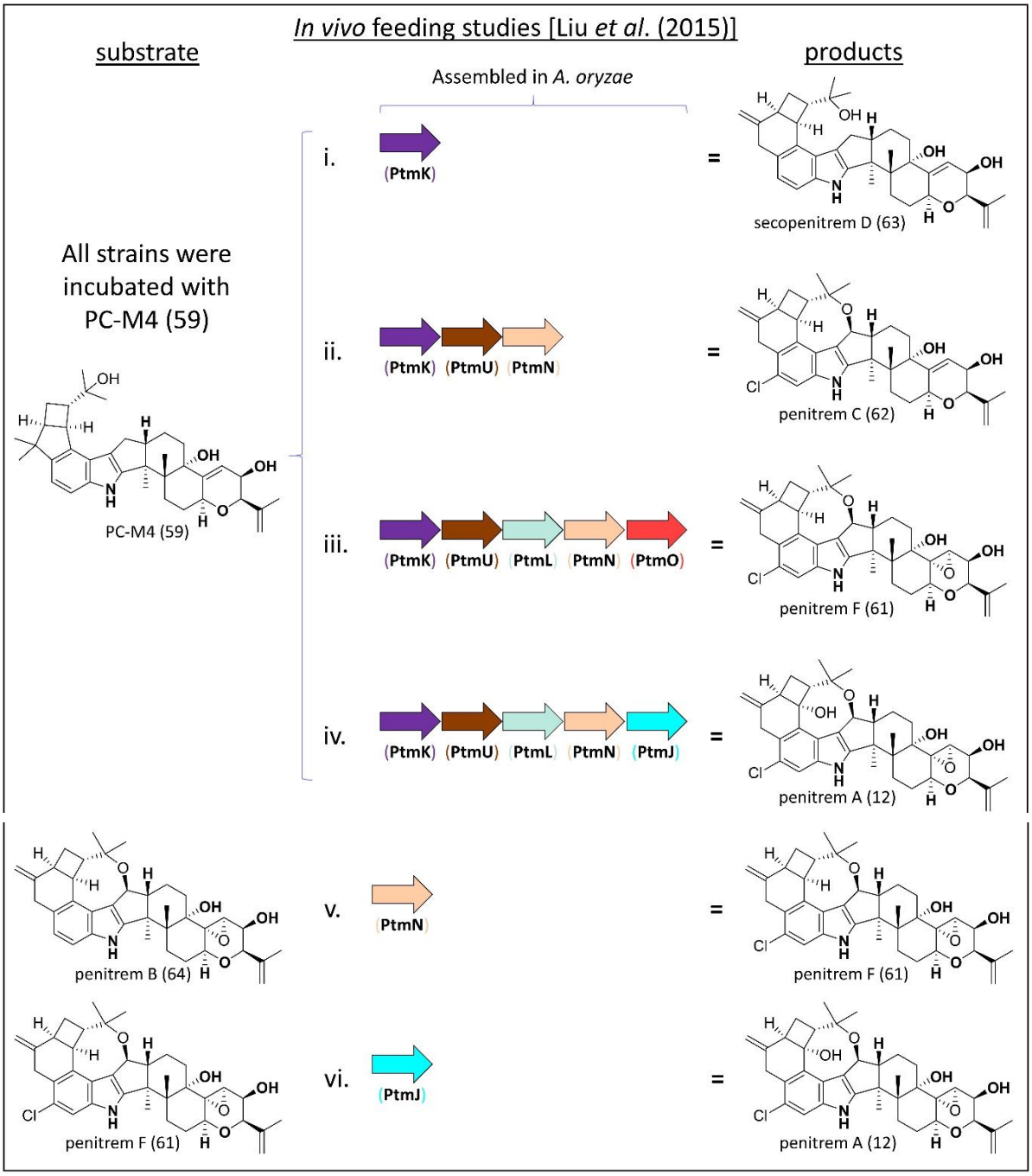


Figure 1.30. Summary of Liu *et al*. (2015) feeding studies used to delineate the biosynthetic steps from PC-M4 (59) to penitrem A (12).[31]

With the final five cytochrome P450 oxidation steps confirmed, Liu *et al*. (2015) turned their attention back to the intricacies of the biosynthetic steps involved in converting 20-prenylpenijanthine A (58) to PC-M4 (59). *In vitro* analysis of recombinant PtmO incubated with 20-prenylpenijanthine A (58) yielded PC-M5 (60) as shown in Figure 1.31, demonstrating that PtmO terminally hydroxylated the C20 prenyl group. To confirm the unusual action of PtmE, they analysed PtmE *in vitro*. Incubation of PtmE

with PC-M5 (60), Mg^{2+}, and dimethylallyl pyrophosphate (8) yielded PC-M4 (59) as shown in Figure 1.31, confirming that just a single enzyme was capable of catalysing the prenylation construction of the highly unusual bicyclo[4.2.0]heptane skeleton of PC-M4 (59).

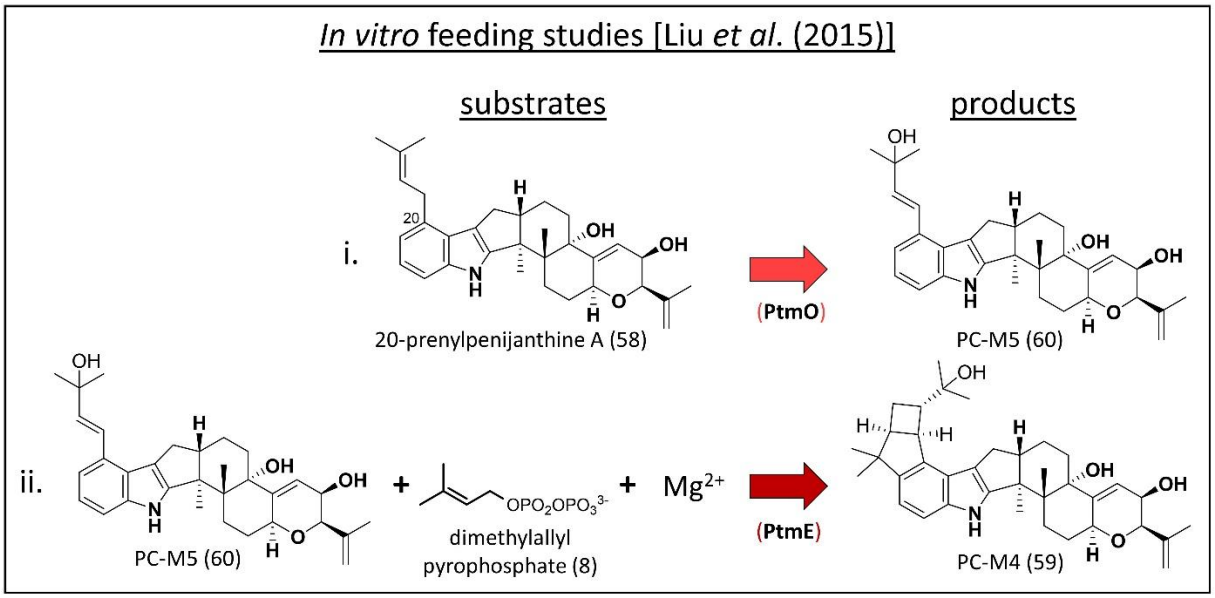


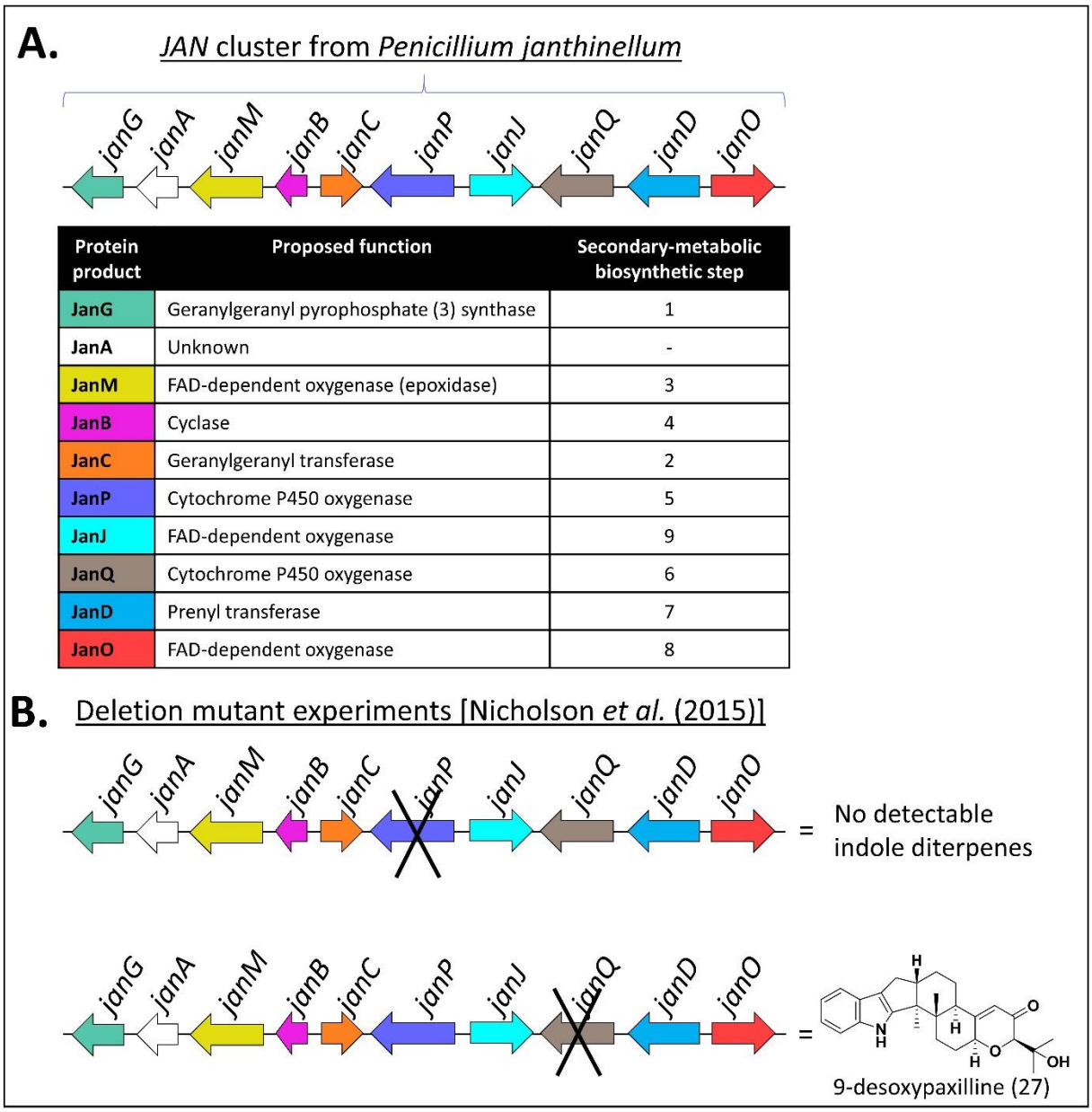
Figure 1.31. Summary of Liu *et al*. (2015) recombinant enzyme analyses used to functionally examine PtmO and PtmE; the biosynthetic steps from 20-prenylpenijanthine A (58) to PC-M4 (59).[31]

Overall, Liu *et al*. (2015) clearly confirmed the function of each of the thirteen genes involved in the biosynthetic pathway required to convert paspaline (11) to penitrem A (12).[31] This work is easily the most comprehensive and impressive analysis of the genes involved in the biosynthesis of an indole diterpene product, as they first-off identified the novel *PTM* gene clusters and then functionally confirmed the identity of every unique gene involved in penitrem biosynthesis.

Identification of the *JAN* gene cluster and functional analysis of the genes

In the same report where Nicholson *et al*. (2015) announced the identification of the *PEN* cluster in *P. crustosum*, they also announced the identification of the *JAN* cluster for janthitremane (i.e. shearinine) biosynthesis in *P. janthinellum*.[32] As with the *PEN* cluster, they used degenerate primers designed to conserved regions of indole

diterpene biosynthetic genes homologous to *paxC* and *paxP* to create a fragment of the putative *JAN* gene cluster that was used to generate radioactively labelled hybridisation probes for cosmid library screening. Results from the cosmid screening were used to generate a contiguous sequence of 68 kb that contained 10 putative genes (*janGAMBCPJQDO*) that were potentially involved in janthitremane biosynthesis (Figure 1.32, A). Notably, the nine *JAN* genes share high amino acid identity compared with their respective *PAX* homologues: JanG (42%), JanC (71%), JanM (67%), JanA (50%), JanB (77%), JanP (71%), JanQ (65%), JanD (66%), JanO (72%). Interestingly, the orientations of all of the genes and the positions of introns are the same in the *PAX* and *JAN* clusters except for the presence of *janJ*. Therefore, Liu *et al*. (2016) speculated that the *PAX* cluster could produce janthitremanes with a paxilline (2) core, but this has not been observed because *paxD* and *paxO* are naturally expressed at low levels in *P. paxilli*.[43]



Protein product	Proposed function	Secondary-metabolic biosynthetic step
JanG	Geranylgeranyl pyrophosphate (3) synthase	1
JanA	Unknown	-
JanM	FAD-dependent oxygenase (epoxidase)	3
JanB	Cyclase	4
JanC	Geranylgeranyl transferase	2
JanP	Cytochrome P450 oxygenase	5
JanJ	FAD-dependent oxygenase	9
JanQ	Cytochrome P450 oxygenase	6
JanD	Prenyl transferase	7
JanO	FAD-dependent oxygenase	8

Figure 1.32. Summary of results from Nicholson *et al*. (2015). The *JAN* cluster is depicted in **A** and the knockout strains with their corresponding indole diterpene porducts are shown in **B**.[32]

To confirm the involvement of these *janP* and *janD* in janthitremane biosynthesis, Nicholson *et al*. (2015) prepared two *P. janthinellum* knockout mutants where *janP* or *janD* was removed.[32] They noted the absence of all indole diterpene compounds in the extract of the Δ*janP*-mutant and increased accumulation of 9-desoxypaxilliine with lack of janthitremane production in the Δ*janD*-mutant (Figure 1.32, B), thereby confirming that both genes were involved in janthitremane biosynthesis. Based on the genes in the *JAN* cluster, Nicholson *et al*. (2015) proposed the most likely biosynthetic pathway

for shearninine A biosynthesis, but full elucidation of individual gene functionality remained unclear until Liu *et al*. (2016) published their findings from their analysis of *janQDOJ* using their *A. oryzae* heterologous expression system.[43]

Liu *et al*. (2016) prepared three *A. oryzae* transformants that contained *janQ*, *janQDO*, or *janQDOJ* and conducted feeding experiments on each strain using 9-desoxypaxilline (27).[43] They found that the *janQ*-transformant readily metabolised 9-desoxypaxilline (27) to paspalinine (38) and that no paxilline (2) was produced indicating that janQ catalysed the C9 hydroxylation and the acetal ring formation of the terminal hydroxide of paspaline (11) to form paspalinine (38) as shown in Figure 1.33, i. Subsequent incubation of 9-desoxypaxilline (27) with the *janQDO*-transformant yielded shearinine K (65) and shearinine A (66) showing that JanD and JanO were responsible for the prenylation of paspalinine (38) to form shearinine K (65) and subsequent oxidative ring closure of shearinine K (65) to form shearinine A (66) as shown in Figure 1.33, ii. Finally, incubation of 9-desoxypaxilline (27) with the *janQDOJ*-transformant yielded sheadinine D showing that JanJ catalysed the hydroxylation of shearinine A (66) to form shearinine D (67) as shown in Figure 1.33, iii.

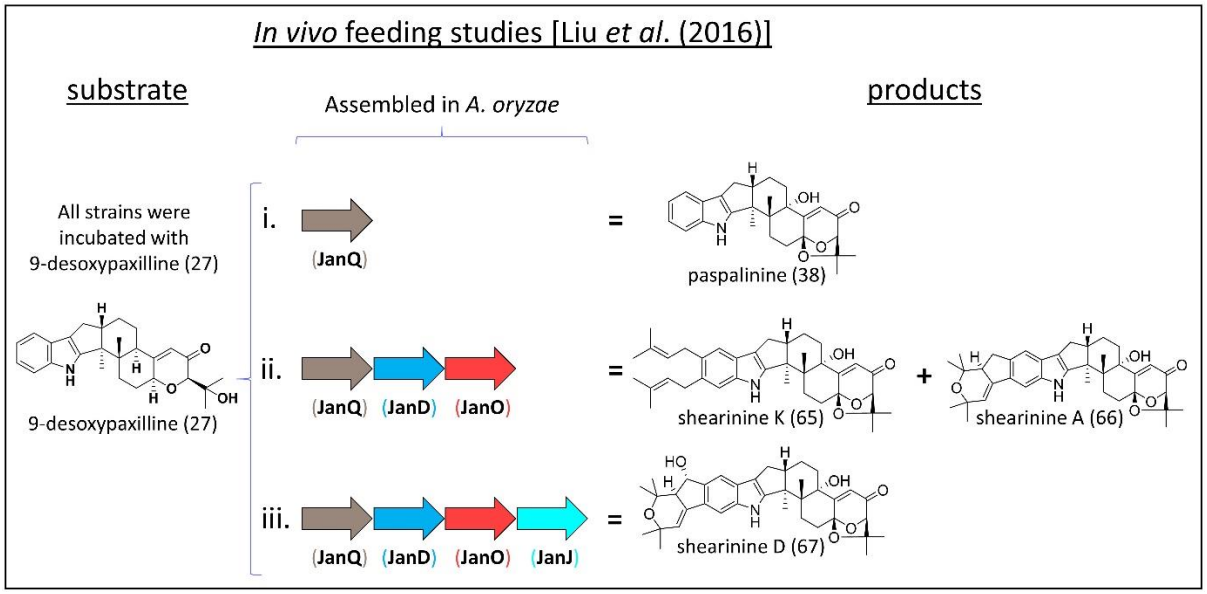


Figure 1.33. Summary of results from heterologous feeding study of *JAN* genes in *A. oryzae* by Liu *et al*. (2016).[43]

To further investigate the detailed mechanism of the *janDO* prenylation and oxidative cyclisation, they analysed the JanD and JanO recombinant enzymes *in vitro*. They

incubated recombinant JanD with paspalinine (38) and paxilline (2) and found that paspalinine (38) was specifically bisprenylated to shearinine K (65) whereas paxilline (2) was unspecifically bisprenylated to form both 21,22-bisprenylpaxilline (33) and 20,21-bisprenylpaxilline (68) as shown in Figure 1.34, i and ii. Additionally, they analysed recombinant PaxD and found that it readily converted paxilline (2) to 21,22-bisprenylpaxilline (33) as shown in Figure 1.15, i) and also converted paspalinine (38) to shearinine K (65) as shown in Figure 1.34, ii. Next, they incubated recombinant JanO with shearinine K (65) and saw rapid complete conversion to shearinine A (66) as a shown in Figure 1.34, iii, indicating that JanO was solely responsible for the oxidative cyclisation to form shearinine A (66). Finally, they conducted a large-scale JanO mediated oxidation of 21,22-bisprenylpaxilline (33) and a coupled JanD-JanO reaction with paxilline (2) as shown respectively in Figure 1.34, v and vi. Interestingly JanO poorly oxidised 21,22-bisprenylpaxilline (33) to 22-protoshearinine (69) and shearinine B (70) but was unable to oxidise or cyclise the 20,21-prenylpaxilline. In accordance with this, incubation of paxilline (2) with JanD-JanO readily yielded 20,21-bisprenylpaxilline (68) and 21,22-bisprenylpaxilline (33) but only produced very small quantities of shearinine B (70) and 22-protoshearinine (69).

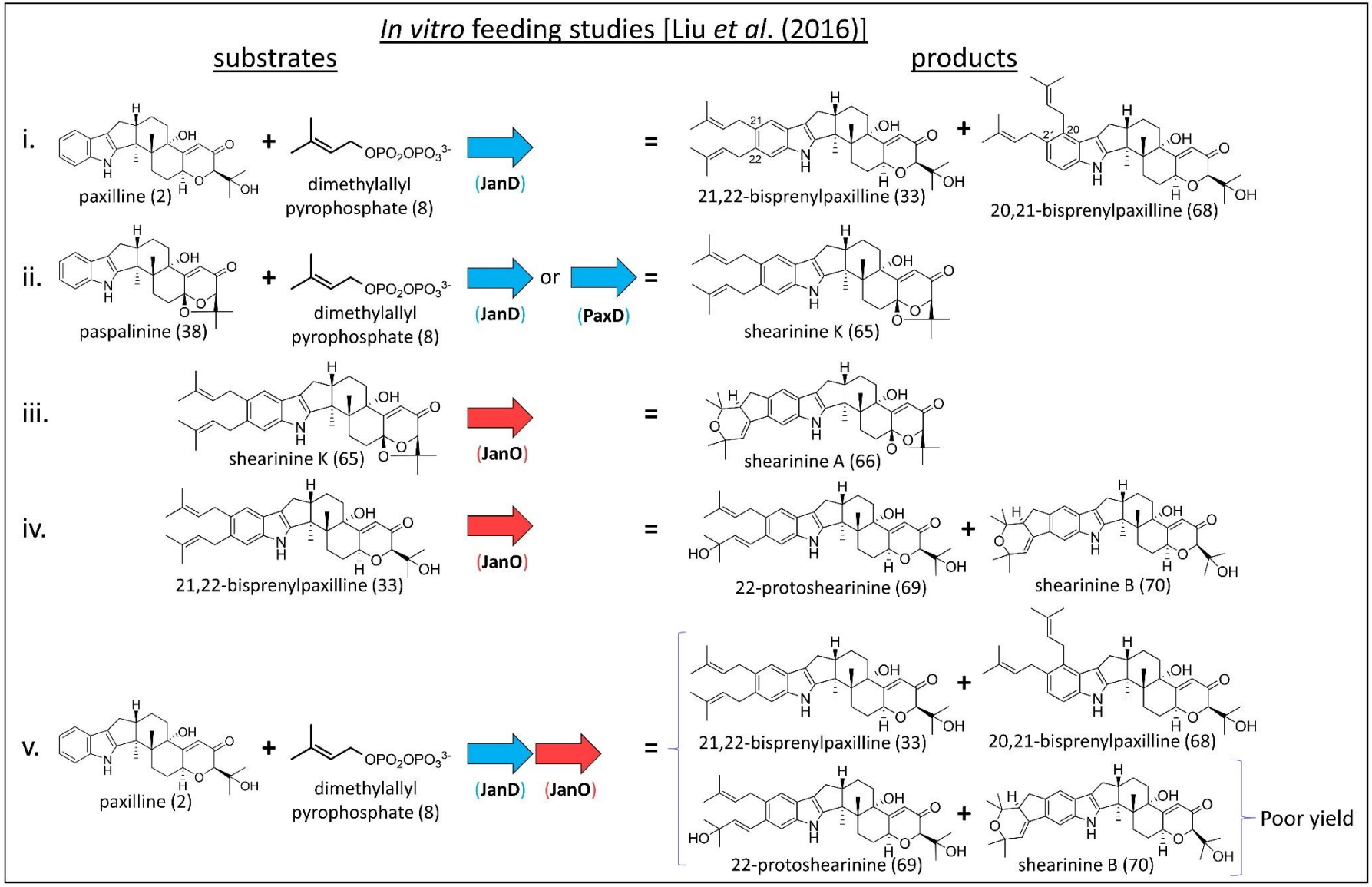


Figure 1.34. Summary of recombinant enzyme substrate studies of JanD and JanO by Liu *et al*. (2016).[43]

The diprenylation-flavoprotein oxidative cyclisation mechanism of JanD and JanO in shearinine biosynthesis provides a third pathway by which the indole diterpene cores are prenylated and oxidised. In contrast to this mechanism, the bicylco[3.2.0]heptane in penitrem biosynthesis is catalysed by a prenyl transferase (PtmE) acting on the hydroxylated-prenyl group of PC-M5 (60), and the bicylic core of lolitrems is formed by a diprenylation-P450-catalysed oxidative cyclisation. Thus, there are three different known mechanisms that fungi use during construction of the bisprenylation-oxidation sequence (Figure 1.35).

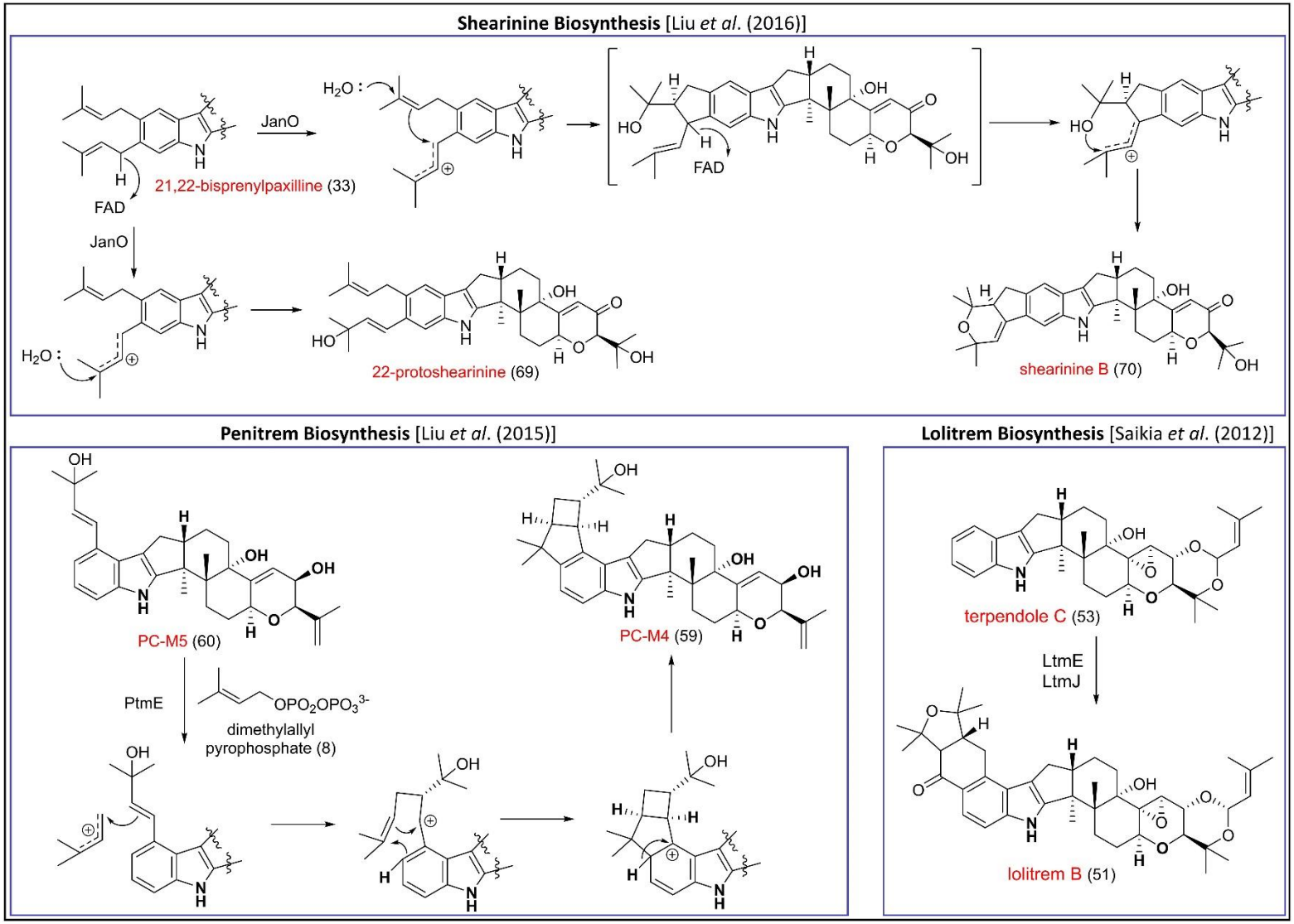
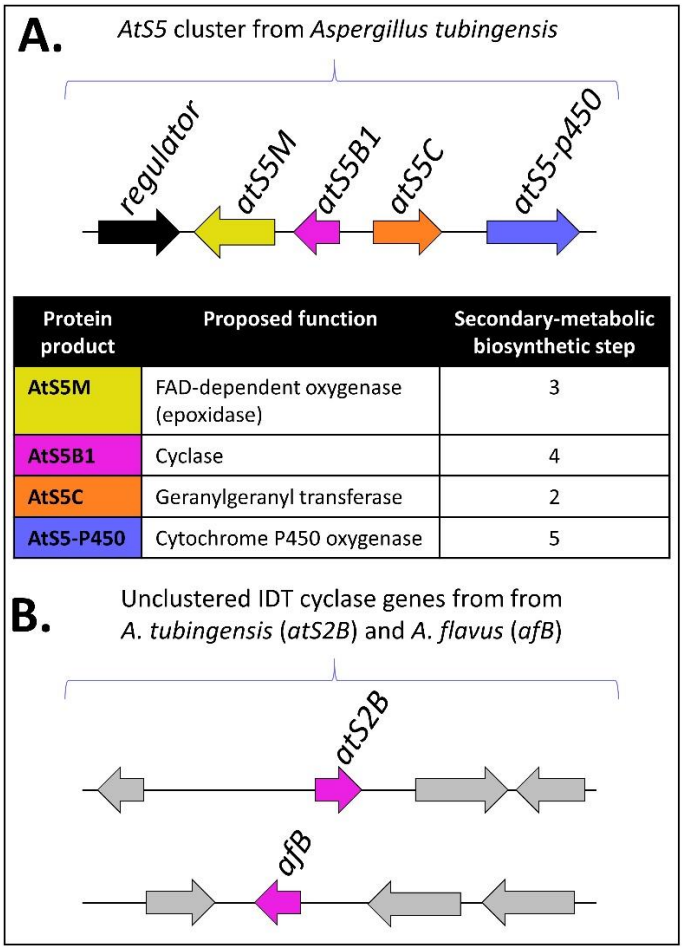


Figure 1.35. Summary of the three different kinds of oxidative cyclisation mechanisms proposed by Saikia *et al*. (2012) for the lolitrems,[29] Liu *et al*. (2015) for the penitrems,[31] and Liu *et al*. (2016) for the shearinines.[43]

Identification and functional analysis of unclustered indole diterpene genes, which led to the identification of another indole diterpene cluster, *ATS5*

The identification and functional confirmation of the various clustered indole diterpene genes established a sequence library that Tang *et al*. (2015) used to selectively probe for novel, unclustered indole diterpene genes.[33] Their studies specifically focused around the identification and functional analysis of three novel indole diterpene cyclases (AtS2B, AtS5B1, and AfB) but further exploration around one of the novel cyclases (Ats5B1) revealed a new indole diterpene cluster, the *ATS5* locus, in *Aspergillus tubingenesis* (Figure 1.36). Rather than use the typical filamentous fungi expression system for their heterologous pathway reconstitution studies, they showed, for the first time, that yeast (*Saccharomyces cerevisiae*) readily accepted indole diterpene genes for production of indole diterpene products.



Protein product	Proposed function	Secondary-metabolic biosynthetic step
AtS5M	FAD-dependent oxygenase (epoxidase)	3
AtS5B1	Cyclase	4
AtS5C	Geranylgeranyl transferase	2
AtS5-P450	Cytochrome P450 oxygenase	5

Figure 1.36. Schematic of AtS5 locus in **A** and the location of the two unclustered indole diterpene cyclases AtS2B and AfB in **B** (Tang *et al*. 2015).[33]

To demonstrate that their yeast expression system worked, Tang *et al*. (2015) first reassembled the core indole diterpene genes (*atmGCM*) from the aflatrem (36) biosynthetic pathway into *S. cerevisiae*.[33] When AtmG and AtmC were co-expressed they unexpectedly saw generation of 17,18-epoxy-3-geranylgeranylindole (15), which they attributed to an endogenous yeast oxygenase (Figure 1.37, i). Upon addition of AtmM to the *atmGC*-mutant, they saw formation of the two expected products shown in Figure 1.37, ii [13,14-epoxy-3-geranylgeranylindole (16) and bisepoxy-3-geranylgeranylindole (17)]. The triple gene transformant provided the epoxy-substrates required for the analysis of various indole diterpene cyclases. Thus, as a proof of concept, the first indole diterpene cyclase they tested was AtmB. When *atmB* was introduced into the *atmGCM*-strain, paspaline (11) was detected as the major product and emindole SB (18) was detected as the minor product (Figure 1.37, iii) confirming that their expression system was suitable for the analysis of other indole diterpene cyclases.

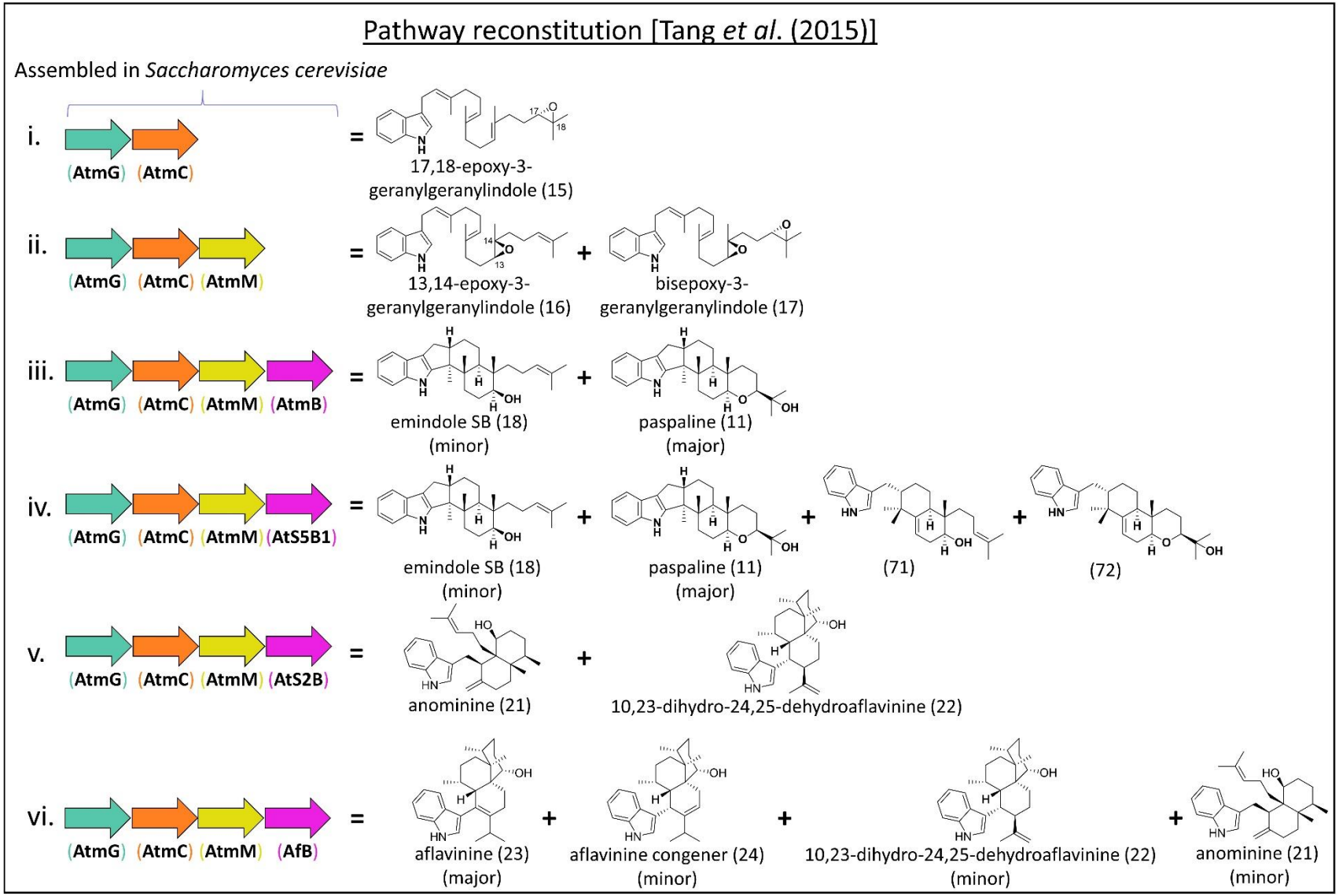


Figure 1.375. Summary of heterologous expression studies used to analyse indole diterpene cyclases [Tang *et al*. (2015)].[33]

Tang *et al*. (2015) chose to focus on identifying indole diterpene cyclases that were responsible for the formation of the Markovnikov-derived indole diterpenes anominine (21) and aflavinine (23) because no indole diterpene cyclases had been identified that gave rise to any Markovnikov-derived indole diterpene products.[33] From Protein BLAST searches on AtmB homologues they generated a phylogenetic tree containing 140 homologous enzymes grouped into five clades. From the phylogenetic tree they systematically chose three indole diterpene cyclases to study; two from *Aspergillus tubingenesis* (AtS2B and AtS5B1) and one from *A. flavus* (AfB). AtS2B and AfB were both from Clade 4 of their tree, which notably contained standalone indole diterpene cyclases, and AtS5B1 was from Clade 5, the same clade as the clustered indole diterpene cyclases that give rise to paspaline (11)-derived indole diterpenes (i.e. AtmB, PaxB, etc.). To examine whether phylogenetically-grouped indole diterpene cyclases had similar cyclisation regioselectivity they first tested AtS5B1. Upon cloning *atS5B1* into the *atmGCM*-strain, they saw formation of the expected paspaline (11) and emindole SB (18) as well as two new products (71) and (72) shown in Figure 1.37, iv. Intriguingly, compound (71) and (72) are derived from a different cyclisation pattern than paspaline (11) and emindole SB (18), which involves a 1,2-methyl migration and concomitant ring expansion to yield the 6-6 bicyclic terpene moiety. Although the cyclisation patterns differ, both involve anti-Markovnikov generated ring closures thereby demonstrating that these phylogenetically-grouped indole diterpene cyclases share similar cyclisation mechanisms.

Tang *et al*. (2015) then turned their attention to the standalone indole diterpene cyclases, AtS2B and AfB. When AtS2B was introduced into the *atmGCM*-strain two new compounds, confirmed by NMR spectroscopy as anominine (21) and 10,23-dihydro-24,25-dehydroaflavinine (22), were produced (Figure 1.37, v). In contrast, when AfB was introduced into the *atmGCM*-strain, aflavinine (23) was produced as the major product and anominine (21), 10,23-dihydro-24,25-dehydroaflavinine (22), and aflavinine cogener (24) were produced as minor products (Figure 1.37, vi). Thus the two standalone indole diterpene cyclases are solely responsible for the cyclisation mechanism that gives rise to anominine (21) and aflavinines.

Finally, Tang *et al*. (2015) further demonstrated the power of the phylogenetically-guided yeast platform by identifying and confirming the function of a P450 (AtS5-P450) that was clustered with AtS5B1. Co-expression of AtS5-P450 in a mutant containing

atmGC, *atS5M* (the *ATS5* cluster homologue of *atmM*), and *atS5B1* yielded three novel terminally-hydroxylated compounds (73), (74) and (75) shown in Figure 1.38, i. Surprisingly, when AtS5-P450 was combinatorially co-expressed with *atmGCM* and *atS2B*, two terminally-hydroxylated anominine (21) two new derivatives were produced, compound (76) and (77) as shown in Figure 1.38, ii. Overall, these results showed that phylogeny-guided pathway reassembly is a promising method for the characterisation of novel genes, which in this case were indole diterpene cyclases and an oxygenase.

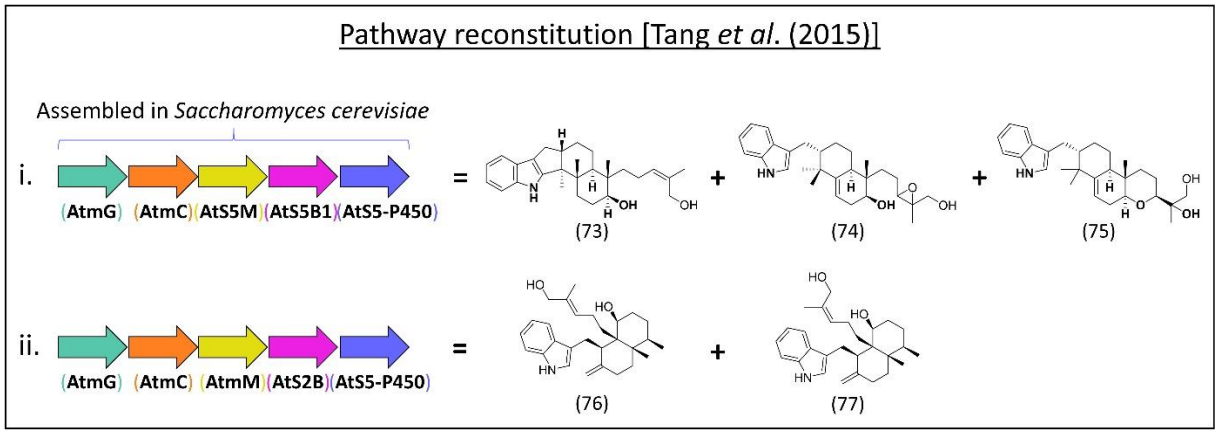


Figure 1.38. Summary of Tang *et al*. (2015) results from heterologous expression studies with AtS5-P450.[33]

Conclusions

This section has explored how the indole diterpenes are biosynthesised and the methods that have been used to identify and functionally analyse the indole diterpene genes. In summary, it is clear that the functional groups around the indole diterpene core structures are derived from the differences in the genes from their host organism. Some of the differences are obvious, as in the specific biosynthetic pathway divergence seen between the hosts that produce paspaline (11)-derived indole diterpenes; where the Eurotiomycetes (*Penicilli and Aspergilli*) use 9-desoxypaxilline (27) as substrate and have β-stereochemistry at C10, and the Sordariomycetes (*C. alba*, *N. lolii*, *E. festucae*) use terpendole E (47) as substrate and have α-stereochemistry at C10. But other differences, like the under-utilization of PaxD and PaxO in *P. paxilli* in comparison to JanD and JanO in *P. janthinellum* remain a bit of a

mystery. Mysteries like this one will be exposed as more information about the genes and gene products is amassed. With the identification of each new indole diterpene gene and functional characterisation of the gene product, comes insight into the key differences in gene sequence that introduce substrate specificity and gene function.

As the databases grow, it becomes easier to envision a future where made-to-order chemical products can be easily engineered by altering genetic information. Components of a chemical's pharmacophore will be alterable to remove toxic moieties and deliver more potent therapeutic properties (i.e. mixing and matching the decorative genes from different organisms may enable creation of made-to-order pseudo-natural products with prescribed bioactivities).

In parallel, approaches will be developed to enhance compound production to enable access to useful compounds that are produced in very small quantities or potentially not at all (e.g. turning on silent genes). This thesis serves as a proof of concept to address this parallel approach and explore how to gain access to the valuable nodulisporic acids that have remained just beyond reach for the past three decades. The following section reviews what is known about the nodulisporic acids to provide a convincing argument for the purpose of pursuing this research project.

Nodulisporic acids: the target for this research

Nodulisporic acid A (1) (Figure 1.39) was first discovered in 1992 by Merck Research Laboratories during an ongoing study to identify unique bioactive compounds.[11] Merck scientists isolated nodulisporic acid A (1), using bioassay-guided isolation, from the fermentation extracts of the endophytic filamentous fungus, *Hypoxylon pulicicidum* (formerly classified as *Nodulisporium* sp.[17] MF 5954, ATCC 74245), which they acquired from the tropical woody plant, *Bontia daphnoides*, in Hawaii.[11] The discovery of the potent bioactive profile associated with nodulisporic acid led to an extensive screening project aimed at identifying additional nodulisporic acid A (1) producers. Twelve new strains of *Nodulisporium* spp. (presumably all *Hypoxylon* spp.) were identified across seven tropical regions on four continents; and in a parallel approach to identify congeners and/or natural analogues of nodulisporic acid A (1) that were inaccessible by chemical derivitization, Merck scientists created a library of chemically mutated *Hypoxylon* strains that overproduced key nodulisporic acid intermediates (i.e. A1, A2, A4, Δ23-A4, B, B1, B2, C, C1, C2, Δ23-C4, D, D1, D2, D3, E, and F; Figure 1.39).[11, 50-53] They screened the bioactive profiles of this nodulisporic acid compound library and found that nodulisporic acid A (1) had the most potent insecticidal properties of all of the naturally occurring nodulisporic acid derivatives. Although nodulisporic acid A (1) had very good *in vitro* and *in vivo* activity against fleas, its potency and pharmacokinetic properties did not justify its development as a drug, leading Merck to launch a medicinal chemistry study to optimise the nodulisporic acid A (1) profile. Merck scientists generated and screened a library of more than 1000 nodulisporic acid A (1) analogues leading to a comprehensive understanding of the nodulisporic acid A (1) pharmacophore and the identification of an ideal drug target, *N*-tert-butyl nodulisporamide (78).[54-55] Details of Merck's studies are discussed below.

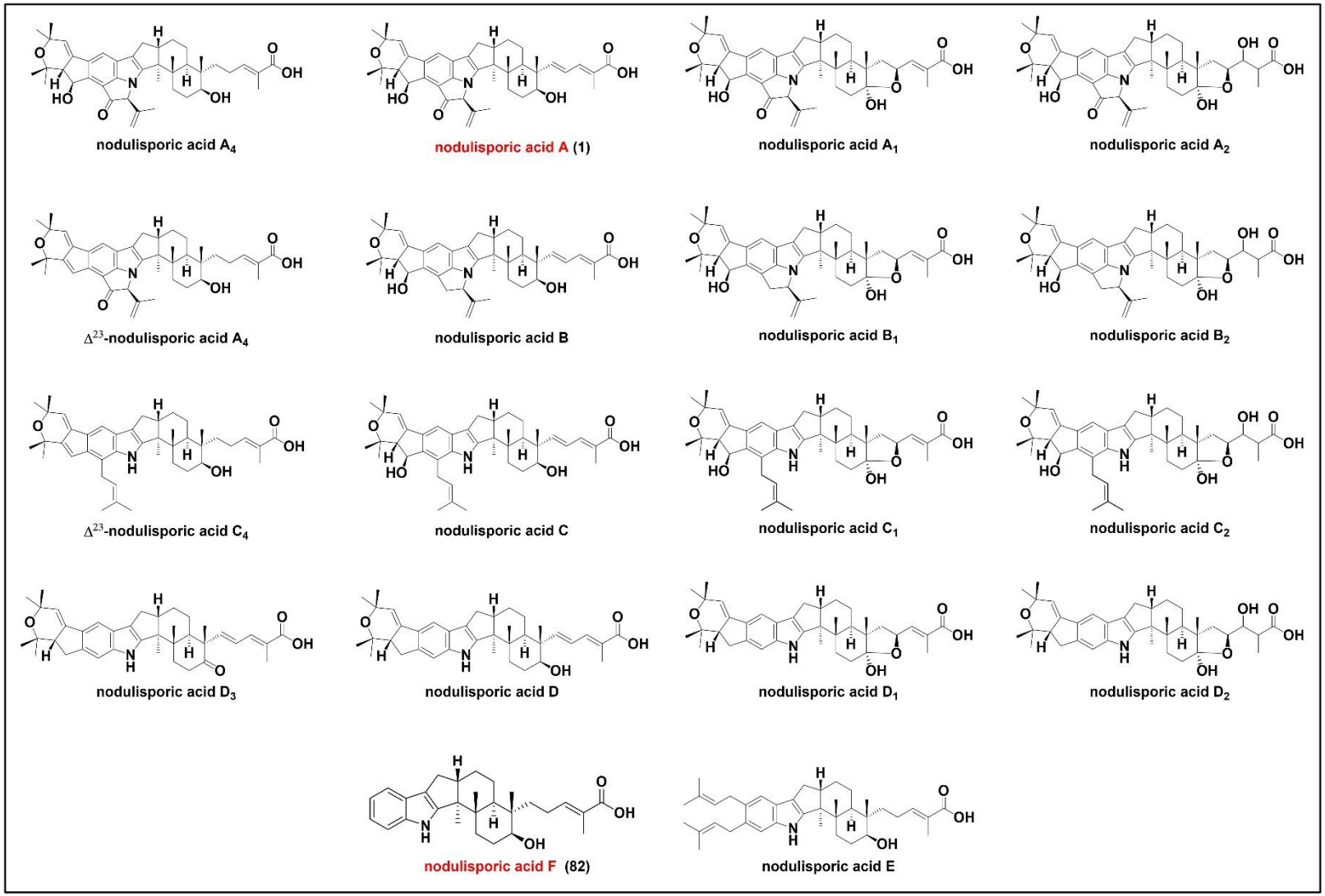


Figure 1.39. Collection of known nodulisporic acids.[11, 50]

The discovery of nodulisporic acid A's (1) insecticidal activity

The insecticidal bioactivity of nodulisporic acid A (1) was first detected by Merck in 1992 during crude natural product extract screening with an *in vitro Lucilia sericata* (blowfly) assay.[56] In parallel to the *L. sericata* assay, Merck employed an *in vivo* bioassay using *Cimex lectularius* (bedbug) infected mice to isolate nodulisporic acid A (1) through evaluations of crude broth fractionations (i.e. infected mice were dosed with fractions of the fungal extract).[57-58] Once nodulisporic acid A (1) was isolated, pure compound was found to have a lethal concentration necessary to kill 50% (LC_{50}) against *L. sericata* of 300-1000 ppb.[56, 58] These results demonstrated that nodulisporic acid A (1) had potential for use as an insecticide because it was more potent than the widely used pyrethroids (~22,000-28,000 ppb).[11] Thus, additional bioassays were completed to analyse nodulisporic acid A's (1) potency to a range of other ectoparasites.

Merck first screened pure nodulisporic acid A (1) in an *in vivo C. lectularius* bioassay, and observed complete parasite paralysis at 0.4 ppm, which, when compared to 5.0 ppm of ivermectin, indicated that nodulisporic acid A (1) could be developed into a useful therapeutic to fight blood-feeding arthropods. To test this theory further, nodulisporic acid A (1) was assayed against the cat flea (*Ctenocephalides felis*) in an artificial membrane flea-feeding apparatus, and again found to be roughly ten-times more potent than ivermectin (i.e. 1 ppm vs 10 ppm).[59] Nodulisporic acid A (1) was further tested against the hard tick (*Dermacentor variabilis*) in a similar *in vivo* bioassay to that used with *C. lectularius*, but no observable activity was observed when mice were treated with a single oral dose of 2.5 mg/kg.[60] Notably, anthelmintic activity was also tested using *in vitro* and *in vivo* assays but no desirable activity was observed.[11]

Nodulisporic acid A (1) selectively targets glutamate gated chloride ion channels

The potency that nodulisporic acid A (1) exhibited toward fleas prompted Merck scientists to examine nodulisporic acid A's (1) mechanism of action. To this end, they developed a series of nodulisporic acid A (1) and avermectin derived radioligands and used them to establish that nodulisporic acid A (1) selectively activates glutamate-gated chloride ion channels that are unique to invertebrates.[61-62] Activation of these channels silences neuronal activity and leads to paralysis, starvation and death of the

insect. Importantly, they found that nodulisporic acid A (1) does not activate glycine- and GABA-gated chloride ion channels, which are found across both invertebrates and vertebrates and are notably activated by ivermectin. Therefore, the ion channel selectivity of nodulisporic acid A (1) imparts it with a unique vertebrate safety profile that confers the ability to orally dose cats and dogs at higher levels than ivermectin, allowing for greater anitparasitic activity and enhanced therapeutic potential.

Nodulisporic acid A (1) is an orally safe, systemically active, ectoparasiticide

In the case of animal therapeutics, oral treatments are preferred over topical as they ensure uniform drug coverage and are not subject to variation in drug effectiveness caused by environmental factors. That said, systemically active oral therapeutics are much harder to come by as they must be nontoxic (i.e. cannot interfere detrimentally with the inner workings of the animal undergoing treatment). Therefore, nodulisporic acid A (1) is an ideal oral therapeutic because it is a systemically active ectoparasiticide that shows no acute mammalian toxicity[59] and does not interfere with ligand gated channels (GABA, kainite, glycine) or sodium and potassium voltage gated channels,[62] which is an essential property of long-lasting therapeutics (i.e. therapeutics dosed on a monthly basis) as they will be exposed to plasma for extended periods of time. Additionally, the unique targeting of nodulisporic acid A (1) to the glutamate-gated chloride ion channels imparts nodulisporic acid A (1) with a novel mechanism of action compared to the widely used pyrethroids, which alter the kinetics of sodium gated ion channels.[63] Therefore, by targeting a different ion channel, nodulisporic acid A (1) circumvents the problem of insecticidal resistance that has developed toward the common insecticides like the pyrethroids,[64] carbamates, and organophosphates.[65-67]

To directly demonstrate nodulisporic acid A's (1) unique vertebrate safety profile, Merck completed *in vivo* bioassays using *C. felis* (flea) infected dogs.[59] Dogs were experimentally infected with *C. felis* and administered nodulisporic acid A (1) orally, one time, at 15 mg/kg in two studies. Upon a 48 hour post-dose examination the nodulisporic acid A (1)-treated dogs harboured 80% and 99.7% fewer fleas than the control dogs. Additional new fleas were placed on dogs four and five days after initial administration of nodulisporic acid A (1) and examined 48 hours later, day six and seven post-dose (study dependent), at which point 50% residual efficacy was

observed. These results led Merck to examine the efficacy of oral administration and the pharmacodynamics of nodulisporic acid A (1) *in vivo*.

Merck scientists examined the efficacy of oral administration by comparing the maximum blood plasma concentration and plasma half-life of nodulisporic acid A (1) when administered orally versus intravenously.[59] They found that the results were similar upon an intravenous administration of 1.5 mg/kg nodulisporic acid A (1) and an oral administration of 15 mg/kg nodulisporic acid A (1), where the maximum blood plasma concentration was respectively 2.2 mg/ml 1 hour post-dose compared to 1.0 mg/ml 8 hours post-dose and the plasma half-life was 23 hours compared to 19.3 hour, indicating that oral administration was viable. Importantly, the pharmacodynamic data from this dog study showed that blood concentrations of less than 0.1 mg/ml nodulisporic acid A (1) were effective at killing ~50% of fleas.

Together these results demonstrated that nodulisporic acid A (1) had potent *in vitro* and *in vivo* activity toward fleas but that its pharmacodynamics profile was not ideal for drug development. Thus, Merck launched a medicinal chemistry effort to design a nodulisporic acid A (1) analogue that had a longer plasma half-life and more potent insecticidal properties.

Medicinal chemistry studies: elucidation of the nodulisporic acid A (1) pharmacophore and development of nodulisporic acid A (1) analogues with more desirable properties

When Merck scientists launched their medicinal chemistry efforts they were unaware of how modifications to different parts of the nodulisporic acid A (1) structure would affect its inherent bioactive properties. As an initial approach they made a mixture of simple modifications to each reactive centre on the nodulisporic acid A (1) scaffold to elucidate the key components of the nodulisporic acid A (1) pharmacophore that were essential for its potent ectoparasiticidal activity. In particular, they altered the oxidation state of the various functional groups (Figure 1.40) and evaluated a variety of dienoic chain modifications by inserting esters, thioesters, and amides.[54]

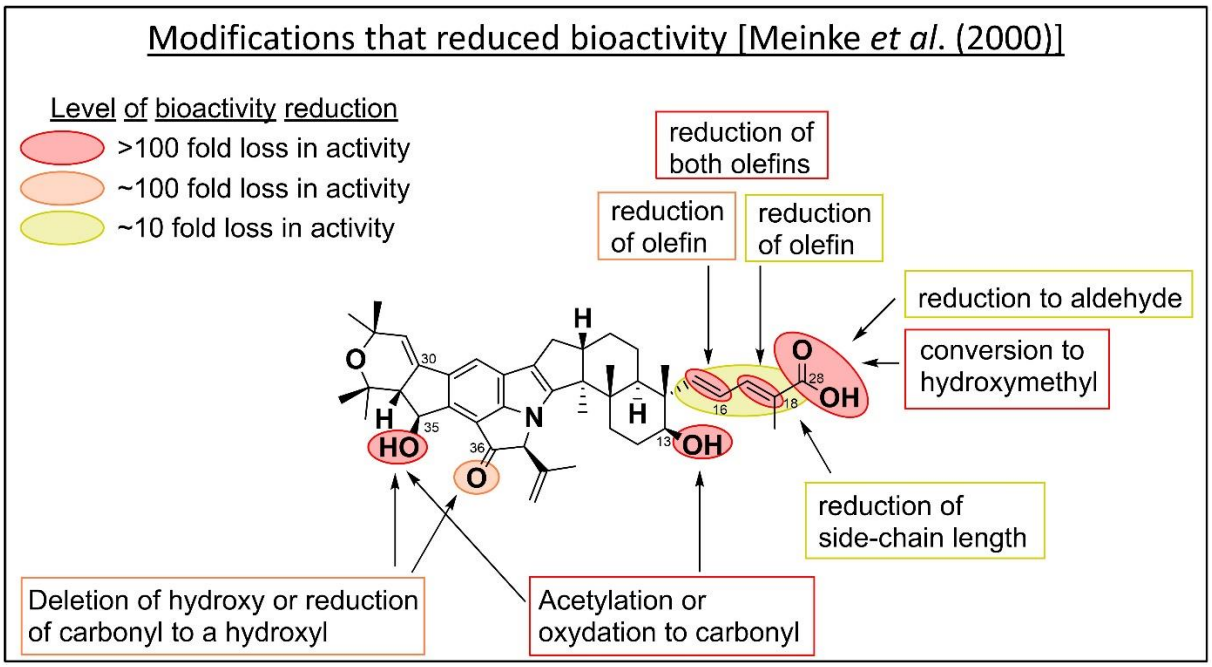


Figure 1.40. Summary of modifications cited in by Meinke *et al*. (2000) that diminished the inherent insecticidal properties of nodulisporic acid A (1).[54]

Results from these initial modification studies revealed that alterations to the core nodulisporic acid A (1) structure were deleterious to its inherent insecticidal properties whereas many modifications to the dienoic acid sidechain enhanced desirable bioactive properties. This led them to characterise the nodulisporic acid A (1) core as "non-permissive" and the dienoic acid side chain as "permissive" to chemical modifications (Figure 1.41).[11]

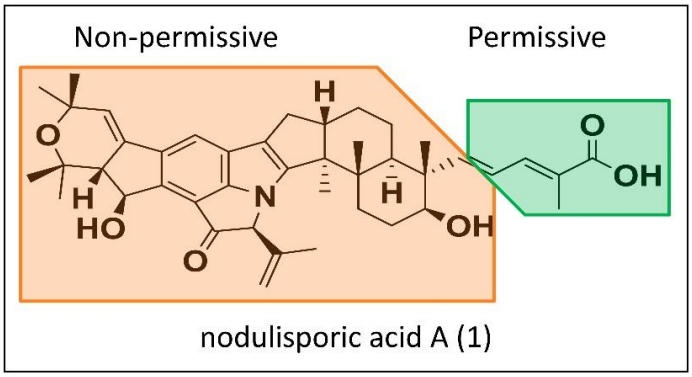


Figure 1.41. Depiction of the non-permissive and permissive regions of nodulisporic acid A (1).[11] All chemical modifications to the non-permissive region were deleterious to the inherent insecticidal activity of nodulisporic acid A (1) whereas some key modifications to the permissive region greatly enhance the insecticidal profile of nodulisporic acid A (1).

Although many of their amide side chain modifications resulted in more potent anti-flea activity (i.e. >1 order of magnitude), there were two key amide modifications to the acidic tail of nodulisporic acid A (1), the mono-substituted $NHC(Me)_2C(O)NMe_2$ compound (79) and the disubstituted *N*-1-[(4-SO_2Me)Piperazinyl] compound (80) shown in Figure 1.42, that enhanced the flea efficacy by two orders of magnitude (i.e. flea $LC_{\geq 80}$ of 0.01 ppm vs 1 ppm).[54] With this in mind, Merck scientists directed their focus to further enhancing the anti-tick and pharmacokinetic properties of nodulisporic acid A (1) by developing more intricate side chain modifications.

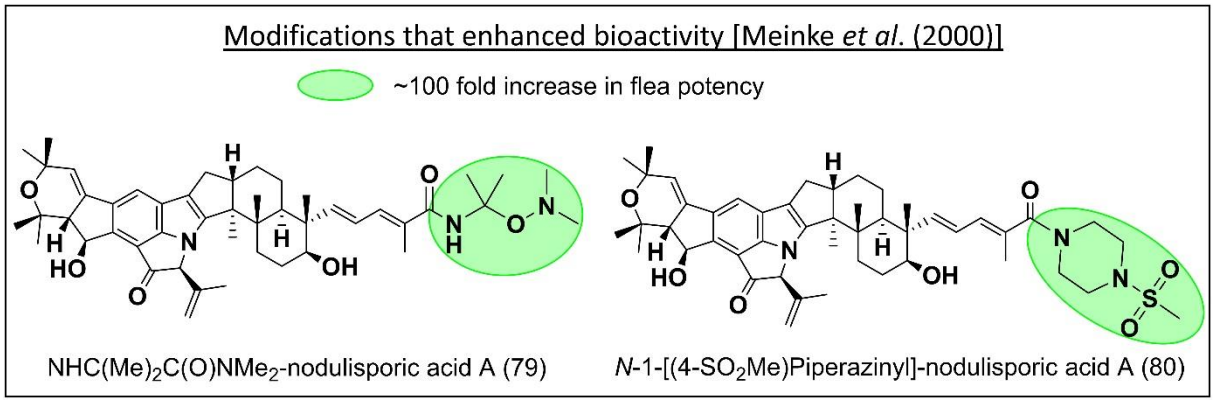


Figure 1.42. Depiction of the two key nodulisporamides that were nearly two orders of magnitude more potent to fleas than nodulisporic acid A (1).[54]

One approach led by Berger *et al*. (2001) was to install heterocylic oxazoles and thiazoles into the 17,18-olefin of the dienoic acid chain of nodulisporic acid A (1).[60] Of the three key heterocycles tested, the dimethylamino oxazole compound (81) shown in Figure 1.43 was the most well balanced as it demonstrated superior *in vivo* flea and tick efficacy to nodulisporic acid A (1) - 1.5 weeks longer protection from fleas and 4 weeks protection from ticks - and had an improved plasma half-life - 2.8 days vs 0.8 days for nodulisporic acid A (1). Unfortunately, the synthesis of the dimethylamino oxazole compound (81) was uneconomical preventing further development of bioactive properties of the heterocylic derivatives.[55]

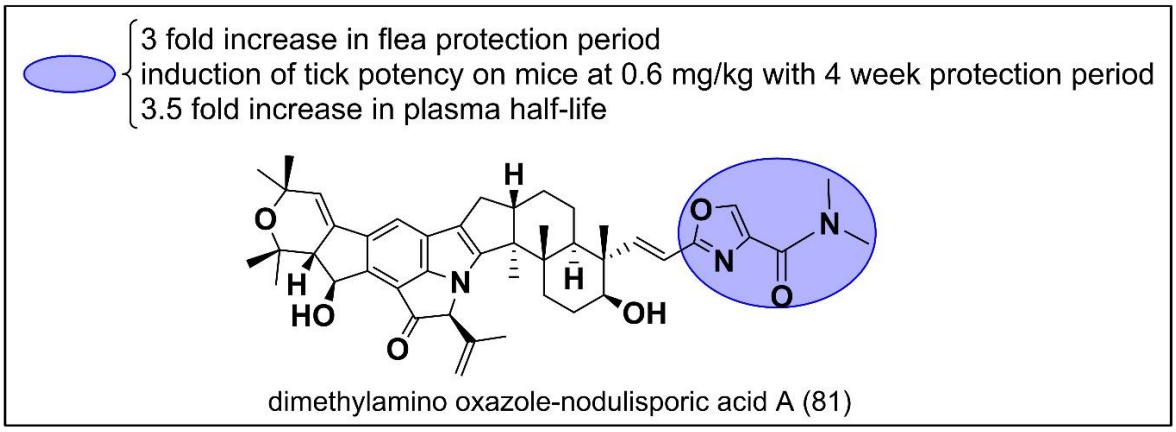


Figure 1.43. Depiction of the key dimethylamino oxazole that installed potent anti-tick activity onto the nodulisporic acid A (1) core [Berger *et al*. (2001)].[60]

With a general understanding of the nodulisporic acid A (1) pharmacophore and the chemical modifications that would likely install desirable properties, Merck scientists pushed forward with their efforts to create a safe, economically viable nodulisporamide derivative that could be dosed monthly for the prevention of fleas and ticks on companion animals. They prepared a library of 335 nodulisporamides and screened them *in vitro* using their artificial flea-feeding assay and *in vivo* using a mouse/bedbug assay.[55] Of the 335 nodulisporamides tested, 66 that had promising properties (i.e. flea LC_{50} ≤ 1.0 ppm and bedbug ED_{50} ≤ 1.0 mg/kg) were analysed in an *in vitro* dog/flea assay, of which twelve were screened more thoroughly in an eight-week dog/flea assay, four-week dog/tick assay, and four-week cat/flea assay. Of the twelve nodulisporamides tested, all of which were synthesised in a single chemical reaction from nodulisporic acid A (1), one, *N*-tert-butyl nodulisporamide (78) shown in Figure 1.44, displayed the most desirable ectoparasitic activity and was selected for drug development.

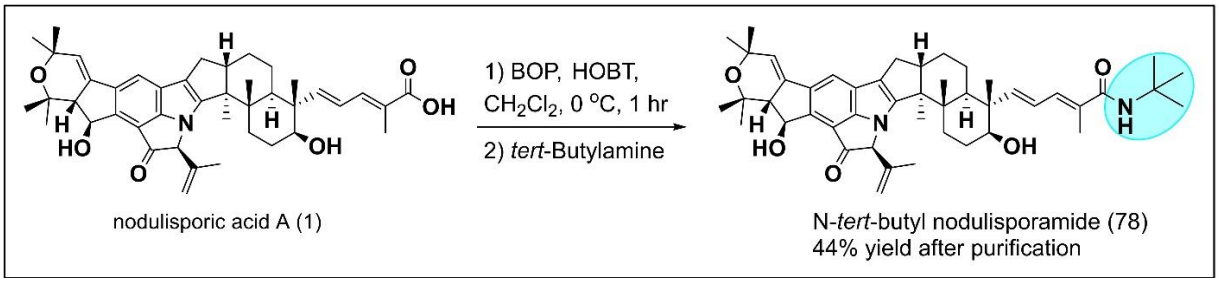


Figure 1.44. *N*-tert-butyl nodulisporamide (78) can be synthesised from nodulisporic acid A (1) using a single chemical reaction [Meinke *et al*. (2009)].[55]

Analysis of the ideal nodulisporic acid A (1) analogue drug target: *N*-tert-butyl nodulisporamide (78)

Flea efficacy of *N*-tert-butyl nodulisporamide (78) in dogs

To validate the development of *N-tert*-butyl nodulisporamide (78) into a drug, Merck scientists conducted additional comparison studies against two widely used topical agents, fipronil and imidacloprid.[55] They conducted a nine-week dog/flea efficacy study where *N-tert*-butyl nodulisporamide (78) was orally dosed at 10 mg/kg, and fipronil and imidacloprid were applied as described in their packaging. Following the single oral dose of *N-tert*-butyl nodulisporamide (78) complete (100%) flea control was observed for six weeks and marginal loss of efficacy was observed in week seven and eight until nearly 50% loss in activity was reached in week nine. Corresponding pharmacokinetic data showed that the flea control diminished when the plasma drug levels decreased to ≤ 0.20 μg/mL. In comparison, fipronil controlled 100% fleas until week five and had >90% efficacy throughout week six to nine and imidacloprid exhibited 100% flea control for four weeks and diminished activity in week five and six. A subsequent test of *N-tert*-butyl nodulisporamide (78) flea efficacy was completed where dogs were dosed with 30 mg/kg and checked for fleas at 61 and 90 days post-dose. At day 61, 100% flea control was observed but by day 90 there was an exceptional loss in efficacy (23% control). These results demonstrated that orally dosed *N-tert*-butyl nodulisporamide (78) competes well against the most widely used topical flea therapeutics.

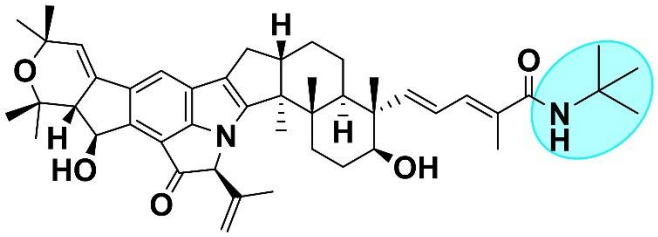
Flea efficacy of *N*-tert-butyl nodulisporamide (78) in cats

From their initial study of fleas on cats, Merck found that cats eliminated *N-tert*-butyl nodulisporamide (78) from their system more quickly than dogs resulting in a shorter half-life of 6.6 days vs 8.7 days and lower maximum concentration of 1.52 ± 1.65 vs 9.8 ± 0.28 μM, and thus a greatly reduced anti-flea therapeutic affect when dosed with 10 mg/kg.[55] Therefore, Merck conducted a second four-week flea challenge on cats given a single oral dose of either 15 mg/kg or 20 mg/kg or a topical dose of fipronil or imidacloprid. Although the cats given 15 mg/kg *N-tert*-butyl nodulisporamide (78) experienced flea breakthrough at three weeks, those dosed with 20 mg/kg were completely protected (100%) from fleas for three weeks with 97% and 94% protection

on week four and five, respectively. By week six the efficacy dropped to 76% where the systemic plasma concentration was 0.2-0.3 μg/mL. In contrast, fipronil and imidacloprid provided nearly complete flea protection for four weeks and showed only moderate loss of efficacy in week five, 94% and 97% respectively. In week six, flea breakthrough was observed for fipronil whereas imidacloprid showed >97% efficacy.

Tick efficacy of *N*-tert-butyl nodulisporamide (78) in dogs

Merck scientists examined the tick ectoparasitic efficacy of *N-tert*-butyl nodulisporamide (78) compared to fipronil over a four-week tick/dog challenge (imidacloprid was not assessed because it does not exhibit anti-tick activity).[55] They found that a 10 mg/kg oral dose of *N-tert*-butyl nodulisporamide (78) was 80%, 78%, and 41% effective against brown dog ticks (*Riphicephalus sanguineus*) on dogs at days 12, 19, and 26 respectively. In comparison, fipronil was 91% and 77% effective at days 19 and 26 showing that a higher oral dose of *N-tert*-butyl nodulisporamide (78) would be required to have competitive advantage over fipronil. Therefore, Merck launched a second four-week tick challenge where dogs were challenged with *R. sanguineus* or Lone Star ticks (*Amblyoma americanum*) and given a single oral dose of 30 mg/kg. Although 100% protection was observed at day 14 for both tick species, the drug efficacy dropped to 85% and 87% at day 21 and 84% and 92% at day 28 for *R. sanguineus* and *A. americanum* respectively. The corresponding pharmacokinetic data showed that tick control diminished when the plasma drug levels decreased to ≤ 1 μg/mL. Notably, these results demonstrated that a single 30 mg/kg oral dose of *N-tert*-butyl nodulisporamide (78) had a greater flea and tick efficacy than fipronil.



N-*tert*-butyl nodulisporamide (78)

Summary of bioactivity testing [Meinke *et al*. (2009)]

- 7 fold increase in flea potency (0.1 ppm LC_{50}) compared to NAA
- 8 fold increase in bedbug potency (0.125 mg/kg ED_{50} on mice) compared to NAA
- 8.5 fold increase in plasma half life of **dogs** dosed with **10** mg/kg compared to NAA
- 7 week flea prevention and induction of tick prevention for 3 weeks at **10** mg/kg dose on **dogs**
- 9 week flea prevention (100%) and 4 week tick prevention (84-92%) of **dogs** dosed with **30** mg/kg
- 3-4 week flea prevention on **cats** dosed with **10** mg/kg
- 5 week flea prevention (94%) of **cats** dosed with **20** mg/kg

Figure 1.45. Summary of bioactive properties of *N*-tert-butyl nodulisporamide (78).[55]

Further examination of the *N*-tert-butyl nodulisporamide (78) pharmacokinetic safety profile

To further analyse the safety profile of *N-tert*-butyl nodulisporamide (78), Merck conducted a series of 102 *in vitro* assays against *N-tert*-butyl nodulisporamide (78) at 10 μM.[55] Of the 102 assays, monoamine uptake was the only one that showed meaningful activity where 90% and 52% inhibition were observed at 10 μM and 1 μM respectively. A subsequent covalent binding study using radiolabelled *N-tert*-butyl nodulisporamide (78) in rat and dog liver microsomes showed <50 pM equiv/mg protein labelling. Importantly, *N-tert*-butyl nodulisporamide (78) was found to bind very tightly to plasma proteins (99.3%), an attribute that likely enhances its drug delivery and mammalian safety profile. A further study looking into [^{14}C] *N-tert*-butyl nodulisporamide (78) drug distribution over time found that at 8 hours postdose the majority of the drug is present in the stomach and gastrointestinal tract whereas at 21 and 28 days postdose the majority of the drug is in the abdominal and altilateral adipose tissue. This distribution study suggests that *N-tert*-butyl nodulisporamide (78) is initially absorbed into the adipose tissue and gradually leeches out, which confers it with a prolonged half-life and enhanced efficacy toward ectoparasites.

Difficulties with nodulisporic acid A (1) production prevent drug development

Although Merck scientists were able to identify an ideal nodulisporic acid A (1) analogue for drug development, they were hampered along the way by severe limitations in the availability of nodulisporic acid A (1); a limitation that has also prevented commercialization of *N-tert*-butyl nodulisporamide (78).[55] Firstly, the structural complexity of nodulisporic acid A (1) has precluded total synthesis of the compound for decades - the mechanisms for the total synthesis of nodulisporic acid F (82)[68] and nodulisporic acid D[69] have been elucidated but total synthesis of nodulisporic acid A (1) has not been achieved.[70] Secondly, the fermentation of nodulisporic acid A (1) from its natural producer, *H. pulicicidum*, is exceptionally difficult where the reported nodulisporic acid A (1) biosynthesis methods require that *H. pulicicidum* be grown for 21 days in complete darkness in highly nutrient rich medium.[44] Notably, personal attempts to isolate nodulisporic acid A (1) from *H. pulicicidum* using reported fermentation methods[44, 50] have been unsuccessful and more than 8 months of attempts at eliciting nodulisporic acid A (1) biosynthesis yielded nothing.

Thus the questions at the foundation of this thesis presents itself: *What secondary-metabolic genes and corresponding enzymatic steps are involved in nodulisporic acid biosynthesis?* In order to answer this question, the *H. pulicicidum* genome was sequenced, a gene cluster responsible for nodulisporcic acid biosynthesis was predicted, and four genes in the cluster were functionally characterised through gene complementation and pathway reconstitution studies using a more robust fungal host, *Penicillium paxilli*.

2. THE GENE CHARACTERISATION TOOLBOX

In this chapter the four main tools used to functionally characterise genes involved in nodulisporiac acid biosynthesis are described. The four tools that make up the gene characterisation toolbox include bioinformatics, DNA assembly, heterologous expression, and compound analysis (Figure 2.1). First, bioinformatics are used to predict the genes that are likely involved in a biosynthetic pathway of interest. Second, the predicted genes are assembled into plasmids for amplification. Third, amplified plasmids are inserted into a new host for heterologous expression in a process called transformation. Fourth, transformants are grown in production cucltures, compounds are extracted and extraction profiles are analysed for compounds of interest.

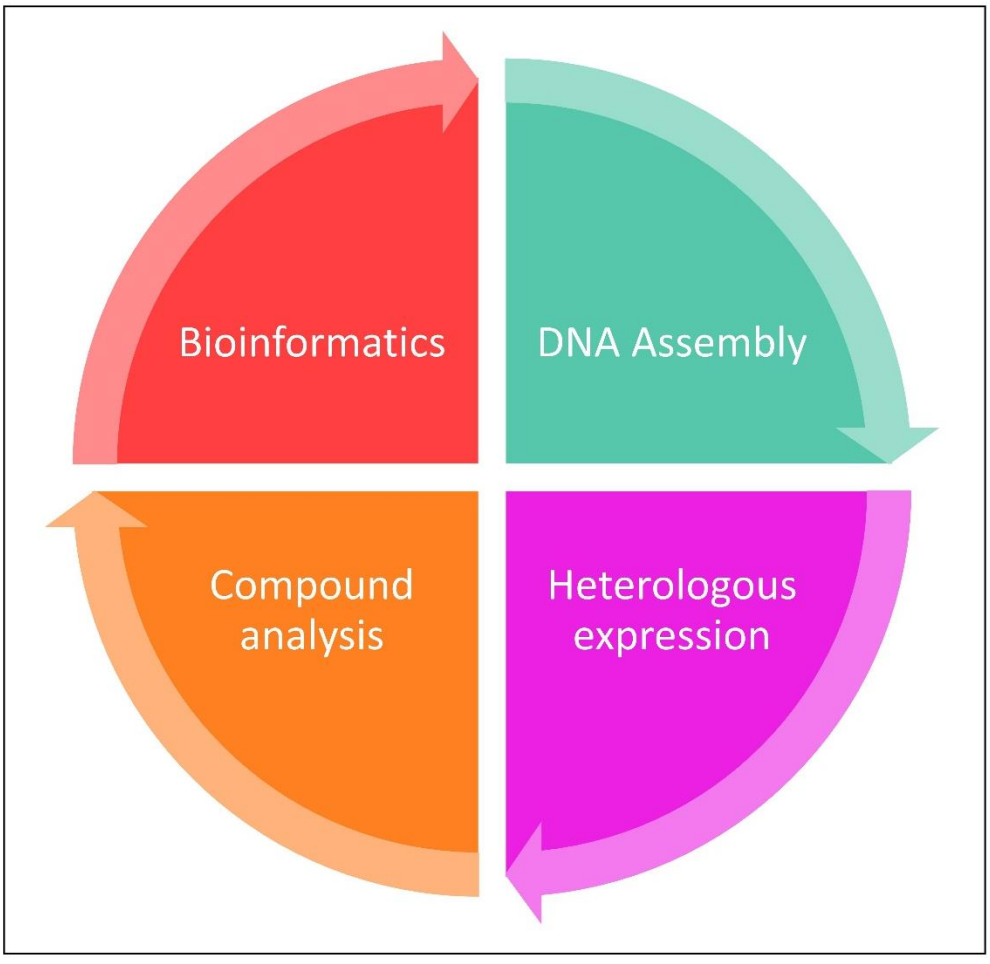


Figure 2.1. Depiction of the main tools in the gene characterisation toolbox used in this thesis.

In order to establish the gene characterisation toolbox used for this thesis, four main components were required:

1. Genes encoding information for the biosynthesis of the compound of interest (i.e. *H. pulicicidum* genes encoding information for nodulisporic acid biosynthesis)
2. A DNA assembly system to combine genetic components [i.e. the Modular Idempotent DNA Assembly System (MIDAS)]
3. A library of expression hosts for confirming gene function (i.e.the library of *P. paxilli* deletion mutants)
4. A means to evaluate compound production (i.e. extraction solvents, purification materials, TLC, LCMS and NMR spectroscopy techniques)

Each of these components are described in detail below.

Bioinformatics: Identification of the *NOD* cluster

In order to understand the molecular mechanisms of the biosynthesis of nodulisporic acids, the genomic DNA sequence of the natural nodulisporic acid producer, *Hypoxylon pulicicidum*, was required. To this end, *H. pulicicidum* (ATCC® 74245™) was obtained from the American Type Culture Collective, DNA was extracted from *H. pulicicidum* mycelia, and full genome sequencing was commissioned (New Zealand Genomics Limited).[2] By comparing the genome sequence to that of genes that have been confirmed to be involved in indole diterpene biosynthesis, a cluster of thirteen potential genes involved in nodulisporic acid biosynthesis was identified (Figure 2.2). Designation of the cluster boundaries was supported by identification of flanking genes (described in table 2.1) that have highly conserved homologues in the genome of a closely related species of *Hypoxylon* (CO27-5),[71] most of which were found at a single locus in the related genome. Notably, the closely related species (CO27-5) lacks the *NOD* gene cluster and does not produce nodulisporic acids. Details ragrding the genome assembly, annotation and cluster identification can be found in **Chapter 8. MATERIALS AND METHODS: gDNA isolation for genome sequencing and transcription unit module amplification** and **Genome sequencing, identification and annotation of the *H. pulicicidum* (ATCC® 74245™) *NOD* gene cluster and GenBank accession numbers**.

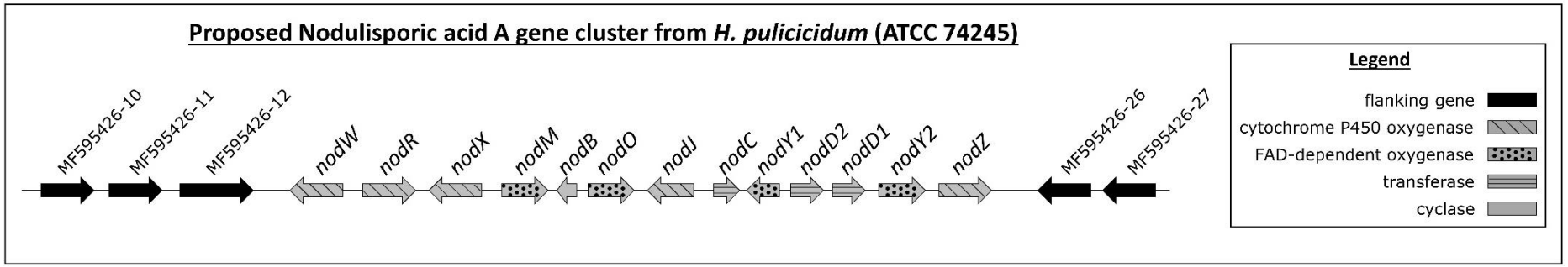


Figure 2.2. Proposed nodulisporic acid gene cluster (*NOD*) discovered in the genome of *H. pulicicidum* (ATCC 74245). Figure has been modified from Van de Bittner *et al*. 2018.[1]

Table 2.1. Functional assignment of predicted genes in and around the nodulisporic acid gene cluster. A *nod* name was assigned to each *H. pulicicidum* gene in the nodulisporic acid A (1) gene cluster [shaded green ( )]. *H. pulicicidum* genes that share homology (> 35% amino acid identity of predicted translational products) with genes found in known indole diterpene pathways are followed by an asterisk (*) and, with the exception of *nodR*, were given letters corresponding to known confirmed genes (e.g. the protein encoded by *nodC* shares 52.8% amino acid identity with the protein product of *paxC*). The genes that do not share homology with known indole diterpene genes were assigned letters that are not shared with any of the confirmed indole diterpene genes. Notably, the cluster contains two sets of paralogous genes (share >40% amino acid identity with each other), which are distinguished by numbers (i.e. *nodD1*/*nodD2* and *nodY1*/n*odY2*). Genes flanking the cluster have been included in the table and are positioned relative to how they flank the *nod* cluster. Closest matches were identified using BLASTp (protein-protein BLAST) against the non-redundant protein sequence database with 'expect threshold' set at 10 and 'word size' set at 6. The BLOSUM62 scoring matrix was applied with a gap opening penalty of 11 and a gap extension penalty of 1 with conditional compositional score matrix adjustment. Table has been modified from Van de Bittner *et al*. 2018.[1]

Gene	Size of encoded protein (aa)	Predicted function [Specific Function]	Most notable BLASTp match		
			Organism	E-value % identity/ %coverage	Protein name and accession number
MF595426-10	500	MFS transporter	*Hypoxylon* sp. CO27-5	0 88% ID/99%	OTA93954
MF595426-11	496	FAD-dependent oxygenase	*Hypoxylon* sp. CO27-5	0 84% ID/96%	OTA93953
MF595426-12	1664	WD40 domain protein	*Hypoxylon* sp. CO27-5	0 36% ID/80%	OTA80149
nodW	608	Cytochrome P450 oxygenase [terminal-C dioxygenase]	*Aspergillus aculeatus*	9.00E-153 44% ID/ 97%	XP_020058732

*nodR**	511	Cytochrome P450 oxygenase	*Penicillium simplicissimum*	3.00E-108 36% ID/ 97%	PtmU/BAU61563
nodX	593	Cytochrome P450 oxygenase	*Hypoxylon* sp. CO27-5	0 62% ID/ 70%	OTA78491
*nodM**	463	FAD-dependent oxygenase [indole diterpene mono-epoxidase]	*Aspergillus flavus*	5.00E-173 55% ID/ 93%	AtmM/Q672V4
*nodB**	243	indole diterpene cyclase [indole diterpene cyclase]	*Penicillium crustosum*	9.00E-119 68% ID/ 99%	PenB/AGZ20190
*nodO**	448	FAD-dependent oxygenase	*Penicillium janthinellum*	2.00E-160 60% ID/ 97%	JanO/AGZ20488
nodJ	514	Cytochrome P450 oxygenase	*Aspergillus clavatus*	3.00E-148 42% ID/ 99%	XP_001270361
*nodC**	326	Geranylgeranyl transferase [Geranylgeranyl transferase]	*Penicillium crustosum*	2.00E-136 66% ID/ 83%	PenC/AGZ20189
nodY1	431	FAD-dependent oxygenase	*Penicillium oxalicum*	2.00E-71 34% ID/ 99%	OxaD/AOC80388
*nodD2**	434	Prenyl transferase	*Penicillium janthinellum*	1.00E-144 48% ID/ 96%	JanD/AGZ20478
*nodD1**	431	Prenyl transferase	*Penicillium janthinellum*	1.00E-155 53% ID/ 94%	JanD/AGZ20478
nodY2	461	FAD-dependent oxygenase	*Aspergillus alliaceus*	3.00E-105 42% ID/ 98%	AspB/P0DOW1
nodZ	477	Cytochrome P450 oxygenase	*Penicillium flavigenum*	7.00E-166 48% ID/ 96%	OQE14847
MF595426-26	535	Conserved hypothetical protein [function not known]	*Hypoxylon* sp. CO27-5	9.00E-139 46% ID/ 94%	OTA93952
MF595426-27	646	Glycosyl hydrolase	*Hypoxylon* sp. CO27-5	0 89% ID/98%	OTA93951

Details of the predicted genes and their proposed function are shown in Table 2.1. Importantly, six of the genes in the cluster (*nodM*, *nodB*, *nodO*, *nodC*, *nodD2*, and *nodD1*) encode protein products with notably high amino acid sequence identities to those of known indole diterpene biosynthetic genes (e.g. >40% to homologues in the *PAX* cluster of *P. paxilli*; see Appendix, Figures 9.2 to 9.4, 9.6, 9.7) indicating the likelihood that this gene cluster was involved in indole diterpene biosynthesis. The functionalities of the protein products of these six homologous genes were predicted to include a geranylgeranyl transferase (NodC), a regioselective flavin adenine dinucleotide-dependent epoxidase (NodM), an indole diterpene cyclase (NodB), two paralogous prenyl transferases (NodD2 and NodD1), and one flavin adenine dinucleotide-dependent cyclo-oxygenase (NodO). The other seven genes encode five cytochrome P450 oxygenases (NodW, NodR, NodX, NodJ and NodZ), and a pair of paralogous flavin adenine dinucleotide-dependent oxygenases (NodY1 and NodY2; see Appendix, Table 9.1). By combining the information about predicted gene functionality with the previously hypothesised biosynthetic pathway of nodulisporic acid biosynthesis[50] twelve of the genes in the cluster were tentatively assigned to their most likely role in the proposed pathway (Figure 2.3).

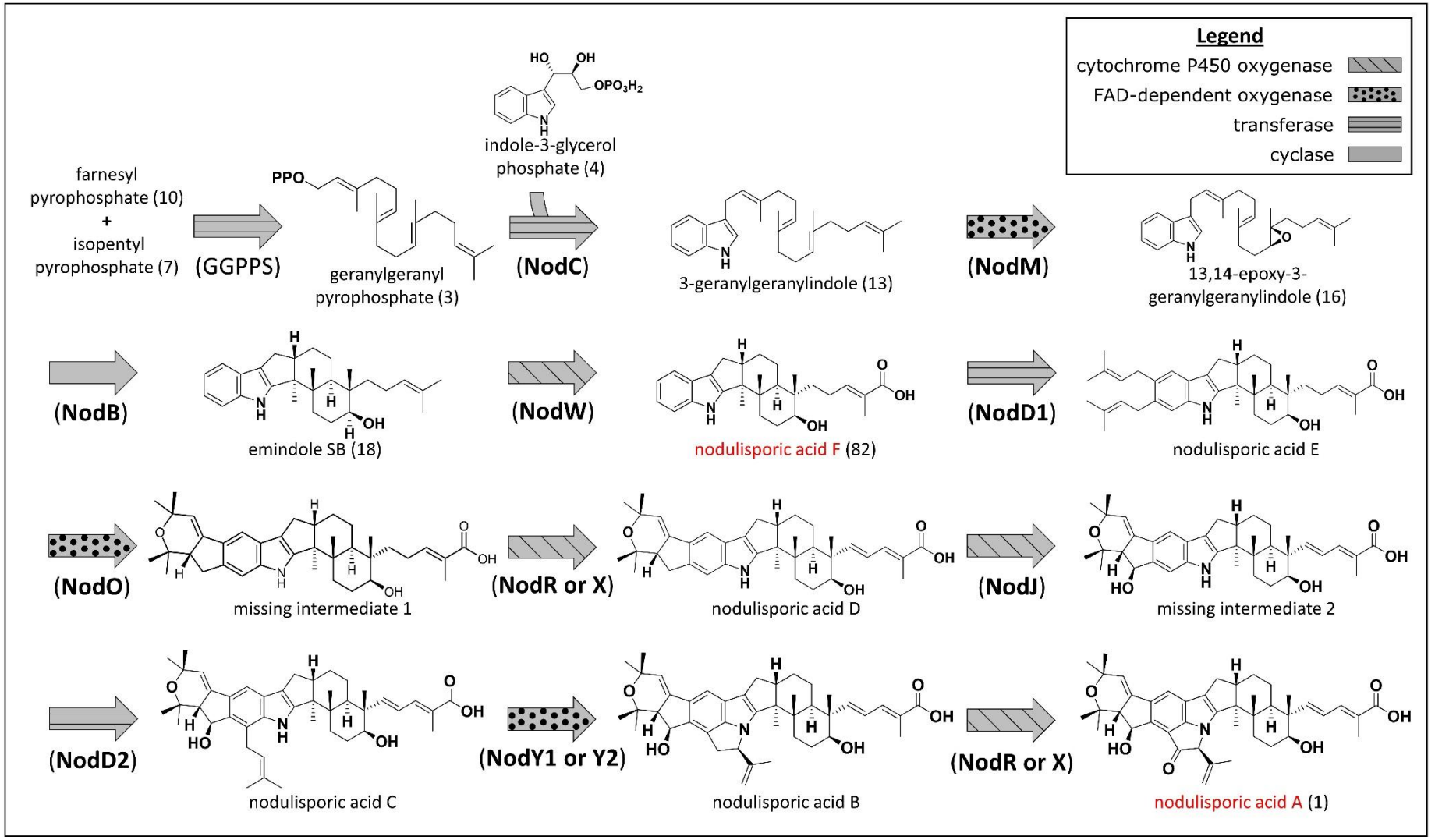
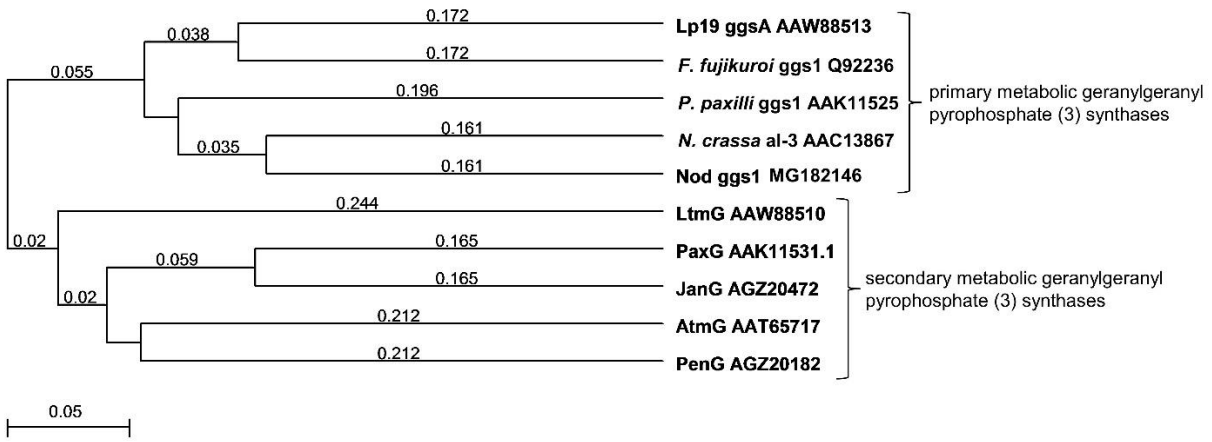


Figure 2.3. Proposed pathway for the biosynthesis of nodulisporic acids. Each arrow represents an enzymatic step and the arrow decoration correspond to enzymatic function. The protein products of gene(s) in the *NOD* cluster have been assigned to their most likely enzymatic step in the proposed pathway. Figure has been modified from Van de Bittner *et al*. 2018.[1]

Intriguingly, the *NOD* cluster lacks a secondary metabolic geranylgeranyl pyrophosphate (3) synthase-encoding gene, which has previously been assigned as the first secondary-metabolic step in indole diterpene biosynthesis.[33, 37-38, 40-41] That said, other indole diterpene gene clusters have been identified that lack a secondary metabolic geranylgeranyl pyrophosphate (3) synthase-encoding gene, like the *TER* gene cluster from *Chaunopycnis alba* (*Tolypocladium album*)[30] and the *AtS5* cluster from *Aspergillus tubingensis*,[33] respectively responsible for terpendole and emindole-analogue biosynthesis. To explore whether the *H. pulicicidum* genome contained a secondary metabolic geranylgeranyl pyrophosphate (3) synthase-encoding gene outside of the cluster, amino acid sequence comparisons of the full *H. pulicicidum* genome against the sequences of functionally confirmed primary and secondary metabolic geranylgeranyl pyrophosphate (3) synthases were completed to generate a phylogenetic tree (Figure 2.4 A) and corresponding similarity matrix (Figure 2.4 B). Searches revealed that only one other geranylgeranyl pyrophosphate (3)-encoding gene (Nod ggs1) was present and that based on its location outside of the *NOD*-cluster and the amino acid sequence homology of its protein product to primary geranylgeranyl pyrophosphate (3) synthases, Nod ggs1 was most likely a primary geranylgeranyl pyrophosphate (3) synthase.

A



B

Enzyme	Lp19 ggsA AAW88513	AtmG AAT65717	*N. crassa* al-3 AAC13867	Nod ggs1	*F. fujikuroi* ggs1 Q92236	PenG AGZ20182	PaxG AAK11531.1	*P. paxilli* ggs1 AAK11525	LtmG AAW88510	JanG AGZ20472
Lp19 ggsA AAW88513	100	36.7	47.9	48.7	58.3	33.7	33.3	42.5	33.2	33.4
AtmG AAT65717	48.7	100	40.1	47	40.5	56.9	48.9	43.8	50.7	50
N. crassa al-3 AAC13867	61.7	54.1	100	60.2	52.7	35.7	36.2	50.3	37.6	36.4
Nod ggs1	61.8	60.7	71.6	100	54	39.9	41	56.4	41.2	40.4
F. fujikuroi ggs1 Q92236	67.7	53.2	66.8	69.9	100	37.9	36.1	45.7	36	37.4
PenG AGZ20182	45.7	71.4	49	53.6	49.4	100	49.2	40.2	47.6	54.9
PaxG AAK11531.1	46	63.2	49.7	54.5	50.3	64.2	100	39.6	43	65.3
P. paxilli ggs1 AAK11525	52.7	60.7	62.8	68.7	58	54.6	57.3	100	41.7	40.8
LtmG AAW88510	46.9	65.5	51.7	55.6	50.2	62.8	59.7	59.1	100	46
JanG AGZ20472	46.7	66.1	49.9	56.3	50.8	67	78.8	56.6	61.9	100

Figure 2.4. Phylogenetic tree (**A**) and similarity matrix (**B**) of geranylgeranyl synthases ('ggs enzymes' are involved in primary metabolism and 'G' enzymes are involved in secondary metabolism). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the enzymes. Figure has been modified from Van de Bittner *et al*. 2018.[1]

DNA assembly: The Modular Idempotent DNA Assembly System (MIDAS)

In order to test the gene functionality of the *NOD* genes through heterologous expression, a method to isolate the predicted *NOD* gene coding sequences and combine them with promoters that would work efficiently in an expression host was required. To this end, a DNA assembly system, termed MIDAS for Modular Idempotent DNA Assembly System, was developed in a collaborative effort with Callaghan Innovation.[3]

MIDAS is a powerful synthetic biology platform with higher efficiency, fidelity and modularity than similar assembling systems. It allows for the seamless, efficient and flexible multigene assembly of tailored combinations of transcription unit modules (e.g. promoters, coding sequences, and terminators) from one or more biosynthetic pathways. MIDAS is composed of three Levels (Figure 2.5) that enable full user-control over the components contained within individual transcription units as well as the direction, orientation and position of transcription units within multigene constructs. MIDAS separates itself from other MoClo systems, in that at the third level of the system, components can be added iteravely into a single construct indefinitely (in theory), rather than having to move into additional levels to make longer, more complex constructs.[3]

In Level-1, transcription unit modules (e.g. promoters, coding sequences and terminators) are cloned into the Level-1 destination vector (pML1) to build defined, fully reusable, sequence-verified module libraries. In Level-2, compatible sets of Level-1 modules are released from pML1 and assembled into full length transcription units in a Level-2 destination vector (pML2) to create a transcription unit library. In Level-3, Level-2 transcription units are sequentially assembled into a Level-3 vector (pML3) to form functional multigene constructs. The complete Level-3 multigene constructs are then transformed into an expression host to test gene functionality.

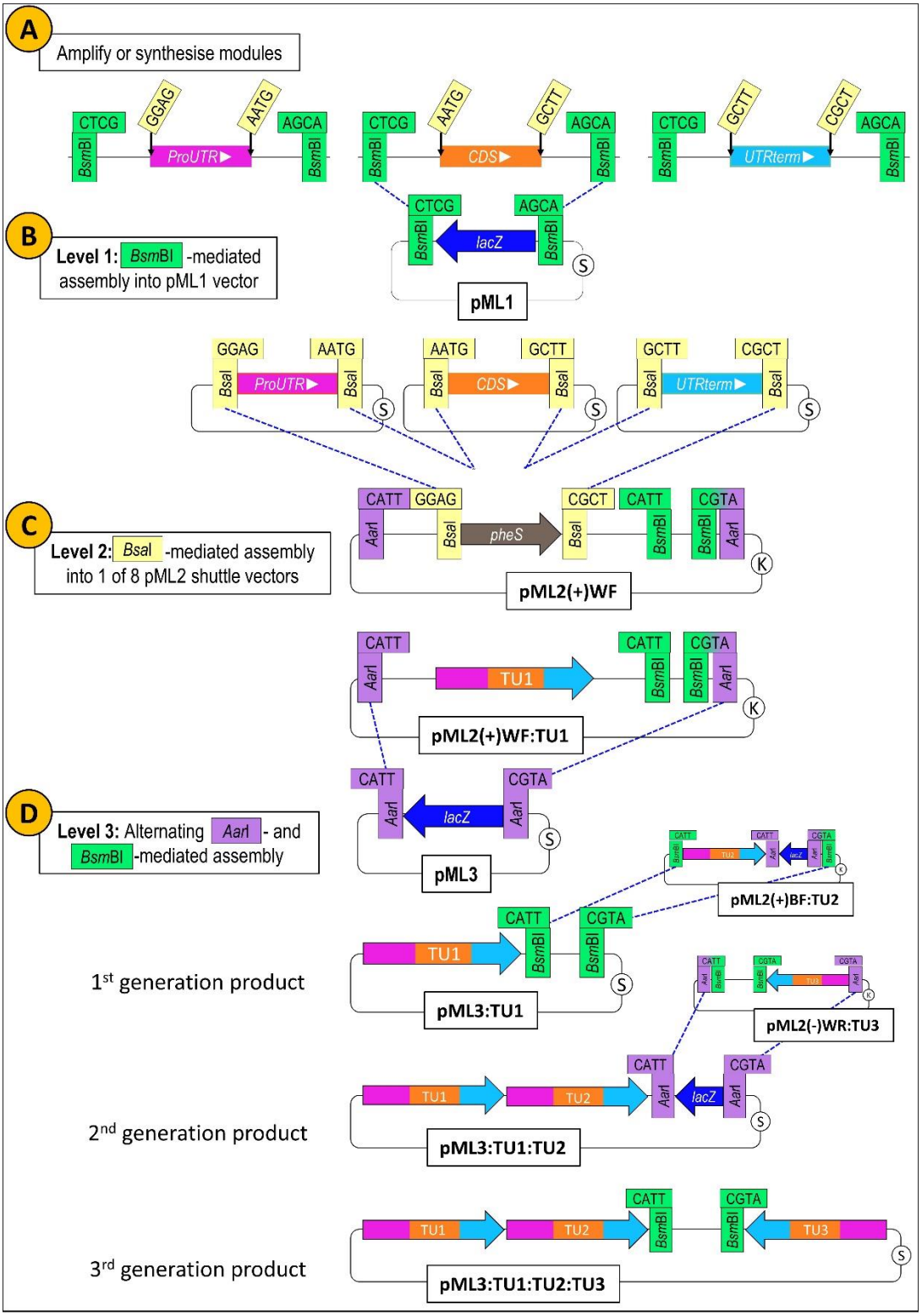


Figure 2.5. Overview of MIDAS. First (**A**), modules such as promoters (ProUTRs), coding sequences (CDSs) and terminators (UTRterms) are amplified or synthesized. Second (**B**), modules are cloned, via *Bsm*BI-mediated restriction-ligation reactions, into the pML1 source vector. Third (**C**), cloned modules are assembled, via *Bsa*I-mediated restriction-ligation reactions, into one of eight pML2 vectors to form full length transcription units (TUs). Fourth (**D**), full length TUs are iteratively assembled, via alternating *Aar*I- and *Bsm*BI-mediated restriction-ligation reactions, into Level-3 vectors to form multigene constructs. Additional pML2 vectors are shown in Figure 8.4.

Assembly at each Level of the MIDAS platform is performed in a one-pot Golden-Gate-type assembly reaction that is mediated by a type IIS restriction enzyme (*Bsm*BI, *Bsa*I or *Aar*I). Each of these enzymes has a unique binding sequence and creates a defined 4-nucleotide overhang (see Figure 2.5) that is used to ensure correct assembly of the transcription unit modules into pML1, transcription unit assembly into pML2, and multigene assemble in pML3. As part of the MIDAS platform, all type IIS restriction sites within original transcription unit module sequences are removed during primer design in a process called domestication. In Chapter 3, a full explanation of each level of the MIDAS platform is provided. For additional details of the MIDAS platform including the success rate of each step of MIDAS cloning, how this was affected by the number of fragments being assembled in a single step, etc. please refer to the van Dolleweerd *et al.* 2018 MIDAS manuscript and its supporting information.[3]

Heterologous expression

Heterologous expression is a powerful bioengineering tool to, among other things, identify the function of genes and reconstruct biosynthetic pathways. It is often used to work towards industrially relevant titres of target compound and especially useful when working with genes that are poorly expressed or highly regulated in their native host, because it imparts the ability to alter the gene expression levels (e.g. by promoter swapping or increasing copy number) and to bypass natural regulatory mechanisms (e.g. specific light, temperature or nutrient source).

At the basis of a heterologous expression system, genes from one organism (e.g. the natural producer of nodulisporic acids, *Hypoxylon pulicicidum*) are transferred to another organism (e.g. *Penicillium paxilli*) that serves as the heterologous expression host and often has some sort of beneficial role (e.g. grows faster and produces more compound).

In the case of this project, heterologous expression was used to confirm gene function of genes involved in nodulisporic acid biosynthesis because the native nodulisporic acid producer, *H. pulicicidum*, produces very low titres of nodulisporic acids and the biosynthetic pathway appears to be highly regulated (i.e. to elicit nodulisporic acid production *H. pulicicidum* must be grown in complete darkness in nutrient rich medium for three weeks).[44, 50] Therefore, performing gene knockout studies on *H. pulicicidum* was not an efficient method to confirm gene function because it would involve long, complicated growth procedures and, due to low compound

production, likely lead to discrepancies around assigning functionality (i.e. a gene knockout may produce compounds at levels below the limit of detection that could lead to false conclusions).

A library of expression hosts for confirming gene function

There are a number of things to consider when making a choice about an expression host for confirming gene function through heterologous expression, and for each type of expression system the ideal host will likely change. In most cases though, an ideal expression host will grow quickly on relatively nutrient poor medium (to reduce time and supply costs), lack undesirable regulatory elements for the target pathway, encode the metabolic machinery required to produce the target compound(s) of interest, produce high titres of target compound(s), and be safe and easy to work with (e.g. does not produce deadly toxins, is susceptible to antibiotics for transformant screening, has morphology that enables easy separation from fermentation broths, etc.).

With these selection criteria in mind, it was determined that the filamentous fungi, *Penicillium paxilli*, which has been extensively used to confirm the function of the genes involved in the biosynthesis of various indole diterpenes,[25-29, 35-36, 38] would be an excellent expression host for nodulisporic acids. *P. paxilli* naturally produces a class of indole diterpenes known as paxillines, the biosynthesis of which has been fully elucidated (Figure 2.6, A).[38] In total, the gene cluster for the biosynthesis of paxillines contains nine genes, of which seven have been functionally analysed and are involved in the biosynthesis of prenylated paxillines (Figure 2.6, B).[35-38, 40, 42] These seven genes encode a geranylgeranyl pyrophosphate (3) synthase (PaxG), a geranylgeranyl transferase (PaxC), a regioselective flavin adenine dinucleotide-dependent epoxidase (PaxM), an indole diterpene cyclase (PaxB), two cytochrome P450 oxygenases (PaxP and PaxQ), and a prenyl transferases (PaxD). The two other genes encode a gene with unknown function (PaxA) and one flavin adenine dinucleotide-dependent cyclo-oxygenase (PaxO).

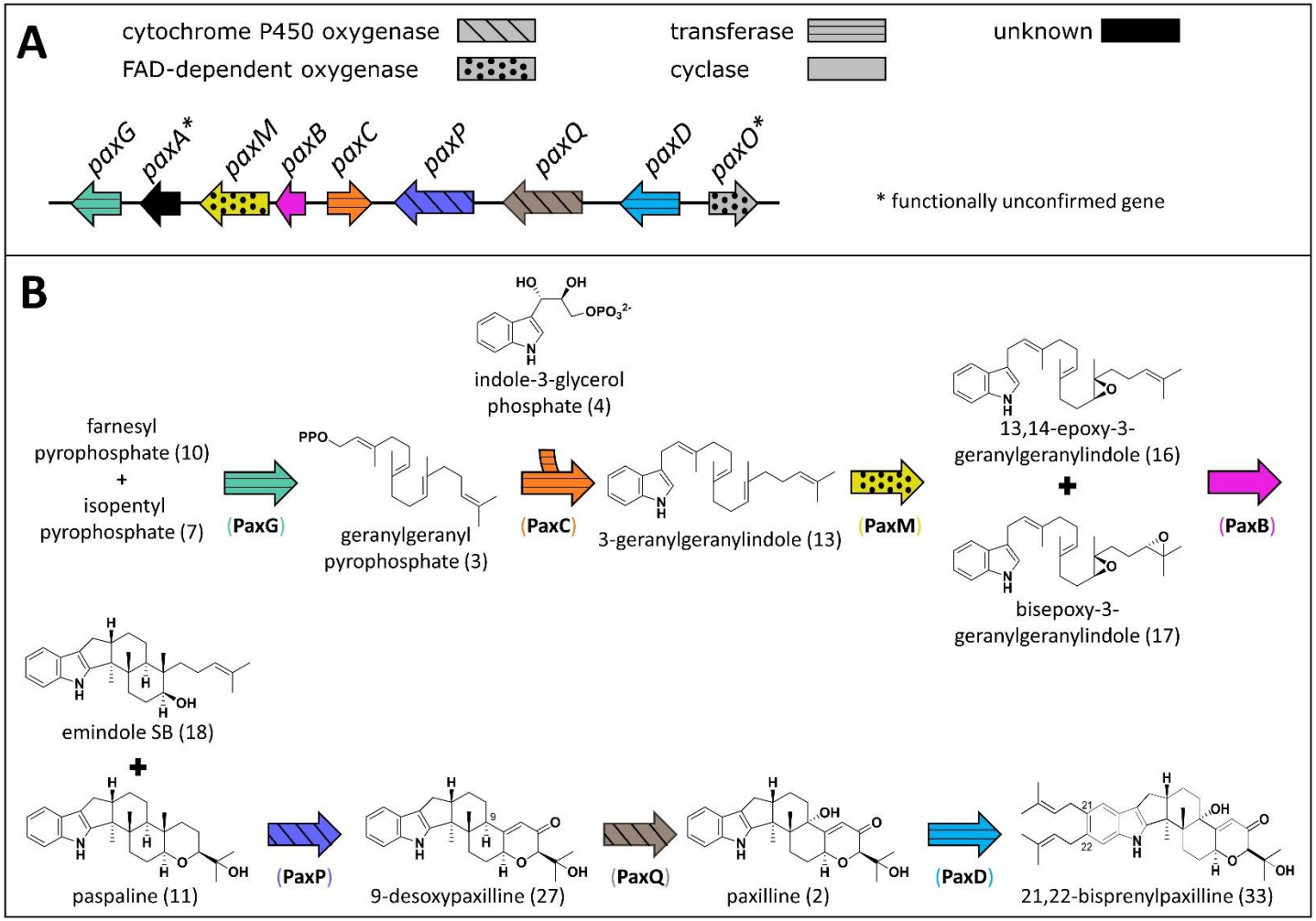


Figure 2.6. Depiction of the identified paxilline (2) gene cluster from *Penicillium paxilli* (**A**) and corresponding paxilline (2) biosynthetic pathway (**B**). The gene cluster contains nine genes, of which six (*paxG,M,B,C,P,Q*) are required for paxilline (2) biosynthesis.[40] Of the other three genes, (*paxA*, *paxD* and *paxO*) only *paxD* has been functionally analysed for its role in converting paxilline (2) to 21-prenylpaxilline (32) and 21,22-bisprenylpaxilline.[42] Arrows represent individual genes and arrow decorations represent gene function. Figure has been modified from Van de Bittner *et al*. 2018.[1]

During the identification and functional analysis of the genes in the *PAX* cluster a series of single and multigene-deletion mutants[24, 35, 37-38] were generated that are unable to synthesise the indole diterpenes normally produced by wild-type *P. paxilli* (strain PN2013[72]). The selective deletion of a gene within the paxilline (2) biosynthetic pathways prevents paxilline (2) biosynthesis at a specific biosynthetic step. For example, a Δ*paxC* deletion mutant is not able to produce any of the metabolites downstream of the PaxC-catalysed step (i.e. condensation of indole-3-glycerolphosphate and geranylgeranyl pyrophosphate (3) to form gernaylgeranylindole). These single gene deletion mutants are excellent platforms for testing gene orthology – i.e. the ability to see if a homologous gene from a different indole diterpene producer has the same function as that of the gene that has been deleted in the *P. paxilli* deletion mutant. For instance, if a *paxC* homologue (e.g. *nodC*) were heterologously expressed in the Δ*paxC P. paxilli* deletion mutant and paxilline (2) were observed in the transformant extract, then that would indirectly indicate that the *paxC* homologue was orthologous to *paxC* because it restored paxilline (2) biosynthesis. In contrast to the single-gene deletion mutants, multigene-deletion mutants are ideal for ensuring that the paxilline (2) biosynthesis genes do not have an influence on heterologously expressed genes and thereby serve as preferred hosts to examine how multiple genes from a single multigene construct interact with one another.[1] Overall, the *P. paxilli* deletion mutant library provides a useful resource for confirming gene function and individually testing whether MIDAS-assembled transcription units contain all the necessary *cis*-acting elements to successfully transcribe and express their coding sequences.

Compound analysis

In order to link genes to biosynthetic steps, a method had to be developed to produce, extract and analyse the compounds of interest. In the case of the heterologous expression system used in this thesis, after genes are introduced into *P. paxilli*, the transformants are grown in production medium for a week and indole diterpenes are extracted with an organic solvent. The extracts are initially screened using TLC to identify transformants that produce compounds not present in the respective *P. paxilli* deletion mutant. Extracts containing novel compounds are then subjected to LCMS analysis to identify compounds by their mass/charge ratio and

fragmentation patterns. If a novel compound is produced for which no standard is available, then the transformant that produced the unknown product is grown on larger scales and the compound of interest is extracted and purified by semi-preparative HPLC for analysis by NMR spectroscopy.

Conclusion

As a whole, these four components: the *NOD* cluster, MIDAS, the *P. paxilli* expression host library and the techniques to analyse indole diterpenes give rise to the gene characterisation toolbox used in this thesis that provided a means to confirm gene functionality and partially reconstruct the nodulisporic acid biosynthetic pathway. The next four chapters exemplify how each of these techniques has been used to confirm the function of individual genes within the *NOD* cluster – thereby demonstrating that this toolbox provides a powerful platform to reconstruct biosynthetic pathways of secondary metabolites.

3. FUNCTIONAL CHARACTERISATION OF NodC

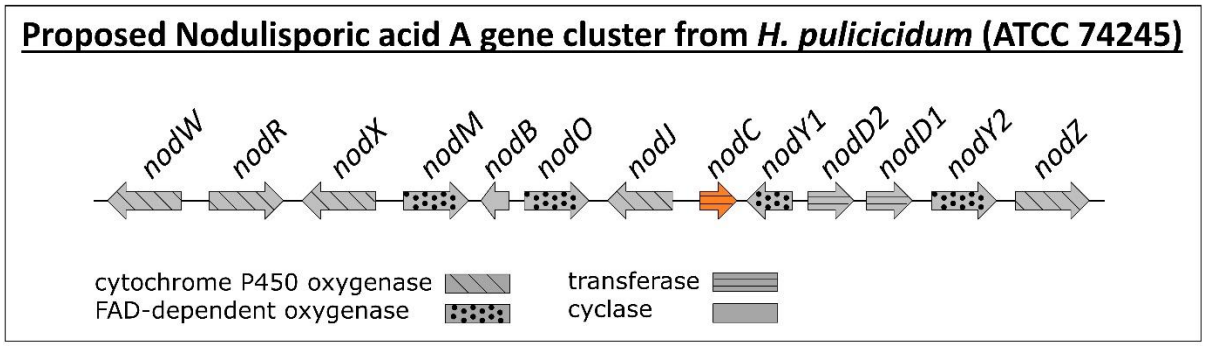


Figure 3.1. Depiction of the predicted nodulisporic acid gene cluster (*NOD*) from *H. pulicicidum* (ATCC® 74245™). The *nodC* gene is highlighted in orange.

As shown in Figure 3.1, the *NOD* gene cluster does not contain a secondary-metabolic geranylgeranyl pyrophosphate (3) synthase-encoding gene, which has been shown to encode the enzyme responsible for catalysing the first secondary-metabolic step in indole diterpene biosynthesis, the condensation of farnesyl pyrophosphate (10) and isopentyl pyrophosphate (7) to form geranylgeranyl pyrophosphate (3) shown in Figure 3.2. Therefore, attention was turned to the second secondary-metabolic step in indole diterpene biosynthesis, the production of geranylgeranylindole from the condensation of geranylgeranyl pyrophosphate (3) and indole-3-glycerol phosphate (4) shown in Figure 3.2.

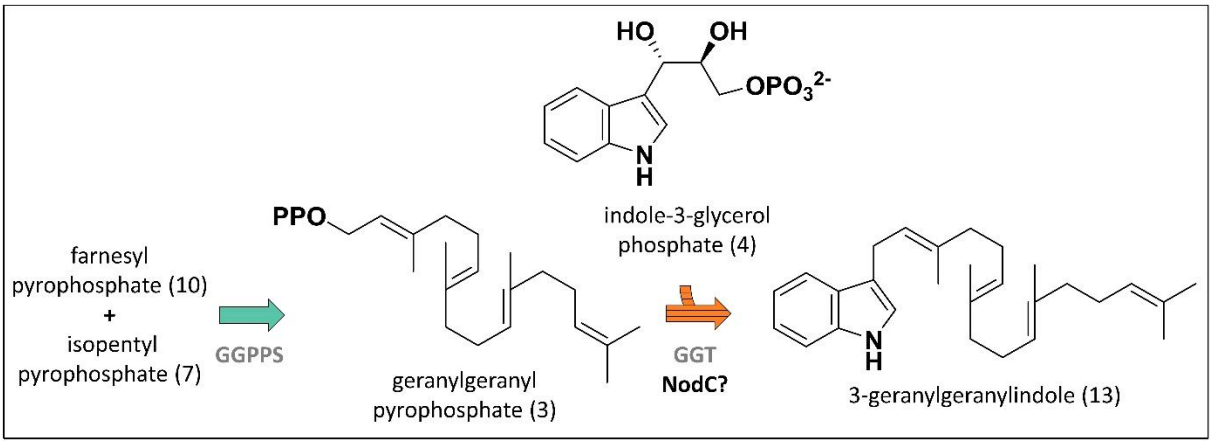


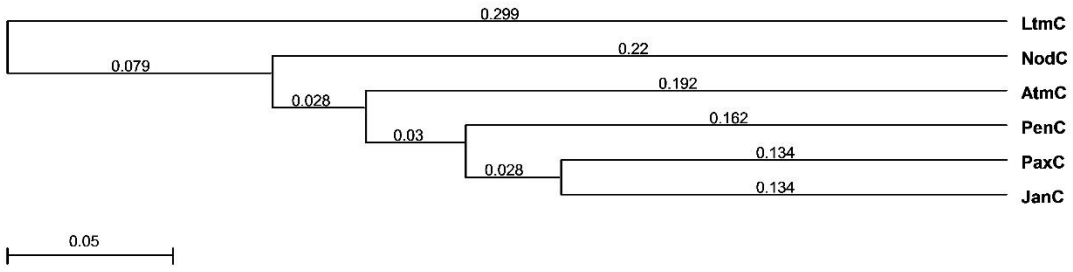
Figure 3.2. Schematic of the first two secondary-metabolic steps in indole diterpene biosynthesis. In step 1 a geranylgeranyl pyrophosphate (3) synthase (GGPPS, turquoise arrow) catalyses the condenation of farnesyl pyrophosphate (10) and isopentyl pyrophosphate to form geranylgeranyl pyrophosphate (3). In step 2, a geranylgeranyl transferase (GGT, orange arrow) catalyses the condensation of geranylgeranyl pyrophosphate (3) and indole-3-glycerol phosphate (4) to produce 3-geranylgeranylindole (13). Notably the *NOD* cluster isolated from *H. pulicicidum* does not contain a GGPPS-encoding gene but it does contain a gene candidate (*nodC*) that is predicted to encode the geranylgeranyl transferase necessary for the biosynthesis of 3-geranylgeranylindole (13). The likely cofactor for GGT enzymes is Mg^{2+} but not all enzymatic mechanisms have been thoroughly examined.

Prediction of *nodC*

It has been well established that production of geranylgeranyl indole is catalysed by a class of indole diterpene geranylgeranyl transferase enzymes that conventionally go by the name “C” (e.g. the PaxC in paxilline (2) biosynthesis).[33, 40] Using sequence homology comparisons of the amino acid sequences from functionally confirmed indole

diterpene geranylgeranyl transferases to the predicted gene products within the *NOD* cluster, a likely geranylgeranyl transferase encoding-gene named "*nodC"* was identified (Figure 3.3, A and B). NodC shares greater than 50% amino acid identity to four other known geranylgeranyl transferases (PaxC, PenC, AtmC, and JanC), the highest being JanC with 55.4% amino acid identity.

A



B

Enzyme	PaxC	NodC	LtmC	AtmC	JanC	PenC
PaxC	100	52.3	35.3	54.3	70.9	60.5
NodC	66.7	100	38.8	55.1	55.4	54.4
LtmC	54	61.7	100	40.2	36.4	39.5
AtmC	69.9	70.5	61.5	100	58.3	63.6
JanC	79.5	70.3	55.5	74.1	100	66.7
PenC	73.4	70.5	59.1	78	79.2	100

Figure 3.3. Phylogenetic tree (**A**) and similarity matrix (**B**) of geranylgeranyl transferases ('C' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the 'C' enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the 'C' enzymes.

Functional analysis of NodC

To assess whether NodC was a functional orthologue of the indole diterpene geranylgeranyl transferases, and establish directly the role of *nodC* in nodulisporic acid A (1) biosynthesis, a series of MIDAS Level-3 plasmids was constructed, which were then introduced into the *P. paxilli* Δ*paxC* deletion mutant strain PN2290. As mentioned in Chapter 2, MIDAS is composed of a three-tiered system (Figure 3.4), termed MIDAS Level-1, Level-2, and Level-3, which employs the use of unique MIDAS destination

vectors for the assembly of transcription unit modules (promoters, coding sequences, and terminators) in Level-1, full length transcription units in Level-2, and multigene assemblies in Level-3.[3]

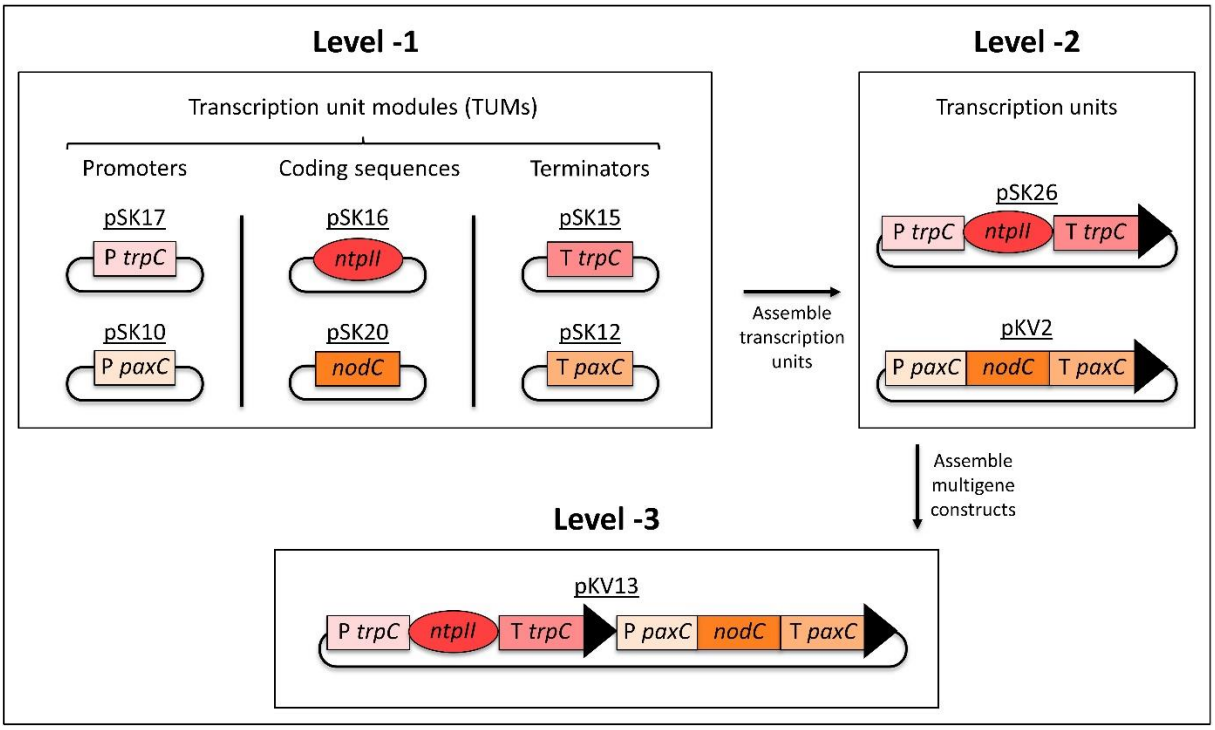


Figure 3.4. Overview of the MIDAS platform for plasmid assembly of pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$). Note: A prefix of P or T respectively represent a promoter or terminator region.

Assembly of the first MIDAS Level-3 *nodC*$_{CDS}$-containing plasmid [pKV13, (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$)] is described in detail to provide a thorough introduction to the MIDAS platform that was used throughout this project to assemble plasmids for heterologous expression studies of genes within the *NOD* cluster. The *nodC*$_{CDS}$-containing plasmid (pKV13) was designed to harbour a geneticin resistance cassette, used to screen transformed colonies, and a *nodC* transcription unit, used to assess the role of *nodC* in nodulisporic acid A (1) biosynthesis. In accordance with the MIDAS vector assembly platform, six Level-1 MIDAS constructs, three for the geneticin resistance cassette [pSK17 (*trpC*$_{ProUTR}$), pSK16 (*ntpII*$_{CDS}$), pSK15 (*trpC*$_{UTRterm}$)] and three for the *nodC* transcription unit [pSK10 (*paxC*$_{ProUTR}$), pSK20 (*nodC*$_{CDS}$), pSK12 (*paxC*$_{UTRterm}$)], were prepared and used to assemble the two MIDAS Level-2 constructs, pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$)

and pKV2 (*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) respectively. An in-depth description of each Level of the MIDAS platform is presented throughout this chapter.

Preparation of MIDAS Level-1 plasmids

Two key steps are involved in the preparation of MIDAS Level-1 plasmids: (1) polymerase chain reaction (PCR) primers are designed and used to generate PCR products, and (2) PCR products are assembled into a Level-1 destination vector and propagated in bacteria. In this project, all PCR products were generated from the genomic DNA (gDNA) of specific fungal species in a fungal library. They were then inserted into the MIDAS Level 1 destination vector (pML1) via a one-pot *Bsm*BI-mediated reaction and cloned in *E. coli* (Stellar). Alternatively, domesticated synthetic modules could be ordered from a DNA synthesis company and directly assembled into the Level-1 destination vector (pML1).

Generation of PCR products

The foundation of the MIDAS Level-1 plasmids relies on incorporation of MIDAS-specific sequences that flank each end of the transcription unit modules (i.e. promoters, coding sequences, and terminators), which can be achieved by appropriately designing PCR primers. Flanking primers are designed to amplify all of the transcription unit modules with two notable characteristics; the first is the *Bsm*BI enzyme-specific recognition sites and corresponding sticky ends that are essential to the correct assembly of PCR fragments into the MIDAS Level-1 destination vector, and the second is the *Bsa*I-mediated sticky ends that encode the '4-nucleotide address system' essential for proper transcription unit module assembly into MIDAS Level-2 destination vectors to form complete transcription units. In addition to the generic transcription unit module flanking primers, additional primers, known as domestication primers, must be designed to remove any transcription unit module internal recognition sites for *Aar*I, *Bsa*I and *Bsm*BI. Transcription unit modules that require domestication are therefore amplified in separate fragments that are reassembled into full transcription unit modules in MIDAS Level-1. For example, the *nodC* coding sequence contains an internal *Bsm*BI restriction site and thus had to be amplified as two fragments (Figure 3.6, B) in order to eliminate the *Bsm*BI site. Upon PCR amplification

of each transcription unit module or fragment thereof, PCR products of expected size were spin-column purified and assessed by agarose gel electrophoresis as shown in Figure 3.5, A.

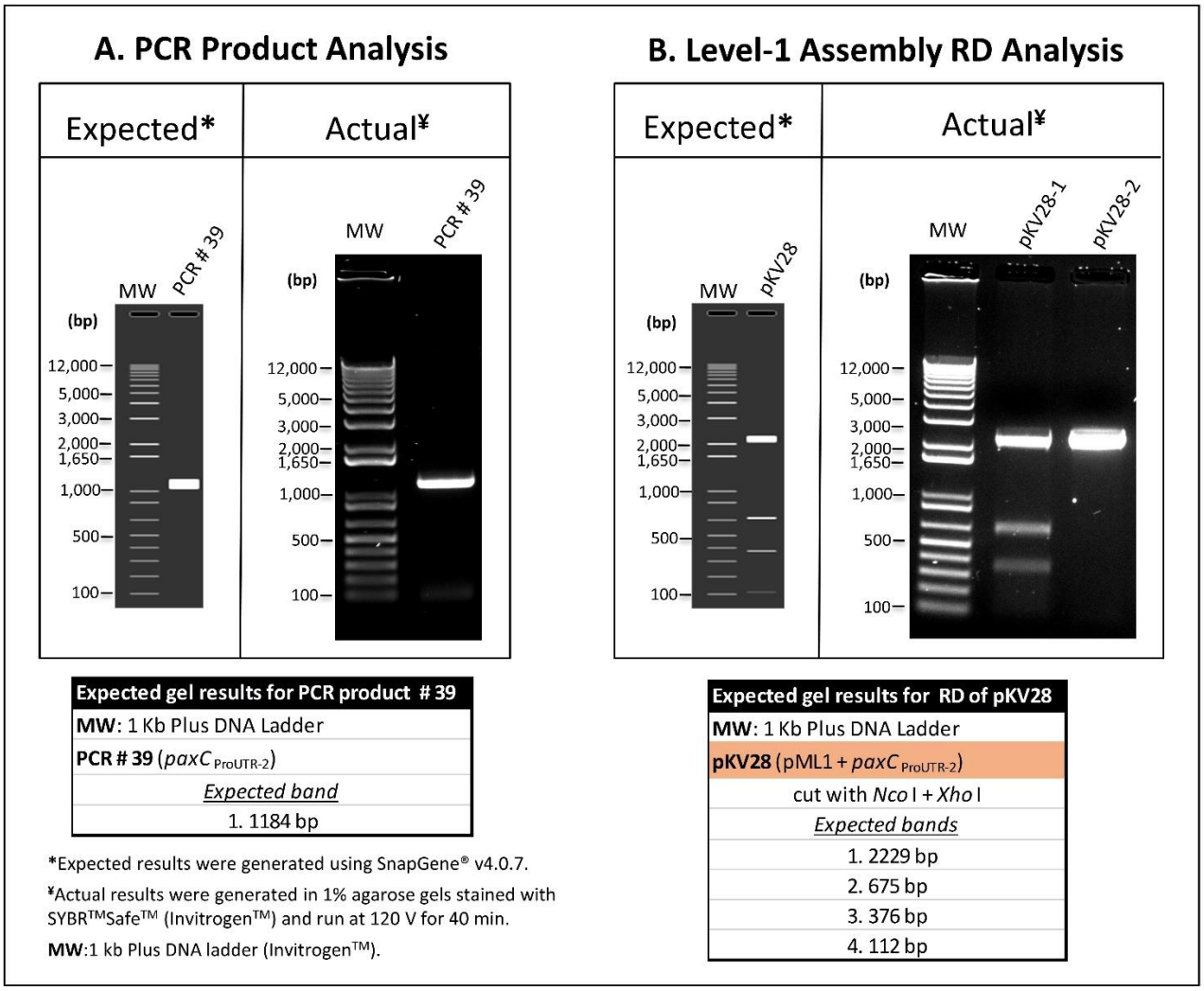


Figure 3.5. An example of the expected and actual gel results of a PCR product (PCR # 39) used to assemble a promoter transcription unit module (*paxC*$_{ProUTR-2}$) in the pML1 MIDAS Level-1 destination vector (**A**) and an example of the expected and actual gel results of the restriction digest of two MIDAS Level-1 plasmids, pKV28-1 and pKV28-2, assembled from PCR #39 (**B**).

Assembly of Level-1 plasmids

PCR products that matched their expected length were assembled into transcription unit modules (i.e. promoters, coding sequences and terminators) in the MIDAS Level-1 plasmid, pML1, by *Bsm*BI-mediated Golden Gate assembly (as illustrated in Figure 3.6). Plasmids were then cloned in *E. coli*. Successful cloning via *Bsm*BI-mediated

Golden Gate assembly resulted in disruption of the *lac*Z gene (note absence of *lac*Z in Figure 3.6 plasmid products) so only white colonies were selected for analysis by restriction digest. As an example, results from the cloning of pKV28 are shown in Figure 3.5, B, where only one of the three clones picked (pKV28-1) was positive by restriction analysis. To ensure that the Level-1 transcription unit modules were free of errors introduced by PCR and that the *Bsm*BI assembly junctions were correct, typically one plasmid with the correct restriction profile (in this case pKV28-1) was sequenced. Once Level-1 transcription unit modules were sequence-verified they were used to assemble the transcription units in MIDAS Level-2.

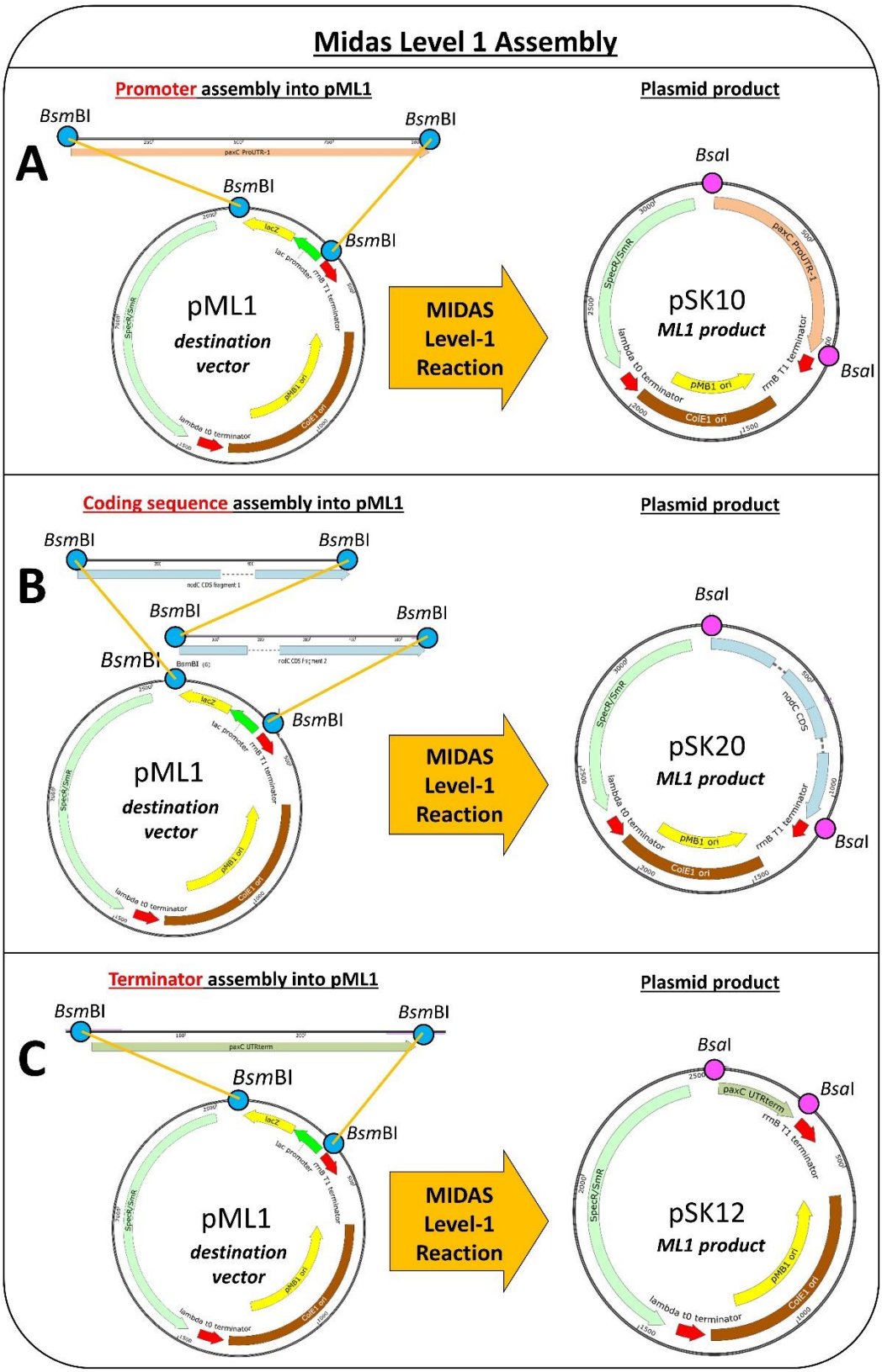


Figure 3.6. Depiction of MIDAS Level-1 assembly of the *paxC*$_{ProUTR}$ transcription unit module, pSK10 (**A**), the *nodC*$_{CDS}$ transcription unit module, pSK20 (**B**), and the *paxC*$_{UTRterm}$ transcription unit module, pSK12 (**C**). All plasmids were assembled using a *Bsm*BI-mediated Golden Gate reaction into the pML1 destination vector to form the plasmid products.

Preparation of MIDAS Level-2 plasmids

MIDAS Level-2 assembly provides the backbone of MIDAS, which implements the unique modularity of transcription unit direction and position within the final MIDAS Level-3 vector. Eight MIDAS Level-2 destination vectors [pML2(+)BF, pML2(-)BF, pML2(+)WF, pML2(-)WF, pML2(+)BR, pML2(-)BR, pML2(+)WR and pML2(-)WR); Figure 8.4] are available for assembly of the MIDAS Level-1 transcription unit modules, but due to the simplicity of testing the function of a single gene, *nodC* as discussed in this chapter, only two of the Level-2 destination vectors (pML2(+)WF and pML2(+)BF) were used to assemble the two Level-2 transcription units [pSK26 ($trpC_{ProUTR}$-***ntpII***$_{CDS}$-$trpC_{UTRterm}$+WF) and pKV2 ($paxC_{ProUTR}$-***nodC***$_{CDS}$-$paxC_{UTRterm}$+BF) respectively] required for the stepwise MIDAS Level-3 assembly of pKV22 ($trpC_{ProUTR}$-***ntpII***$_{CDS}$-$trpC_{UTRterm}$) and pKV13 ($trpC_{ProUTR}$-***ntpII***$_{CDS}$-$trpC_{UTRterm}$_$paxC_{ProUTR}$-***nodC***$_{CDS}$-$paxC_{UTRterm}$).

Each MIDAS Level-2 plasmid [pSK26 ($trpC_{ProUTR}$-***ntpII***$_{CDS}$-$trpC_{UTRterm}$+WF) and pKV2 ($paxC_{ProUTR}$-***nodC***$_{CDS}$-$paxC_{UTRterm}$+BF)] was assembled in a 1-pot *Bsa*I-mediated Golden Gate-type assembly reaction containing three MIDAS Level-1 plasmids and one Level-2 destination vector. For example, Figure 3.7 shows the MIDAS Level-2 assembly of pKV2 ($paxC_{ProUTR}$-***nodC***$_{CDS}$-$paxC_{UTRterm}$+BF) from three MIDAS Level-1 plasmids, pSK10 ($paxC_{ProUTR}$), pSK20 ($nodC_{CDS}$) and pSK12 ($paxC_{UTRterm}$), and one MIDAS Level-2 destination vector, pML2+BF. Correct assembly of the Level-1 transcription unit modules into the full length Level-2 transcription unit is dependent upon the sticky ends accounted for during primer design known as the MIDAS Level-2 '4-nucleotide address system' (Figure 3.7, red and blue sequences). Upon assembly, Level-2 plasmids were transformed into competent *E. coli* cells and typically three colonies were picked for propagation and plasmid purification. Purified plasmids were subjected to restriction analysis and, for these initial plasmids, sequence analysis. It became apparent that the sequence verification of Level-2 plasmids was unnecessary and this step was redundant as the Level-1 sequence verification already ensured that no errors were present within sequences of the transcription unit modules. Therefore, restriction analysis became the preferred method to verify proper assembly of the Level-2 transcription units prior to MIDAS Level-3 assembly.

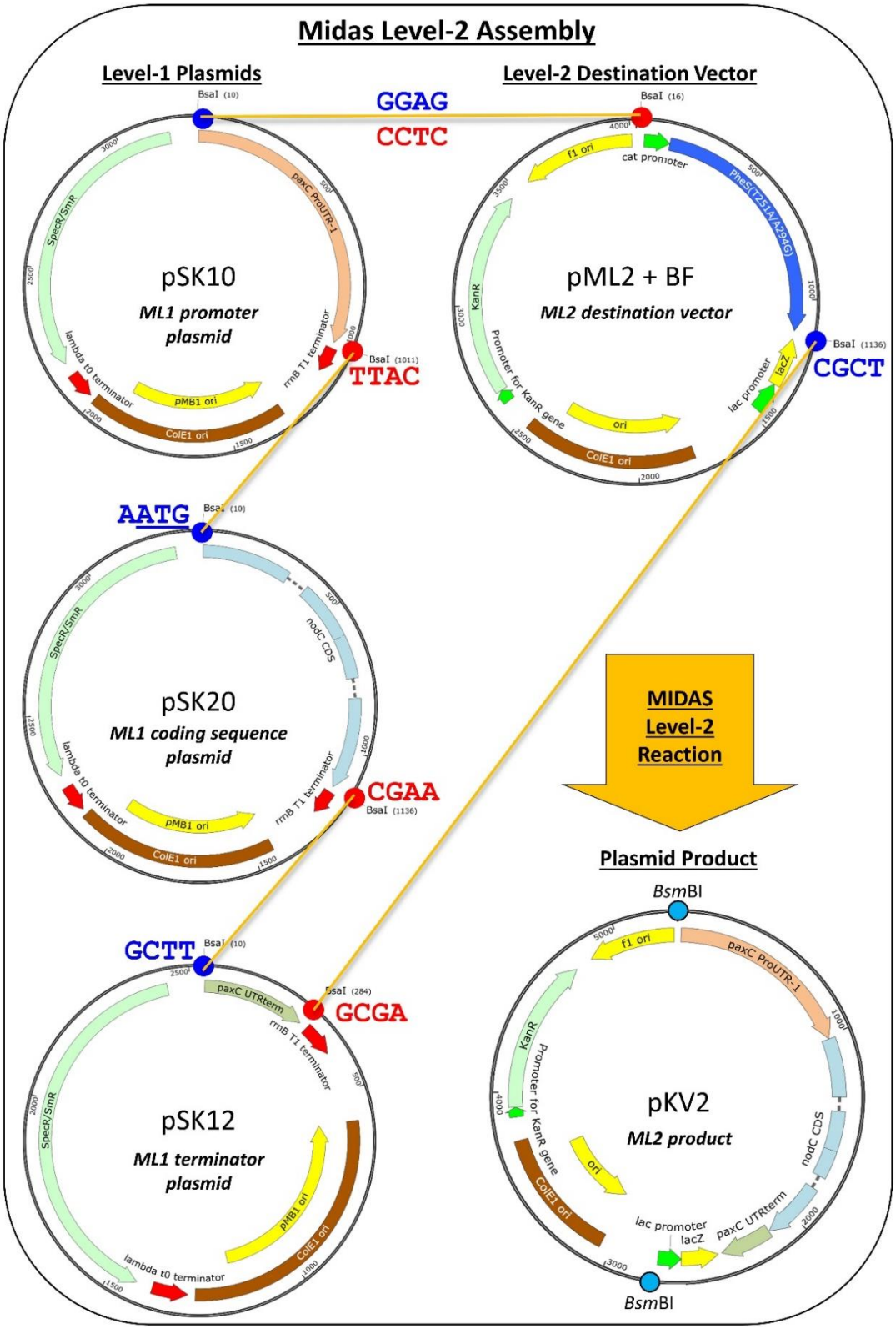


Figure 3.7. Depiction of MIDAS Level 2 Assembly using three MIDAS Level-1 plasmids [pSK10 (*paxC*$_{ProUTR}$), pSK20 (*nodC*$_{CDS}$) and pSK12 (*paxC*$_{UTRterm}$)] and one MIDAS Level-2 destination vector (pML2+BF) to create a MIDAS Level-2 plasmid product [pKV2 (*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$+BF)]. The reaction was carried out in a single tube and the sticky ends created upon *Bsa*I digestion are depicted in blue (leading DNA strand, 5' to 3') and red (complementary DNA strand, 3' to 5'). The *nodC*$_{CDS}$ ATG start site is underlined in the leading strand overhang of pSK20 (*nodC*$_{CDS}$).

The white Level-2 destination vectors impart *Aar*I recognition sites that flank the inserted transcription unit [e.g. Figure 3.8 pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$+WF)], whereas the blue Level-2 destination vectors impart *BsmB*I recognition sites that flank the inserted transcription unit [e.g. Figure 3.7 and Figure 3.8, pKV2 (*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$+BF)]. The *Aar*I and *Bsm*BI recognition sites are used in the stepwise assembly of the Level-3 plasmids, where the first Level-3 assembly reaction into pML3.1 is always mediated by *Aar*I and the second is always mediated by *Bsm*BI as shown respectively in A and B of Figure 3.8 (Note: an additional Level-3 destination vector could be synthesised to be mediated first by a *Bsm*BI reaction, but this was not required for this work). Therefore, the first transcription unit that is placed into the Level-3 destination vector (pML3.1) must be inserted into a white (W) Level-2 destination vector. In the case of this project, introducing geneticin resistance (encoded by the *nptII*$_{CDS}$) into all of the *Penicillium paxilli* knockout strains intended for use in heterologous expression experiments, with the exception of strain PN2458 (Δ*paxB*) (see Table 8.5 in Materials and Methods), would allow easy screening of transformed colonies that contained desired gene(s) of interest. To this end, the first Level-3 assembly reaction was designed to create a Level-3 destination plasmid [pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$) shown in Figure 3.8, A] that contained a *ntpII*$_{CDS}$ transcription unit for geneticin resistance; and thus the *ntpII*$_{CDS}$ cassette was assembled into a white Level-2 destination vector (pML2(+)WF) to create the Level-2 plasmid product pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$+WF).

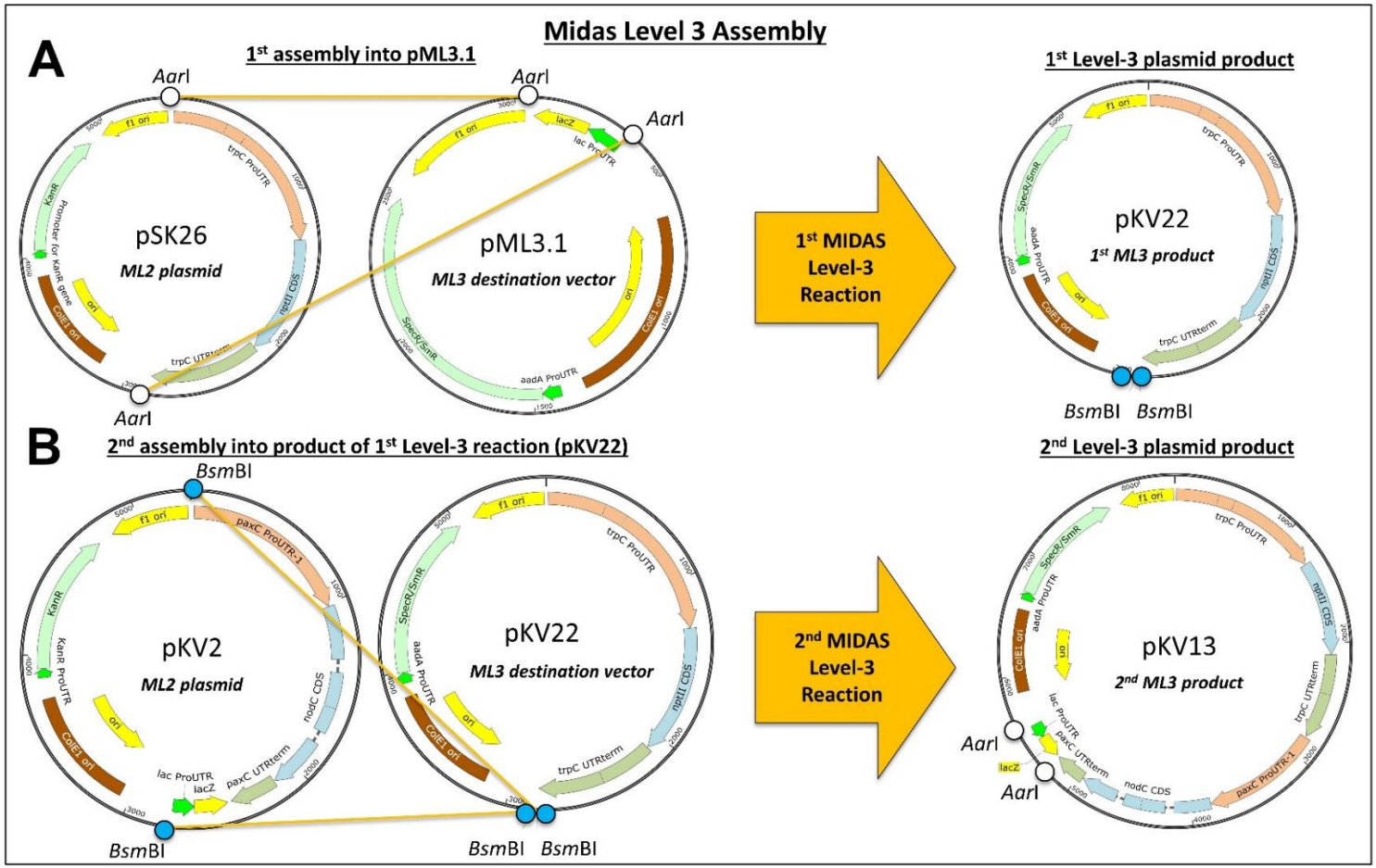


Figure 3.8. Depiction of the 2-step MIDAS Level-3 Assembly used to insert two transcription units, the *nptII*$_{CDS}$ cassette from pSK26 and the *nodC*$_{CDS}$ cassette from pKV2, into a single MIDAS Level-3 plasmid [pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$)]. In the first assembly reaction (**A**), the *nptII*$_{CDS}$ cassette from pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$+WF) was inserted into the pML3.1 destination vector using an *Aar*I-mediated Golden Gate assembly reaction to create the 1st Level-3 plasmid product, pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$). In the second assembly reaction (**B**), the *nodC*$_{CDS}$ cassette from pKV2 was inserted into the 1st Level-3 plasmid product, pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$), using a *Bsm*BI-mediated Golden Gate assembly reaction to create the 2nd Level-3 plasmid product, pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$).

Assembly of a white Level-2 plasmid product [e.g. pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$+WF)] into a Level-3 destination vector imparts *Bsm*BI recognition sites into the Level-3 plasmid product [e.g. pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$)] shown in Figure 3.8, A]. Therefore, the next transcription unit that is assembled into the Level-3 plasmid product [e.g. pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$)] must be assembled in a blue Level-2 destination vector (e.g. pML2(+)BF). Accordingly, the *nodC*$_{CDS}$ transcription unit was assembled into a blue Level-2 destination vector (pML2(+)BF) to create the Level-2 plasmid product pKV2 (*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$+BF) shown in Figure 3.8. Since direction (forward or reverse as indicated respectively by the F and R in the Level-2 destination vectors) or orientation (positive orientation of ProUTR-CDS-UTRterm► versus negative orientation of ◄UTRterm-CDS-ProUTR as indicated respectively by the (+) and (-) of the Level-2 destination vectors) of the *nptII*$_{CDS}$ transcription unit and the *nodC*$_{CDS}$ transcription unit were not seemingly paramount to successful expression of either gene, both transcription units were assembled into (+)F Level-2 destination vectors.

Preparation of MIDAS Level-3 plasmids

MIDAS Level-3 assembly reactions are used to create the plasmids necessary for fungal transformation. A combination of *Aar*I-mediated Golden Gate assembly reactions followed by *Bsm*BI-mediated Golden Gate assembly reactions are used to assemble transcription units from Level-2 plasmids into Level-3 plasmids in a stepwise fashion until the final plasmid product, used for fungal transformation, is assembled. For example, the Level-3 plasmid pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) was assembled stepwise using two MIDAS Level-3 assembly reactions (Figure 3.8, A and B); the first being the assembly of pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$) from the Level-2 plasmid pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$+WF) into the Level-3 destination vector pML3.1 (Figure 3.8, A), and the second being the assembly of pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) from the Level-2 plasmid pKV2 (*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$+BF) into the Level-3 destination vector pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$) (Figure 3.8, B).

Following transformation of the assembled products into *E. coli*, three transformants analysed by restriction analysis all contained the correctly assembled insert. One of the properly assembled Level-3 plasmid products was used as either a destination

vector for the next MIDAS Level-3 assembly if more genes needed to be added [as done with pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$)] or for fungal transformation if it contained all of the desired genes [as done with pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$)].

Control plasmids

To control for any problems that may have arisen from the MIDAS plasmid assembly system, a MIDAS Level-3 control plasmid, pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) shown in Figure 3.9, was assembled alongside pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$). The pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) control plasmid was designed exactly as pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) with the exception of the coding sequence in the second transcription unit being *paxC*$_{CDS}$ in pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) instead of *nodC*$_{CDS}$ as described in pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$). Therefore, pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) was assembled from the same Level-3 destination vector [pKV22 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$)] as pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) (shown in Figure 3.8, B), but with a different Level-2 plasmid, pKV24 (*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$+BF), which contained the *paxC*$_{CDS}$ transcription unit. In addition to this first control plasmid [pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$)], a second control plasmid, pPN1783 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) (plasmid map unavailable), which harbours the native *paxC*$_{CDS}$ transcription unit and was used to restore paxilline (2) production in the *P. paxilli* Δ*paxC* deletion mutant strain (PN2290) during the initial paxilline (2) biosynthetic pathway confirmation experiments,[38] was obtained from Prof. Barry Scott's lab.

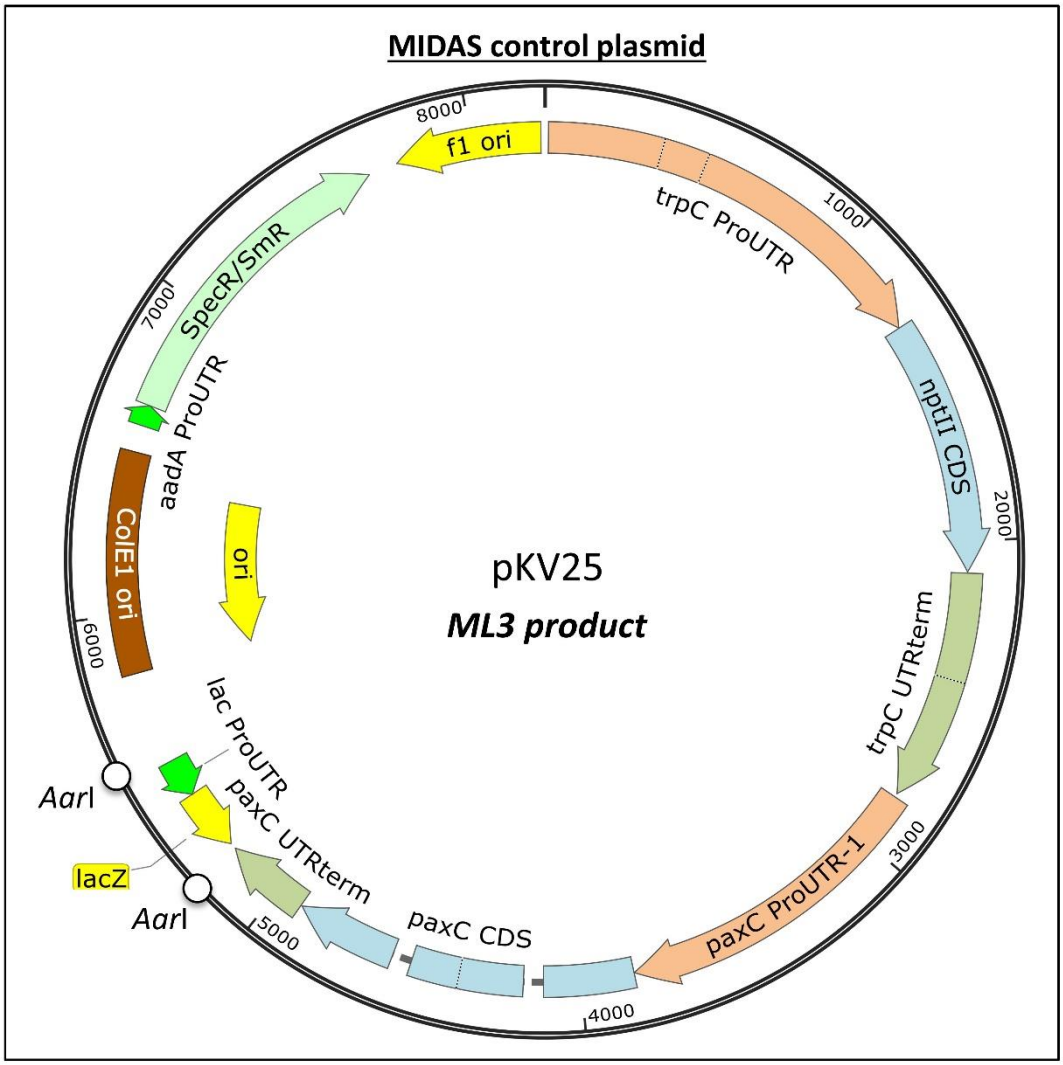


Figure 3.9. Map of the pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) MIDAS control plasmid used during the pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$):PN2290(Δ*paxC*) fungal transformation experiment. Notes: SpecR/SmR imparts spectinomycin resistance, lacZ is used for blue/white selective screening, ColE1 ori is the bacterial origin of replication and F1 ori is the fungal origin of replication.

Fungal transformations

The completed MIDAS Level-3 plasmids pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) and pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) provided the DNA required for the first fungal transformation experiment. Protoplasts of the *P. paxilli* Δ*paxC* deletion mutant strain (PN2290), which does not produce paxilline (2) as shown in lane 2 of Figure 3.10, were transformed with the three plasmids [pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$), pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$), and pPN1783 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-

trpC$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$)] individually to test for the ability to restore paxilline (2) production. Contrary to expectation, neither of the MIDAS Level-3 plasmids [pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) and pKV25(*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$)] restored paxilline (2) production (Figure 3.10, lanes 4-11 and 12-17 respectively) but the positive control plasmid, pPN1783 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$), did (Figure 3.10, lane 3), thereby indicating that the lack of paxilline (2) production seen in the MIDAS transformants was likely not due to the transformation procedure, but rather a problem within the MIDAS constructs.

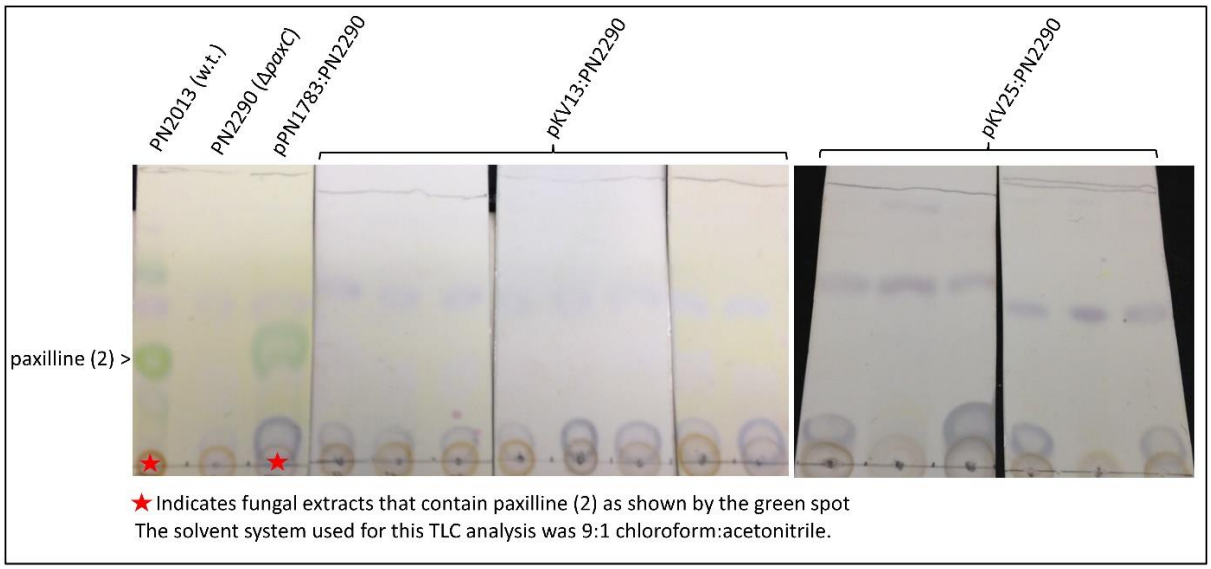


Figure 3.10. TLC results of fungal extracts from the individual transformation of *P. paxilli* Δ*paxC* deletion mutant PN2290 with three plasmids [pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$), pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$), and pPN1783 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$)]. Results from the fungal extracts of the *P. paxilli* wild-type (w.t.) strain PN2013 and the *P. paxilli* Δ*paxC* deletion mutant PN2290 have been included as positive and negative controls respectively.

Resolution of MIDAS construct problem

Direct comparison of the sequences of the MIDAS *paxC*$_{CDS}$ construct [pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$)] to pPN1783 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) showed that the *paxC*$_{ProUTR}$ promoter region in pPN1783 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) (termed *paxC*$_{ProUTR-2}$ in Figure 3.11) contained an additional 144 bp upstream of the *paxC*$_{ProUTR}$ promoter region used in pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) (termed *paxC*$_{ProUTR-1}$ in Figure 3.11).

These extra 144 bp were not accounted for in the MIDAS constructs because the *paxC*$_{ProUTR}$ promoter region was designed based on annotation of the paxilline (2) biosynthetic pathway located on GenBank, which, as demonstrated, contains designated promoter regions that have not been functionally confirmed.

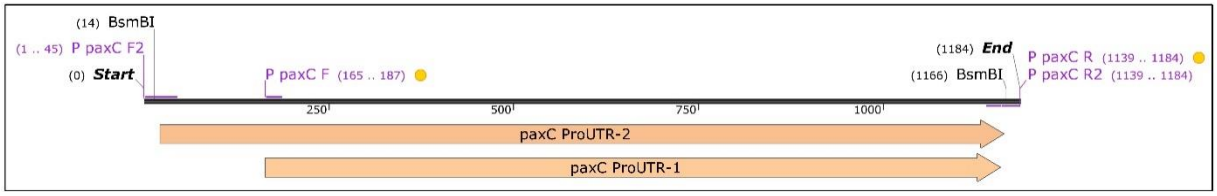


Figure 3.11. Direct comparison of the *paxC*$_{ProUTR-1}$ promoter region used in pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) and pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) to the *paxC*$_{ProUTR-2}$ promoter region used in pPN1783 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$). Notably *paxC*$_{ProUTR-1}$, which is 144 bp shorter than *paxC*$_{ProUTR-2}$, did not promote the expression of *nodC*$_{CDS}$ or *paxC*$_{CDS}$ whereas the *paxC*$_{ProUTR-2}$ did promote the expression of *paxC*$_{CDS}$.

Since these results suggested that there was a problem with the promoter region used in pKV13 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) and pKV25 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$), and it was apparent that the *trpC*$_{ProUTR}$ promoter region worked from the colony growth of the transformed MIDAS constructs in the presence of geneticin, the modularity of the MIDAS system was used to create a MIDAS Level-3 plasmid, pKV27 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodC***$_{CDS}$-*trpC*$_{UTRterm}$) shown in Figure 3.12, that harboured *nodC*$_{CDS}$ driven by *trpC*$_{ProUTR}$ promoter region.

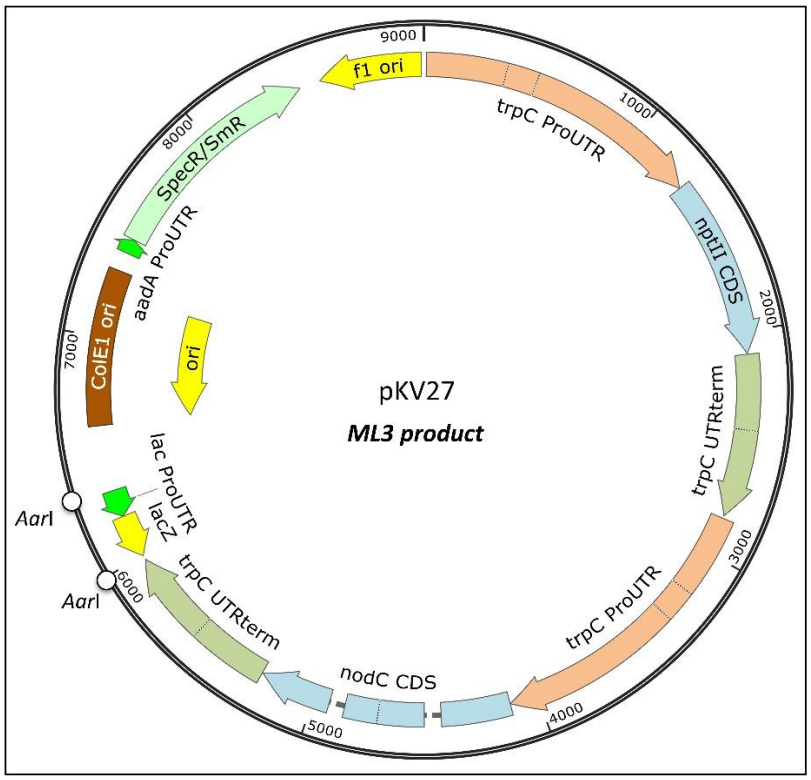


Figure 3.12. Plasmid map of pKV27 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodC***$_{CDS}$-*trpC*$_{UTRterm}$) showing the use of the *trp*C$_{ProUTR}$ promoter region to drive expression of the *nodC*$_{CDS}$ coding sequence. Notes: SpecR/SmR imparts spectinomycin resistance, lacZ is used for blue/white selective screening, ColE1 ori is the bacterial origin of replication and F1 ori is the fungal origin of replication.

Successful complementation of *nodC* into the Δ*paxC P. paxilli* deletion mutant

Transformation of *P. paxilli* Δ*paxC* deletion mutant strain (PN2290) with pKV27 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodC***$_{CDS}$-*trpC*$_{UTRterm}$) restored paxilline (2) production in at least six of ten transformants as determined by TLC analysis (Figure 3.13, lanes 1-5 and 7). Some of the pKV27 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodC***$_{CDS}$-*trpC*$_{UTRterm}$) transformants produce paxilline (2) and some do not because of variations in integration site (i.e. some plasmid DNA may insert into a particular portion of the genome that prevents compound production) and partial integration (i.e. only a portion of the plasmid has integrated). Notably, the production of paxilline (2) was much lower in the *nodC*$_{CDS}$-containing (pKV27) transformants (lanes 1-5 and 7) than in the *paxC*$_{CDS}$-containing (pPN1783) control transformants (lanes 8 and 9). One possible explanation for this was promoter choice because the *nodC*$_{CDS}$ coding sequence was driven by the *trpC*$_{ProUTR}$ promoter region and the *paxC*$_{CDS}$ coding sequence was driven by the *paxC*$_{ProUTR}$ promoter region. In the next chapter, "**NodM, The key epoxidase,"**

a study was designed to test how promoter choice affected compound production and conclusions can be found in the chapter's subsection: "***A note about promoter choice***".

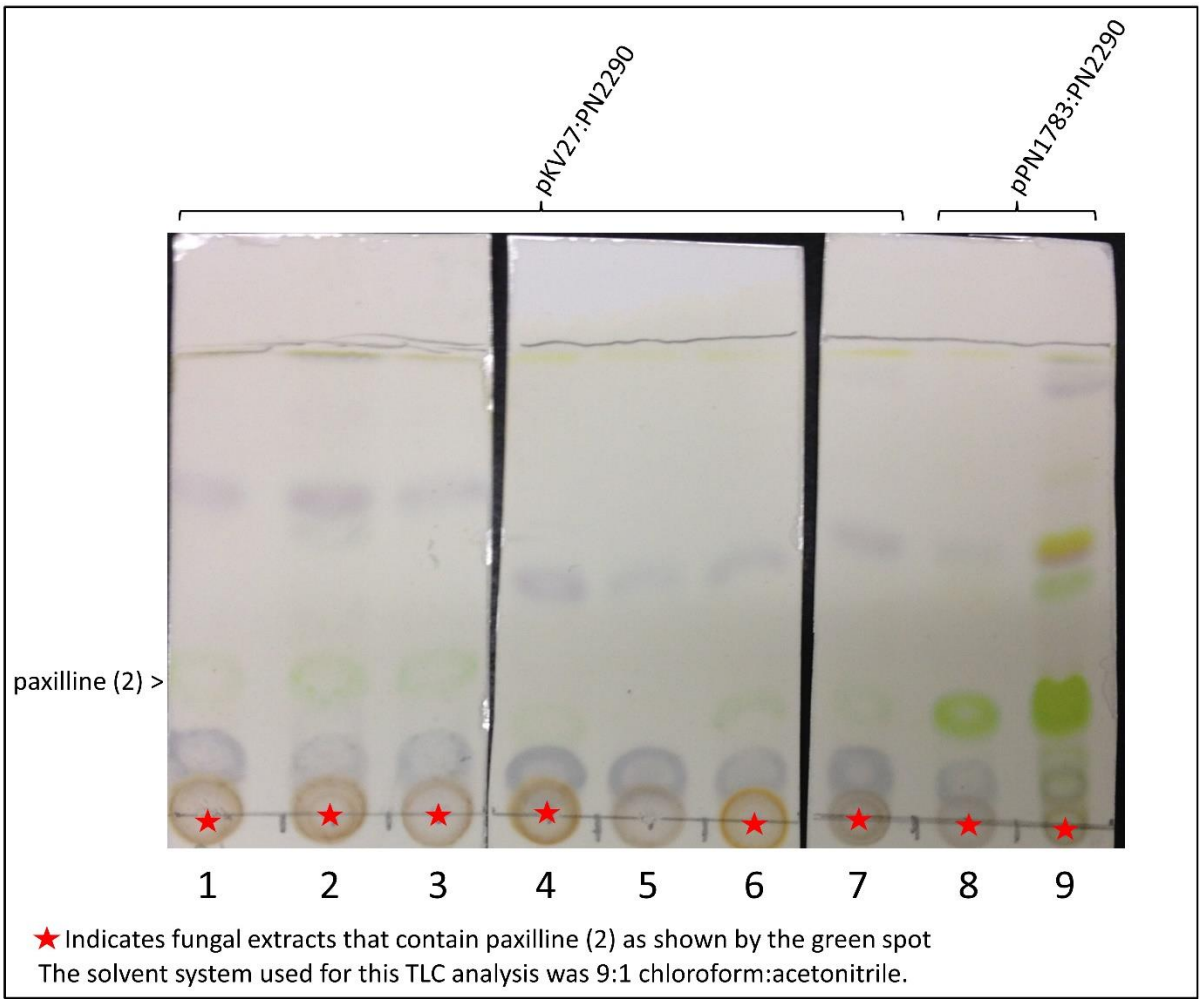


Figure 3.13. TLC results from seven of ten fungal extracts from the transformation of *P. paxilli* Δ*paxC* knockout strain PN2290 with pKV27 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-*nodC*$_{CDS}$-*trpC*$_{UTRterm}$) are shown in lanes 1-7. Results from two of three extracts of the transformation of *P. paxilli* Δ*paxC* knockout strain PN2290 with the positive control plasmid pPN1783 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxC*$_{ProUTR}$-*paxC*$_{CDS}$-*paxC*$_{UTRterm}$) are shown in lanes 8 and 9. Note that the TLC results shown are from a second round of TLC analysis and that extracts that are not shown (e.g. three of ten *nodC*$_{CDS}$-containing pKV27 transformants) did not contain visible indole diterpenes in the first analysis.

LCMS confirmation of paxilline (2) production

The production of paxilline (2) by the *nodC*$_{CDS}$-containing transformants was confirmed by LCMS. In particular, a new compound was present, that was not present in the extract of the Δ*paxC* deletion mutant strain PN2290 (Figure 3.14, trace i), whose

retention time (5.2 min) and high resolution-mass spectra ($[M+H]^+$ = 436.2485 *m/z*) corresponded to that of the paxilline (2) standard (Figure 3.14 trace ii).

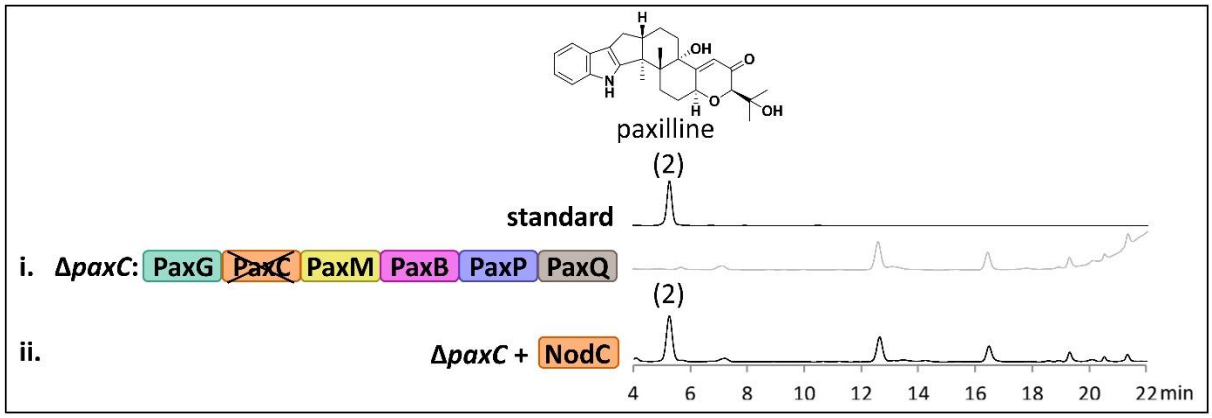


Figure 3.14. HPLC analysis (271 nm) of fungal extracts from the *P. paxilli* Δ*paxC* deletion mutant strain, PN2290 (trace i), and the *nodC*$_{CDS}$-containing pKV27 transformed *P. paxilli* Δ*paxC* deletion mutant strain (pKV27:PN2290) expressing the NodC enzyme from *H. pulicicidum* (trace ii). Extracted ion chromatograms are located in the Appendix (Figure 9.8).

Conclusions

Restoration of paxilline (2) biosynthesis by the introduction of the *nodC*$_{CDS}$ coding sequence into the Δ*paxC* deletion mutant strain (PN2290) indirectly demonstrated that NodC is a functional orthologue of PaxC, giving the first indication that the MIDAS system was a viable method for gene cluster reconstitution and corroborating the hypothesis that the *NOD* cluster located in the *H. pulicicidum* genome is responsible for nodulisporic acid A (1) biosynthesis. With this, attention was turned to the gene likely responsible for the expression of the protein product that catalysed the next step in the proposed nodulisporic acid biosynthetic pathway, the epoxidation of 3-geranylgeranylindole (13).

4. NodM, THE KEY EPOXIDASE

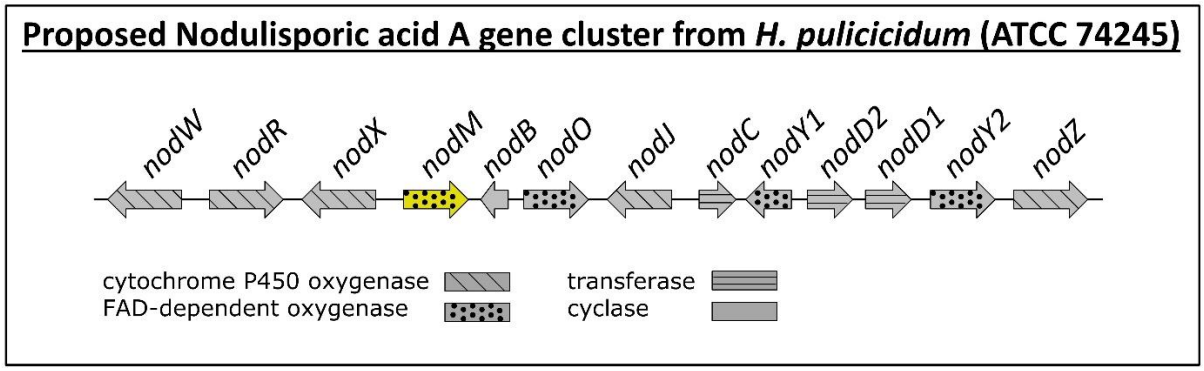


Figure 4.1. Depiction of the predicted nodulisporic acid gene cluster (*NOD*) from *H. pulicicidum* (ATCC® 74245™). The *nodM* gene is highlighted in yellow.

The indole diterpene epoxidases

In step three of indole diterpene biosynthesis the geranylgeranyl tail of geranylgeranylindole, the product of the previous NodC-catalysed reaction, gets primed for cyclisation by epoxidation. Three different types of epoxidation of the geranylgeranyl tail of 3-geranylgeranylindole (13) have been described in indole diterpene biosynthetic pathways and are shown in Figure 4.2. In the first type, the C13-C14 alkene is epoxidised to form 13,14-epoxy-3-geranylgeranylindole (16); in the second, both the C13-C14 and C17-C18 alkenes are epoxidised to form bisepoxy-3-geranylgeranylindole (17); and in the third, the C17-C18 alkene is epoxidised to form 17,18-epoxy-3-geranylgeranylindole (15). Each epoxidation primes the indole diterpene scaffold for cyclisation in a variety of unique indole diterpene cores. In other words, this epoxidation serves as the key step that gives rise to the first branches of the complex indole diterpene biosynthetic tree.

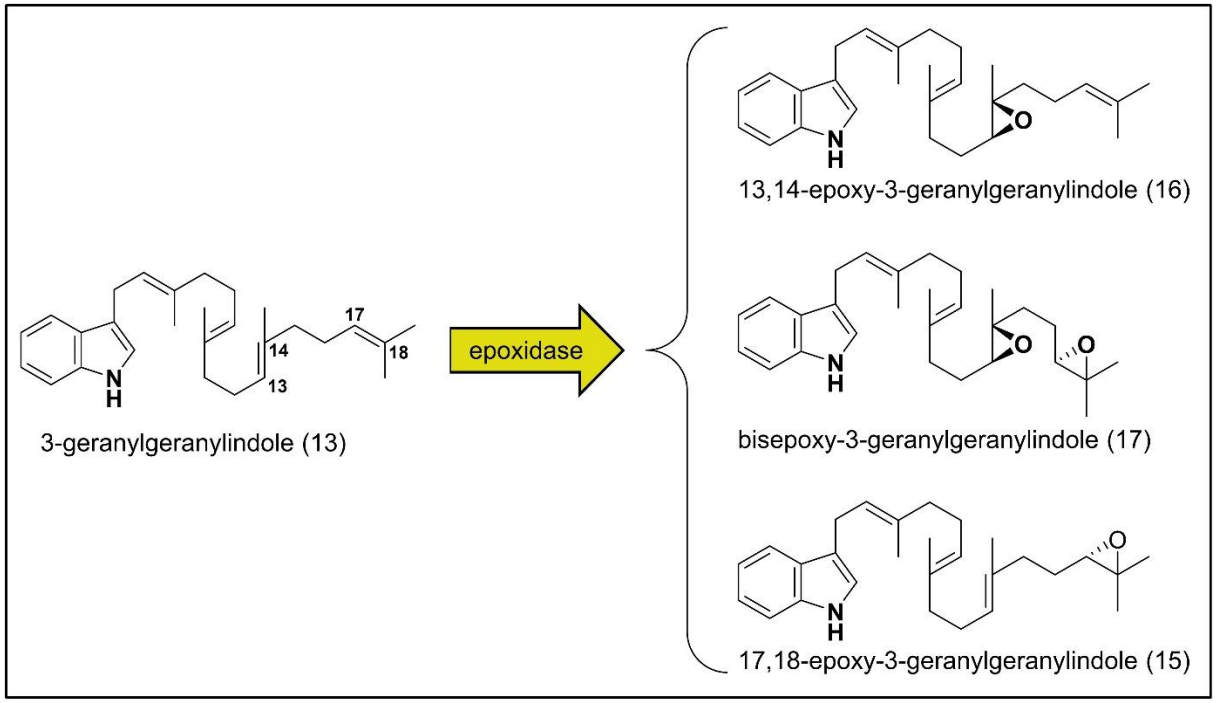
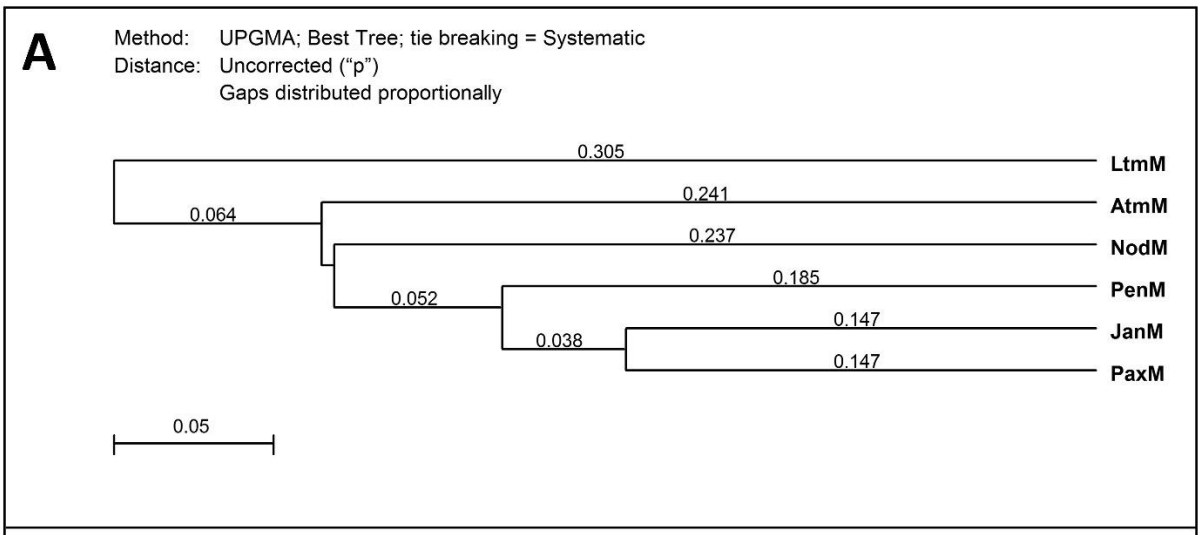


Figure 4.2. Depiction of the third secondary-metabolic step of indole diterpene biosynthesis. In this step, 3-geranylgeranylindole (13) is epoxidised into one or more epoxidised products: 13,14-epoxy-3-geranylgeranylindole (16), bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15). The type of epoxidase (yellow arrow) is species dependant and determines what epoxidised product(s) is/are formed. The predicted cofactors for these epoxidases are NADPH and O_2 but the majority of the *in vitro* enzyme characterisations have not been completed.

Identifying the *NOD* epoxidase

For the biosynthesis of nodulisporic acid A (1) it was predicted that 13,14-epoxy-3-geranylgeranylindole (16) is the dedicated geranylgeranylindole-epoxide intermediate (Figure 2.3) because all of the nodulisporic acids have a pentacyclic skeleton containing the key C17-C18 alkene moiety. By comparing the amino acid sequences of the functionally confirmed indole diterpene epoxidases to the gene products in the NOD cluster, one gene product, NodM, was identified that shares >57% amino acid identity with five other indole diterpene epoxidases (Figure 4.3; 57.2% to LtmM, 63.4% AtmM, 65.8% PaxM, 66% PenM, and 67% JanM). As seen in the NodC similarity screening, the enzyme from the *JAN* cluster, JanM, had the highest sequence similarity (67%) to NodM. Intriguingly, all of the functionally characterised NodM homologues catalyse both the single epoxidation reaction of geranylgeranylindole to form 13,14-epoxy-3-geranylgeranylindole (16) as well as the double epoxidation to form bisepoxy-3-geranylgeranylindole (17).[26, 33, 40] Therefore, if NodM only catalysed the single epoxidation of geranylgeranyl indole to form 13,14-epoxy-3-geranylgeranylindole (16), as hypothesised, then it would be the first report of a functionally confirmed enzyme of this kind.



B

Enzyme	LtmM	NodM	AtmM	PenM	JanM	PaxM
LtmM	100	37.3	38.2	36.8	36.9	37.9
NodM	57.2	100	48.3	51.2	52.9	48.6
AtmM	56.3	63.4	100	48.3	47.9	48.9
PenM	54.8	66	62.8	100	61.6	60.6
JanM	54.7	67	61.6	75.1	100	66.7
PaxM	56.6	65.8	64.4	74	79.9	100

Figure 4.3. Phylogenetic tree (**A**) and similarity matrix (**B**) of indole diterpene epoxidases ('M' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the 'M' enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the 'M' enzymes.

Functional analysis of *nodM*

To assess the epoxidative capacity of NodM (i.e. how many and what type of epoxidations it can catalyse), to establish directly its role in nodulisporic acid A (1) biosynthesis, and to explore how promoter choice affected compound biosynthesis, two MIDAS Level 3 vectors shown in Figure 4.4 [pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-*nodM*$_{CDS}$-*trpC*$_{UTRterm}$) and pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$)] were designed that respectively contained the *nodM*$_{CDS}$ driven by the *trpC*$_{ProUTR}$ promoter region from *Aspergillus nidulans* or driven by the *paxM*$_{ProUTR}$ promoter region from *P. paxilli* and inserted them into the Δ*paxM P. paxilli* deletion mutant strain (PN2257). It was hypothesised that

rather than restore paxilline (2) production, a *P. paxilli* strain that specifically produced emindole SB (18) would be generated.

MIDAS assembly

The MIDAS assembly of two plasmid constructs [pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$) and pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$)] used for functional characterisation of the *nodM*$_{CDS}$ are depicted in Figure 4.4. A total of six Level-1 MIDAS vectors encoding transcription unit modules for the *trpC*$_{ProUTR}$ promoter region (pSK17), *paxM*$_{ProUTR}$ promoter region (pSK4), *nodM*$_{CDS}$ coding sequence (pSK18), *ntpII*$_{CDS}$ coding sequence (pSK16), *trpC*$_{UTRterm}$ terminator region (pSK15), and *paxM*$_{UTRterm}$ terminator region (pSK6) were used in the assembly of these two *nodM*$_{CDS}$-containing Level-3 vectors. While Level-1 MIDAS vectors containing the *trpC*$_{ProUTR}$ promoter region (pSK17), *nptII*$_{CDS}$ coding sequence (pSK16) and *trpC*$_{UTRterm}$ terminator region (pSK15) were assembled during the functional analysis of *nodC*, the other three Level 1 MIDAS vectors encoding the *paxM*$_{ProUTR}$ promoter region (pSK4), *nodM*$_{CDS}$ coding sequence (pSK18), and *paxM*$_{UTRterm}$ terminator region (pSK6) were newly assembled into individual Level-1 destination vectors. The six Level-1 vectors were used to assemble three unique transcription units in individual Level-2 MIDAS vectors pSK26 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$+WF), pKV33 (*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$+BF) and pKV57 (*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$+BF). The final two Level-3 MIDAS vectors for use in fungal transformation pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$) and pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$) were respectively generated from the individual assembly of either pKV33 (*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$+BF) or pKV57 (*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$+BF) into the Level-3 MIDAS vector pKV22 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$), which was generated during *nodC* functionality testing and contains the geneticin resistance (*nptII*) cassette.

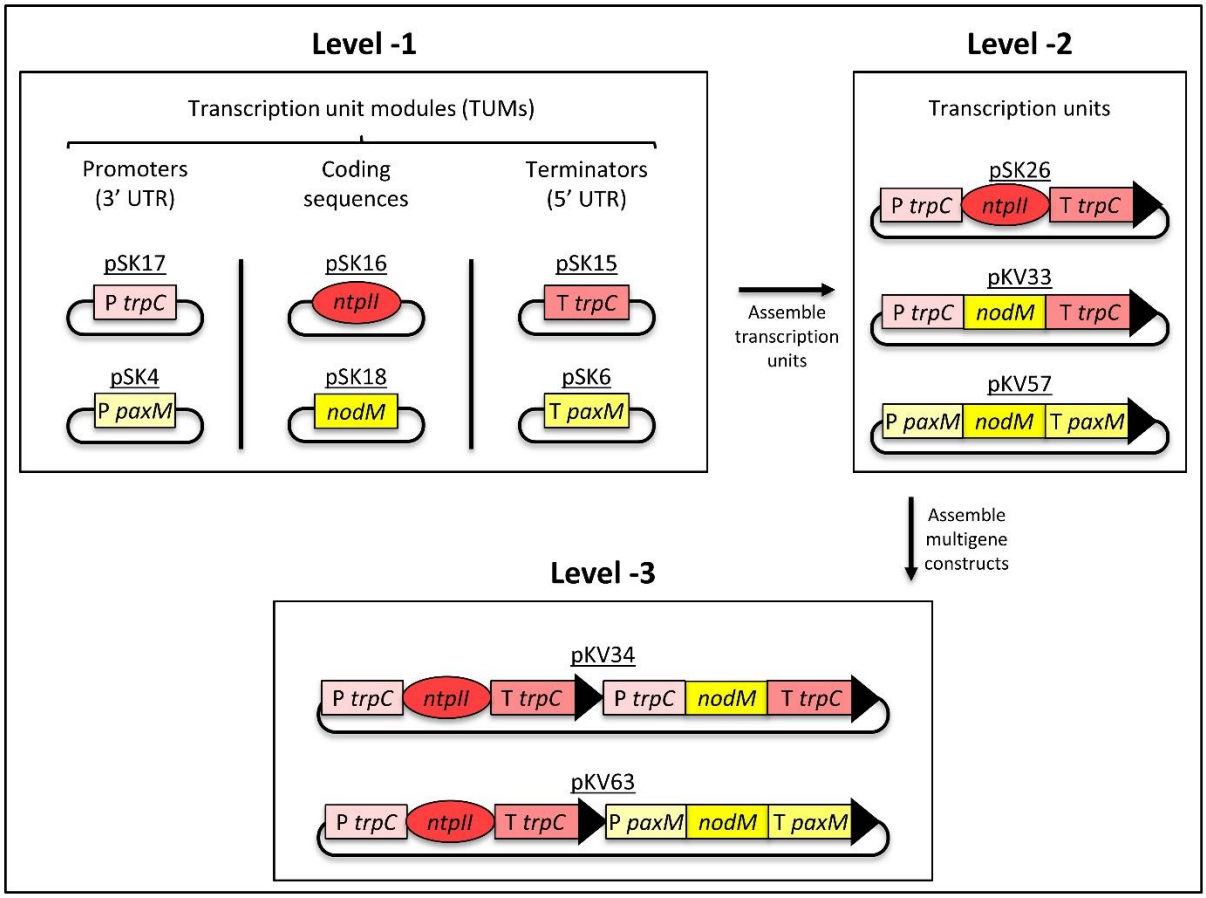


Figure 4.4. Depiction of the assembly of MIDAS vectors pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$) and pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$) used to confirm the function of the *nodM*$_{CDS}$ coding sequence.

Fungal transformations

The two Level-3 *nodM*-containing vectors, pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$) and pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$), were each individually inserted into protoplasts of the Δ*paxM* deletion mutant strain PN2257, and a parallel transformation with vector pPN1857 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$) was used as a positive control. The pPN1857 vector harbours the native *paxM* transcription unit and was used to restore paxilline (2) production during initial paxilline (2) biosynthetic pathway confirmation experiments.[38] Six transformants were selected from the pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$) transformation, ten were selected from the pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$) transformation, and a total of eight

transformants were selected from two pPN1857 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$) control transformations. The 24 transformants were grown in production medium and analysed by TLC and LCMS as shown in Figures 4.5 and 4.6. As expected, the initial TLC screening suggested that paxilline (2) was produced by seven of eight *paxM*$_{CDS}$-containing control transformants (Figure 4.5, lanes: 7-10 and 21-23) and none of the *nodM*$_{CDS}$-containing transformants (Figure 4.5, lanes: 1-6 and 11-20), and that at least eight of sixteen *nodM*$_{CDS}$-containing transformants [two of six pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$) and five of ten pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$)] produced a novel indole diterpene compound (Figure 4.5, lanes: 4, 5, 11, 13, 15, 16, 20).

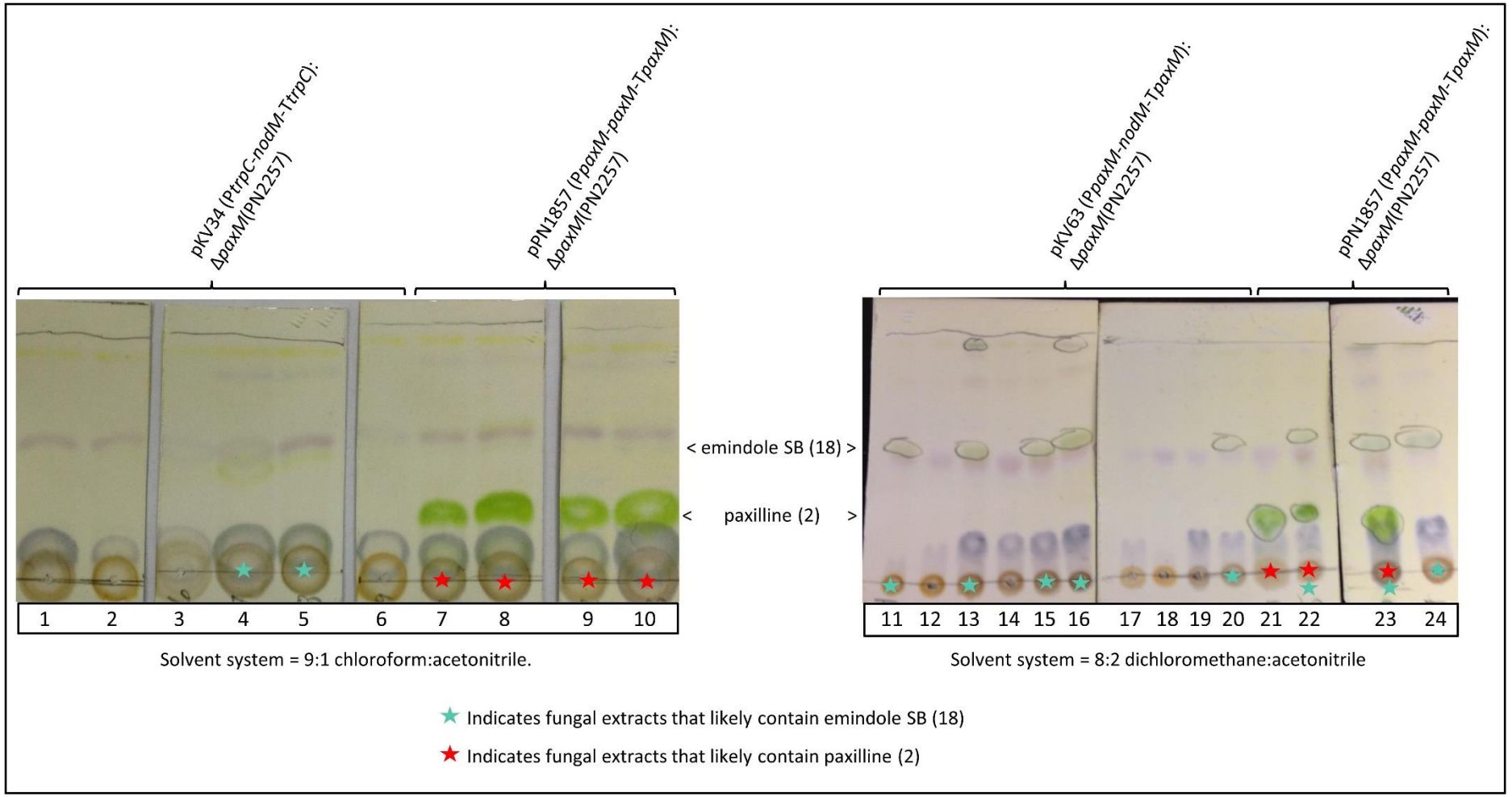


Figure 4.5. TLC results of fungal extracts from the individual transformation of *P. paxilli* Δ*paxM* deletion mutant PN2257 with three plasmids (pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$), pPN1857 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$), and pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$)).

A note about promoter choice

Importantly, in these experiments it was observed that there was greater compound production when the *nodM*$_{CDS}$ coding sequence was driven by the *paxM*$_{proUTR}$ promoter region (pKV63) than when driven by the *trpC*$_{proUTR}$ promoter region (pKV34) (Figure 4.5, lane 4 and 5 vs. lanes 11, 13, 14, 15). Although this difference was not quantified by LCMS, the visual inspection by TLC clearly implied that promoter choice greatly influenced compound production. It is especially important to consider promoter choice during host optimisation because there are multiple hurdles to overcome *viz*. negative feedback loops and overexpression, which could lead to a decrease in cell performance and potentially result in cell death.[73]

Novel compound identification

To identify the novel compound that was produced by the *nodM*$_{CDS}$-containing transformants, a high producing strain (as determined by visual inspection of TLCs) from the pKV34 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodM***$_{CDS}$-*trpC*$_{UTRterm}$) transformants (Figure 4.5, lane 5) and the pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$) transformants (Figure 4.5, lane 13) were analysed by LCMS. Evaluation of the spectra revealed that, based on the retention time (19.3 min) and mass (high resolution-mass spectra $[M+H]^+$ = 406.3109 *m/z*, corresponding to $C_{28}H_{39}NO$), the novel compound was emindole SB (18). Notably, no paxilline (2) or paspaline (11) were present in the either of the transformant extracts

[results only shown for the the pKV63 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$) transformant, Figure 4.6, trace ii].

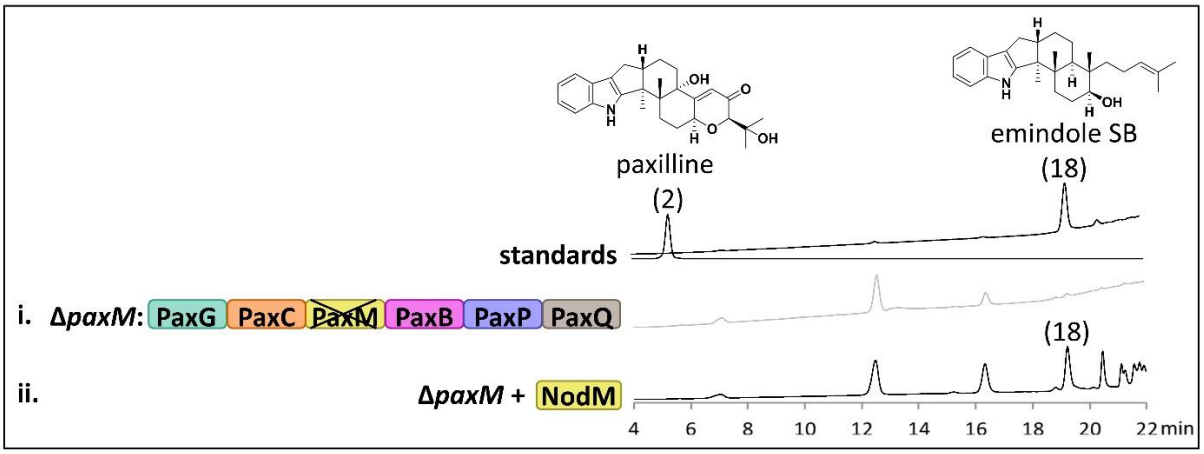


Figure 4.6. HPLC analysis (271 nm) of fungal extracts from the *P. paxilli* Δ*paxM* deletion mutant strain, PN2257 (trace i), and the *P. paxilli* Δ*paxM* deletion mutant strain [pKV63 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$):PN2257] expressing the NodM enzyme from *H. pulicicidum* (trace ii). Extracted ion chromatograms are located in the Appendix (Figure 9.9).

Conclusions

Explanation for emindole SB (18) production

It has previously been shown that *paxB*, a gene present in the Δ*paxM P. paxilli* mutant strain, is capable of cyclising both 13,14-epoxy-3-geranylgeranylindole (16) and bisepoxy-3-geranylgeranylindole (17) to produce emindole SB (18) and paspaline (11) respectively (see Figure 4.7).[40] Thus, these results indirectly confirmed the hypothesis that NodM was only capable of a single epoxidation at the C13-C14 position of 3-geranylgeranylindole (13) to produce 13,14-epoxy-3-geranylgeranylindole (16) because the absence of paspaline (11) or paxilline (2) production demonstrate that this enzyme was not capable of producing the double epoxidised product bisepoxy-3-geranylgeranylindole (17). NodM is the first functionally confirmed epoxidase that performs only one epoxidation of 3-geranylgeranylindole (13) to form 13,14-epoxy-3-geranylgeranylindole (16). By solely epoxidising the C13-C14 alkene of geranylgeranylindole, NodM ensures that the only available substrate for cyclisation is 13,14-epoxy-3-geranylgeranylindole (16); thereby establishing the foundation of nodulisporic acid biosynthesis.

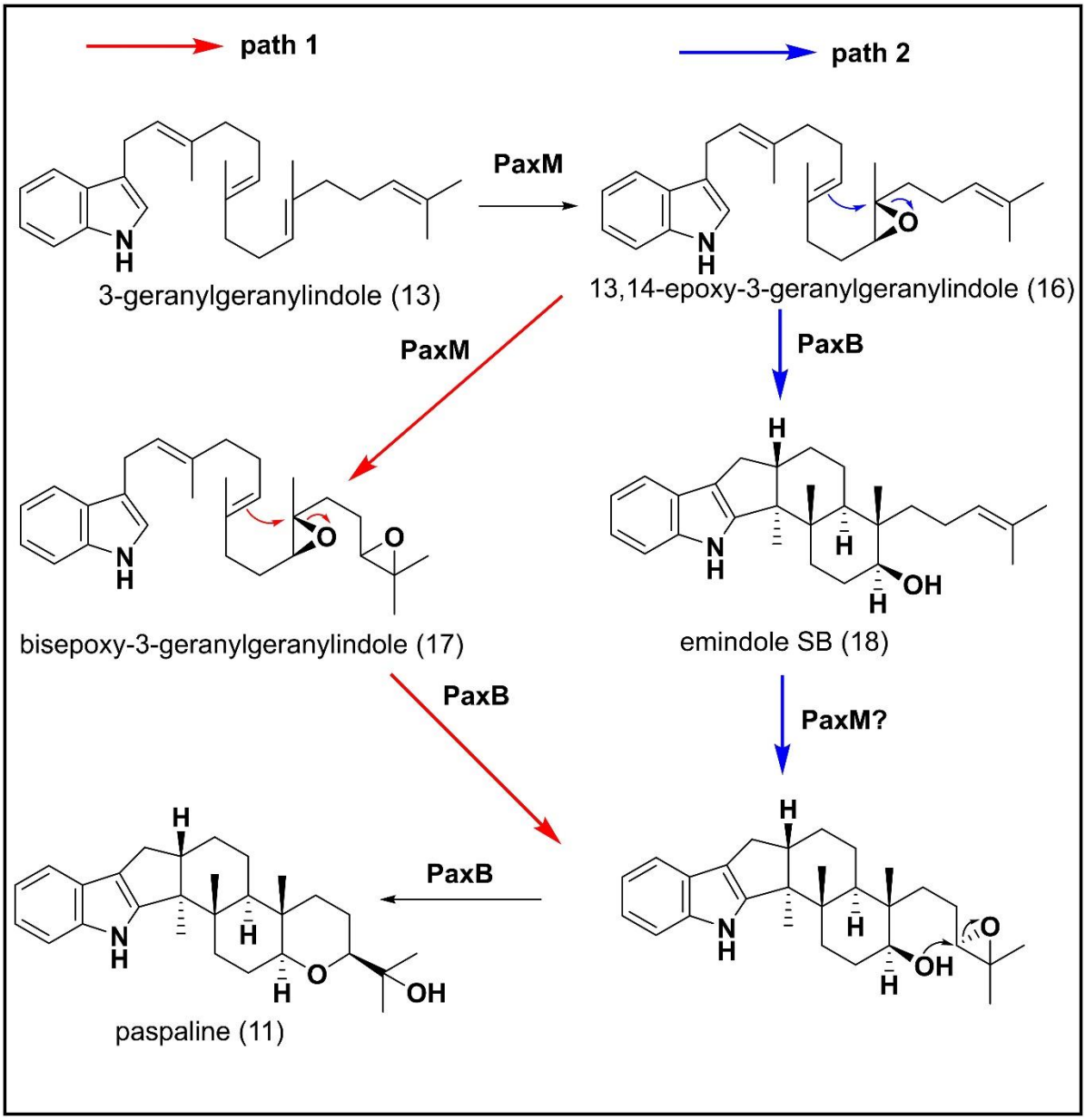


Figure 4.7. Two-path stepwise epoxidation scheme of 3-geranylgeranylindole (13) involving the epoxidase (PaxM) and the cyclase (PaxB) proposed by Tagami *et al*. (2013).[40]

5. NodB, A PROMISCUOUS INDOLE DITERPENE CYCLASE

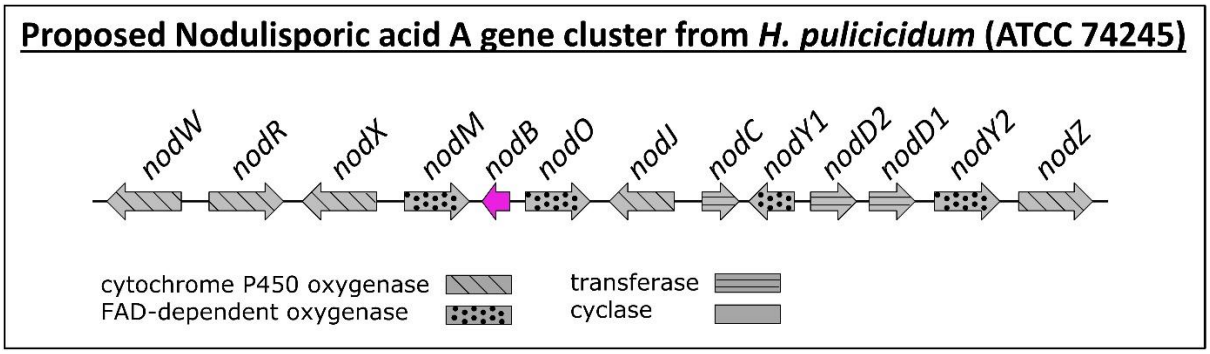


Figure 5.1. Depiction of the predicted nodulisporic acid gene cluster (*NOD*) from *H. pulicicidum* (ATCC® 74245™). The *nodB* gene is highlighted in pink.

Indole diterpene cyclases are notoriously promiscuous

The indole diterpene cyclases are the most elaborate group of enzymes in indole diterpene biosynthesis as they have an extremely diverse capacity to cyclise the three epoxide products [13,14-epoxy-3-geranylgeranylindole (16), bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15)] into a variety of distinct core compounds (Figure 5.2). The cyclases produce a range of ring closures, often coupled to sigmatropic shifts, which result in the addition of 2-4 rings to the epoxidised-geranylgeranyl tail of the epoxide substrate. To date, every functionally confirmed indole diterpene cyclase has shown promiscuity around its ability to cyclise more than one of the three epoxide substrates [13,14-epoxy-3-geranylgeranylindole (16), bisepoxy-3-geranylgeranylindole (17) and 17,18-epoxy-3-geranylgeranylindole (15)].[33, 40] While their promiscuity leads to the delivery of a wide array of compounds, the cyclases usually deliver only one major cyclised core that is used in one specific biosynthetic pathway of the native indole diterpene producer. For example, the cyclase PaxB, used in the paxilline (2) biosynthetic pathway of *P. paxilli*, shows substrate promiscuity around its ability to cyclise 13,14-epoxy-3-geranylgeranylindole (16) into emindole SB (18) and bisepoxy-3-geranylgeranylindole (17) into paspaline (11); of which paspaline (11) is the major product and dedicated substrate for paxilline (2) biosynthesis.

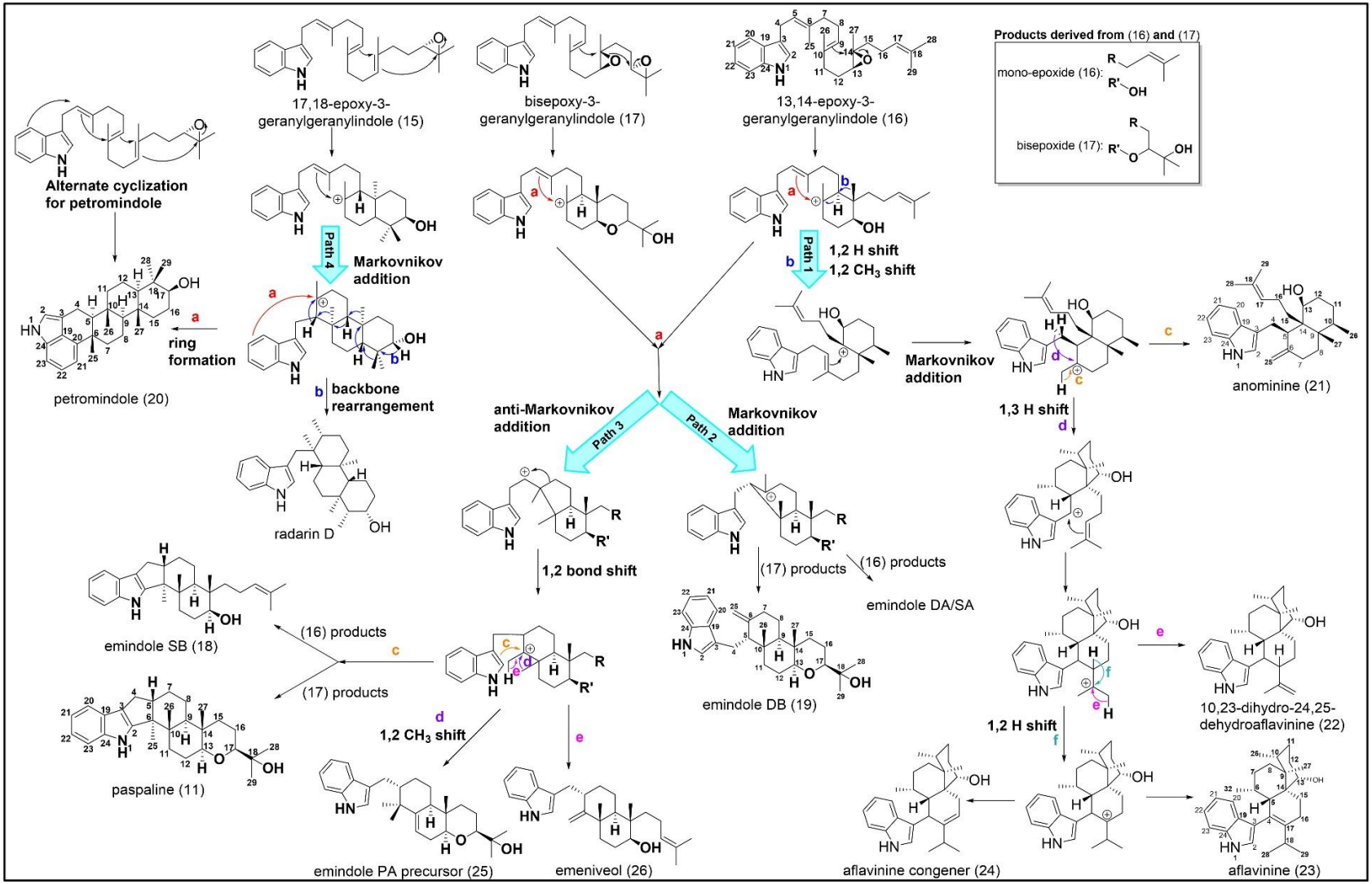
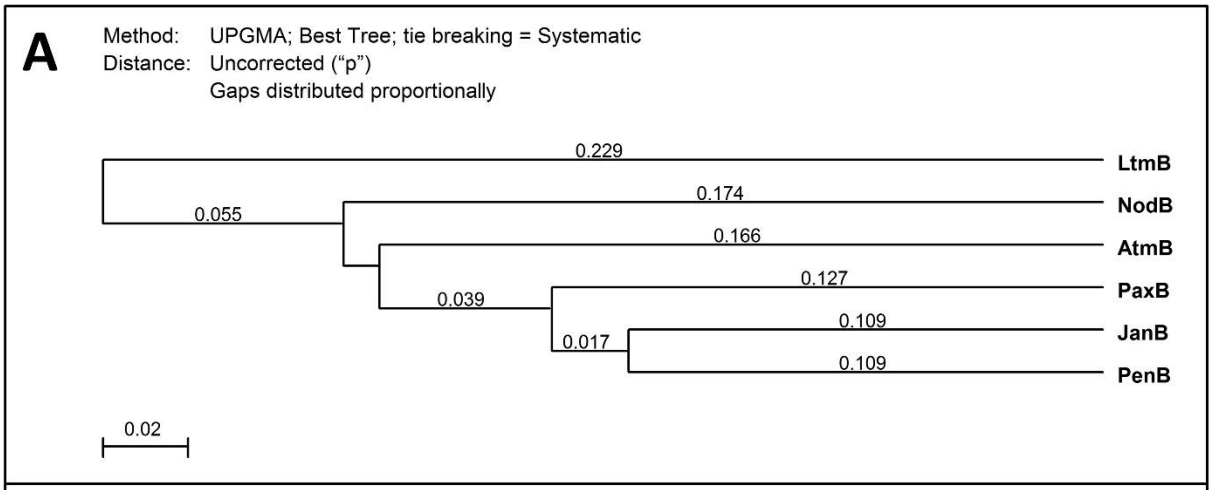


Figure 5.2. Depiction of the pathway diversity generated by indole diterpene cyclases. Indole diterpene cyclases from different species have the capacity to cyclise the epoxidised substrates [(15), (16), and (17)] in different ways using either a Markovnikov reaction mechanism or an anti-Markovnikov reaction mechanism often coupled with select sigmatropic shifts.

Identifying the *NOD* cyclase

In terms of nodulisporic acid biosynthesis, emindole SB (18) was predicted to be the dedicated pentacyclic core intermediate (Figure 2.3). Using amino acid sequence comparisons of the functionally confirmed indole diterpene cyclases to the gene products in the *NOD* cluster a likely indole diterpene cyclase, NodB, that shares >48% amino acid identity with five other indole diterpene cyclases (Figure 5.3; 48.8% LtmB, 63% PaxB, 64.2% AtmB, 65.4% JanB, 67.9% PenB) was identified. As with the previous steps, the MIDAS platform was used to generate plasmids for expression and functional characterisation of *nodB*.



B

Enzyme	PaxB	NodB	LtmB	AtmB	JanB	PenB
PaxB	100	63	49.6	62.1	77	72.4
NodB	78.2	100	48.8	64.2	65.4	67.9
LtmB	65.6	63.5	100	48.8	51.6	52
AtmB	77	78.2	65.2	100	67.5	70.8
JanB	87.2	77.8	66.4	79.8	100	78.2
PenB	87.7	80.2	66.4	82.3	86.4	100

Figure 5.3. Phylogenetic tree (**A**) and similarity matrix (**B**) of indole diterpene epoxidases ('B' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the 'B' enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the 'B' enzymes.

Functional analysis of *nodB*

Initial approach

Initially, the functionally characterisation of *nodB* was attempted using a single gene heterologous expression experiment of a $nodB_{CDS}$-containing transcription unit in a *P. paxilli* Δ*paxB* deletion mutant (PN2458). To this end, the $nodB_{CDS}$ coding sequence, driven by the $trpC_{ProUTR}$ promoter region from *A. nidulans*, was assembled, along with the hygromycin resistance cassette encoded by the *hph* transcription unit, into a Level 3 MIDAS vector pKV36 (hph_{ProUTR}-**hph_{CDS}**-$hph_{UTRterm}$_$trpC_{ProUTR}$-**$nodB_{CDS}$**-$trpC_{UTRterm}$) and introduced into the *P. paxilli* Δ*paxB* deletion mutant strain PN2458. Contrary to expectations, all of the extracts from ten transformants did not contain any detectable indole diterpenes as shown in Figure 5.4, lane 1-10.

Notably, no control plasmid with the native *paxB* transcription unit was available to confirm that its complementation of the *P. paxilli* Δ*paxB* deletion mutant strain PN2458 would restore paxilline (2) biosynthesis. In part, this was due to the fact that *P. paxilli* Δ*paxB* deletion mutant strain PN2458 was never transformed with the native *paxB* transcription unit, as was done with the *P. paxilli* Δ*paxC* and Δ*paxM* deletion mutants using pPN1783 and pPN1857 respectively.[38] Careful examination of the *PAX* cluster revealed that a portion of the $paxM_{ProUTR}$ promoter region (~750 bp) overlaps with the $paxB_{CDS}$ coding sequence. Therefore, when the *P. paxilli* Δ*paxB* deletion mutant was created, it may have also incurred a *paxM* deletion. Rather than explore the intricacies of the potential problems within the genome of the *P. paxilli* Δ*paxB* deletion mutant (PN2458), a different approach involving the generation of larger multigene constructs was used to analyse the function of *nodB*.

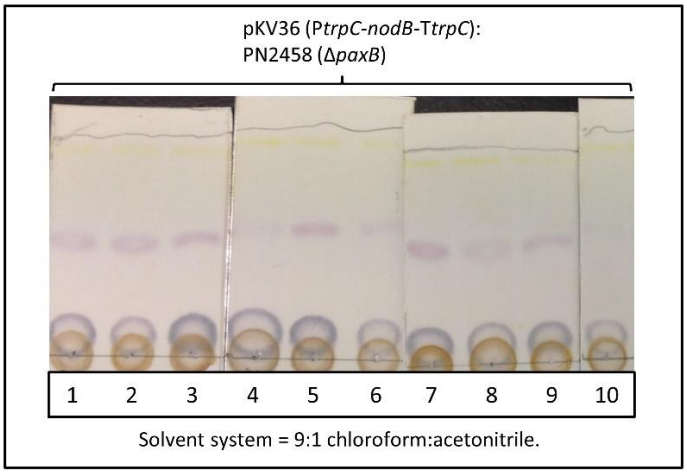


Figure 5.4. TLC results of fungal extracts from ten transformants of plasmid pKV36 (*hph*$_{ProUTR}$-***hph***$_{CDS}$-*hph*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodB***$_{CDS}$-*trpC*$_{UTRterm}$) into the *P. paxilli* Δ*paxB* deletion mutant PN2458.

Assembly of multigene constructs

As opposed to the single indole diterpene transcription unit that was used in the initial attempt to elucidate the function of *nodB*, the multigene approach required the assembly of four indole diterpene transcription units that encoded the *paxG*$_{CDS}$, *nodC*$_{CDS}$, *nodM*$_{CDS}$ and *nodB*$_{CDS}$. Rather than use the MIDAS Level 2 vectors that were assembled during the functional analyses of *nodC*, *nodM* and the initial *nodB* approach for assembly of the multigene construct, new MIDAS Level 2 vectors were designed so that the position and direction of each transcription unit in the final multigene construct resembled that found in the native *PAX* and *NOD* gene clusters as shown in Figure 5.5. The final multigene construct contained five transcription units encoding the geneticin resistance cassette [pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$+WF)], *paxG* [pSK21 (*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$+BR)], *nodM* [pSK28 (*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$+WF)], *nodB* [pSK29 (*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$+BR)] and *nodC* [pSK60 (*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$+WF)] that were respectively assembled into the final construct pSK66 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) in an iterative approach (Figure 5.6).

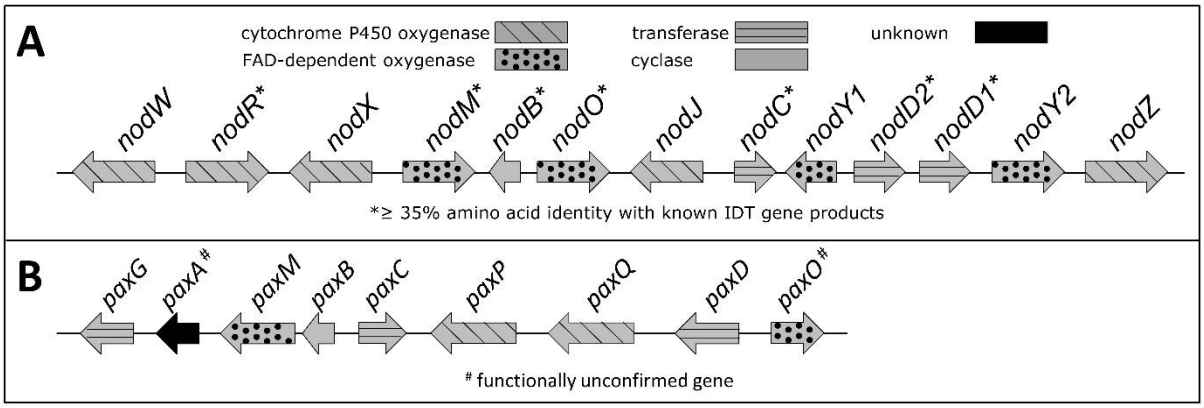


Figure 5.5. Figure of the *NOD* cluster (**A**) and the *PAX* cluster (**B**) showing the order and direction of each gene (arrows).

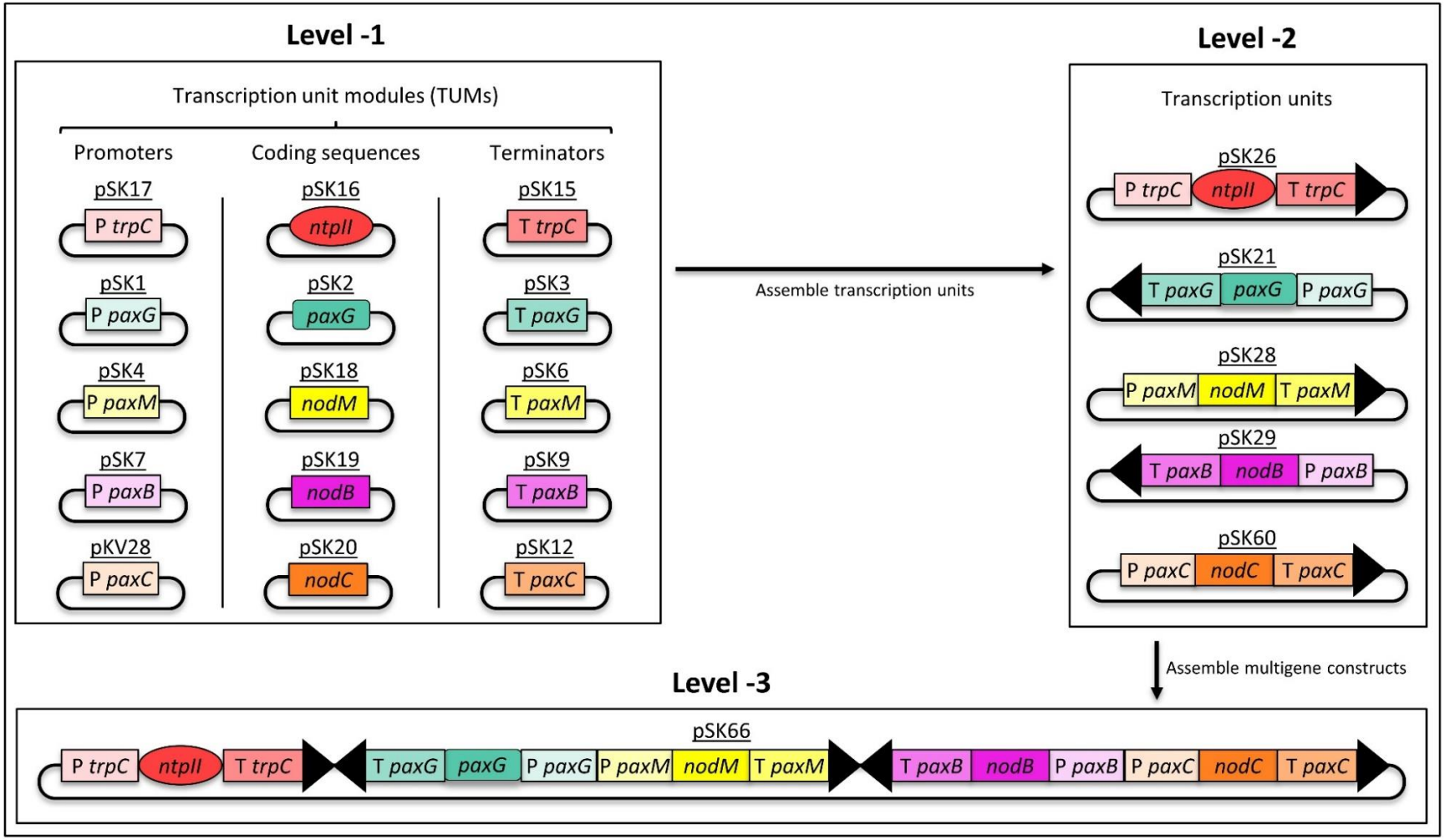


Figure 5.6. Depiction of the MIDAS approach used to assemble the multigene construct pSK66 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$). In Level-1, the plasmids encoding transcription unit modules (promoters, coding sequences, terminators) are shown. In Level-2 the plasmids that encode the five full length transcription units are shown –where the arrow represents the direction of transcription. In Level-3, the final multigene plasmid (pSK66) that was used for fungal transformation is depicted.

The high percentage amino acid identity that NodB shared with the functionally confirmed indole diterpene cyclases suggested that NodB was capable of similar catalytic function and was therefore also likely to be capable of performing two cyclisations; however, only one type of cyclised indole diterpene core, that corresponding to the core of the nodulisporic acids, has ever been isolated from the native nodulisporic acid producer, *H. pulicicidum*. One possible explanation is that only one epoxidised intermediate, 13,14-epoxy-3-geranylgeranylindole (16), is provided by NodM, as was demonstrated during the functional confirmation of NodM, and thus only one cyclised product is formed. Alternatively, NodB may only be able to cyclise one epoxide intermediate and thus NodM and NodB would work in conjunction to establish the foundation of nodulisporic acid biosynthesis. To test whether NodM solely established the foundation of nodulisporic acid production or whether NodM and NodB worked in conjunction NodB would have to be presented with an alternative epoxide product, bisepoxy-3-geranylgeranylindole (17) and/or 17,18-epoxy-3-geranylgeranylindole (15), and extracts would have to be examined for alternative cyclic core compounds [e.g. paspaline (11)]. Hence, a second multigene construct was assembled to do just that.

The second multigene construct, pKV74 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$), was designed to specifically incorporate a *paxM* transcription unit and a *nodB* transcription unit into the same Level-3 vector (Figure 5.7, Level-3). The *paxM* transcription unit was used in this second multigene construct because it catalyses the production of bisepoxy-3-geranylgeranylindole (17) (Figure 5.8), the ideal substrate to test the promiscuity of NodB because all of the functionally confirmed cyclases that NodB is highly similar to are capable of cyclising bisepoxy-3-geranylgeranylindole (17) to paspaline (11).[33, 40] Therefore, it was hypothesised that paspaline (11) would be produced if NodB were provided with bisepoxy-3-geranylgeranylindole (17) (Figure 5.8).

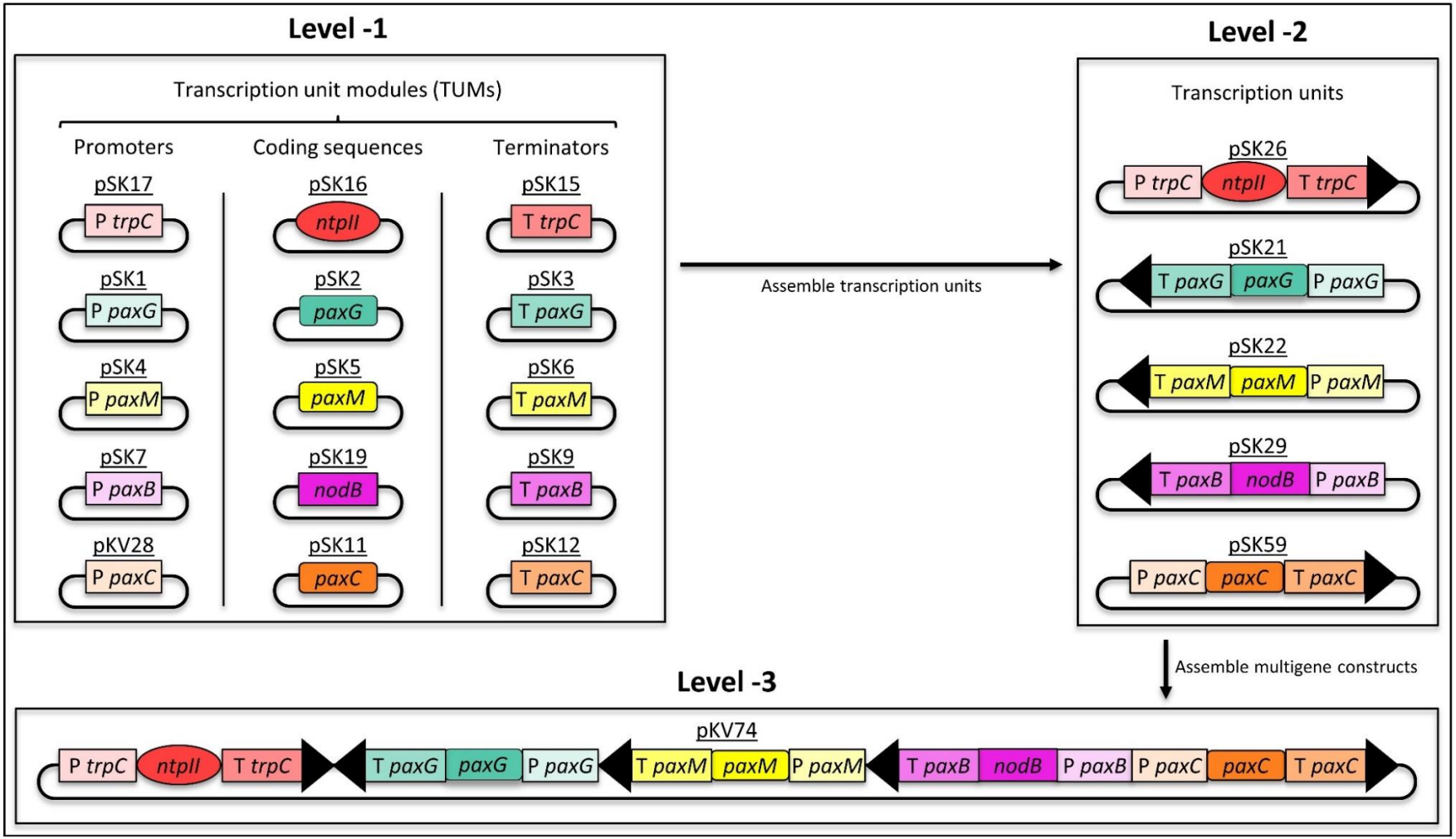


Figure 5.7. Depiction of the MIDAS approach used to assemble the multigene construct pKV74 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$). In Level-1, the plasmids encoding transcription unit modules (promoters, coding sequences, terminators) are shown. In Level-2 the plasmids that encode the five full length transcription units are shown –where the arrow represents the direction of transcription. In Level-3, the final multigene plasmid (pKV74) that was used for fungal transformation is depicted.

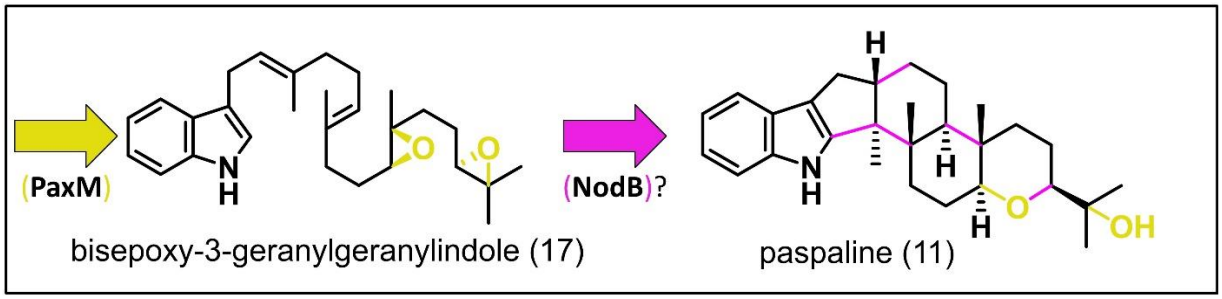


Figure 5.8. Is NodB capable of cyclising bisepoxy-3-geranylgeranylindole (17) to paspaline (11)?

In addition to the *paxM* and *nodB* transcription units, the second multigene construct, pKV74 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$), also contained the geneticin resistance cassette and the *paxG* and *paxC* transcription units (Figure 5.4, Level-3). Thus, as with the first multigene construct [pSK66 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$)], the second [pKV74 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$)] also contained a total of five transcription units, three that were from the same Level 2 MIDAS vectors as used in the first multigene construct (pSK66): the geneticin resistance cassette [pSK26 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$+WF)], *paxG* [pSK21 (*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$+BR)] and *nodB* [pSK29 (*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$+BR)], and two that were not and encoded *paxC* [pSK59 (*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$+WF)] and *paxM* [pSK22 (*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$+WR)]. Due to the modular nature of MIDAS and the iterative addition of transcription units to the final Level-3 construct, the Level-3 MIDAS construct that contained the geneticin resistance cassette and the *paxG* transcription unit [pSK34 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$)], assembled during construction of the first multigene construct (pSK66), were used as the first destination vector for this second multigene construct (pKV74).

Transformations of the multigene constructs

Upon completion of the multigene assembly of pSK66 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) and pKV74 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$), each multigene construct was individually introduced into protoplasts of the *P. paxilli* strain PN2250 (CY2) that contains a deletion of the entire *PAX* locus. Fifteen transformants were picked from the pSK66 transformation and ten transformants were picked from the pKV74 transformation. TLC analysis of the pSK66 transformant extracts indicated that eleven of fifteen produced an indole diterpene compound with the same retention time as that of emindole SB (18) as shown in Figure 5.9, lanes 2, 4 to 8, 10, 13-16. In contrast, no obvious indole diterpene compounds were visible from the TLC analysis of the nine pKV74 transformants (Figure 5.9, lanes 17 to 26). To confirm the production of emindole SB (18) by the pSK66 transformants and to assess if any indole diterpenes were present in the pKV74 transformants that may have been below the limit of detection of the TLC analysis, three LCMS samples were prepared from the extracts of each group of transformants and analysed.

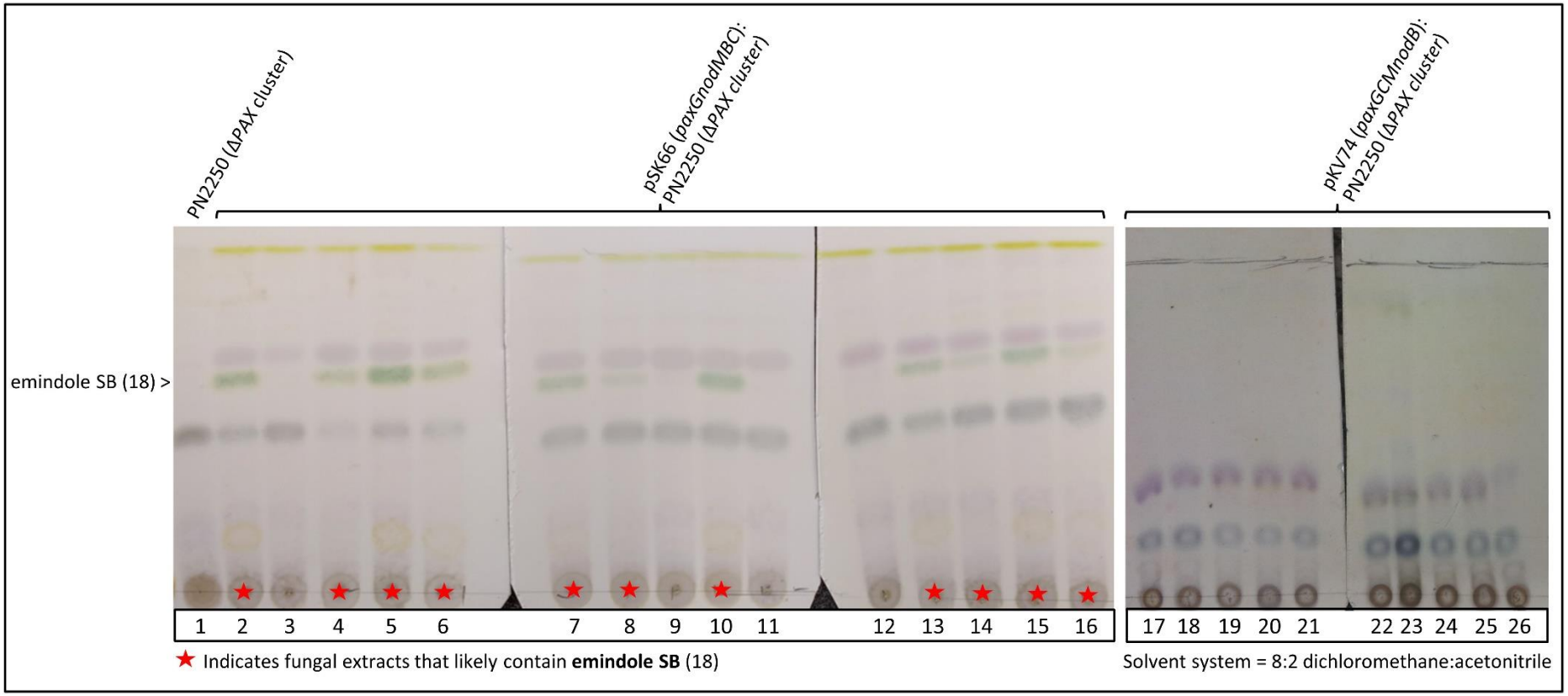


Figure 5.9. TLC results of fungal extracts from the individual transformation of *P. paxilli* Δ*PAX*-cluster deletion mutant PN2250 with two multigene constructs (pSK66 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) and pKV74 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$)). As a negative control, lane 1 contains extract from the *P. paxilli PAX*-cluster deletion mutant PN2250.

Analysis of LCMS results

As expected, LCMS analysis revealed that the pSK66 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$) transformants were indeed producing emindole SB (18) (high resolution-mass spectra 406.3109 *m/z*; Figure 5.10, trace ii) – indicating that NodB was capable of cyclising 13,14-epoxy-3-geranylgeranylindole (16) to form emindole SB (18). Additionally, paspaline (11) was identified in the extracts of the pKV74 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***paxM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***paxC***$_{CDS}$-*paxC*$_{UTRterm}$) transformants (high resolution-mass spectra $[M+H]^+$ = 422.3055 *m/z*; Figure 5.10, trace iii.)– showing that NodB was also capable of cyclising bisepoxy-3-geranylgeranylindole (17) to form paspaline (11).

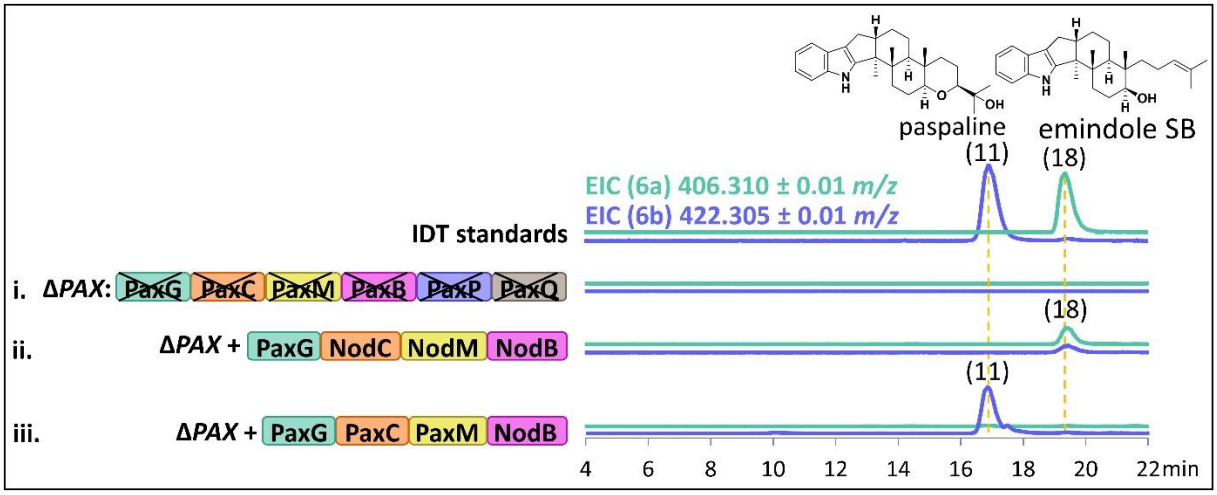


Figure 5.10. Mass spectrometry extracted ion chromatogram traces corresponding to the $[M+H]^+$ ion of emindole SB (18) (406.310 ± 0.01 m/z) in turquoise and paspaline (11) (422.305 ± 0.01 m/z) in purple. Indole diterpene (IDT) standards [emindole SB (18) and paspaline (11)] are depicted at the top, traces from the full Δ*PAX*-cluster deletion mutant (strain PN2250) are depicted in **i**., traces from the *paxG*, *nodCMB* (pSK66) transformant are depicted in **ii.**, and traces from the *paxGCM*, *nodB* (pKV74) transformant are depicted in **iii**. Notably, there is a peak associated to $[M+H]^+$ of 422.305 in trace **ii.** That has same retention time as emindole SB (18) but no compounds with these properties have been characterized.

Conclusions

As suggested by NodB's high level of similarity compared with functionally characterised indole diterpene cyclases, it was confirmed that NodB possesses the same promiscuous capability in that it is able to catalyse the cyclisation of both 13,14-epoxy-3-geranylgeranylindole (16) and bisepoxy-3-geranylgeranylindole (17) to respectively form emindole SB (18) and paspaline (11). Therefore, NodB is a functional orthologue of PaxB and the other functionally confirmed indole diterpene cyclases that share high amino acid sequence identity. This shows that NodB and NodM do not work in conjunction to streamline nodulisporic acid biosynthesis; rather, by only epoxidising geranylgeranylindole at the C14-C15 alkene, NodM prevents the formation of shunt products [e.g. paspaline (11)] by NodB and is thereby solely responsible for streamlining nodulisporic acid biosynthesis.

6. NodW, THE FUNNEL TO NODULISPORIC ACIDS

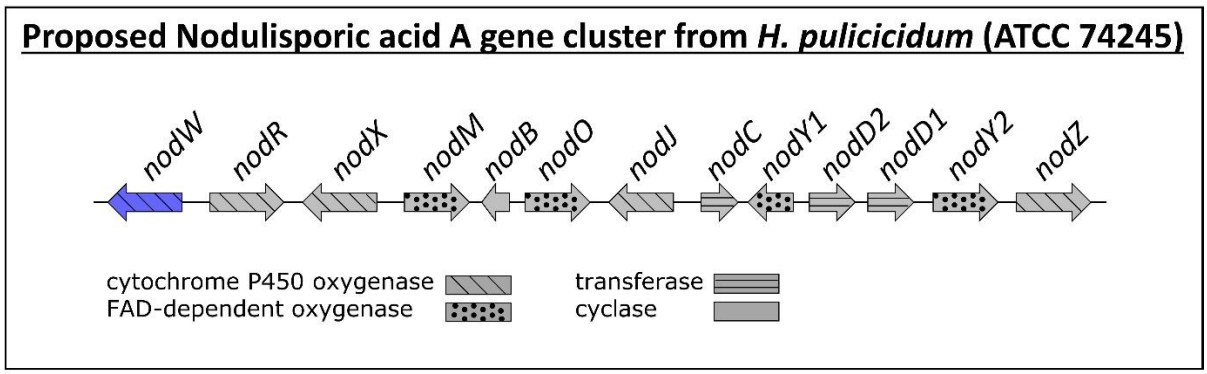


Figure 6.1. Depiction of the predicted nodulisporic acid gene cluster (*NOD*) from *H. pulicicidum* (ATCC® 74245™). The *nodW* gene is highlighted in purple.

Review of post cyclisation oxidations by cytochrome P450 oxygenases in indole diterpene biosynthesis

After cyclisation (i.e. step four of indole diterpene biosynthesis), structurally complex indole diterpenes are typically oxidised in a two-step manner by two separate cytochrome P450 oxygenases (Figure 6.2). For example, in the case of paxilline (2) biosynthesis by *P. paxilli*, the cyclised core, paspaline (11), is oxidised first by PaxP to form 9-desoxypaxilline (27) and then by PaxQ to form paxilline.[36, 40] This same pattern is not only seen in the biosynthesis of paxilline-derived indole diterpenes like penitrems (*P. simplicissimum*),[31] but also in the biosynthesis of non-paxilline-derived indole diterpenes like the aflatrems (*A. flavus*),[28, 41] shearinines (*Penicillium janthinellum*),[32, 43] terpendoles (*Chaunopycnis alba*)[30] and lolitrems (*Epichloë festucae*).[29]

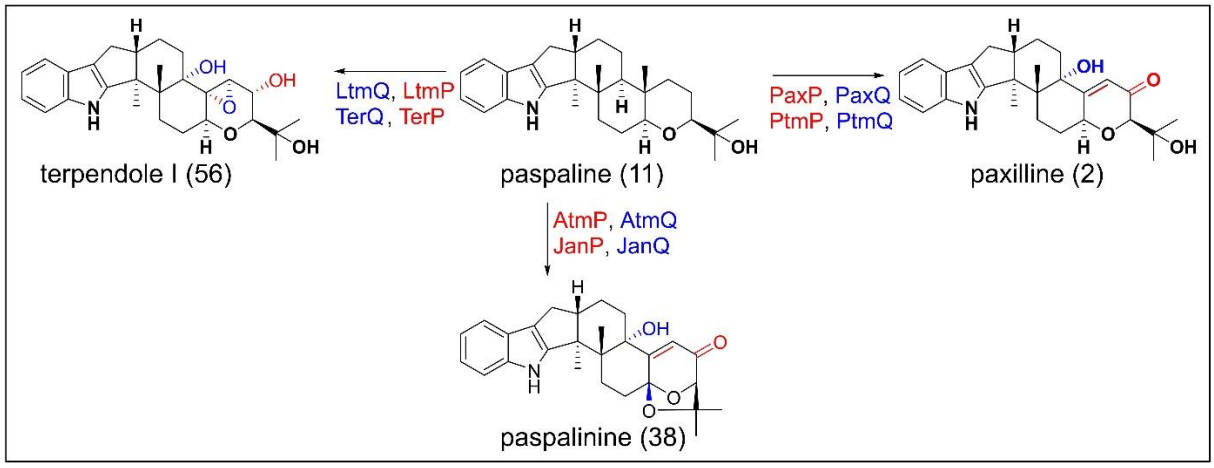


Figure 6.2. Depiction of the typical two-step oxidation that occurs in indole diterpene biosynthesis post-cyclisation.[28-29, 31-32, 36, 40-41, 43]

The *NOD* cluster lacks highly homologous cytochrome P450 oxygenases

When considering the biosynthesis of nodulisporic acids it was predicted that the cyclised core compound, emindole SB (18), was oxidised at one of its two terminal methyl carbons to form nodulisporic acid F (82) (Figure 2.3). The oxidation of a non-functional terminal carbon to form the carboxylic acid moiety of nodulisporic acid F (82) is consistent with the involvement of a cytochrome P450 oxygenase, but no indole diterpene cytochrome P450 oxygenases have ever been functionally confirmed that

perform this type of oxidation and thus it was unclear whether only one cytochrome P450 oxygenase was responsible for installing the carboxylic acid moiety or whether multiple cytochrome P450 oxygenases were involved. Additionally, because no other indole diterpene cytochrome P450 oxygenases had been identified with this type of functionality, sequence homology screening could not be used to delineate what cytochrome P450 oxygenase(s) within the five cytochrome P450 oxygenases in the *NOD* gene cluster (*nodW*, *nodR*, *nodX*, *nodJ*, and *nodZ*) were responsible for this oxidation (Figure 6.3). Therefore, an experiment was designed to test all five of the cytochrome P450 oxygenases for their ability to catalyse the oxidation of emindole SB (18) to form nodulisporic acid F (82).

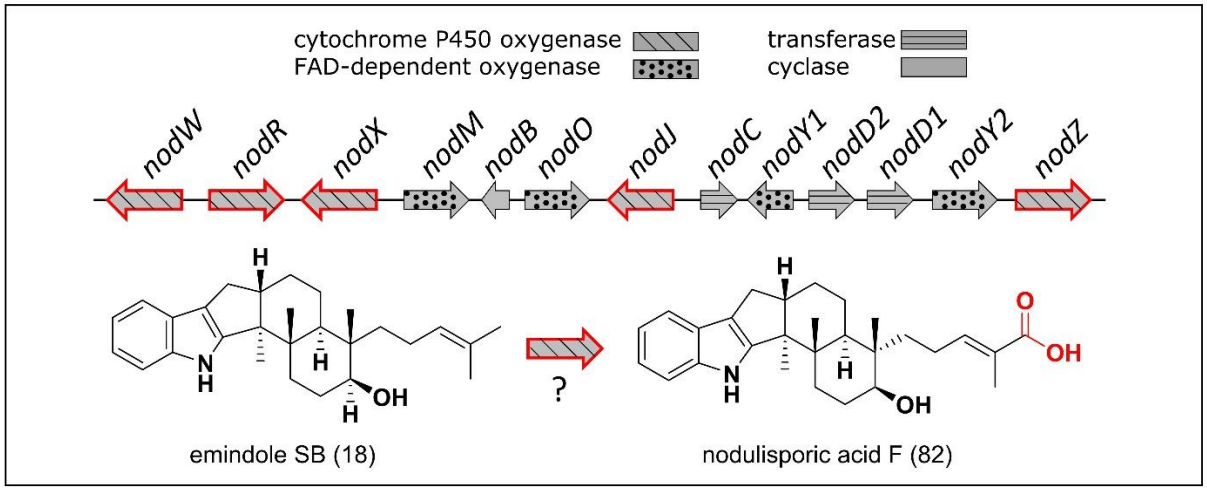


Figure 6.3. Depiction of the *NOD* cluster with the five-predicted cytochrome P450 oxygenases outlined in red and a schematic of the oxidation of emindole SB (18) to form nodulisporic acid F (82).

P450 analysis: An approach to functionally confirm non-homologous genes

During the gene functionality analysis of *nodM* it was confirmed that introduction of *nodM* into the *P. paxilli* Δ*paxM* deletion mutant resulted in production of emindole SB (18), the substrate required for functional analysis of the *NOD* cluster cytochrome P450 oxygenases. In addition, it was demonstrated that when *nodM* was driven by the $paxM_{ProUTR}$ promoter region it produced more emindole SB (18) than when it was driven by the $trpC_{ProUTR}$ promoter region. Therefore, in order to test the five *NOD*

cytochrome P450 oxygenases, each of the genes predicted to encode a cytochrome P450 oxygenase was assembled into a separate multigene construct [pKV64 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$), pKV65 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodR***$_{CDS}$-*trpC*$_{UTRterm}$), pKV66 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodX***$_{CDS}$-*trpC*$_{UTRterm}$), pKV67 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodJ***$_{CDS}$-*trpC*$_{UTRterm}$), and pKV68 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodZ***$_{CDS}$-*trpC*$_{UTRterm}$)] that contained the *nodM*$_{CDS}$ coding sequence driven by the *paxM*$_{ProUTR}$ promoter region. Due to the versatility of the MIDAS system, the same Level-3 geneticin resistance cassette-containing/*nodM*-containing destination vector [pKV63 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$)] was used for the assembly of each cytochrome P450 oxygenase fungal-transformation vector. Since the destination vector (pKV63) required insertion of a white level 2 vector, five white MIDAS Level-2 vectors, each containing a separate *NOD* cytochrome P450 oxygenase coding sequence were assembled with the *trpC*$_{ProUTR}$ promoter and *trpC*$_{UTRterm}$ terminator regions. The Level-2 vectors were designed to incorporate each cytochrome P450 oxygenase coding sequence into the final destination vector in the same orientation that it has in the *NOD* cluster. Therefore, *nodW*, *nodR* and *nodJ* were assembled into white forward Level-2 destination vectors and *nodX* and *nodZ* were assembled into white reverse Level-2 destination vectors (Figure 6.4).

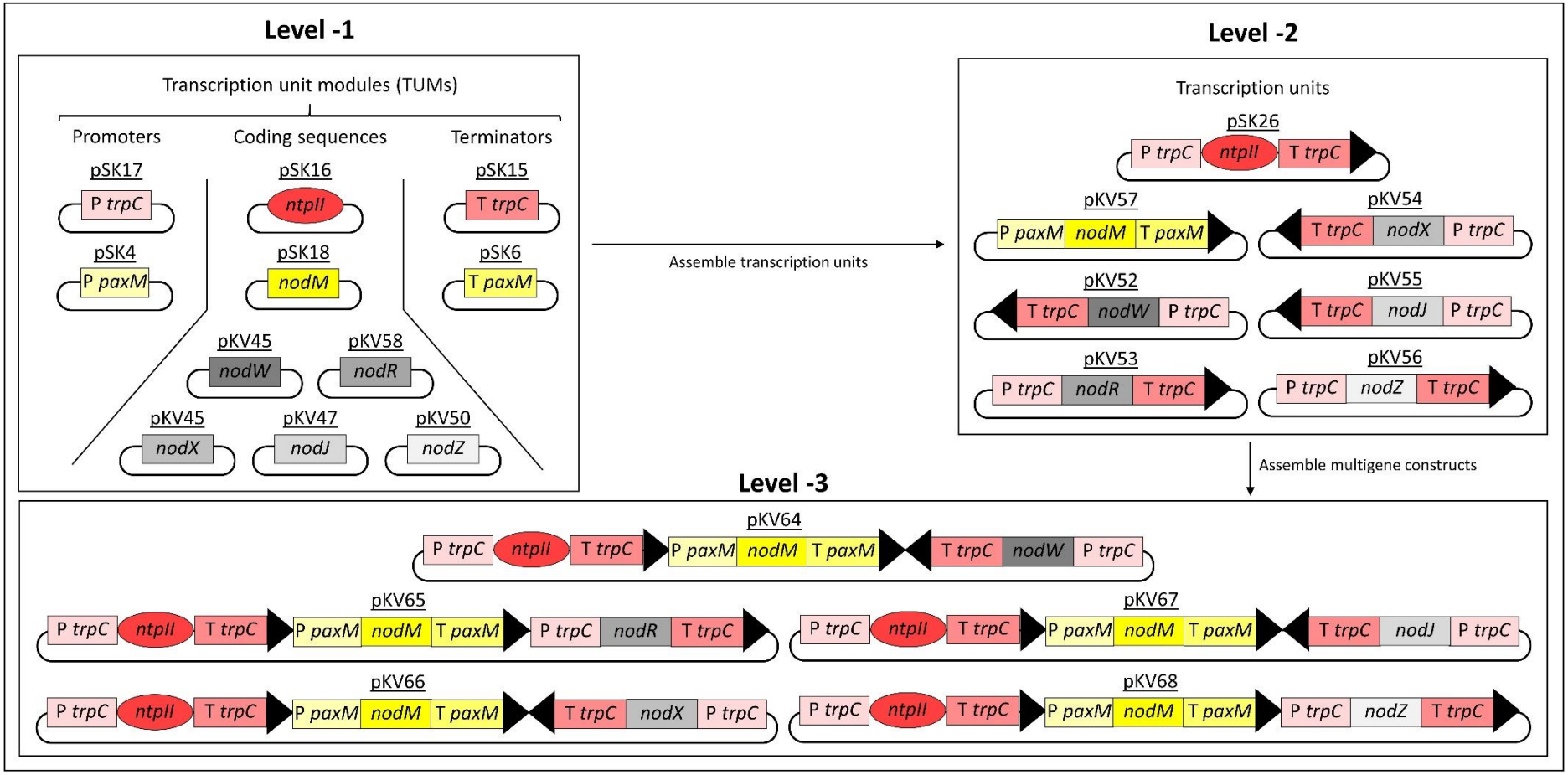


Figure 6.4. Depiction of the MIDAS approach used to assemble the five multigene constructs (pKV64, pKV65, pKV66, pKV67, and pKV68) used to identify the cytochrome P450 involved in the oxidation of emindole SB (18) to form nodulisporic acid F (82). In Level-1, the plasmids encoding transcription unit modules (promoters, coding sequences, terminators) are shown. In Level-2 the plasmids that encode the five full length transcription units are shown –where the arrow represents the direction of transcription. In Level-3, the final multigene plasmids that was used for fungal transformation are depicted.

Transformation of *P. paxilli* with P450-encoded vectors

Protoplasts of the *P. paxilli* Δ*paxM* deletion mutant strain PN2257 were transformed individually with each of the five multigene P450-encoding Level-3 vectors. Ten transformants from each transformation were picked and fermented in production medium. Initial TLC analysis of all of the transformant extracts showed that emindole SB (18) was present in in a selection of extracts from every transformation (Figure 6.5) – showing that *nodM* had been successfully incorporated and expressed thereby suggesting that the cytochrome P450 oxygenases were also successfully incorporated into the genomes of these transformants. Since the TLC analysis showed that there was a wide range of variation in compound production of the transformants, a selection of three extracts from each transformation was chosen for further LCMS analysis.

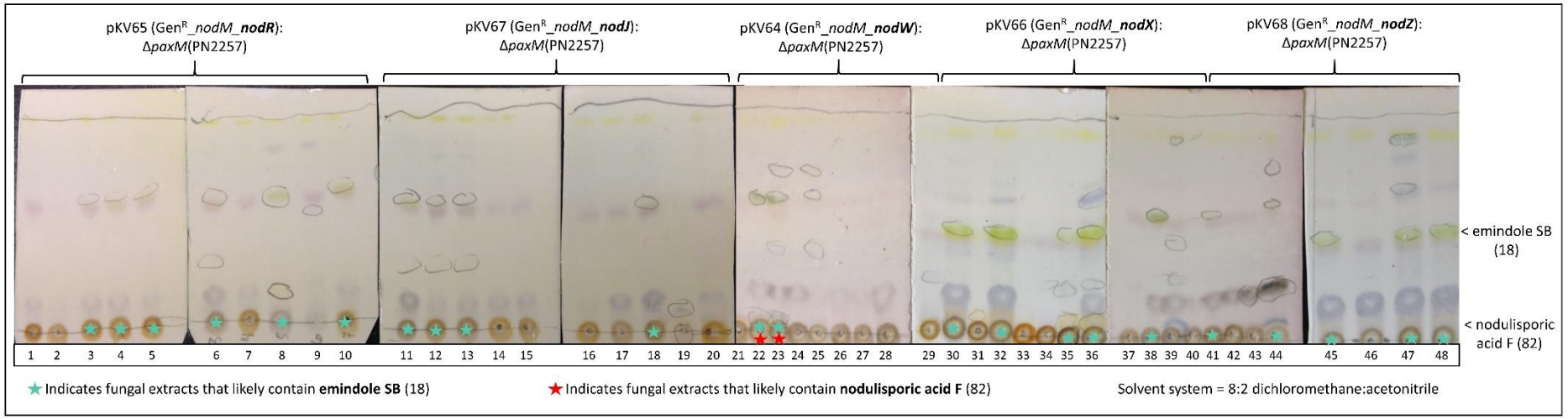


Figure 6.5. TLC results of fungal extracts from the individual transformations of *P. paxilli* Δ*paxM* deletion mutant PN2257 with five multigene constructs [pKV64 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$), pKV65 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodR***$_{CDS}$-*trpC*$_{UTRterm}$), pKV66 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodX***$_{CDS}$-*trpC*$_{UTRterm}$), pKV67 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodJ***$_{CDS}$-*trpC*$_{UTRterm}$), and pKV68 (*trpC*$_{ProUTR}$-*ntpII*$_{CDS}$-*trpC*$_{UTRterm}$_*paxM*$_{ProUTR}$-*nodM*$_{CDS}$-*paxM*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodZ***$_{CDS}$-*trpC*$_{UTRterm}$)].

Identification of nodulisporic acid F (82)

The LCMS analysis of the selection of extracts tested from each transformation revealed that emindole SB (18) was present in transformants from every *nodM*/P450 transformation (Figure 6.6, traces iii.-vii.). Importantly, a novel indole diterpene compound with a mass corresponding to that of nodulisporic acid F (82) (high resolution-mass spectra $[M+H]^+$ = 436.2870 *m/z*, r.t. 5.6 min) was present in the extracts of only the *nodW*-harbouring transformant (Figure 6.6, trace iii.). The identity of the novel compound was confirmed by NMR spectrometry as nodulisporic acid F (82). A table of the ^{1}H and ^{13}C NMR assignments are shown in the Appendix in Table 9.9 and the corresponding 1-dimensional and 2-dimensional NMR spectra are shown in Figures 9.22 to 9.25. The production of nodulisporic acid F (82) by the the *nodW*-harbouring transformant indicates that a single cytochrome P450 oxygenase (NodW) catalyses the oxidation of the terminal carbon on emindole SB (18) to form the reactive carboxylicic acid handle of nodulisporic acid F (82).

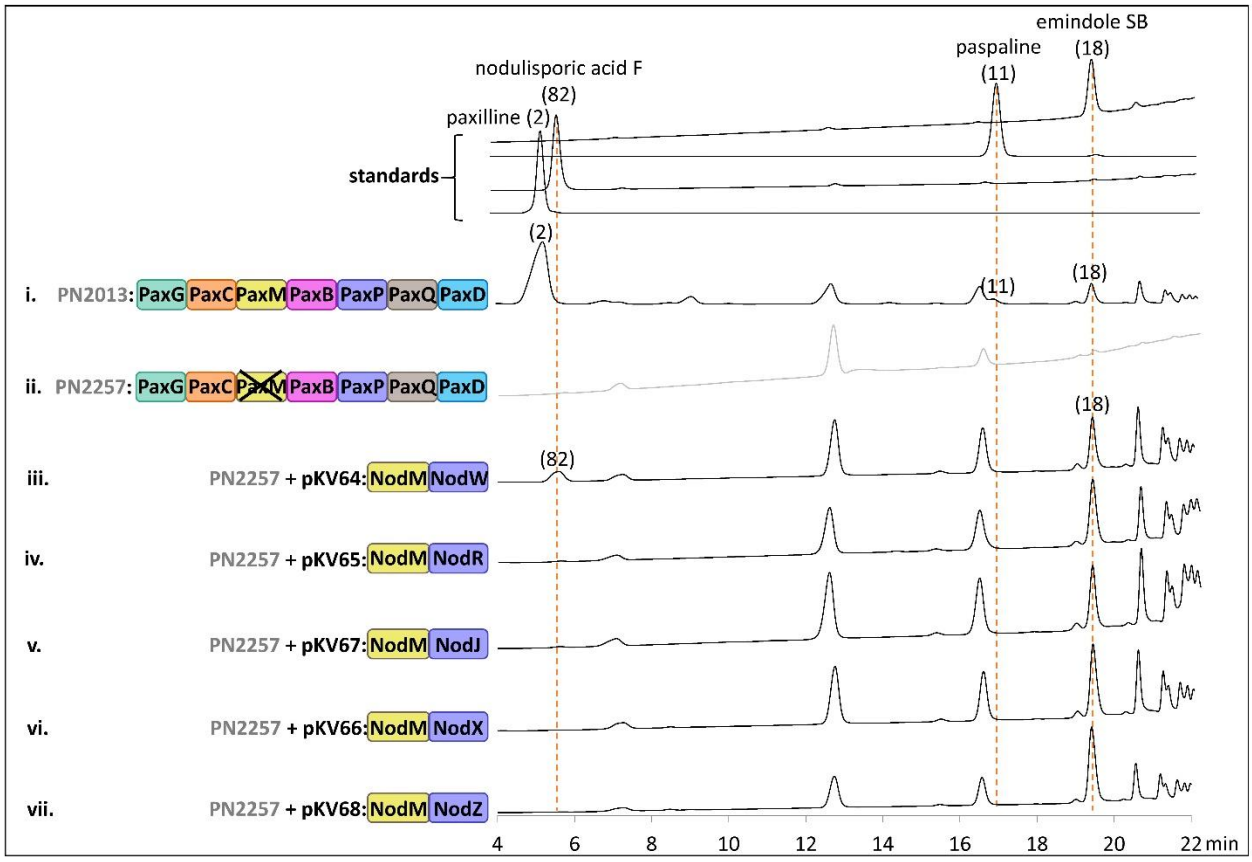


Figure 6.6. HPLC analysis (271 nm) of fungal extracts from the *P. paxilli* wild-type strain (PN2013) (trace i.), *P. paxilli* Δ*paxM* deletion mutant strain, PN2257 (trace ii.), pKV64 transformed *P. paxilli* Δ*paxM* deletion mutant strain (pKV64:PN2257) expressing the NodM and NodW enzymes from *H. pulicicidum* (trace iii.), pKV65 transformed *P. paxilli* Δ*paxM* deletion mutant strain (pKV65:PN2257) expressing the NodM and NodW enzymes from *H. pulicicidum* (trace iv.), pKV67 transformed *P. paxilli* Δ*paxM* deletion mutant strain (pKV67:PN2257) expressing the NodM and NodJ enzymes from *H. pulicicidum* (trace v.), pKV66 transformed *P. paxilli* Δ*paxM* deletion mutant strain (pKV66:PN2257) expressing the NodM and NodX enzymes from *H. pulicicidum* (trace vi.), and pKV68 transformed *P. paxilli* Δ*paxM* deletion mutant strain (pKV68:PN2257) expressing the NodM and NodZ enzymes from *H. pulicicidum* (trace vii.). Extracted ion chromatograms are located in the Appendix (Figures 9.8 to 9.11).

Confirmation that five genes are required for nodulisporic acid biosynthesis

We were aware that the Δ*paxM* deletion mutant strain PN2257 contained decoration genes (i.e. *paxP*, *paxQ*, *paxD*, and *paxO*) that could have influenced to the production of nodulisporic acid F (82) and wanted to ensure that none of these genes were involved in

its biosynthesis. Therefore, to establish that only five genes were involved in nodulisporic acid F (82) biosynthesis a six-gene construct encoding *ntpII*, *paxG*, *nodC*, *nodM*, *nodB* and *nodW* [pSK68 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$), Figure 6.7] was assembled and inserted it into the entire Δ*PAX*-cluster deletion mutant strain PN2250. With the exception of the promoter region for *nodW*, native *PAX* promoter regions were used to drive their respective genes or gene homologues for *paxG*, *nodC*, *nodM* and *nodB*. As the *trpC*$_{ProUTR}$ promoter region was used in the initial testing of *nodW*, and the strength of the promoter regions in the promoter library had not yet been assessed, the *trpC* promoter region was selected to drive *nodW* in pSK68. Transformation of the *P. paxilli PAX* deletion mutant strain PN2250 with this six-gene construct (pSK68) potentially established nodulisporic acid F (82) production in at least two of six transformants, as confirmed by TLC analysis (Figure 6.8). Subsequent LCMS analysis revealed that nodulisporic acid F (82) was only present in one of the transformant extracts and that it was only present in very low amounts (Figure 6.9). Therefore, an attempt was made to enhance nodulisporic acid F (82) biosynthesis by changing the promoter region used to drive the *nodW*$_{CDS}$ coding sequence.

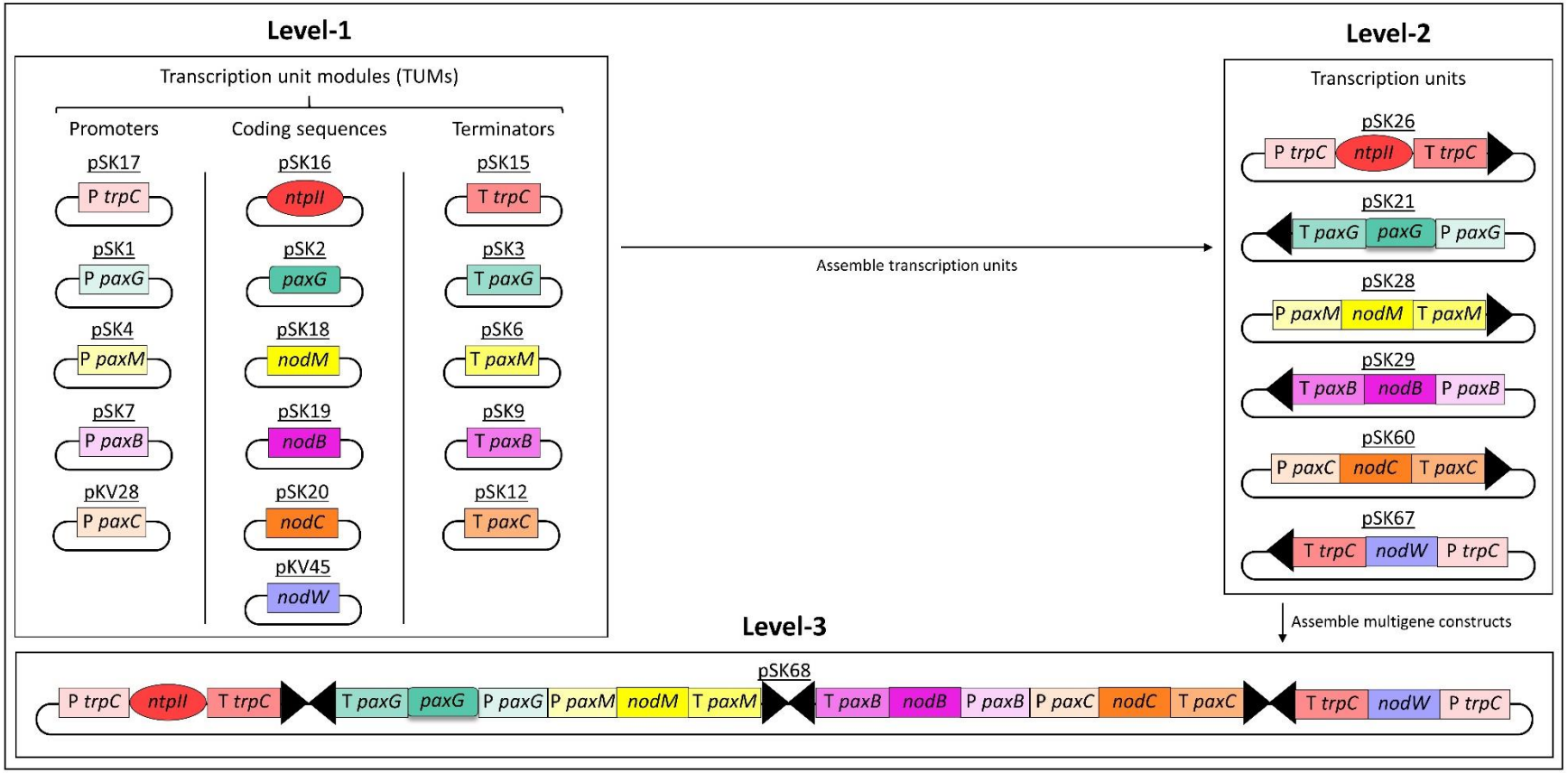


Figure 6.7. Depiction of the MIDAS approach used to assemble the multigene construct pSK68 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$). In Level-1, the plasmids encoding transcription unit modules (promoters, coding sequences, terminators) are shown. In Level-2 the plasmids that encode the five full length transcription units are shown –where the arrow represents the direction of transcription. In Level-3, the final multigene plasmid (pSK68) that was used for fungal transformation is depicted.

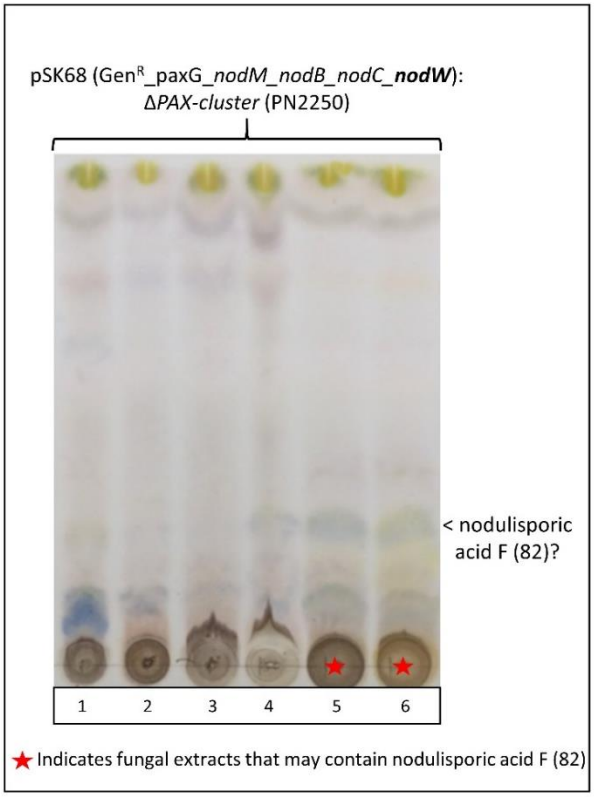


Figure 6.8. TLC results of fungal extracts from the transformation of *P. paxilli* Δ*PAX*-cluster deletion mutant PN2250 with the multigene construct pSK68 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$).

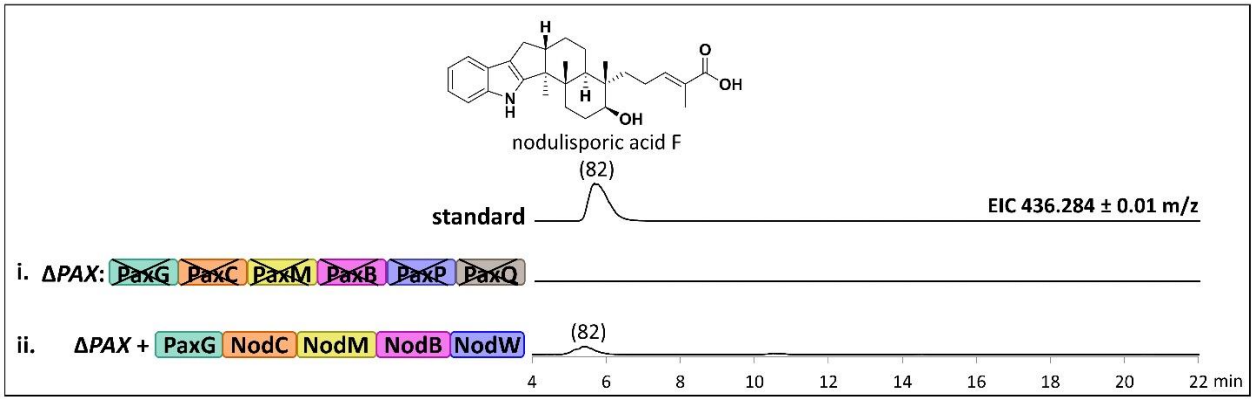


Figure 6.9. Mass spectrometry extracted ion chromatogram (EIC) traces corresponding to the $[M+H]^+$ ion of nodulisporic acid F (82) (436.284 ± 0.01 m/z). The EIC trace of the nodulisporic acid F (82) standard is depicted at the top, the EIC of extract from the full Δ*PAX*-cluster deletion mutant (strain PN2250) is depicted in trace **i**., and the EIC of extract from the *paxG*, *nodCMBW* (pSK68) transformant is depicted in **ii**.

Enhancing nodulisporic acid F (82) biosynthesis

To try to enhance the biosynthesis of nodulisporic acid F (82) a new six-gene construct was assembled that directly resembled pSK68 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$_*trpC*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$) with the exception of the promoter region used to drive the *nodW*$_{CDS}$ coding sequence (Figure 6.10). In this new six-gene construct, pSK81 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$_*janD*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$), *nodW* was driven by the *janD*$_{ProUTR}$ promoter region. Upon assembly, pSK81 was inserted into protoplasts of the entire *PAX* deletion mutant strain PN2250. The TLC analysis of fifteen transformants (Figure 6.11) showed that at least six (lanes 1, 3, 4, 8, 10, and 12) produced nodulisporic acid F (82) and that nodulisporic acid F (82) was produced in higher concentrations than that of the pSK68 transformants (See Figure 6.8 for analysis of pSK68). Subsequent LCMS analysis of the pSK81 transformants (Figure 6.12) also showed much higher concentrations of nodulisporic acid F (82) than were found in the pSK68 transformants (the nodulisporic acid F (82) in the extracts of the pSK68 transformants was not observed in the UV trace and only detectable by the mass spectrometer) - confirming that promoter swapping of a single gene in a multi-gene construct could greatly enhance compound biosynthesis.

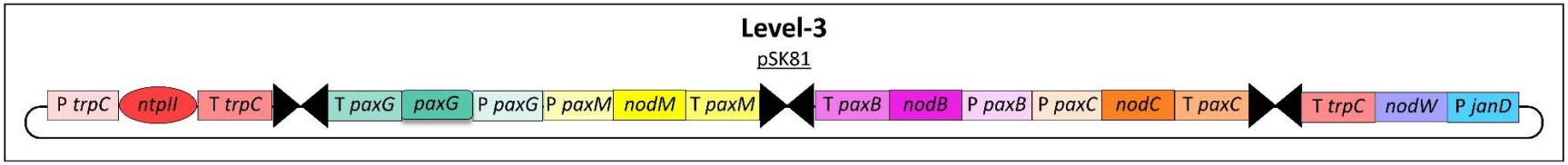


Figure 6.10. Representation of the six-gene construct pSK81 used to test how promoter region swapping (the *trpC*$_{ProUTR}$ promoter region in pSK68 to the *janD*$_{ProUTR}$ promoter region in pSK81) of a single gene (*nodW*) in this construct would affect biosynthesis of nodulisporic acid F (82).

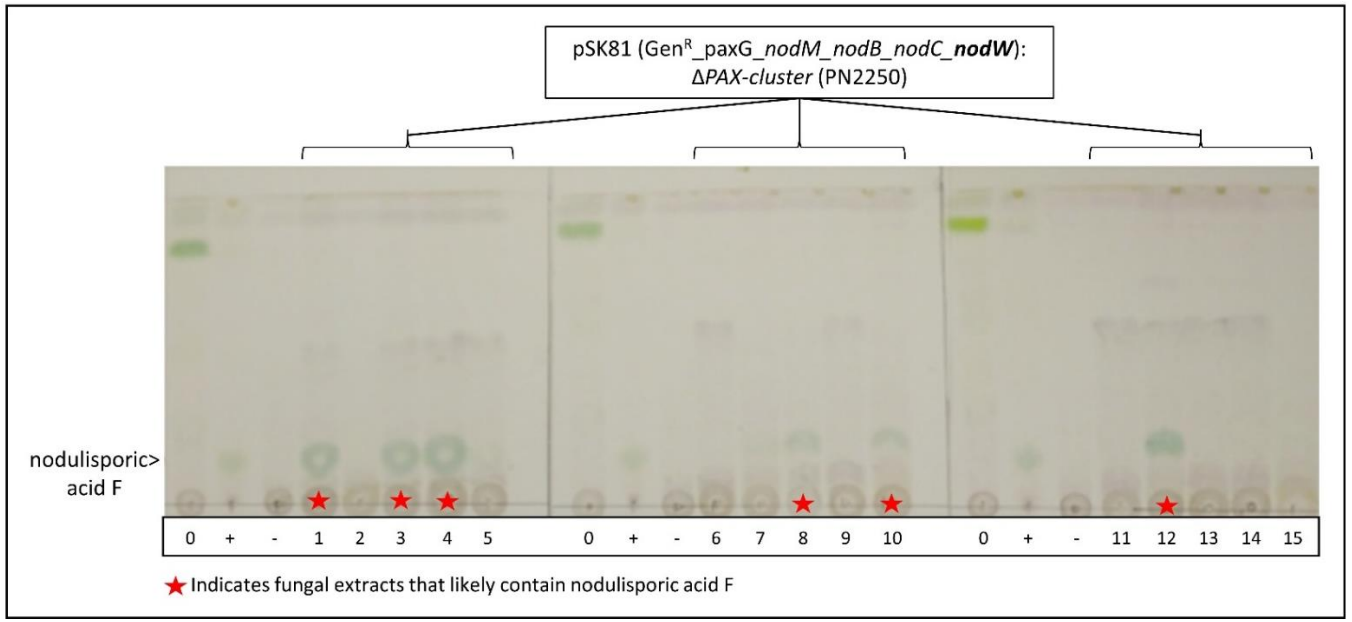


Figure 6.11. TLC results of fungal extracts from the transformation of *P. paxilli* Δ*PAX*-cluster deletion mutant PN2250 with the multigene construct pSK81 (*trpC*$_{ProUTR}$-***ntpII***$_{CDS}$-*trpC*$_{UTRterm}$_*paxG*$_{ProUTR}$-***paxG***$_{CDS}$-*paxG*$_{UTRterm}$_*paxM*$_{ProUTR}$-***nodM***$_{CDS}$-*paxM*$_{UTRterm}$_*paxB*$_{ProUTR}$-***nodB***$_{CDS}$-*paxB*$_{UTRterm}$_*paxC*$_{ProUTR}$-***nodC***$_{CDS}$-*paxC*$_{UTRterm}$_*janD*$_{ProUTR}$-***nodW***$_{CDS}$-*trpC*$_{UTRterm}$). As controls, extract from the *P. paxilli* wildtype strain is shown in lanes with a (0), the nodulisporic acid F (82) standard is shown in lanes with a (+), and extract from the *P. paxilli* Δ*PAX*-cluster deletion mutant PN2250 is shown in lanes with a (-).

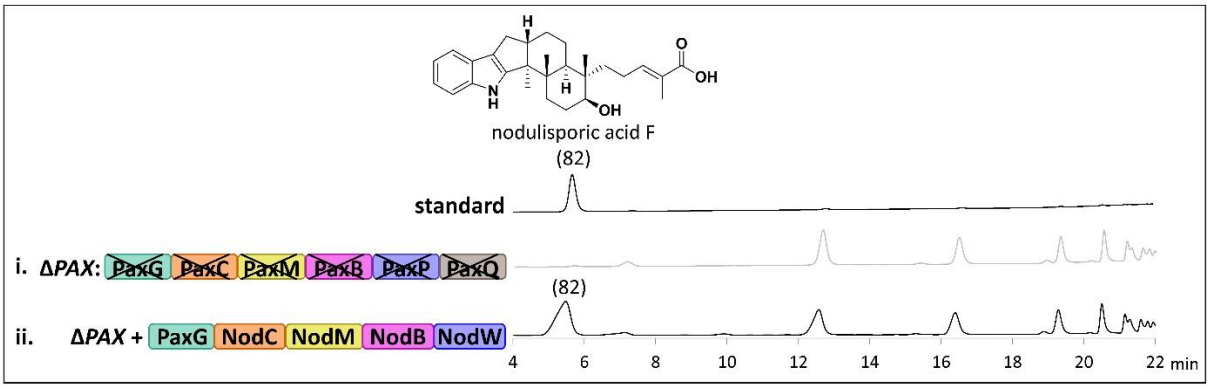


Figure 6.12. HPLC analysis (271 nm) of fungal extracts from the *P. paxilli* Δ*PAX*-cluster deletion mutant strain, PN2250 (trace i), and one of the pSK81 (*paxG*, *nodCMBW*) transformants (trace ii.). The top trace corresponds to the nodulisporic acid F (82) standard. Extracted ion chromatograms are located in the Appendix (Figure 9.11).

Conclusions

Overall, this work confirms that five secondary metabolic genes (*paxG*, *nodC*, *nodM*, *nodB* and *nodW*) are involved in the heterologous biosynthesis of nodulisporic acid F (82) by *P. paxilli* (Figure 6.13). Importantly, previous studies of the nodulisporic acid pharmacophore established that the carboxylic acid handle of nodulisporic acids is readily modified using chemical synthesis to enhance the bioactivity and insecticidal specificity of the natural compounds.[11, 54-55, 60] Thus, the identification of *nodW* provides access to a key modifiable core structure, nodulisporic acid F (82), for exploration of novel indole diterpenes. Additionally, the functional confirmation of the $nodW_{CDS}$ coding sequence establishes a new subclass of indole diterpene cytochrome P450 oxygenases and will facilitate the identification of other oxygenases that may modify the indole diterpene cores in a similar but discrete manner. Furthermore, it was demonstrated that simple promoter region exchange of a single coding sequence within a multi-gene construct can dramatically enhance the biosynthesis of nodulisporic acid F (82). Most important of all though is that nodulisporic acid F (82) was produced in one week, in the light and in relatively nutrient poor medium showing that heterologous expression is a powerful method to bypass the regulatory mechanisms found in the native nodulisporic acid producer, *H. pulicicidum*.

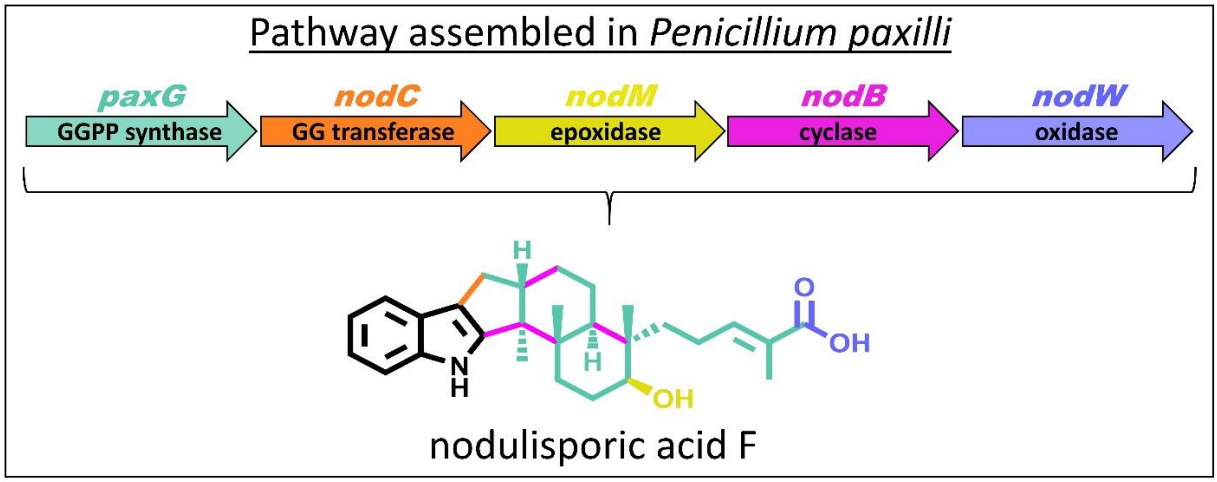


Figure 6.13. Summary of the biosynthetic genes involved in the heterologous biosynthesis of nodulisporic acid F (82) by *P. paxilli* (Van de Bittner *et al*. 2018).[1]

7. CONCLUSIONS

Key findings and results

In summary, the genome of *Hypoxylon pulicicidum* (ATCC® 74245™)[2] has been sequenced and sequence homology comparisons have been used to locate a cluster of thirteen genes, named the *NOD* cluster, which as described in this thesis, is responsible for nodulisporic acid biosynthesis. Using the efficient gene reassembly of the Modular Idempotent DNA Assembly System (MIDAS)[3] in combination with heterologous expression in a high producing indole diterpene species, *Penicillium paxilli*, the functions of four of the thirteen genes (*nodC*, *nodM*, *nodB* and *nodW*) within this cluster have been confirmed. These four genes encode the enzymes that catalyse the second through fifth secondary-metabolic steps of nodulisporic acid biosynthesis shown in Figure 7.1: *nodC* encodes the geranylgeranyl transferase (NodC) that catalyses the condensation of geranylgeranyl pyrophosphate (3) and indole-3-glycerol phosphate (4) to form 3-geranylgeranylindole (13), *nodM* encodes a flavin adenine dinucleotide monooxygenase (NodM) that catalyses the mono-epoxidation of 3-geranylgeranylindole (13) at the C13-C14 alkene to form 13,14-epoxy-3-geranylgeranylindole (16), *nodB* encodes an indole diterpene cyclase (NodB) that catalyses the cyclisation of 13,14-epoxy-3-geranylgeranylindole (16) into emindole SB (18), and *nodW* encodes a cytochrome P450 oxygenase (NodW) that catalyses the oxidation of emindole SB (18) to form nodulisporic acid F (82). Of these four genes, two – *nodM* and *nodW* – had functions that were previously unreported and are specifically involved in streamlining nodulisporic acid biosynthesis.

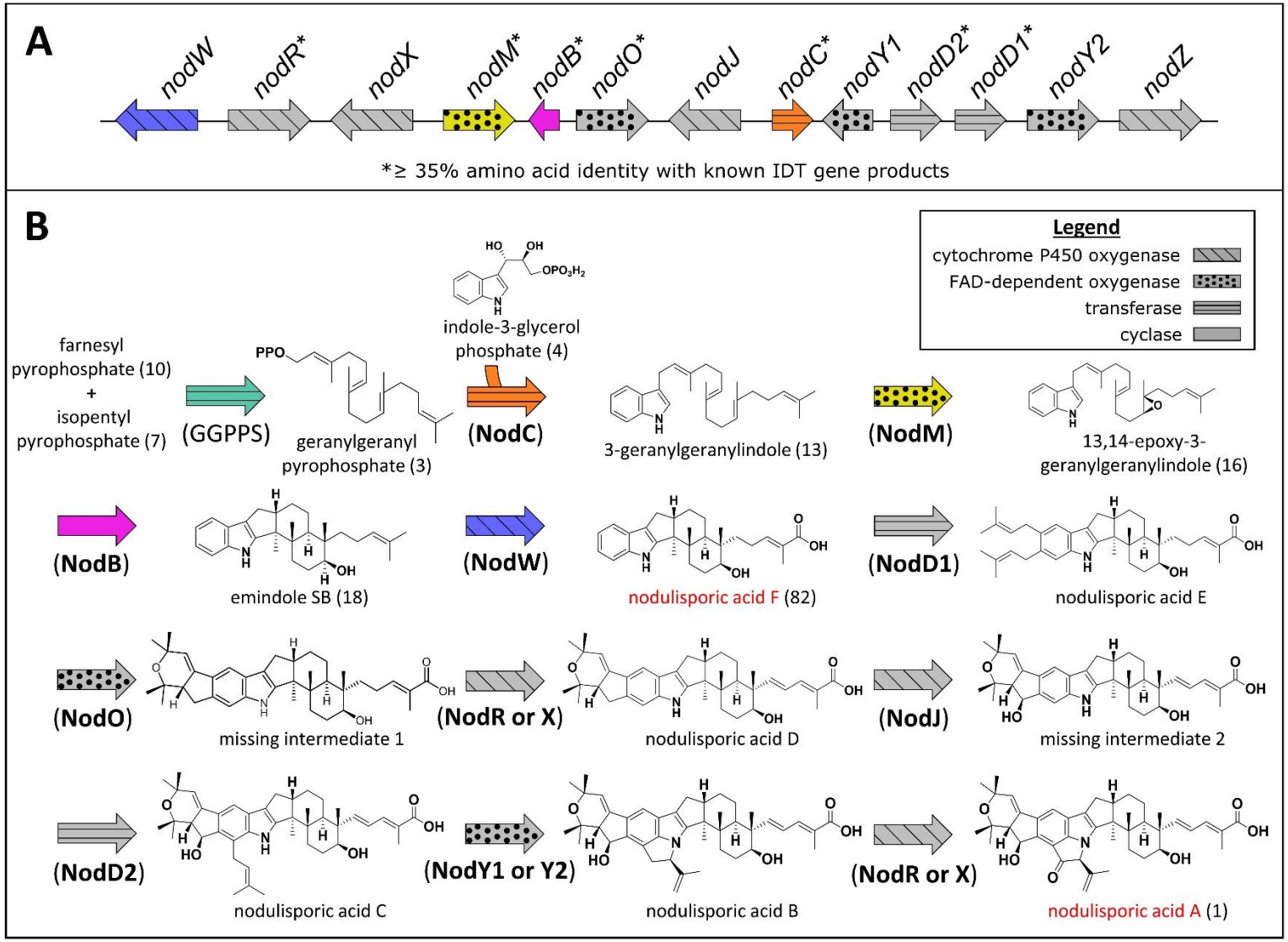


Figure 7.1. The nodulisporic acid gene cluster from *Hypoxylon pulicicidum* (**A**) and the predicted nodulisporic acid A (1) biosynthetic pathway (**B**). The coloured genes (arrows) have been functionally characterised.

By introducing *nodM* into the *P. paxilli* Δ*paxM* deletion mutant it was demonstrated that NodM specifically epoxidised 3-geranylgeranylindole (13) at the C13-C14 alkene and not at the C17-C18 alkene, a characteristic that is not shared with any of its functionally characterised homologues [*paxM* (*P. paxilli*)[40], *atmM* (*A. flavus*),[33] and *ltmM* (*E. festucae*)[26]]. In turn, by epoxidizing 3-geranylgeranylindole (13) at the one specific location essential for nodulisporic aicd biosynthesis, *nodM* prevents the diversion of resources into shunt products (e.g. paspaline) that are not essential to the nodulisporic acid pathway. In a similar manner, it was demonstrated that *nodW* installs emindole SB (18) with the distinctive carboxylic acid moiety of nodulisporic acids to form nodulisporic acid F (82). As nodulisporic acid F (82) is the first nodulisporic acid in the pathway, this oxidation step acts as a 'funnel' for the biosynthesis of nodulisporic acids. An oxidation of this kind on an indole diterpene core has, until now, never been functionally characterised, and thus this discovery establishes a new subclass of indole diterpene cytochrome P450 oxygenases that will facilitate the identification of other oxygenases that may modify indole diterpene cores in a similar but discrete manner.

Notably, the cluster lacks a gene that encodes a secondary-metabolic geranylgeranyl pyrophosphate (3) synthase to catalyse the first secondary metabolic step of nodulisporic acids, the production geranylgeranyl pyrophosphate (3). The absence of a secondary-metabolic geranylgeranyl pyrophosphate (3) synthase provides a possible explanation as to why *H. pulicicidum* produces nodulisporic acids in such low yield. The low production is a problem for both resolving gene function and useful generation of the compounds. Heterologous expression has provided a way to overcome both of these problems as it was demonstrated that nodulisporic acid F (82) can be produced in a third of the time (one week in *P. paxilli* vs. 3 weeks in *H. pulicicidum*)[50] and without using the strict growth protocols required to elicit nodulisporic acid production in *H. pulicicidum* (i.e. *H. pulicicidum* cultures cannot be exposed to the light and required relatively nutrient-rich medium, while *P. paxilli* cultures can be exposed to light and grown in relatively nutrient-poor medium). Additionally, by exchanging promoter regions of the *NOD* gene coding sequences it was demonstrated that nodulisporic acid biosynthesis can be enhanced, further providing evidence that heterologous expression is a promising means to gain access to valuable compounds.

Future work: Functional analysis of the remaining *NOD* genes

From here, efforts will be dedicated to confirming the function of the remaining genes within the *NOD* cluster involved in the biosynthesis of nodulisporic acid A (1). The biosynthetic pathway of nodulisporic acid F (82) that was elucidated as part of this thesis provides a foundation for the relatively straightforward identification of the "decoration" steps to make nodulisporic acid A (1). Already, using gene product homology comparisons (see Appendix Figures 9.1 to 9.7, Table 9.1), each gene has been ascribed to its most likely step within the biosynthesis of nodulisporic acid A (1) as shown in Figure 7.1, and functionality testing will be carried out using a bottom-up and top-down approach. The bottom-up approach will follow the previous gene functional analysis studies in this thesis, where multigene constructs will be assembled and introduced into either the Δ*paxM P. paxilli* deletion mutant strain PN2257 or the Δ*PAX*-cluster *P. paxilli* deletion mutant strain PN2250. For the top-down approach, a MIDAS-modified bacterial artificial chromosome has been developed that will be used to assemble and introduce the entire *NOD* cluster into the Δ*PAX*-cluster *P. paxilli* mutant strain PN2250.

Along side the functional characterisation of genes involved in the nodulisporic acid A (1) biosynthetic pathway, efforts will be directed toward quantifying titres and optimising the expression host system to produce the highest possible yields of nodulisporic acid A (1). Many steps will be involved in yield-optimisation and some promising places to explore include the exchange of promoter regions, combinatorial anaylsis of different promoters, controlling integration sites, varying the copy number of specific genes, experimenting with different host species, trialling different fermentation conditions, development of quantitative assays for all intermediates in the pathway, and work to understand flux bottlenecks and engineer improved titres using classical metabolic engineering approaches.

Overall, work in this thesis demonstrates the power of using a gene characterisation platform to functionally characterise genes and identify how valuable compounds are produced.

What is the benefit of this work and what will be done with all that has been learned?

As indole diterpenes are well known to have potent insecticidal[7-10, 74-78] and anti-cancer properties,[13, 30, 79-82] they are ideal for application in both the agri-technology and pharmaceutical industries. Fundamentally, this thesis has focused around the development of a powerful biosynthetic platform for the functional analysis of genes to enable commercialization of one valuable indole diterpene, nodulisporic acid A (1), for use as an animal health therapeutic; but that is only the tip of the iceberg. Over 200 indole-diterpene compounds have been identified from biological samples so far, representing different natural combinations of genes with overlapping and distinct gene-encoded chemistry. Approximately 40 secondary metabolic indole diterpene genes have been characterised, which naturally deliver around 80 unique indole diterpene products. The chemical diversity within the indole diterpenes arises from the role of different gene products that modify one general starting block into multiple bioactive products.

The functional confirmation of genes within biosynthetic pathways provides access to natural products not only for derivatisation by synthetic chemical approaches but also through derivatisation by gene recombination approaches. Therefore, upon completion of the functional characterisation of each gene in the nodulisporic acid A (1) biosynthetic pathway, analogues will be generated by mixing-and-matching functionally confirmed genes from different indole diterpene pathways. MIDAS will be used to combine genes from different biosynthetic pathways into Level-3 vectors that will be heterologously expressed in appropriate *P. paxilli* deletion mutants (Figure 7.2). As an example, a theoretical non-natural product is depicted in Figure 7.2 that could be biosynthesised by mixing two decoration genes (*ltmE* and *ltmJ*) from the lolitrem biosynthetic pathway of *Epicloë festucae* with the genes from the nodulisporic acid pathway.

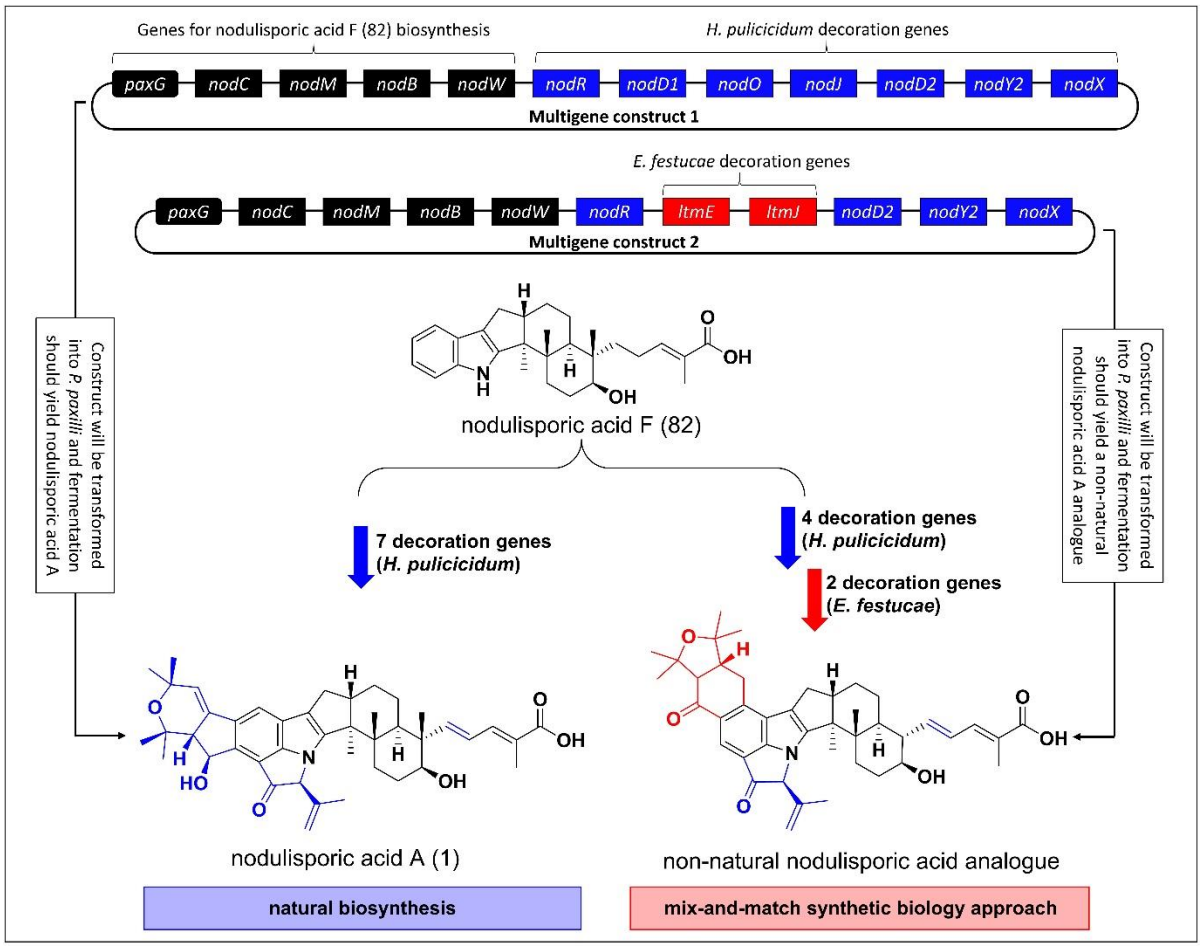


Figure 7.2. Depiction of two theoretical MIDAS-generated multigene constructs for the biosynthesis of a natural product or a non-natural analogue. Multigene construct 1 would be used for the biosynthesis of the natural product nodulisporic acid A (1) and Multigene construct 2 would be used to create a non-natural nodulisporic acid A (1) analogue through a mix-and-match synthetic biology approach.

The opportunity for a metabolic engineering business

As mentioned in a recent review, major efforts are being put forth toward metabolic engineering of microbes for the sustainable production of biofuels, novel chemicals, pharmaceuticals and nutraceuticals.[83] Taken as a whole, the work presented in this thesis provides an opportunity create a metabolic engineering business based around providing natural products and their non-natural derivatives for use across many industries. Accordingly, the general business model would follow a pipeline similar to that presented in Figure 7.3, where biosynthetic gene clusters will be identified, genes will be annotated, MIDAS will be used to assemble the genes into expression vectors, genes will be heterologously expressed, transformants will be chemotyped for novel compounds, novel compounds will be characterised and microbial factories will be

optimised for market. As with any business model, there are risks associated along every step of the pipeline from not being able to locate essential biosynthetic genes to producing chemicals that have no commercial value. Many of these risks can be mitigated by targeting compounds that are clearly valuable (i.e. have useful chemical properties) and belong to a family of compounds with well characterised biosynthetic pathways (i.e. homologous genes sequences are available to identify new genes using sequence comparisons). The prospects are exciting as the tools are now available to exploit efficiently the nearly infinite pool of natural products that nature provides.

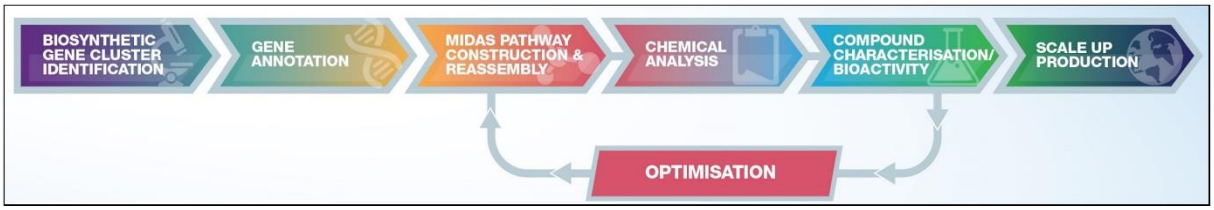


Figure 7.3. Depiction of business model for microbial factories.

8. MATERIALS AND METHODS

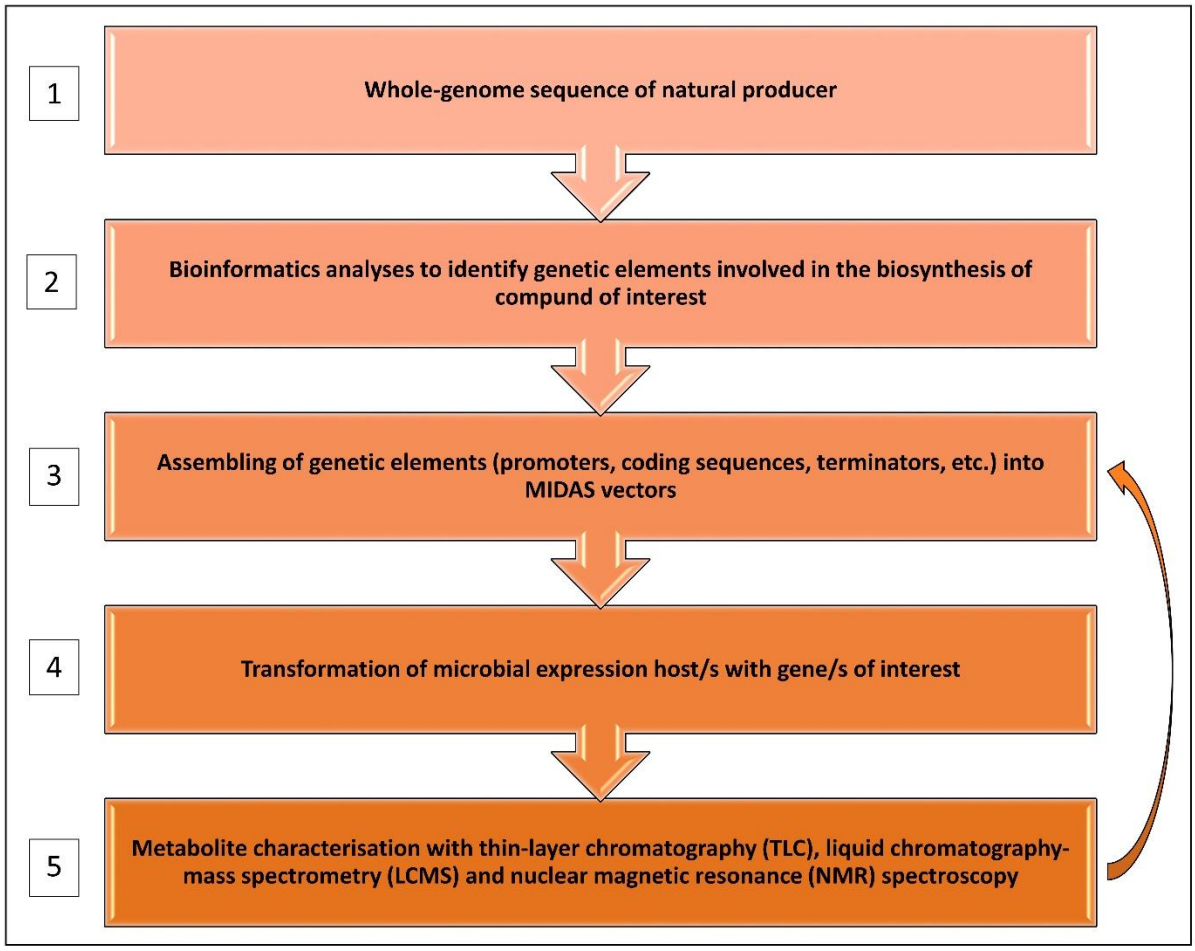


Figure 8.1. Overview of methods used in this thesis.

All of the below methods were described in Van de Bittner *et al*. (2018)[1] or van Dolleweerd *et al*. (2018)[3] and have been listed here for convenience. Some of the methods have been slightly modified from their published versions.

gDNA isolation for genome sequencing and transcription unit module amplification

Initially, genomic DNA for genome sequencing and transcription unit module amplification by PCR was isolated from *Penicillium paxilli* strain ATCC® 26601™ (PN2013)[72] and *Hypoxylon pulicicidum* strain ATCC® 74245™ [84] according to Byrd *et al*.[85] with modifications. Sterile 2.4% (w/v) Difco™ potato dextrose broth (Becton, Dickinson and Company, Maryland, U.S.A.) in Milli-Q® water was prepared in 25 mL aliquots in 125 mL Erlenmeyer flasks and inoculated with 5×10^6 spores or ~1 cm^2 freshly ground mycelia (for non-sporulating strains). Cultures were incubated for 2-4 days at 22°C with shaking (200 rpm). The fermentation broth was filtered through a sterile nappy liner and the mycelia were rinsed three times with sterile water. Mycelia were transferred to a sterile 15 mL centrifuge tube and flash frozen in liquid nitrogen for lyophilization for 24-48 hours. 15-20 mg freeze dried mycelia was placed in a mortar with liquid nitrogen and ground into a powder. The ground mycelia were transferred into a 2 mL tube and resuspended in 1 mL extraction buffer (150 mM EDTA, 50 mM Tris-HCl, and 1% (w/v) sodium lauroyl sarcosine). 1.6 mg proteinase K was added to the tube and contents were incubated at 37º C for 30 min. The tube was centrifuged at 13,000 rpm for 10 min and the supernatant was transferred to a fresh 2 mL tube. 500 µL phenol and 500 µL chloroform were added to the tube and the contents were mixed by vortex before centrifugation for 10 min at 13,000 rpm. The aqueous phase was transferred to a fresh 2 mL tube and washed two more times with 500 µL phenol and 500 µL chloroform as previously described. The aqueous phase was then transferred to a fresh 2 mL tube and washed (vortexed and centrifuged at 13,000 rpm for 10 min) with 1 mL chloroform. The aqueous phase was transferred to a fresh 2 mL tube and mixed with 1 mL chilled isopropanol. The DNA was precipitated overnight at 20º C and pelleted at 13,000 rpm for 10 min. The supernatant was discarded and the DNA was resuspended in 1 mL 1 M NaCl. The tube was incubated for 10 min at room temperature and then centrifuged at 13,000 rpm for 10 min to pellet polysaccharides. The supernatant was transferred to a fresh tube and mixed with 1 mL isopropanol. The tube was incubated at room temperature for 10 min and DNA was pelleted by centrifugation at 13,000 rpm for 10 min. The supernatant was discarded and 1 mL chilled 70% ethanol was added to the pellet without resuspension. The tube was centrifuged for 2 min at 13,000 rpm and the supernatant was discarded. The tube was

centrifuged for 1 min at 13,000 rpm and residual 70% ethanol was pipetted off. The pellet was air dried at room temperature, resuspended in 50 µL Milli-Q® water and stored at -20° C.

Alternatively, genomic DNA was isolated using the ZR Fungal/Bacterial DNA MicroPrep™ Kit from Zymo Research. First, sterile 2.4% (w/v) DifcoTM potato dextrose broth (Becton, Dickinson and Company, Maryland, U.S.A.) in Milli-Q® water was prepared in 25 mL aliquots in 125 mL Erlenmeyer flasks and inoculated with 5×106 spores or ~1 cm2 freshly ground mycelia (for non-sporulating strains). Cultures were incubated for 2-4 days at 22°C with shaking (200 rpm). The fermentation broth was filtered through a sterile nappy liner and the mycelia were rinsed three times with sterile water. Mycelia were transferred to a 1.5 mL Safe-Lock tube containing 750 µL Lysis Solution (ZR Fungal/Bacterial DNA MicroPrepTM, Zymo Research) and ~100 µL 1.0 mm zirconium oxide beads (Next Advance). Samples were placed in a Storm 24 Bullet Blender (Next Advance) and homogenised at speed 8 for 5 min. Tubes were centrifuged at 10,000 x g for 1 min and 400 µL supernatant was transferred to a Zymo-SpinTM IV Spin Filter in a Collection Tube and ultra-pure DNA was purified as detailed in the ZR Fungal/Bacterial DNA MicroPrepTM kit (Zymo Research). DNA was eluted in Milli-Q® water rather than the elution buffer provided with the ZR Fungal/Bacterial DNA MicroPrepTM kit (Zymo Research) and stored at -20° C.

Genome sequencing, identification and annotation of the *H. pulicicidum* (ATCC® 74245™) *NOD* gene cluster and GenBank accession numbers

Genomic DNA was sequenced on an Illumina MiSeq instrument by New Zealand Genomics Limited (NZGL) obtaining approximately 49-fold genome coverage. Reads were dynamically trimmed using the SolexaQA++ package to their longest fragment such that base call error rates did not exceed P = 0.01 and the minimum length of paired end reads was 25 bp. *De novo* assembly was performed using SPAdes Genome Assembler (version 3.5.0) with the default parameters using a kmer range of 39-127, and scaffolding was performed using SSPACE version 1.10 and GapFiller version 3.0.

Bioinformatic analyses were performed using BLAST and Fgenesh gene finding software[86] to facilitate identification and manual annotation of the gene cluster that is

responsible for nodulisporic acid biosynthesis (Chapter 2, Table 2.1). The annotated nucleotide sequence of the *H. pulicicidum* (ATCC74245) *NOD* gene cluster has been deposited with GenBank[2] with accession number MG182145. The accession number for the *H. pulicicidum* (ATCC74245) primary geranylgeranyl pyrophosphate (3) synthase is MG182146. Preparation of similarity matrices and phylogenetic trees for cluster identification.

All similarity matrices and phylogenetic trees (See Appendix, Figures 9.1 to 9.7 and Table 9.1) were constructed in MacVector 10.0.2 and based on pairwise and multiple sequence alignments generated using ClustalW (v1.83). Alignments were performed with the following parameters: open gap penalty = 10, extend gap penalty = 0.1, and using the Gonnet similarity matrix. Additional parameters for multiple alignments were: delay divergent = 40% and gap distance = 8. Phylogenetic trees were calculated using UPGMA best tree with systematic tie breaking. Distance was uncorrected and gaps were distributed proportionally.

Molecular biology reagents

Restriction endonucleases were purchased from New England Biolabs, except AarI, which was purchased from Thermo Fisher Scientific. T4 DNA Ligase, 10× T4 DNA Ligase buffer and 10 mM ATP were from New England Biolabs. Primers and gBlocks were synthesised by Integrated DNA Technologies. Other synthetic polynucleotides were synthesised by Epoch Life Science, Inc. Plasmids pBB528 and pBB535 were a gift from Bernd Bukau (Addgene plasmids #27390 and #27392, respectively). Kits for purification of plasmid DNA and PCR products using spin-column protocols were purchased from Macherey-Nagel. All PCRs for the construction of the MIDAS source, shuttle and destination vectors, and for amplification of MIDAS modules were performed using Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs). Construction of the MIDAS vectors is described below. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Calbiochem, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) from PanReac AppliChem, and 4-chloro-DL-phenylalanine (4CP) from Sigma. Antibiotics used in this work were; Geneticin (G418,

from Sigma), kanamycin (PanReac AppliChem) and spectinomycin (Gold Biotechnology).

Bacterial strains

Routine growth of *Escherichia coli* was performed at 37°C in LB broth. Chemically competent *E. coli* HST08 Stellar cells (Clontech Laboratories, Inc.) were used for routine transformations and maintenance of all plasmids (including source, shuttle and destination vectors, and all plasmids assembled at Level-1, Level-2, and Level-3).

PCR

All PCRs for the construction of the MIDAS source, shuttle and destination vectors, and for amplification of MIDAS modules were performed using Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs). Reactions were completed in 200 µL tubes typically containing 10 µL of 5X Phusion Buffer, 1 µL 10 mM dNTPs, 2 µL template (e.g. 20 ng genomic DNA), 2.5 µL 10 µM forward primer, 2.5 µL 10 µM reverse primer, 31.5 µL Milli-Q® water, and 0.5 µL Phusion polymerase for a total volume of 50 µL. Reactions were carried out in a DNA Engine PTC-200 Peltier Thermal Cycler (Bio-Rad) using the following parameters: 2 minutes at 98º C, 35 cycles of (30 seconds at 98° C, 30 seconds at n-2º C (where n is the lowest primer melting temperature) and 90 seconds at 72°C), and 5 minutes at 72°C followed by an infinite hold at 4º C.

Construction of MIDAS vectors

The initial plasmids produced during construction of the MIDAS vectors were constructed in a modular fashion by *Bsa*I-mediated Golden Gate assembly of PCR-generated DNA fragments containing terminal, convergent *Bsa*I recognition sites (Tables 8.1, 8.2 and 8.3 below). Since all MIDAS source, shuttle and destination vectors were required to be domesticated (i.e. made free of recognition sites for *Aar*I, *Bsa*I and *Bsm*BI), the PCR fragments were chosen to correspond to specific, functional plasmid structures (e.g., origin of replication, resistance markers, etc.), and by taking into account the number of internal Type IIS sites to be removed. Typically, 1–2 µL of each purified PCR fragment, 1 µL of *Bsa*I (20 U/µL), 1 µL of T4 DNA Ligase (400 U/µL)

and 2 µL of 10× T4 DNA Ligase buffer in a total reaction volume of 20 µL were incubated overnight at 37°C. Construction of the plasmids is described in detail below.

Construction of the pML1 source vector.

A precursor to source vector pML1 was generated using a four-way *Bsa*I-mediated Golden Gate reaction between PCR fragments CV-161, CV-162, CV-74 and CV-152 (Table 8.1). Following Golden Gate assembly, an aliquot of each reaction was transformed into *E. coli* HST08 Stellar competent cells by heat shock and spread onto LB plates supplemented with 50 µg/mL spectinomycin, 1 mM IPTG and 50 µg/mL X-Gal. Blue colonies were screened by restriction enzyme digestion of purified plasmid DNA, and confirmed by sequencing.

Table 8.1. PCR primers used in the construction of the pML1 source vector. Type IIS restriction enzyme recognition sites are colour-coded for *Bsa*I (ggtctc) and *Bsm*BI (cgtctc), with overhangs shaded grey. *Nco*I (ccatgg) and *Xho*I (ctcgag) cloning sites are shown underlined. Spec^R denotes spectinomycin resistance.

PCR name (Size, bp)	Plasmid fragment	Fragment description	Template (source/ reference)	Primer name (length): Primer sequence (5' to 3')
CV-161 (1125)	*Bsa*I[ACCG]-P_{aadA}-Spec^R-[GAAT]*Bsa*I	Bacterial spectinomycin resistance gene driven by the *aadA* promoter	pBB535[87]	**cvd2016-01-13a** (41-mer): cgctcacggtctcaACCGgacgtcgatatccggatgaaggc
				cvd2016-01-13b (45-mer): tgaacgaggtctcaATTCttatttgccgactaccttggtgatctc
CV-162 (591)	*Bsa*I[GAAT]-f1*ori*-[GTGT]*Bsa*I	Bacteriophage f1 origin of replication	pET-28a(+) (Novagen)	**cvd2016-01-13c** (55-mer): gatgagttggtctcaGAATtaattcatgagcggatacatatttgaatgtatttag
				cvd2015-11-18d (35-mer): gaggaacggtctccACACtggcgaatgggacgcgc
CV-74 (372)	*Bsa*I[GTGT]-*lacZα*-[GGCA]*Bsa*I	Wild type *lacZα* fragment, driven by the *lac* promoter.	*E. coli* K12 ER2925 (New England Biolabs)	**cvd2015-05-28b** (47-mer): ctttccggtctcaGTGTccatggttattaccaggcaaagcgccattc
				cvd2015-05-28k (54-mer): gacaggtttggtctcgTGCCctcgagcagctggcgcaacgcaattaatgtgagt
CV-152 (1070)	*Bsa*I[GGCA]-pMB1*ori*-[ACCG]*Bsa*I	Bacterial pMB1 plasmid origin of replication.	pUC19 (Clontech Laboratories, Inc.)	**cvd2015-11-18a** (39-mer): cacattaaggtctctGGCAtcactgcccgctttccagtc
				cvd2015-11-18b (50-mer): tgattaggtctcgCGGTctgtcagaccaagtttactcatatatactttag
CV-81 (367)	*Nco*I-[CTCG]*Bsm*BI-*lacZα*-*Bsm*BI[AGAC]-*Xho*I	Fragment containing the pML1 Golden Gate cloning cassette, i.e., *lacZα* (driven by the *lac* promoter) flanked by divergent *Bsm*BI recognition sites.	*E. coli* K12 ER2925 (New England Biolabs)	**cvd2015-05-28d** (50-mer): cagctttcccatggtCTCGtgagacgttattaccaggcaaagcgccattc
				cvd2015-05-28l (46-mer): tcccgacctcgagGTCTagagacggcgcaacgcaattaatgtgagt

The 329 bp *Nco*I-*Xho*I fragment of the pML1 precursor was replaced, by restriction enzyme digestion and ligation, with the 346 bp *Nco*I-*Xho*I fragment of PCR CV-81, which contains a *lacZα*-based Golden Gate cloning cassette ([CTCG]*Bsm*BI-*lacZα*-*Bsm*BI[AGAC]). The resultant plasmid is the Level-1 source vector, pML1 (Figure 8.2).

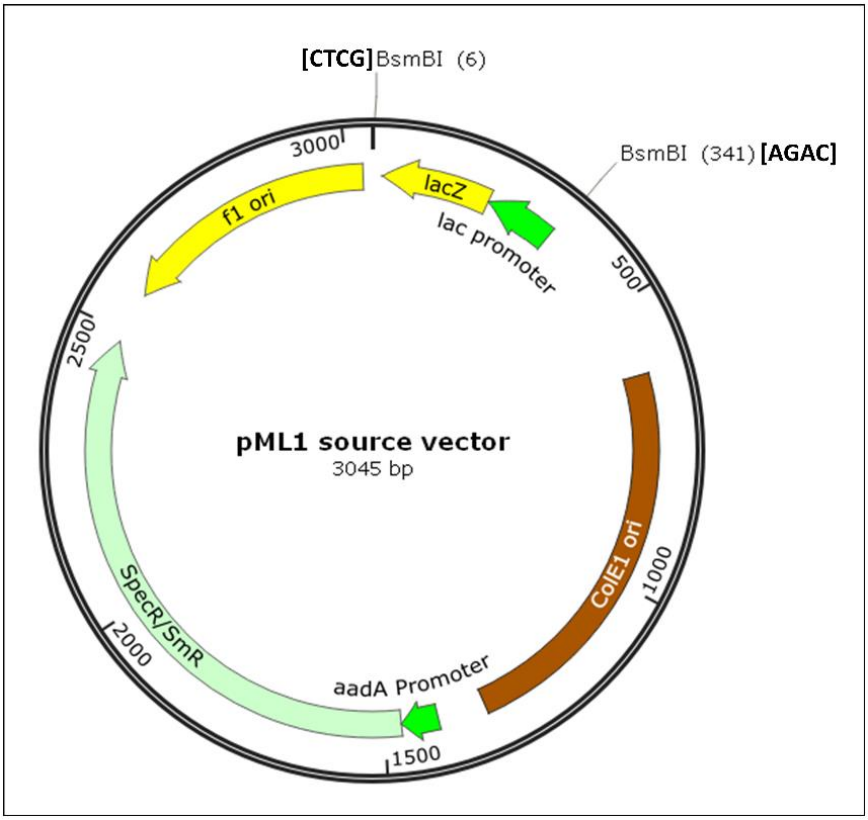


Figure 8.2. pML1 source vector.

Construction of pML2 shuttle vectors.

A precursor to the pML2 shuttle vectors was generated using a four-way *Bsa*I-mediated Golden Gate reaction between PCR fragments CV-153, CV-154, CV-77 and CV-152 (Table 8.2). Following Golden Gate assembly, an aliquot of each reaction was transformed into *E. coli* HST08 Stellar competent cells by heat shock and spread onto LB plates supplemented with 75 µg/mL kanamycin, 1 mM IPTG and 50 µg/mL X-Gal. Blue colonies were screened by restriction enzyme digestion of purified plasmid DNA, and confirmed by sequencing.

Table 8.2. PCR primers used in the construction of the pML2 shuttle vectors. The *Bsa*I restriction enzyme recognition sites are colour-coded (ggtctc), with overhangs shaded grey. *Nco*I (ccatgg) and *Xho*I (ctcgag) cloning sites are shown underlined. Kan^R denotes kanamycin resistance.

PCR name (Size, bp)	Plasmid fragment	Fragment description	Template (source/ reference)	Primer name (length): Primer sequence (5' to 3')
CV-153 (583)	*Bsa*I[ACCG]-Kan^R-[TTTG]*Bsa*I	Bacterial kanamycin resistance gene driven by the *kan* promoter	pET-28a(+) (Novagen)	**cvd2015-11-18c** (40-mer): agaagatccggtctccACCGctacggggtctgacgctcag
				cvd2015-06-17a (38-mer) (domestication primer): gattgcggtctccCAAAcgaaatacgcgatcgctgtt
CV-154 (963)	BsaI[TTTG]-Kan^R-f1*ori*-[GTGT]		pET-28a(+) (Novagen)	**cvd2015-06-17b** (37-mer) (domestication primer): atcgcggtctcgTTTGgctcaggcgcaatcacgaat
				cvd2015-11-18d (35-mer): gaggaacggtctccACACtggcgaatgggacgcgc
CV-77 (363)	*Bsa*I[GTGT]-*lacZα*-[GGCA]*Bsa*I	Wild type *lacZα* fragment, driven by the *lac* promoter.	*E. coli* K12 ER2925 (New England Biolabs)	**cvd2015-05-22b** (47-mer): ctttccggtctcaGTGTccatggttattaccaggcaaagcgccattc
				cvd2015-05-22g (46-mer): cccgacggtctcgTGCCctcgaggcgcaacgcaattaatgtgagtt
CV-152 (1070)	*Bsa*I[GGCA]-pMB1*ori*-[ACCG]*Bsa*I	Bacterial pMB1 plasmid origin of replication.	pUC19 (Clontech Laboratories, Inc.)	**cvd2015-11-18a** (39-mer): cacattaaggtctctGGCAtcactgcccgctttccagtc
				cvd2015-11-18b (50-mer): tgattaggtctcgCGGTctgtcagaccaagtttactcatatatactttag
CV-78 (376)	*Bsa*I[CATT]AarI-*lacZα*-AarI[CGTA]*Bsa*I	Wild type *lacZα* fragment, driven by the *lac* promoter.	*E. coli* K12 ER2925 (New England Biolabs)	**cvd2015-05-22h** (53-mer): agccagcggtctcaCATTgcttgcaggtgttattaccaggcaaagcgccattc
				cvd2015-05-22i (52-mer): caggtttcggtctcgTACGgcgggcaggtggcgcaacgcaattaatgtgagt
CV-129 (140)	[GGAG]*Bsa*I-P$_{cat}$►*Bsa*I[TAAA]	Chloramphenicol acetyltransferase (*cat*) promoter	pBB528[87]	**cvd2015-10-20a** (45-mer): ccacctGGAGaggagaccaagctttgatcggcacgtaagaggttcc
				cvd2015-10-20b (42-mer): cctttccggtctctTTTAgcttccttagctcctgaaaatctc
CV-141 (1029)	BsaI[TAAA]-PheS(T251A/A294G)►*-[CGCT]BsaI	PheS negative selection marker	Synthetic gene (Epoch Life Science, Inc).	**cvd2015-10-20c** (39-mer): agggaaaggtctcaTAAAatgtcacatctcgcagaactg
				cvd2015-11-03a (54-mer): tgcctggAGCGaggaggaccaagcttttattatttaaactgtttgaggaaacgcag

The pML2 precursor (Figure 8.3) was digested with *Nco*I + *Xho*I and ligated to gBlocks digested with the same enzymes. The eight gBlocks (Table 8.3) are classified into two groups — one group comprising four “White” gBlocks (ML2(+)WF, ML2(+)WR, ML2(−)WF and ML2(−)WR) and a second group composed of four “Blue” gBlocks (ML2(+)BF, ML2(+)BR, ML2(−)BF or ML2(−)BR). Ligations were transformed into *E. coli* Stellar and spread onto LB plates supplemented with 75 µg/mL kanamycin, 1 mM IPTG and 50 µg/mL X-Gal. White colonies were analysed by restriction enzyme digestion of purified plasmid DNA, and confirmed by sequencing. The resultant eight plasmids were designated “gBlock precursor plasmids”, comprising four “White” gBlock precursor plasmids and four “Blue” gBlock precursor plasmids.

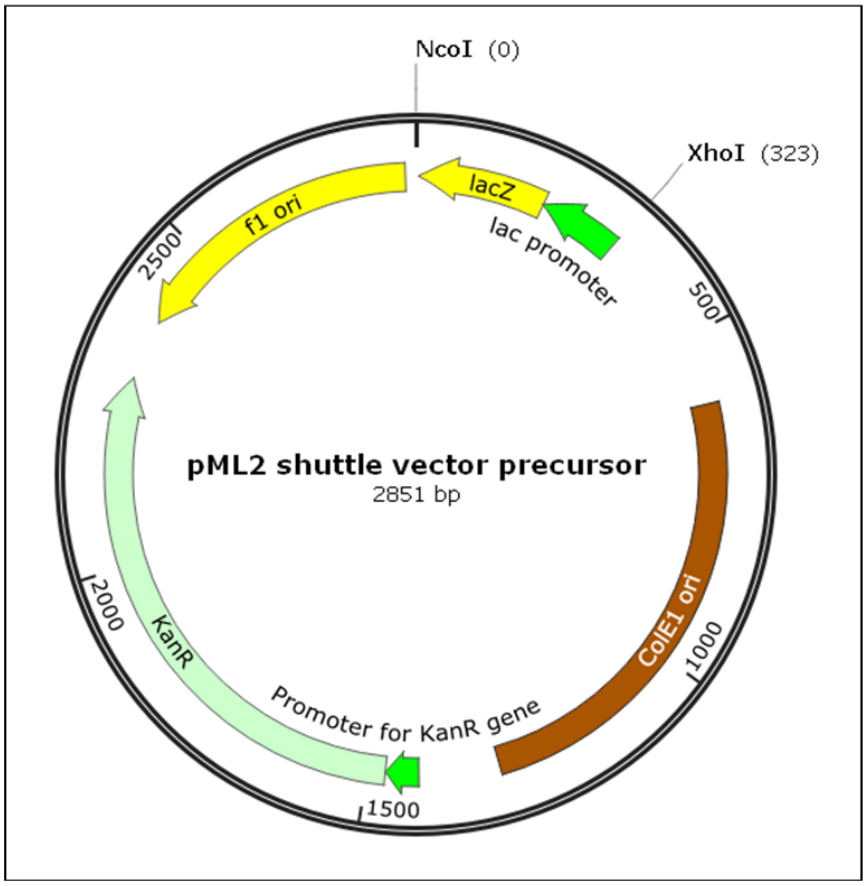


Figure 8.3. pML2 shuttle vector precursor.

Table 8.3. gBlock sequences used for the construction of the pML2 shuttle vectors. gBlocks are synthesised as double stranded DNA; only the top strand (5' to 3') is shown in the table. *Nco*I (ccatgg) and *Xho*I (ctcgag) cloning sites are shown underlined. Type IIS restriction enzyme recognition sites are colour-coded for *Aar*I (cacctgc), *Bsa*I (ggtctc) and *Bsm*BI (cgtctc), with overhangs shaded grey.

	gBlock name	gBlock sequence (5' to 3')
"White" gBlocks	**ML2(+)WF**	gaccccgtaggtgtccatggcacctgcagctcattggagagagaccttaagcttgcagcagcggtctctcgc tcattagagacggccggattgcggccgctaaccggccgtctcacgtaatatgcaggtgctcgagggcatcaa at
	ML2(+)WR	gaccccgtaggtgtccatggcacctgcagctcattagcgagagaccttaagcttgcagcagcggtctctctc ccattagagacggccggattgcggccgctaaccggccgtctcacgtaatatgcaggtgctcgagggcatcaa at
	ML2(−)WF	gaccccgtaggtgtccatggcacctgcagctcattagagacggccggattgcggccgctaaccggccgtctc acgtaggagagagaccttaagcttgcagcagcggtctctcgctcgtaatatgcaggtgctcgagggcatcaa at
	ML2(−)WR	gaccccgtaggtgtccatggcacctgcagctcattagagacggccggattgcggccgctaaccggccgtctc acgtaagcgagagaccttaagcttgcagcagcggtctctctcccgtaatatgcaggtgctcgagggcatcaa at
"Blue" gBlocks	**ML2(+)BF**	gaccccgtaggtgtccatggcgtctcacattggagagagaccttaagcttgcagcagcggtctctcgctcat tatatgcaggtggccggattgcggccgctaaccggccacctgcagctcgtaagagacgctcgagggcatcaa at
	ML2(+)BR	gaccccgtaggtgtccatggcgtctcacattagcgagagaccttaagcttgcagcagcggtctctctcccat tatatgcaggtggccggattgcggccgctaaccggccacctgcagctcgtaagagacgctcgagggcatcaa at
	ML2(−)BF	gaccccgtaggtgtccatggcgtctcacattatatgcaggtggccggattgcggccgctaaccggccacctg cagctcgtaggagagagaccttaagcttgcagcagcggtctctcgctcgtaagagacgctcgagggcatcaa at
	ML2(−)BR	gaccccgtaggtgtccatggcgtctcacattatatgcaggtggccggattgcggccgctaaccggccacctg cagctcgtaagcgagagaccttaagcttgcagcagcggtctctctcccgtaagagacgctcgagggcatcaa at

A synthetic polynucleotide sequence encoding the Thr^{251}Ala/Ala^{294}Gly double mutant of the *E. coli pheS*, as described by Miyazaki,[88] was supplied by Epoch Life Science, Inc. The coding sequence of the *pheS* gene was amplified (PCR CV-141) and placed under the control of the chloramphenicol acetyltransferase (*cat*) promoter by *Bsa*I-mediated Golden Gate assembly with PCR CV-129. The *Bsa*I was heat-inactivated and the assembled *pheS* gene was then ligated into each of the eight gBlock precursor plasmids generated in the previous step that had also been treated with *Bsa*I. Ligations were transformed by heat shock into *E. coli* HST08 Stellar competent cells, spread onto LB plates supplemented with 75 µg/mL kanamycin, and colonies analysed by restriction enzyme analysis and sequencing. For the four “White” gBlock precursor plasmids, ligation of the assembled *pheS* gene resulted in the four “White” Level-2 shuttle vectors: **pML2(+)WF**, **pML2(+)WR**, **pML2(−)WF** and **pML2(−)WR** (Figure 8.4).

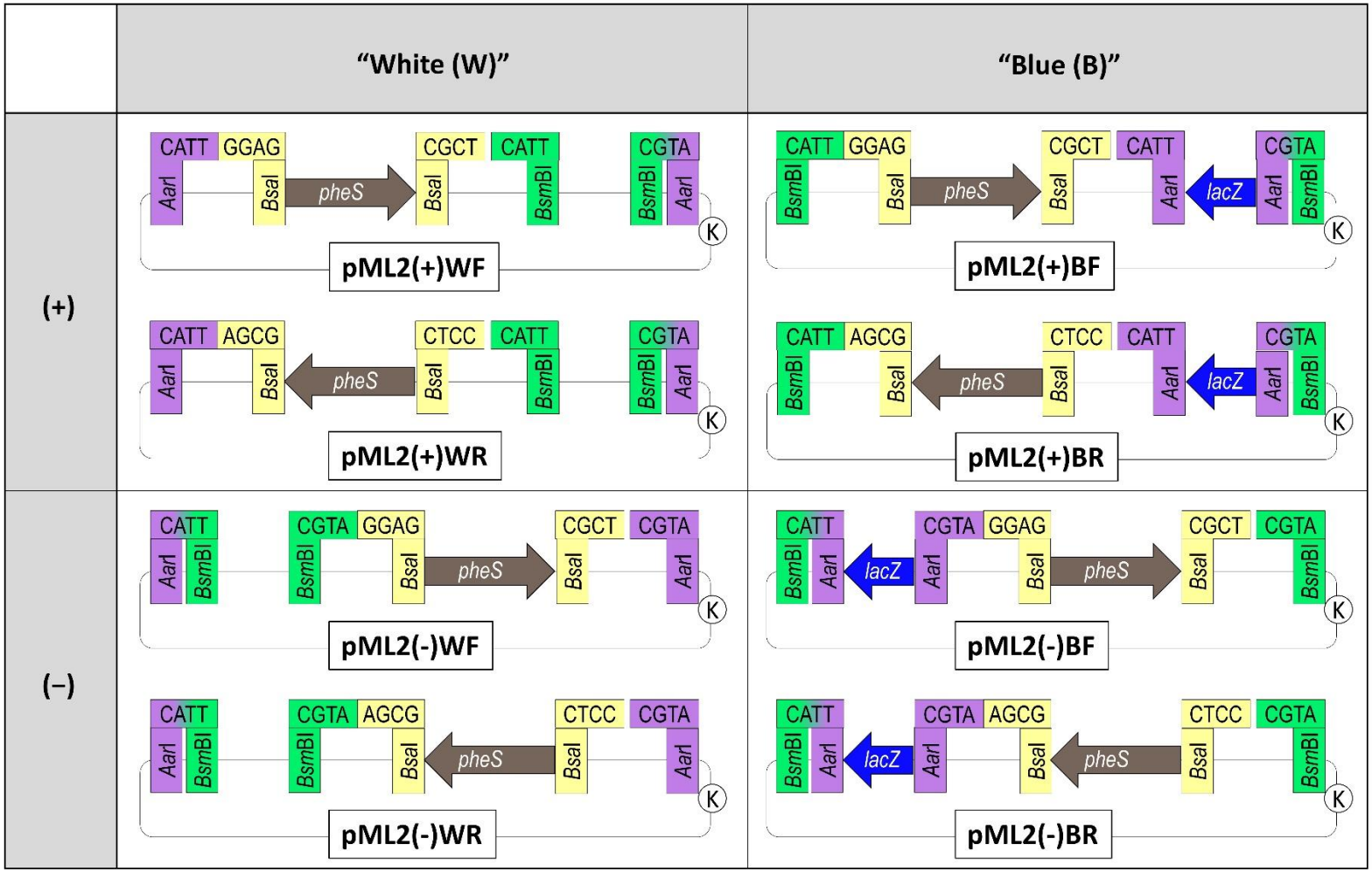


Figure 8.4. Overview of MIDAS cassettes. The structure of the MIDAS cassette in each of the eight pML2 shuttle vectors is depicted schematically. Within each MIDAS cassette, the dashed box shows the Golden Gate cloning cassette (comprised of divergent *Bsa*I sites flanking a pheS negative selection marker) used for Level-2 assembly of TUs.

To produce the four “Blue” Level-2 shuttle vectors, the four “Blue” gBlock precursor plasmids harbouring the assembled *pheS* gene from the previous step, were digested with *Aar*I and ligated to *Bsa*I-digested PCR CV-78, resulting in the four “Blue” Level-2 shuttle vectors: **pML2(+)BF**, **pML2(+)BR**, **pML2(−)BF** and **pML2(−)BR** (Figure 8.4). Ligations were transformed into *E. coli* HST08 Stellar, spread onto LB plates supplemented with 75 µg/mL kanamycin, 1 mM IPTG and 50 µg/mL X-Gal. Blue colonies were analysed by restriction analysis and sequencing.

Functionality of the mutant *pheS* gene in each of the eight pML2 shuttle vectors was confirmed by transforming 0.1 µg of each plasmid into *E. coli* HST08 Stellar competent cells and spreading transformation mixes onto LB plates supplemented with 1.25 mM 4CP and 75 µg/mL kanamycin, and onto LB plates supplemented with 75 µg/mL kanamycin only (no 4CP). In each case, no colonies were observed when 4CP-containing plates were spread with undiluted transformation mixes (giving a calculated efficiency of $<3.25\times10^4$ Kan^R CFU/µg), while $>1.3\times10^6$ Kan^R CFU/µg plasmid DNA was calculated for each vector following growth on plates devoid of 4CP, indicating high counter-selection efficiency.

Construction of the pML3 destination vector.

Plasmid pML3 (Figure 8.5) was generated using a four-way *Bsa*I-mediated Golden Gate reaction between PCR fragments CV-161, CV-162, CV-79 and CV-152 (Table 8.4). Following Golden Gate assembly, an aliquot of each reaction was transformed into *E. coli* HST08 Stellar competent cells by heat shock and spread onto LB plates supplemented with 50 µg/mL spectinomycin, 1 mM IPTG and 50 µg/mL X-Gal. Blue colonies were screened by restriction enzyme digestion of purified plasmid DNA, and confirmed by sequencing.

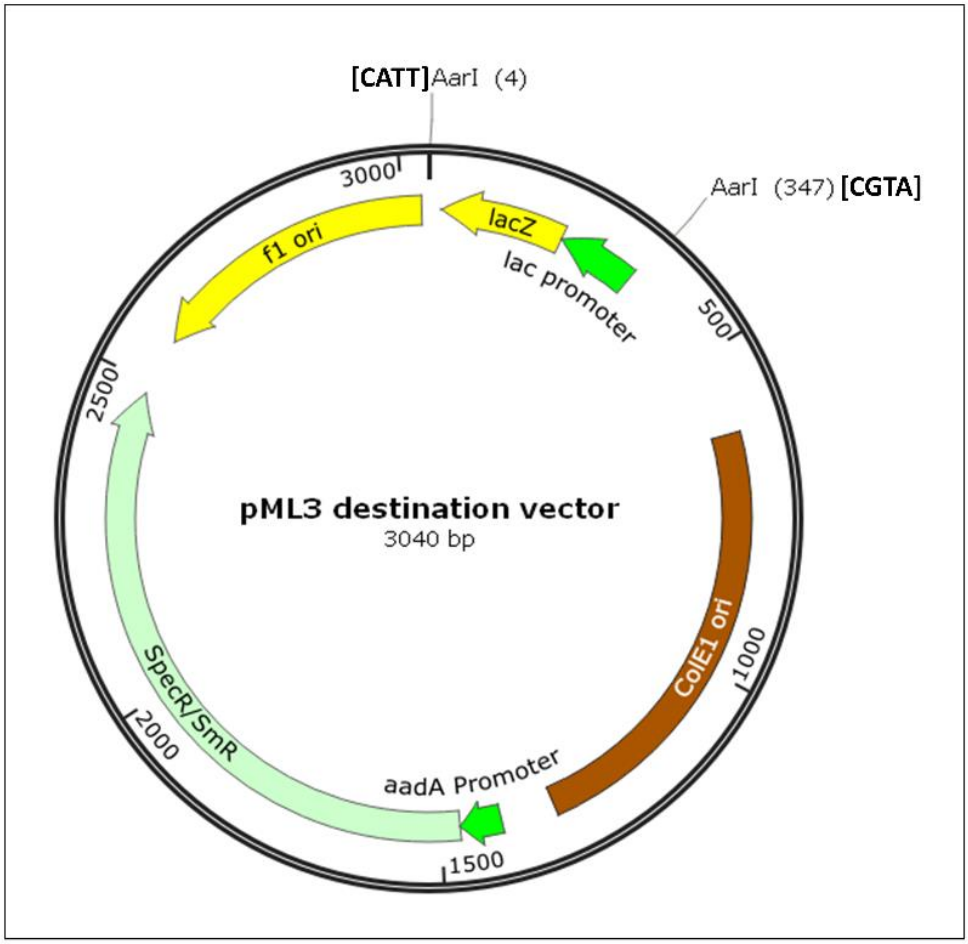


Figure 8.5. pML3 destination vector.

Table 8.4. PCR primers used in the construction of the pML3 destination vector. Type IIS restriction enzyme recognition sites are colour-coded for *Bsa*I (ggtctc) and *Aar*I (cacctgc), with overhangs shaded grey. Spec^R denotes spectinomycin resistance.

PCR name (Size, bp)	Plasmid fragment	Fragment description	Template (source/ reference)	Primer name (length): Primer sequence (5' to 3')
CV-161 (1125)	*Bsa*I[ACCG]-P_{aadA}-Spec^R-[GAAT]*Bsa*I	Bacterial spectinomycin resistance gene driven by the *aadA* promoter	pBB535[87]	**cvd2016-01-13a** (41-mer): cgctcacggtctcaACCGgacgtcgatatccggatgaaggc
				cvd2016-01-13b (45-mer): tgaacgaggtctcaATTCttatttgccgactaccttggtgatctc
CV-162 (591)	*Bsa*I[GAAT]-f1*ori*-[GTGT]*Bsa*I	Bacteriophage f1 origin of replication	pET-28a(+) (Novagen)	**cvd2016-01-13c** (55-mer): gatgagttggtctcaGAATtaattcatgagcggatacatatttgaatgtatttag
				cvd2015-11-18d (35-mer): gaggaacggtctccACACtggcgaatgggacgcgc
CV-79 (384)	*Bsa*I[GTGT][CATT]AarI-*lacZα*-AarI[CGTA][GGCA]*Bsa*I	Fragment containing the pML3 Golden Gate cloning cassette, i.e., *lacZα* (driven by the *lac* promoter) flanked by divergent *Aar*I recognition sites.	*E. coli* K12 ER2925 (New England Biolabs)	**cvd2015-05-22d** (58-mer): actccagcggtctcaGTGTCATTgcttgcaggtgttattaccaggcaaagcgccattc
				cvd2015-05-22j (55-mer): cgacagggtctcgTGCCTACGgcggcaggtggcgcaacgcaattaatgtgagt
CV-152 (1070)	*Bsa*I[GGCA]-pMB1*ori*-[ACCG]*Bsa*I	Bacterial pMB1 plasmid origin of replication.	pUC19 (Clontech Laboratories, Inc.)	**cvd2015-11-18a** (39-mer): cacattaaggtctctGGCAtcactgcccgctttccagtc
				cvd2015-11-18b (50-mer): tgattaggtctcgCGGTctgtcagaccaagtttactcatatatactttag

Golden Gate assembly reactions

One-pot Golden Gate (restriction-ligation) reactions were used to assemble transcription unit modules at Level-1, transcription units at Level-2 and multigene plasmids at Level-3. Detailed methods for each of these reactions are given below.

Preparation of MIDAS constructs

MIDAS constructs for use in heterologous expression studies were prepared as described below. Accordingly, coding sequences of the *H. pulicicidum* genes of interest were amplified (see Appendix, Table 9.2 for primers) and cloned into the MIDAS Level-1 destination vector, pML1 (Table 9.3). At MIDAS Level-2, the cloned coding sequences were placed under the control of heterologous promoter and transcriptional terminator modules to generate full-length transcription units, transcription units (Table 9.4), which were then used to generate the multigene plasmids (Table 9.5).

Primer design.

As each transcription unit module is defined by its flanking prefix and suffix nucleotides, these module-specific bases effectively form an address system for each transcription unit module and they determine its position and orientation within the assembled transcription unit. The developers of MoClo and GoldenBraid2.0 have already worked in concert to develop a common syntax or set of standard addresses (referred to as ‘fusion sites’ in the MoClo system and ‘barcodes’ in GoldenBraid2.0) for a wide variety of transcription unit modules to facilitate part exchangeability for plant expression,[89] and this standard is also adopted here for MIDAS-based assembly of transcription units for expression in filamentous fungi (Figure 8.6).

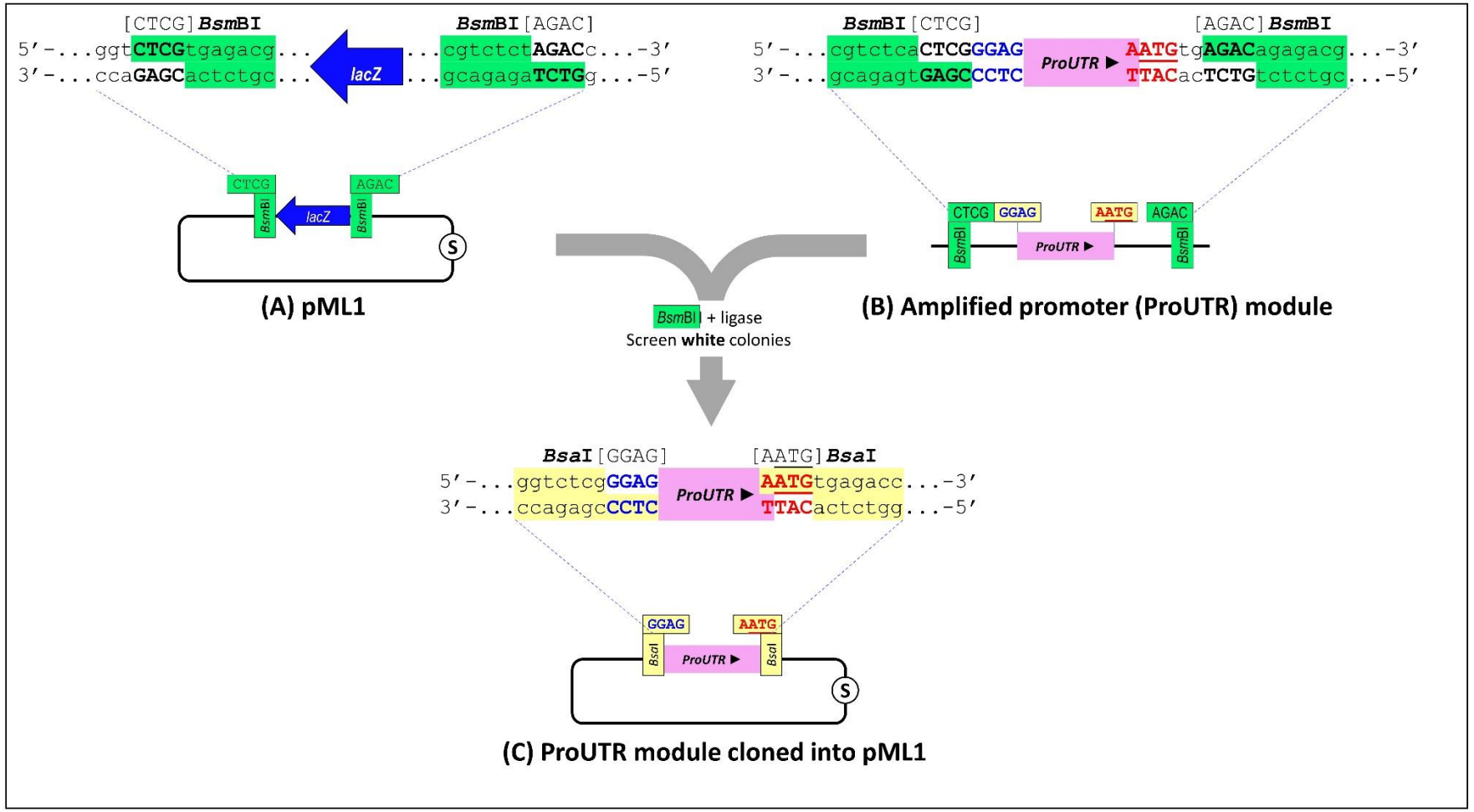


Figure 8.6. Principle of MIDAS Level-1 cloning. (**A**) The Golden Gate cloning cassette (*lacZ*α flanked by divergent *Bsm*BI sites) of the pML1 source vector. (**B**) A PCR product containing a promoter (ProUTR) module flanked by convergent *Bsm*BI sites. The 4 bp sequences GGAG and AATG represent, respectively, the prefix and suffix signature unique to ProUTR modules (as mentioned, prefix and suffix signatures differ for coding sequences (AATG and GCTT) and terminators [UTRterm <GCTT and CGCT>]). Following a *Bsm*BI-mediated Golden Gate reaction between pML1 and the amplified ProUTR module, a plasmid consisting of the ProUTR module cloned into pML1 is obtained (**C**), and the cloned module is now flanked by convergent *Bsa*I recognition sites.

Considerations for the design of PCR primers for amplifying promoters (ProUTR), coding sequence and terminator (UTRterm) transcription unit modules are shown in the Appendix, Table 9.2. Since the module-specific bases are incorporated as part of the PCR primer sequences (and not as part of the pML1 source vector), MIDAS is completely adaptable to the addition of new parts (i.e., new fusion site combinations) and part types. For example, using the module address standard described by Patron *et al*.,[89] ProUTR parts can be broken down into smaller parts, fusion proteins can be constructed, and *N*- or *C*-terminal tags can be added to coding sequences.

Furthermore, to enable subsequent (i.e., Level-2) assembly of full-length transcription units, each transcription unit module is designed to be flanked by four module-specific, user-defined nucleotides at the 5' end (prefix), and four, different, module-specific, user-defined nucleotides at the 3' end (suffix), which are included as part of the PCR primer sequences.

Protocols for MIDAS Level-1 module cloning

PCR-amplified modules were purified using spin-column protocols and cloned into the MIDAS Level-1 plasmid, pML1, by *Bsm*BI-mediated Golden Gate assembly. Typically, 1–2 µL (approximately 50–200 ng) of pML1 plasmid DNA from a miniprep was mixed with 1–2 µL of each purified PCR fragment, 1 µL of *Bsm*BI (10 U/µL), 1 µL of T4 DNA Ligase (400 U/µL) and 2 µL of 10× T4 DNA Ligase buffer in a total reaction volume of 20 µL. Reactions were incubated at 37°C for >4 hours and an aliquot (typically 2–3 µL) was transformed into 50 µL of *E. coli* HST08 Stellar competent cells by heat shock at 42° C. Following a recovery period (addition of 100 µL SOC medium and incubation at 37° C and 200-250 rpm for 1 hour), aliquots of the transformation mix were spread onto LB agar plates supplemented with 50 µg/mL spectinomycin, 1 mM IPTG and 50 µg/mL X-Gal. Plates were incubated overnight at 37°C, and white colonies were chosen for analysis.

Protocols for MIDAS Level-2 transcription unit assembly

Using the modules cloned at Level-1, full-length transcription units were assembled into MIDAS Level-2 plasmids by *Bsa*I-mediated Golden Gate assembly. Typically, 40 fmol of pML2 plasmid DNA was mixed with 40 fmol of plasmid DNA of each Level-1

entry clone, 1 µL of *Bsa*I-HF (20 U/µL), 1 µL of T4 DNA Ligase (400 U/µL) and 2 µL of 10× T4 DNA Ligase buffer in a total reaction volume of 20 µL. Reactions were incubated at 37°C for >4 hours and an aliquot and transformed into *E. coli* HST08 Stellar competent cells as described under ***Protocols for MIDAS Level-1 module cloning*** and spread onto LB agar plates containing 75 µg/mL kanamycin and 1.25 mM 4CP. Following overnight incubation at 37°C, colonies were picked for analysis.

Protocols for MIDAS Level-3 multigene assembly

Full-length transcription units assembled at Level-2, were used to create multigene assemblies in the Level-3 destination vector by alternating Golden Gate assembly using either *Aar*I (for transcription units cloned into pML2 “White” shuttle vectors) or *Bsm*BI (for transcription units cloned into pML2 “Blue” shuttle vectors). Typically, 40 fmol of Level-3 destination vector plasmid DNA was mixed with 40 fmol of Level-2 entry clone plasmid DNA. For *Bsm*BI-mediated assemblies, 1 µL of *Bsm*BI (10 U/µL), 1 µL of T4 DNA Ligase (400 U/µL) and 2 µL of 10× T4 DNA Ligase buffer, was added to the plasmid DNA in a final reaction volume of 20 µL. Two different reaction conditions were tested for *Aar*I-mediated assemblies; (i) conventional Golden Gate reactions performed in T4 DNA Ligase buffer and (ii) reactions performed in *Aar*I restriction enzyme buffer supplemented with ATP. For *Aar*I-mediated reactions performed in T4 DNA Ligase buffer, 1 µL of *Aar*I (2 U/µL), 0.4 µL of 50× oligonucleotide (25 µM, supplied with the enzyme), 1 µL of T4 DNA Ligase (400 U/µL) and 2 µL of T4 DNA Ligase buffer was added to the plasmid DNA, in a final reaction volume of 20 µL. For *Aar*I-mediated reactions carried out in restriction enzyme buffer, 1 µL of *Aar*I (2 U/µL), 0.4 µL of 50× oligonucleotide (25 µM, supplied with the enzyme), 2 µL of 10× Buffer *Aar*I (supplied with the enzyme), 1 µL of T4 DNA Ligase (400 U/µL) and 2 µL of 10 mM ATP was added to the plasmid DNA, in a final reaction volume of 20 µL. All Level-3 reactions (*Bsm*BI- and *Aar*I-mediated) were incubated 37°C for >4 hours and reactions were transformed into *E. coli* HST08 Stellar competent cells as described under ***Protocols for MIDAS Level-1 module cloning*** and spread onto LB agar plates supplemented with 50 µg/mL spectinomycin, 1 mM IPTG and 50 µg/mL X-Gal. Plates were incubated overnight at 37°C. For *Aar*I-mediated assembly reactions, white colonies were chosen for analysis while, for *Bsm*BI-mediated assembly reactions, blue colonies were screened.

Penicillium paxilli PAX gene deletion mutants

Penicillium paxilli and *Hypoxylon pulicicidum* strains used in this study are shown in Table 8.5. Gene deletion mutants of the *PAX* cluster have been described in the literature; the Δ*paxG*,[37] Δ*paxM*[38] and Δ*paxC*[38] single gene deletion mutants and the CY2 full *PAX* cluster deletion mutant[35] of *P. paxilli* (strains PN2662, PN2257, PN2290, and PN2250 respectively) are unable to synthesise the indole diterpenes emindole SB (18), paspaline (11) and paxilline (2) normally produced by wild type *P. paxilli* (strain PN2013). These mutants provide a useful resource for confirming gene function and individually testing whether MIDAS-assembled transcription units contain all the necessary *cis*-acting elements to successfully transcribe and express their coding sequences. Thus, production of novel compounds (e.g. nodulisporic acids) or restoration of emindole SB (18), paspaline (11) and/or paxilline (2) production would indicate successful reconstitution of the deletion mutant and show, in turn, the function(s) of the transformed gene(s).

Table 8.5. Table of fungal species used in this thesis.

***Hypoxylon pulicicidum* (*Nodulisporium* sp.) strain**	**Description**	**Indole diterpene phenotype**	**Source** [reference]
		Nodulisporic acid A (1)	
ATCC® 74245™	Wild type	+	ATCC®[84]

***Penicillium paxilli* strain**	**Description**	**Indole diterpene phenotype**		**Source** [reference]
		Paspaline (11)	**Paxilline (2)**	
PN2013 (ATCC®26601™)	Wild type	+	+	Barry Scott, Massey University[72]
PN2250 (CY2)	PN2013/Deletion of entire *PAX* locus (*ΔPAX*); Hyg^R	−	−	Barry Scott, Massey University[35]
PN2257	PN2013/Δ*paxM*::P_{glcA}-*hph*-T_{trpC}; Hyg^R	−	−	Barry Scott, Massey University[38]
PN2290	PN2013/Δ*paxC*::P_{trpC}-*hph*; Hyg^R	−	−	Barry Scott, Massey University[38]
PN2458	PN2013/Δ*paxB*::P_{trpC}-*nptII*-*trpC*; Gen^R	−	−	Barry Scott, Massey University[38]

Media and reagents used for fungal work

CDYE (Czapex-Dox/Yeast extract) medium with trace elements was made with deionised water and contained 3.34% (w/v) Czapex-Dox (Oxoid Ltd., Hampshire, England), 0.5% (w/v) yeast extract (Oxoid Ltd., Hampshire, England), and 0.5% (v/v) trace element solution. For agar plates, Select agar (Invitrogen, California, USA) was added to 1.5% (w/v).

Trace element solution was made in deionised water and contained 0.004% (w/v) cobalt(II) chloride hexahydrate (Ajax Finechem, Auckland, New Zealand), 0.005% (w/v) copper(II) sulfate pentahydrate (Scharlau, Barcelona, Spain), 0.05% (w/v) iron(II) sulfate heptahydrate (Merck, Darmstadt, Germany), 0.014% (w/v) manganese(II) sulfate tetrahydrate, and 0.05% (w/v) zinc sulfate heptahydrate (Merck, Darmstadt, Germany). The solution was preserved with 1 drop of 12 M hydrochloric acid.

Regeneration (RG) medium was made with deionised water and contained 2% (w/v) malt extract (Oxoid Ltd., Hampshire, England), 2% (w/v) D(+)-glucose anhydrous (VWR International BVBA, Leuven, Belgium), 1% (w/v) mycological peptone (Oxoid Ltd., Hampshire, England), and 27.6% sucrose (ECP Ltd. Birkenhead, Auckland, New Zealand). Depending on whether the medium was to be used for plates (1.5% RGA) or overlays (0.8% RGA), Select agar (Invitrogen, California, USA) was added to 1.5% or 0.8% (w/v), respectively.

Protoplast preparation

The preparation of fungal protoplasts for transformation was according to Yelton *et al*. 1984[90] with modifications. Five 25 mL aliquots of CDYE medium with trace elements, in 100 mL Erlenmeyer flasks, were inoculated with 5×10^6 spores and incubated for 28 hours at 28°C with shaking (200 rpm). The fermentation broth from all five flasks was filtered through a sterile nappy liner and the combined mycelia were rinsed three times with sterile water and once with OM buffer (10 mM Na_2HPO_4 and 1.2 M $MgSO_4.7H_2O$, brought to pH 5.8 with 100 mM $NaH_2PO_4.2H_2O$). Mycelia were weighed, resuspended in 10 mL of filter-sterilised Lysing Enzymes solution (prepared by resuspending Lysing Enzymes from *Trichoderma harzianum* (Sigma) at 10 mg/mL in OM buffer) per gram of mycelia, and incubated for 16 hours at 30°C with shaking at 80 rpm. Protoplasts were filtered through a sterile nappy liner into a 250 mL

Erlenmeyer flask. Aliquots (5 mL) of filtered protoplasts were transferred into sterile 15 mL centrifuge tubes and overlaid with 2 mL of ST buffer (0.6 M sorbitol and 0.1 M Tris-HCl at pH 8.0). Tubes were centrifuged at 2600 x *g* for 15 minutes at 4°C. The white layer of protoplasts that formed between the OM and ST buffers in each tube was transferred (in 2 mL aliquots) into sterile 15 mL centrifuge tubes, gently washed by pipette resuspension in 5 mL of STC buffer (1 M sorbitol, 50 mM Tris-HCl at pH 8.0, and 50 mM $CaCl_2$) and centrifuged at 2600 x *g* for 5 minutes at 4°C. The supernatant was decanted off and pelleted protoplasts from multiple tubes were combined by resuspension in 5 mL aliquots of STC buffer. The STC buffer wash was repeated three times until protoplasts were pooled into a single 15 mL centrifuge tube. The final protoplast pellet was resuspended in 500 µL of STC buffer and protoplast concentration was estimated with a hemocytometer. The protoplast stock was diluted to give a final concentration of 1.25×10^8 protoplasts per mL of STC buffer. Aliquots of protoplasts (100 µL) were used immediately for fungal transformations and excess protoplasts were preserved in 8% PEG solution (80 µL of protoplasts were added to 20 µL of 40% (w/v) PEG 4000 in STC buffer) in 1.7 mL micro-centrifuge tubes and stored at −80°C.

Transformation of *P. paxilli*

Fungal transformations — modified from Vollmer and Yanosfsky 1986[91] and Oliver *et al*. 1987[92] — were carried out in 1.7 mL micro-centrifuge tubes containing 100 µL (1.25×10^7) protoplasts, either freshly prepared in STC buffer, or stored in 8% PEG solution (as described above). A solution containing 2 µL of spermidine (50 mM in H_2O), 5 µL heparin (5mg/mL in STC buffer), and 5 µg of plasmid DNA (250 µg/mL) was added to the protoplasts and, following incubation on ice for 30 minutes, 900 µL of 40% PEG solution (40% (w/v) PEG 4000 in STC buffer) was added. The transformation mixture was incubated on ice for a further 15–20 minutes, transferred to 17.5 mL of 0.8% RGA medium (prewarmed to 50°C) in sterile 50 mL tubes, mixed by inversion, and 3.5 mL aliquots were dispensed onto 1.5% RGA plates. Following overnight incubation at 25°C, 5 mL of 0.8% RGA (containing sufficient G418 to achieve a final concentration of 150 µg per mL of solid medium) was overlaid onto each plate. Plates were incubated for a further 4 days at 25°C and spores were picked from individual colonies and streaked onto CDYE agar plates supplemented with 150 µg/mL G418.

Streaked plates were incubated at 25°C for a further 4 days. Spores from individual colonies were suspended in 50 µL of 0.01% (v/v) triton X-100 and 5 × 5 µL aliquots of the spore suspension was transferred onto new CDYE agar plates supplemented with 150 µg/mL G418. Sporulation plates were incubated at 25°C for 4 days and spore stocks were prepared as follows. Colony plugs from the sporulation plates were suspended in 2 mL of 0.01% (v/v) triton X-100, and 800 µL of suspended spores were mixed with 200 µL of 50% (w/v) glycerol in a 1.7 mL micro-centrifuge tube. Spore stocks were used to inoculate 50 mL of CDYE medium, flash frozen in liquid nitrogen and stored at −80°C.

Fungal transformations and analysis of indole diterpene chemotypes

In a series of complementation or pathway reconstitution experiments, a selection of the Level-3 plasmids produced in this work were transformed into *P. paxilli* strains harbouring appropriate genetic backgrounds (Table 8.5). Indole diterpene chemotypes [emindole SB (18), paspaline (11), paxilline (2), and/or nodulisporic acid F (82)] of G418-resistant fungal transformants were determined by an initial TLC screen of mycelial extracts and confirmed by LCMS and NMR. For these purposes, paspaline (11) and paxilline (2) reference standards were prepared from extracts of wild type *P. paxilli* (strain PN2013) and a nodulisporic acid F (82) reference standard was prepared from a pKV64(*nodM*/*nodW*):Δ*paxM*(PN2257) *P. paxilli* transformant (sKV170) by semi-preparative HPLC. The HPLC peaks were analysed by high-resolution mass spectrometry, which identified $[M+H]^+$ masses of 406.3109 *m/z* at 19.3 minutes, 422.3055 *m/z* at 17.1 minutes, 436.2485 *m/z* at 5.3 minutes and 436.2870 *m/z* at 5.6 minutes corresponding to the masses of emindole SB (18) (calc. $[M+H]^+$ 406.310 *m/z*), paspaline (11) (calc. $[M+H]^+$ 422.305 *m/z*), paxilline (2) (calc. $[M+H]^+$ 436.248 *m/z*) and nodulisporic acid F (82) (calc. $[M+H]^+$ 436.284 *m/z*) respectively (see Appendix Figures 9.8 to 9.11). NMR spectroscopy was used to confirm the structures of the purified reference standards (see Appendix Tables 9.7 to 9.9 and Figures 9.12 to 9.25).

Indole diterpene production and extraction

Fungal transformants were grown in 25 mL of CDYE medium with trace elements for 7 days at 28°C in shaker cultures (≥200 rpm), in 125 mL Erlenmeyer flasks capped

with cotton wool. Mycelia were isolated from fermentation broths by filtration through nappy liners, transferred to 50 mL centrifuge tubes (Lab Serv®, Thermo Fisher Scientific) and indole diterpenes were extracted by vigorously shaking the mycelia (≥200 rpm) in 2-butanone for ≥45 minutes.

Normal phase TLC

The 2-butanone supernatant (containing extracted indole diterpenes) was used for TLC analysis on solid phase silica gel 60 aluminum plates (Merck). Indole diterpenes were chromatographed with 9:1 chloroform:acetonitrile or 8:2 dichloromethane:acetonitrile and visualised with Ehrlich's reagent (1% (w/v) *p*-dimethylaminobenzaldehyde in 24% (v/v) HCl and 50% ethanol).

Reverse phase LCMS

Samples were prepared for LCMS from selected transformants that were identified by TLC. Accordingly, a 1 mL sample of the 2-butanone supernatant (containing extracted indole diterpenes) was transferred to a 1.7 mL micro-centrifuge tube and the 2-butanone was evaporated overnight. Contents were resuspended in 100% acetonitrile and filtered through a 0.2 µm membrane into a LCMS vial. LCMS samples were chromatographed on a reverse phase Thermo Scientific Accucore 2.6 µm C18 (50 × 2.1 mm) column attached to an UltiMate® 3000 Standard LC system (Dionex, Thermo Fisher Scientific) run at a flow rate of 0.200 mL/minute and eluted with aqueous solutions of acetonitrile containing 0.01% formic acid using a multistep gradient method (Table 8.6). Mass spectra were captured through in-line analysis on a maXis™ II quadrupole-time-of-flight mass spectrometer (Bruker).

Table 8.6. Multistep acetonitrile gradient used for LCMS analysis of fungal extracts.

Time (minutes)	% (v/v) of acetonitrile + 0.01% (v/v) formic acid
0	50
1	50
15	70
20	95
25	95
28	50
38	50

Large scale indole diterpene purification for NMR spectroscopy analysis

Fungal transformants that produced high levels of novel indole diterpenes were grown in ≥1 L of CDYE medium with trace elements, as described above under "***Indole diterpene production and extraction***". Mycelia were pooled into 1 L Schott bottles containing stir bars. 2-butanone was added and indole diterpenes were extracted overnight with stirring (≥700 rpm). Extracts were filtered through Celite® 545 (J. T. Baker®) and dry loaded onto silica with rotary evaporation for crude purification by silica column prior to a final purification by semi-preparative HPLC. A 1 mL aliquot of crude extract was injected onto a semi-preparative reversed phase Phenomenex 5 µm C18(2) 100 Å (250 × 15 mm) column attached to an UltiMate® 3000 Standard LC system (Dionex, Thermo Fisher Scientific) run at a flow rate of 8.00 mL/minute. Multistep gradient methods were optimised for the purification of different sets of indole diterpenes. The purity of each indole diterpene was assessed by LCMS and the structure was identified by NMR spectroscopy.

NMR spectroscopy

Paxilline and paspaline (11) samples were prepared in deuterated chloroform and the nodulisporic acid F (82) sample was prepared in deuterated methanol. Proton (^{1}H) and carbon (^{13}C) NMR spectra were recorded on an Agilent 400–MR instrument operating for ^{1}H NMR at 400 MHz and 100 MHz for ^{13}C NMR. All chemical shifts are quoted on the δ-scale in ppm using residual solvent as an internal standard. ^{1}H and ^{13}C spectra were assigned using correlation spectroscopy (COSY), heteronuclear single quantum correlation (HSQC) spectroscopy, and heteronuclear multiple bond

correlation (HMBC) spectroscopy. The nodulisporic acid F (82) HMBC spectra was captured on a Varian Unity Inova 600-MR instrument operating for ^{1}H NMR spectroscopy at 600 MHz and 150 MHz for ^{13}C NMR spectroscopy. Spectra were acquired with VNMRJ 3.2 or VNMRJ 4.2 acquisition software and processed with VNMRJ 3.2, VNMRJ 4.2, or MestReNova processing software. All NMR spectra can be found in the Appendix (Tables 9.7 to 9.9 and Figures 9.12 to 9.25).

9. APPENDIX

Additional Notes:

Introduction: Additional radiography feeding studies:

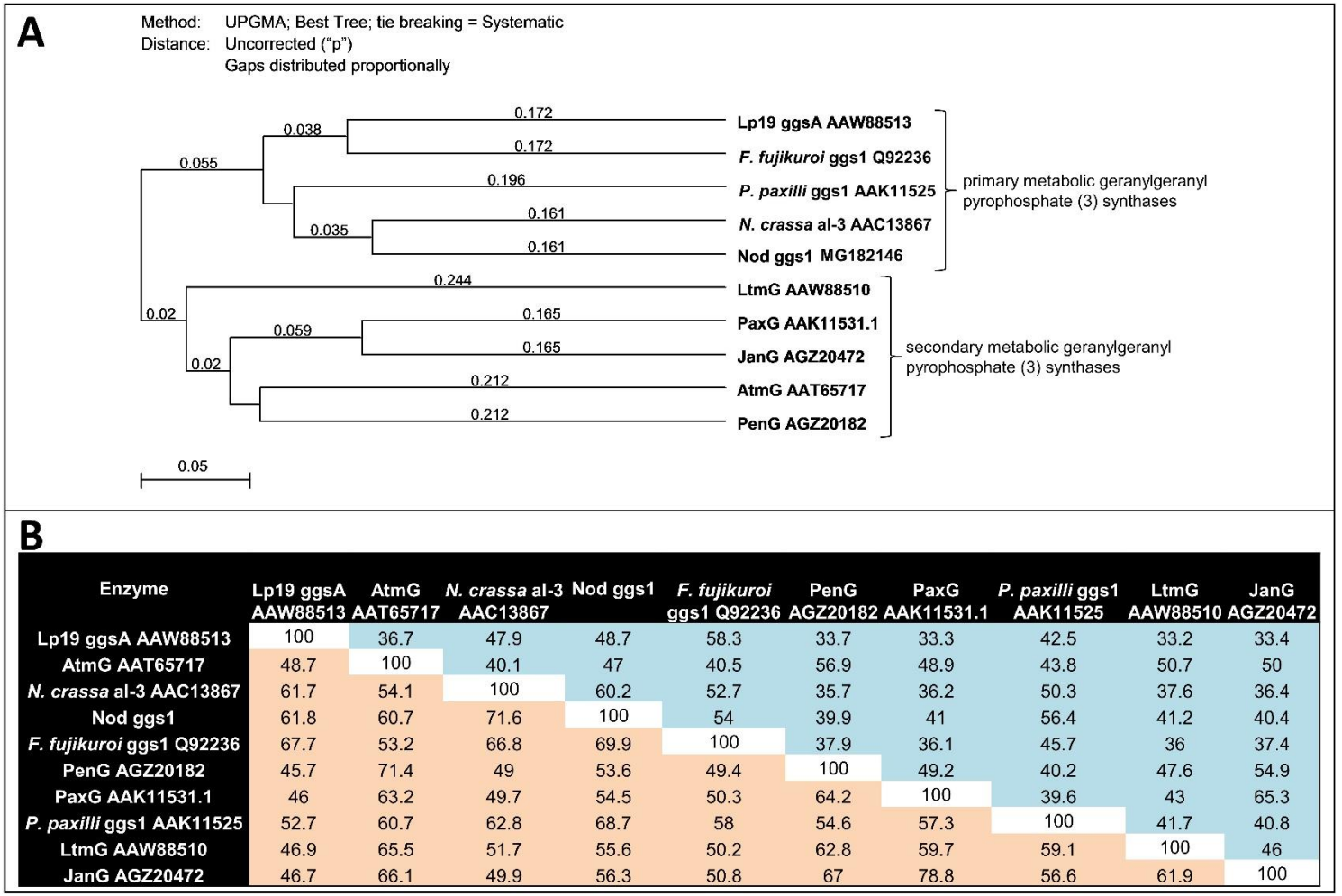
[14C]-paxilline was efficiently incorporated into the major indole diterpenes produced by *P. janczewskii* (penitrem A, E, and pennigritrem) and *P. janthinellum* (janthitrems B and C).[93]

α-[14C]-paxitriol and β-[14C]-paxitriol feeding on *P. janczewskii* and *P. janthinellum* showed incorporation of β-[14C]-paxitriol into indole diterpene end products. A small amount of janthitrem B became radiolabelled from α-[14C]-paxitriol indicating some isomerase activity in *P. janthinellum* but no penitrems were radiolabelled from α-[14C]-paxitriol in *P. janczewskii*. There was clear incorporation of β-[14C]-paxitriol and [14C]-paxilline into janthitrem D but none into janthitrem A.[94] β-10-*O*-acetyl[14C]paxilline was incorporated into janthitrem B in small amounts and janthitrem C in large amounts but most of the substrate remained unutilised in contrast to the β-[14C]-paxitriol and [14C]-paxilline, which were fully utilised – indicating that β-10-O-acetyl[14C]paxilline is unlikely a direct intermediate. From results, suggest that paxilline (2) is likely reduced to β-[14C]-paxitriol during janthitrem and penitrem biosynthesis.[93]

Chapter 3: Testing of the *paxC* promoter region (*paxC*$_{\text{ProUTR-2}}$)

To confirm that the *paxC* promoter region was functional with the additional 144 bp, MIDAS primers were eventually redesigned to include the missing base pairs in a new *paxC* promoter region (*paxC*$_{\text{ProUTR-2}}$, see Chapter 3, Figure 3.11) and two new Midas Level-3 plasmids were created harbouring *paxC* driven by *paxC*$_{\text{ProUTR-2}}$ (pKV30) and *nodC* driven by *paxC*$_{\text{ProUTR-2}}$ (pKV32). These two plasmids (pKV30 and pKV32) were individually transformed into the *P. paxilli* Δ*paxC* deletion mutant strain (PN2290) and chemical analysis of the transformant extracts revealed that paxilline (2) biosynthesis had been restored (Results not shown) thereby confirming that the *paxC*$_{\text{ProUTR-2}}$ was functional.

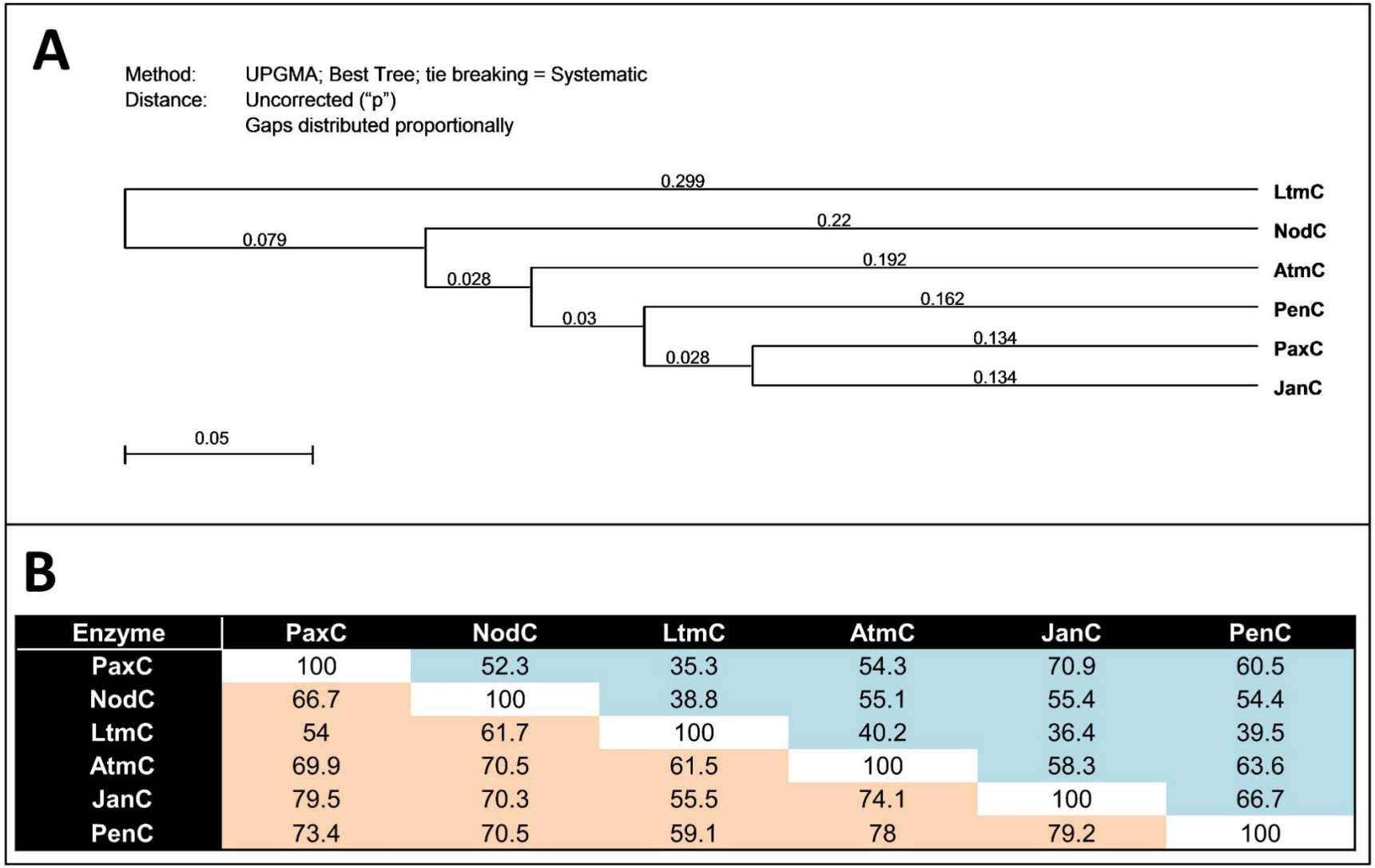
Bioinformatics analyses to identify the potential genetic elements involved in the biosynthesis of nodulisporic acids



B

Enzyme	Lp19 ggsA AAW88513	AtmG AAT65717	*N. crassa* al-3 AAC13867	Nod ggs1	*F. fujikuroi* ggs1 Q92236	PenG AGZ20182	PaxG AAK11531.1	*P. paxilli* ggs1 AAK11525	LtmG AAW88510	JanG AGZ20472
Lp19 ggsA AAW88513	100	36.7	47.9	48.7	58.3	33.7	33.3	42.5	33.2	33.4
AtmG AAT65717	48.7	100	40.1	47	40.5	56.9	48.9	43.8	50.7	50
N. crassa al-3 AAC13867	61.7	54.1	100	60.2	52.7	35.7	36.2	50.3	37.6	36.4
Nod ggs1	61.8	60.7	71.6	100	54	39.9	41	56.4	41.2	40.4
F. fujikuroi ggs1 Q92236	67.7	53.2	66.8	69.9	100	37.9	36.1	45.7	36	37.4
PenG AGZ20182	45.7	71.4	49	53.6	49.4	100	49.2	40.2	47.6	54.9
PaxG AAK11531.1	46	63.2	49.7	54.5	50.3	64.2	100	39.6	43	65.3
P. paxilli ggs1 AAK11525	52.7	60.7	62.8	68.7	58	54.6	57.3	100	41.7	40.8
LtmG AAW88510	46.9	65.5	51.7	55.6	50.2	62.8	59.7	59.1	100	46
JanG AGZ20472	46.7	66.1	49.9	56.3	50.8	67	78.8	56.6	61.9	100

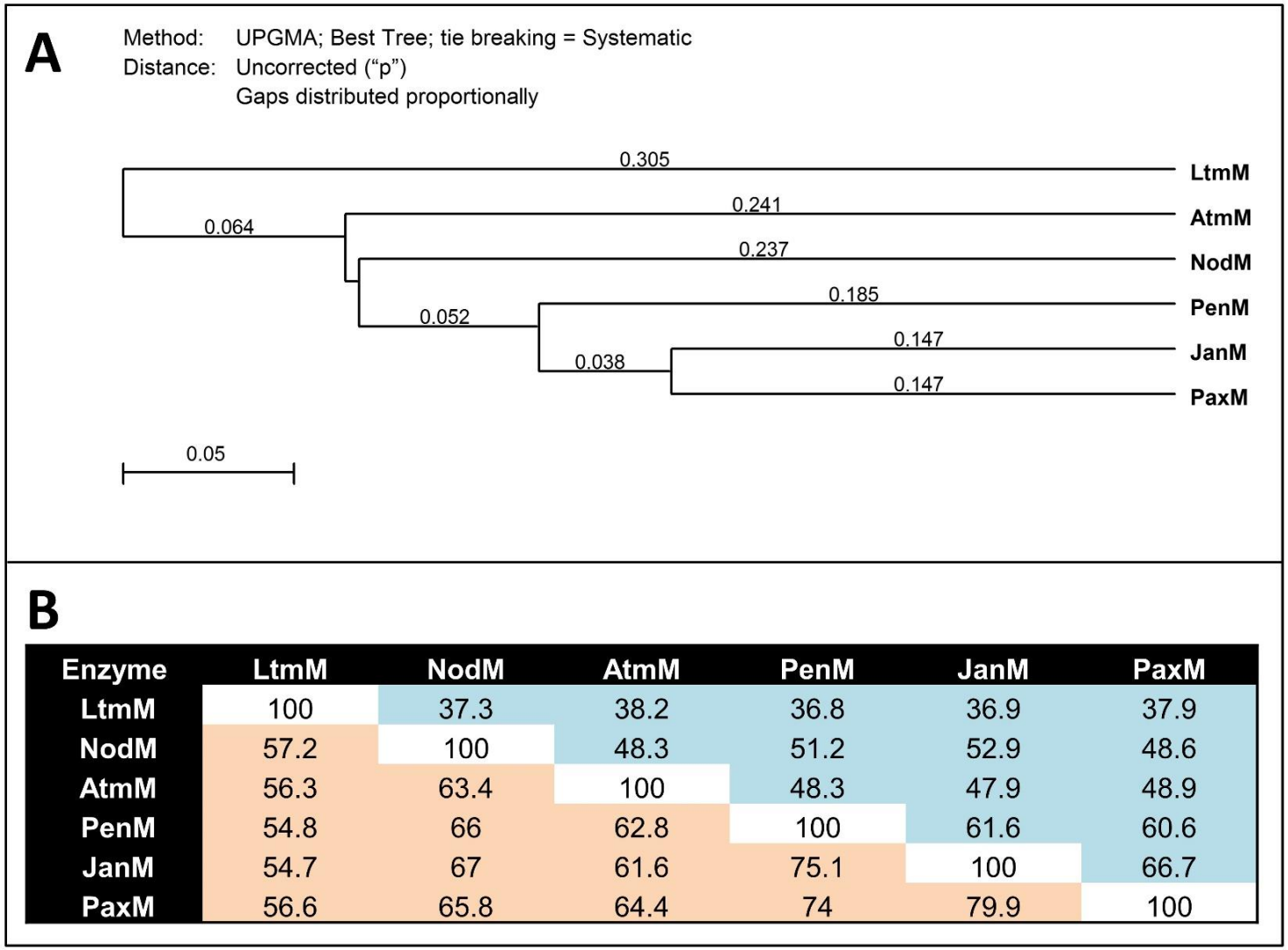
Figure 9.1. Phylogenetic tree (**A**) and similarity matrix (**B**) of geranylgeranyl synthases ('ggs enzymes' are involved inprimary metabolism and 'G' enzymes are involved in secondary metabolism). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the enzymes. Figure has been modified from Van de Bittner *et al*. 2018.[1]



B

Enzyme	PaxC	NodC	LtmC	AtmC	JanC	PenC
PaxC	100	52.3	35.3	54.3	70.9	60.5
NodC	66.7	100	38.8	55.1	55.4	54.4
LtmC	54	61.7	100	40.2	36.4	39.5
AtmC	69.9	70.5	61.5	100	58.3	63.6
JanC	79.5	70.3	55.5	74.1	100	66.7
PenC	73.4	70.5	59.1	78	79.2	100

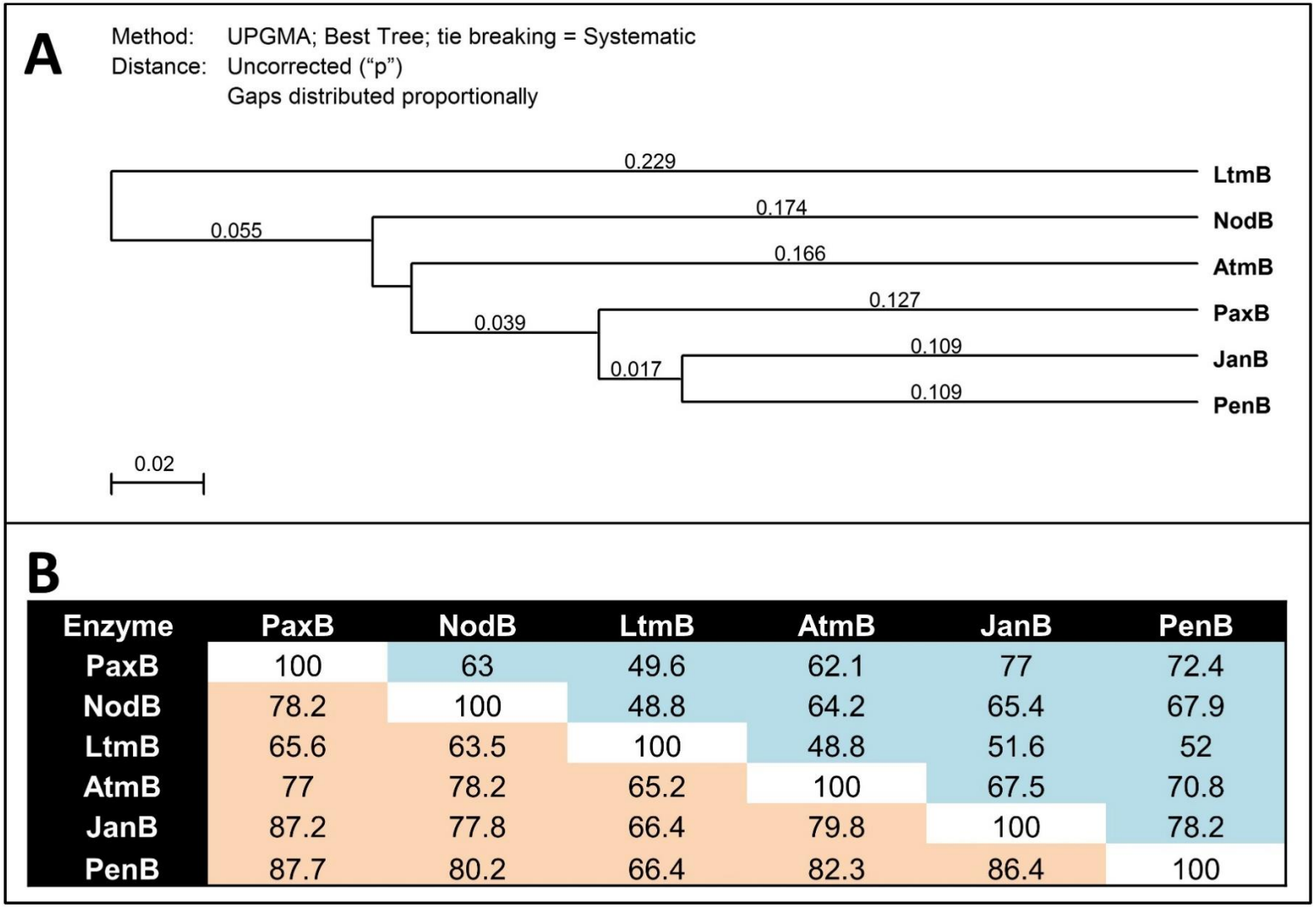
Figure 9.2. Phylogenetic tree (**A**) and similarity matrix (**B**) of geranylgeranyl transferases ('C' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the 'C' enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the 'C' enzymes.



B

Enzyme	LtmM	NodM	AtmM	PenM	JanM	PaxM
LtmM	100	37.3	38.2	36.8	36.9	37.9
NodM	57.2	100	48.3	51.2	52.9	48.6
AtmM	56.3	63.4	100	48.3	47.9	48.9
PenM	54.8	66	62.8	100	61.6	60.6
JanM	54.7	67	61.6	75.1	100	66.7
PaxM	56.6	65.8	64.4	74	79.9	100

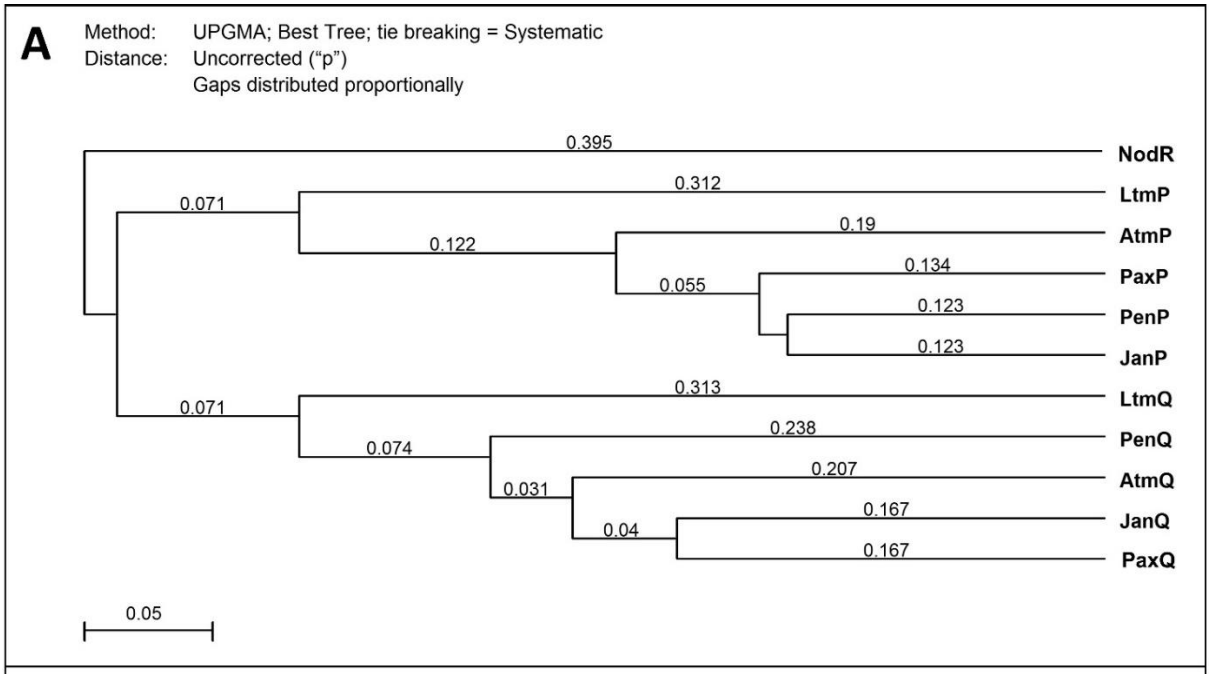
Figure 9.3. Phylogenetic tree (**A**) and similarity matrix (**B**) of indole diterpene epoxidases ('M' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the 'M' enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the 'M' enzymes.



B

Enzyme	PaxB	NodB	LtmB	AtmB	JanB	PenB
PaxB	100	63	49.6	62.1	77	72.4
NodB	78.2	100	48.8	64.2	65.4	67.9
LtmB	65.6	63.5	100	48.8	51.6	52
AtmB	77	78.2	65.2	100	67.5	70.8
JanB	87.2	77.8	66.4	79.8	100	78.2
PenB	87.7	80.2	66.4	82.3	86.4	100

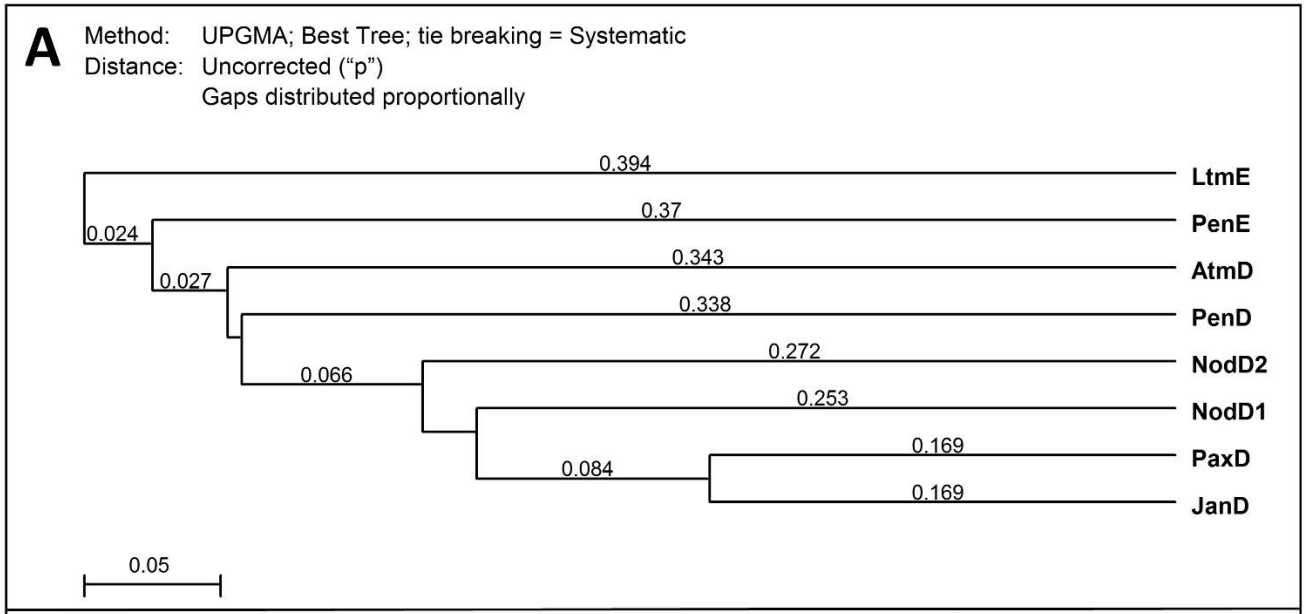
Figure 9.4. Phylogenetic tree (**A**) and similarity matrix (**B**) of indole diterpene epoxidases ('B' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the 'B' enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the 'B' enzymes.



B

Enzyme	AtmP	AtmQ	PenQ	PenP	JanQ	JanP	LtmQ	LtmP	NodR	PaxQ	PaxP
AtmP	100	20.9	21.9	62.1	21.7	61.8	20.9	35.2	16.8	21.8	60.7
AtmQ	39.3	100	48.4	21.4	55.9	21.3	35.2	18.4	20	53.9	21.7
PenQ	39.8	65	100	21.1	50.9	23.3	34.6	16.2	18	51.4	24
PenP	78.4	40.9	36.7	100	22	74.6	20.4	35.2	20.1	22.5	70.6
JanQ	39.9	70.8	66.2	40.7	100	23	33.6	19.2	19.4	64.5	22.7
JanP	75.3	40.6	38	86.2	41.1	100	20.5	33.7	21.3	23.4	75
LtmQ	36.9	52.8	52.7	37.8	54.2	37.8	100	17.4	19.2	35.1	21.8
LtmP	55.1	35.9	35.9	55.5	37.2	53.5	34.4	100	17.3	19.2	35.6
NodR	37.6	36.7	34.7	39	37.5	40.8	35.5	36.7	100	20.3	19.9
PaxQ	40.4	69.1	67.4	41.1	76.9	41.1	52.8	37	37.2	100	23.4
PaxP	75.5	40.7	39.3	85.2	42	86.2	39.2	56	40.8	41.3	100

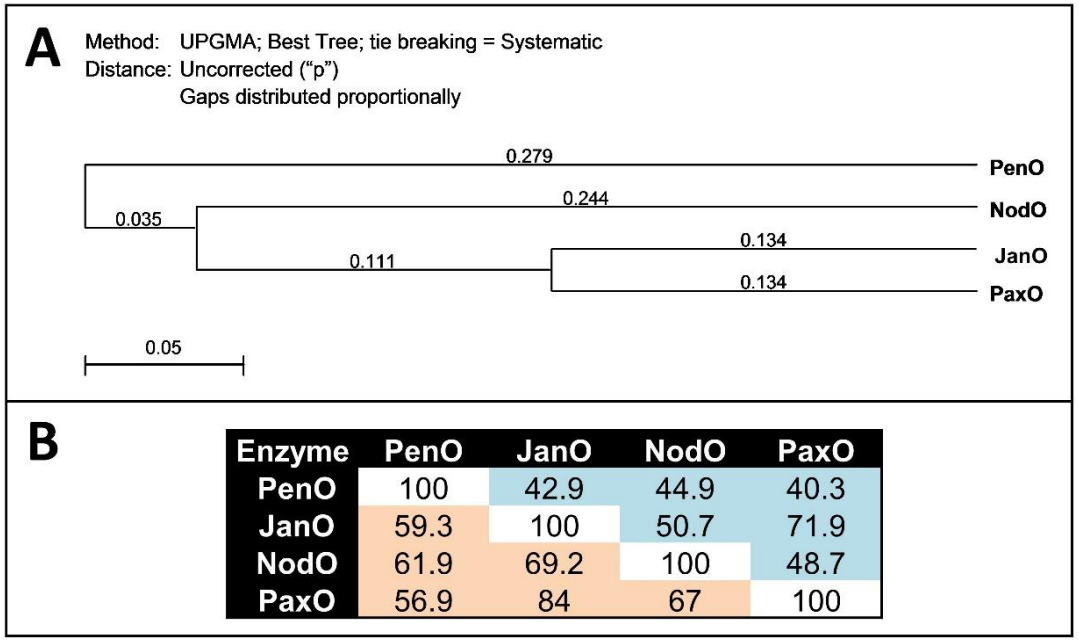
Figure 9.5. Phylogenetic tree (**A**) and similarity matrix (**B**) of indole diterpene cytochrome P450 oxygenases ('P' and 'Q' enzymes compared to NodR). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the enzymes.



B

Enzyme	NodD2	PaxD	NodD1	JanD	AtmD	LtmE	PenD	PenE
NodD2	100	42.3	44.7	45	31.6	11.3	32.6	23.3
PaxD	61.9	100	44.9	65.8	31.3	11.4	31.4	24.3
NodD1	60.7	63.6	100	49.2	29.2	11.2	29.7	22.7
JanD	63.1	80.6	65.6	100	30.5	11.7	31.9	25.2
AtmD	49.6	49.4	47.6	50.2	100	11.2	28.6	25.2
LtmE	20	20.7	21	21.6	19.5	100	10.8	11.6
PenD	54.4	53.5	50.7	51.9	48.1	20.9	100	24.3
PenE	41.1	40.5	40.2	41.3	37.4	21.7	41.3	100

Figure 9.6. Phylogenetic tree (**A**) and similarity matrix (**B**) of indole diterpene prenyl transferases ('D' and 'E' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the enzymes.



Enzyme	PenO	JanO	NodO	PaxO
PenO	100	42.9	44.9	40.3
JanO	59.3	100	50.7	71.9
NodO	61.9	69.2	100	48.7
PaxO	56.9	84	67	100

Figure 9.7. Phylogenetic tree (**A**) and similarity matrix (**B**) of indole diterpene flavin adenine dinucleotide-dependent oxidative cyclases ('O' enzymes). The phylogenetic tree depicts the difference in % identity scores for the amino acid residues of the enzymes. The similarity matrix depicts the % identity scores (blue shaded regions ( )) and the % similarity scores (orange shaded regions ( )) for the amino acid residues of the enzymes.

Table 9.1. Similarity matrix of NodY enzymes. Blue shaded regions ( ) represent % identity scores and orange shaded regions ( ) represent % similarity scores for amino acid residues.

Enzyme	NodY2	NodY1
NodY2	100	41.1
NodY1	60	100

Assembling of genetic elements (5'UTRs, CDS, 3'UTRs, etc.) into MIDAS vectors

Table 9.2. PCR primers for amplification of transcription unit modules (TUMs). The forward and reverse PCR primers used for amplification of TUMs (i.e. promoters [ProUTR], coding sequences [CDSs], and terminators [UTRterm]) are listed. Primers used to amplify TUM fragments for domestication purposes (i.e. removal of internal sites for *Aar*I, *Bsa*I or *Bsm*BI) are shaded in orange ( ). The template for amplification of *nod* CDSs was genomic DNA from *Hypoxylon pulicicidum* strain ATCC® 74245™.[84] The template for amplification of *pax* gene TUMs was genomic DNA from *Penicillium paxilli* strain ATCC® 26601™ (PN2013) [Accession HM171111].[72] The PCR products used to produce the *trpC* ProUTR module, *nptII* CDS module (conferring resistance to geneticin (G418)), and *trpC*$_{UTRterm}$ module were all amplified from plasmid pII99.[95] The *Bsm*BI recognition sites are colour coded (cgtctc), with the overhangs generated following *Bsm*BI cleavage shown by the grey shading ( ). The 5' (prefix) and 3' (suffix) nucleotide bases, which flank each TUM and form the basis of the address system for each of the MIDAS modules, are shown in blue and red respectively.

TUM	Primer name	Primer sequence (5' to 3')
***Hypoxylon pulicicidum* primers**		
nodW		
nodW$_{CDS}$	P4502 frag 1 F	cgatgtacgtctcaCTCGAATGactttagctattttaggcatcagttgcc
	P4502 frag 1 R	actgctcgtctcaACTCccgctgcgagccgct
	P4502 frag 2 F	acgtacgtctccGAGTccggtcctggtggagtgatc
	P4502 frag 2 R	gacctttcgtctctGTCTcaAAGCctaagttatgcccagatatttccag
nodR		
nodR$_{CDS}$	nodP frag 1 F	cgatgtacgtctcaCTCGAATGtttgatattgattttggcattctatttcc
	nodP frag 1 R	gtacagcgtctcaTTGAgtctccttcaggacgctgtc
	nodP frag 2 F	gtcactcgtctcaTCAAcgcctaaaacctatgcaaatc
	nodP frag 2 R	gacctttcgtctctGTCTcaAAGCttacagttgaatctcttctttccgc
nodX		
nodX$_{CDS}$	NA-03_frag1_F	cgatgtacgtctcaCTCGAATGgagtcaagtgcggataatacag
	NA-03_frag1_R	actgacgtctcgGAACaaggttgcgtggtcgctatc
	NA-03_frag2_F	gtatcgcgtctctGTTCctcatgtcccccatgc
	NA-03_frag2-R	actgacgtctcgTCCCaccttgccatttttctgg
	NA-03_frag3-F	gtatcgcgtctctGGGActgtatgaggtcattagggaccc
	NA-03_frag3-R	actgacgtctcgGCCAgcggtcgcttggg
	NA-03_frag4-F	gtatcgcgtctccTGGCcgggctgaaggtg
	NA-03_frag4-R	actgacgtctcgGCGTagatctgccggcc
	NA-03_frag5-F	gtatcgcgtctctACGCcgtgagattgtgagagatgagg
	NA-03_frag5_R	gacctttcgtctctGTCTcaAAGCctaatctgcgatggaccctgac

nodM		
*nodM*CDS	nodM frag1 F	cgatgtacgtctcaCTCGAATGtctacccctgagttcaagg
	nodM frag1 R	cagtcacgtctcaACGCctctcaagaacgatgtgggaaattc
	nodM frag2 F	gtgcatcgtctcaGCGTagtgtaatcgcaccagag
	nodM frag2 R	gacctttcgtctctGTCTcaAAGCctatgaagcgatgtctctaatatggagtaac
nodB		
*nodB*CDS	nodB F	cgatgtacgtctcaCTCGAATGgatggattcgatcgttccaatg
	nodB R	gacctttcgtctctGTCTcaAAGCttattgagccttccgcgcattg
nodJ		
*nodJ*CDS	P4501 frag 1 F	cgatgtacgtctcaCTCGAATGgaacttatcgttattatcataactctcg
	P4501 frag 1 R	actgaccgtctcaCGGTgtctgtatggaacttgcaag
	P4501 frag 2 F	gtatcgcgtctctACCGgcaggggagttcatggtga
	P4501 frag 2 R	gacctttcgtctctGTCTcaAAGCttacccaactagcagtggaaacttc
nodC		
*nodC*CDS	nodC frag1 F	cgatgtacgtctcaCTCGAATGtccttaggtttacagtgcttgg
	nodC frag1 R	cattgacgtctcgGTCAcgtcgccaaaccagcga
	nodC frag2 F	gtcacgcgtctcttGACggcctcactagctttcc
	nodC frag2 R	gacctttcgtctctGTCTcaAAGCtcaatgcgtaagatcgagtttctccttttct
nodZ		
*nodZ*CDS	P4503 frag 1 F	cgatgtacgtctcaCTCGAATGataacagcatccacttccg
	P4503 frag 1 R	actgtccgtctcaGAGTccaatcaggacgcaactaatcatc
	P4503 frag 2 F	actgcacgtctctACTCccatggtgtcgagatgag
	P4503 frag 2 R	ctacgtcgtctctCGCTtctctggccacaaaacggaaag
	P4503 frag 3 F	ctgactcgtctctAGCGtttcacacagaggggaaag
	P4503 frag 3 R	gacctttcgtctctGTCTcaAAGCtaggtatccaccacacggttc
***Penicillium paxilli* primers**		
paxG		
*paxG*ProUTR	PpaxG F	cgatgtacgtctcaCTCGGGAGattcacgacctgtgactagtcaa
	PpaxG R	gacctttcgtctctGTCTcaCATTggcgtcgaacttgatgaagttttc
*paxG*CDS	paxG frag1 F	cgatgtacgtctcaCTCGAATGtcctacatccttgcagaag
	paxG frag1 R	cttctacgtctcgTACTgttctaatcgtgcttggtg
	paxG frag2 F	gcacgacgtctccAGTAcaggtgctagaagatgacgttgac
	paxG frag2 R	aggcgccgtctccACCAatctctttcaatcttgcttgttgga
	paxG frag3 F	gattgacgtctcttGGTgacccccgcgcctt
	paxG frag3 R	gtcgaccgtctcttTTCCctagtatattggaagctccccg
	paxG frag4 F	tccaatcgtctcgGGAAaccctaagtcgacttagtgcg
	paxG frag4 R	gacctttcgtctctGTCTcaAAGCttaaactcttccttttctcattagtaggg
*paxG*UTRterm	TpaxG F	cgatgtacgtctcaCTCGGCTTtcaatcgtgctgcatttctctt
	TpaxG R	gacctttcgtctctGTCTcaAGCGtcactcccgagcaatattgct
paxC		
*paxC*ProUTR-1	P paxC F	cgatgtacgtctcaCTCGGGAGtgaccatgatgcgtattaacggc
	P paxC R	gacctttcgtctctGTCTcaCATTcaaaatgggacctacaccctgaa
*paxC*ProUTR-2	PpaxC F2	cgatgtacgtctcaCTCGGGAGacaacaaaaagatcagccaatgg
	PpaxC R2	gacctttcgtctctGTCTcaCATTaaaatgggacctacaccctgaa
*paxC*CDS	paxC frag1 F	cgatgtacgtctcaCTCGAATGggcgtagcaggga

	paxC frag1 R	cattgacgtctccACGGcgccagacaaggga
	paxC frag2 F	cccttgcgtctcgCCGTgacggagtcaatgggttc
	paxC frag2 R	gacctttcgtctctGTCTcaAAGCtcatgccttcaggtcaagcttc
*paxC*UTRterm	TpaxC F	cgatgtacgtctcaCTCGGCTTttggccttgtgaaatatgggactac
	TpaxC R	gacctttcgtctctGTCTcaAGCGatctctgtcatgtcggatatcagat
paxM		
*paxM*ProUTR	PpaxM F	cgatgtacgtctcaCTCGGGAGgttgttggcatgggagtaggat
	PpaxM R	gacctttcgtctctGTCTcaCATTggtttctgaatcttaaagatacatgaaaag
*paxM*CDS	paxM frag1 F	cgatgtacgtctcaCTCGAATGgaaaaggccgagtttcaag
	paxM frag1 R	tgacaacgtctcgTCCAtcgaataaagcgttgacttgc
	paxM frag2 F	acgcttcgtctcaTGGActcactattgtcacaatccatggaaaag
	paxM frag2 R	gacctttcgtctctGTCTcaAAGCttaaacttgaagaaaataaaacttcagggcac
*paxM*UTRterm	TpaxM frag1 F	cgatgtacgtctcaCTCGGCTTaccattggagcaatttttggttttc
	TpaxM frag1 R	gttcgccgtctcgACTCgattgcttgtgggtct
	TpaxM frag2 F	acaagccgtctccGAGTccagccagcgaacttg
	TpaxM frag2 R	gacctttcgtctctGTCTcaAGCGttttggcttacttcagtttaactgttttg
paxB		
*paxB*ProUTR	PpaxB F	cgatgtacgtctcaCTCGGGAGaaggctgtgttggagagaatc
	PpaxB R	gacctttcgtctctGTCTcaCATTagtttctaaggttgacgtgggaaaaag
*paxB*CDS	paxB F	cgatgtacgtctcaCTCGAATGgacggttttgatgtttcccaa
	paxB R	gacctttcgtctctGTCTcaAAGCtcaatttgctttttcggcccgcttatgc
*paxB*UTRterm	TpaxB F	cgatgtacgtctcaCTCGGCTTtcggcagttgagggtgaaac
	TpaxB R	gacctttcgtctctGTCTcaAGCGggttaacaatgaggaacgatgaacag
Additional primers		
janD		
*janD*ProUTR	P janD F	cgatgtacgtctcaCTCGGGAGcttacccgtccttttttctcaacg
	P janD R	gacctttcgtctctGTCTcaCATTgtttaagatgatcttgagcttgatgaagtg
trpC		
*trpC*ProUTR	PtrpC frag1 F	cgatgtacgtctcaCTCGGGAGgaattcatgccagttgttcccag
	PtrpC frag1 R	cgatgtacgtctcaGCTTggccgactcgctg
	PtrpC frag2 F	caccttcgtctccAAGCagacgtgaagcaggacgg
	PtrpC frag2 R	cgatgtcgtctcgCAGAccattgcacaagcctc
	PtrpC frag3 F	gacctttcgtctcgTCTGcgcatggatcgctgc
	PtrpC frag3 R	gacctttcgtctctGTCTcaCATTatcgatgcttgggtagaataggtaag
*trpC*UTRterm	T trpC frag1 F	cgatgtacgtctcaCTCGGCTTgatccacttaacgttactgaaatcatcaaac
	T trpC frag1 R	gacctttcgtctctCTGCttgatctcgtctgccga
	T trpC frag2 F	cgatgtacgtctcaGCAGatcaacggtcgtcaaga
	T trpC frag2 R	gacctttcgtctctGTCTcaAGCGtctagaaagaaggattacctctaaacaagtgt
nptII		
*nptII*CDS	ntpII F	cgatgtacgtctcaCTCGAATGattgaacaagatggattgcacg
	ntpII R	gacctttcgtctctGTCTcaAAGCctcagaagaactcgtcaagaaggc

Table 9.3. MIDAS Level-1 plasmid library: Assembly of transcription unit modules (i.e. promoters [ProUTR], coding sequences [CDSs], and terminators [UTRterm]) in pML1. This table represents the MIDAS level-1 transcription unit modules that were used to assemble MIDAS level-2 transcription units (Table 9.4). The 4 base prefixes and suffixes (5' to 3') that flank each transcription unit module are shown at the top of the table to highlight the sequences used to bind the transcription unit modules together to make MIDAS level-2 transcription units. These 4 base flanking regions are depicted in the primer table (Table 9.2) in blue (forward addresses) and red (reverse addresses).

[GGAG]		[AATG]		[GCTT]	[CGCT]
ProUTR modules		**CDS modules**		**UTRterm modules**	
Plasmid name	**Description**	**Plasmid name**	**Description**	**Plasmid name**	**Description**
pSK1	*paxG*$_{\text{ProUTR}}$	pKV45	*nodW*$_{\text{CDS}}$	pSK3	*paxG*$_{\text{UTRterm}}$
pSK10	*paxC*$_{\text{ProUTR-1}}$	pKV58	*nodR*$_{\text{CDS}}$	pSK12	*paxC*$_{\text{UTRterm}}$
pKV28	*paxC*$_{\text{ProUTR-2}}$	pKV46	*nodX*$_{\text{CDS}}$	pSK6	*paxM*$_{\text{UTRterm}}$
pSK4	*paxM*$_{\text{ProUTR}}$	pKV59	*nodM*$_{\text{CDS}}$	pSK9	*paxB*$_{\text{UTRterm}}$
pSK7	*paxB*$_{\text{ProUTR}}$	pSK18		pSK15	*trpC*$_{\text{UTRterm}}$
pSK17	*trpC*$_{\text{ProUTR}}$	pSK19	*nodB*$_{\text{CDS}}$		
pLB9	*janD*$_{\text{ProUTR}}$	pKV47	*nodJ*$_{\text{CDS}}$		
		pSK20	*nodC*$_{\text{CDS}}$		
		pKV50	*nodZ*$_{\text{CDS}}$		
		pSK2	*paxG*$_{\text{CDS}}$		
		pSK11	*paxC*$_{\text{CDS}}$		
		pSK5	*paxM*$_{\text{CDS}}$		
		pSK16	*nptII*$_{\text{CDS}}$		

Table 9.4. MIDAS Level-2 plasmid library: Assembly of transcription units (TUs) in pML2 destination vectors. This table represents the construction of the MIDAS level-2 TUs that were used to assemble MIDAS level-3 multi-gene plasmids (Table 9.5) for heterologous expression studies. TUs are described by the coding sequence (CDS) they contain (in the TU column) and the Level-1 entry clones used for TU assembly are depicted in their corresponding promoter (ProUTR), CDS, or terminator (UTRterm) columns. The column labelled pML2 destination vector describes the type of MIDAS Level-2 plasmid that was used to assemble each TU. The names of the Level-2 entry plasmids produced are shown in the blue shaded (■) column and the TU orientation, determined by the pML2 destination vector, is shown by the arrowhead (► for forward (F) and ◄ for reverse (R) destination vector) in the Level-2 entry clone description.

TU	Level-1 entry clones used for TU assembly			pML2 destination vector	Level-2 entry clones	
	ProUTR	CDS	UTRterm		Name	Description
nodW	pSK17	pKV45	pSK15	pML2(+)WR	pKV52	◄(T_{trpC}-*nodW*-P_{trpC}):pML2(+)WR
				pML2(+)BR	pSK67	◄(T_{trpC}-*nodW*-P_{trpC}):pML2(+)BR
	pLB9				pSK80	◄(T_{trpC}-*nodW*-P_{janD}):pML2(+)BR
nodR	pSK17	pKV58	pSK15	pML2(+)WF	pKV53	(P_{trpC}-*nodR*-T_{trpC})►:pML2(+)WF
nodX	pSK17	pKV46	pSK15	pML2(+)WR	pKV54	◄(T_{trpC}-*nodX*-P_{trpC}):pML2(+)WR
nodM	pSK17	pSK18	pSK15	pML2(+)BF	pKV33	(P_{trpC}-*nodM*-T_{trpC})►:pML2(+)BF
	pSK4	pKV59	pSK6		pKV57	(P_{paxM}-*nodM*-T_{paxM})►:pML2(+)BF
		pSK18		pML2(+)WF	pSK28	(P_{paxM}-*nodM*-T_{paxM})►:pML2(+)WF
nodB	pSK17	pSK19	pSK15	pML2(+)BF	pKV35	(P_{trpC}-*nodB*-T_{trpC})►:pML2(+)BF
	pSK7		pSK9	pML2(+)BR	pSK29	◄(T_{paxB}-*nodB*-P_{paxB}):pML2(+)BR
nodJ	pSK17	pKV47	pSK15	pML2(+)WR	pKV55	◄(T_{trpC}-*nodJ*-P_{trpC}):pML2(+)WR
nodC	pSK17	pSK20	pSK15	pML2(+)BF	pKV26	(P_{trpC}-*nodC*-T_{trpC})►:pML2(+)BF
	pKV28		pSK12	pML2(+)WF	pSK60	(P_{paxC-2}-*nodC*-T_{paxC})►:pML2(+)WF
	pSK10			pML2(+)BF	pKV2	(P_{paxC-1}-*nodC*-T_{paxC})►:pML2(+)BF
nodZ	pSK17	pKV50	pSK15	pML2(+)WF	pKV56	(P_{trpC}-*nodZ*-T_{trpC})►:pML2(+)WF
paxG	pSK1	pSK2	pSK3	pML2(+)BR	pSK21	◄(T_{paxG}-*paxG*-P_{paxG}):pML2(+)BR
paxC	pSK10	pSK11	pSK12	pML2(+)BF	pKV24	(P_{paxC-1}-*paxC*-T_{paxC})►:pML2(+)BF
	pKV28			pML2(+)WF	pSK59	(P_{paxC-2}-*paxC*-T_{paxC})►:pML2(+)WF
paxM	pSK4	pSK5	pSK6	pML2(+)WR	pSK22	◄(T_{paxM}-*paxM*-P_{paxM}):pML2(+)WR
nptII	pSK17	pSK16	pSK15	pML2(+)WF	pSK26	(P_{trpC}-*nptII*-T_{trpC})►:pML2(+)WF

Table 9.5. MIDAS Level-3 plasmid library: Multigene assemblies in pML3. This table shows the Level-2 entry clone and Level-3 destination vectors used to construct the Level-3 product plasmids. The number of level 3 assembly reactions used to create the level-3 plasmid is indicated by number in the step column. The name and description of each Level-2 entry clone and the Level-3 destination vector that the Level-2 entry clones were assembled into are shown in their respective columns. The enzyme (*Aar*I or *Bsm*BI) used to mediate the Level-3 reaction is listed in the Golden Gate reaction column. The name (blue shaded ( ) column), description and size of plasmids produced during each cycle of Level-3 assembly are shown in their respective Level-3 product plasmid columns. In the Level-3 product plasmid descriptions, transcription units are annotated with the name of the coding sequence they contain and transcription unit orientation is shown by the arrowhead.

Step	Level-2 entry clone		Destination vector	Golden Gate reaction	Product Level-3 plasmid		
	Name	Description			Name	Description	Plasmid size (kb)
1	pSK26	(P_{trpC}-*nptII*-T_{trpC})►:pML2(+)WF	pML3	*Aar*I	pKV22	pML3:*nptII*►	5.6
2	pKV2	(*P_{paxC-1}-nodC-T_{paxC})*►:pML2(+)BF	pKV22	*Bsm*BI	pKV13	pML3:*nptII*►:*nodC*►	8.3
2	pKV24	(*P_{paxC-1}-paxC-T_{paxC})*►:pML2(+)BF	pKV22	*Bsm*BI	pKV25	pML3:*nptII*►:*paxC*►	8.2
2	pKV26	(P_{trpC}-*nodC*-T_{trpC})►:pML2(+)BF	pKV22	*Bsm*BI	pKV27	pML3:*nptII*►:*nodC*►	9.0
2	pKV33	(P_{trpC}-*nodM*-T_{trpC})►:pML2(+)BF	pKV22	*Bsm*BI	pKV34	pML3:*nptII*►:*nodM*►	9.5
2	pKV35	(P_{trpC}-*nodB*-T_{trpC})►:pML2(+)BF	pKV22	*Bsm*BI	pKV36	pML3:*nptII*►:*nodB*►	8.0
2	pKV57	(*P_{paxM}-nodM-T_{paxM})*►:pML2(+)BF	pKV22	*Bsm*BI	pKV63	pML3:*nptII*►:*nodM*►	9.4
3	pKV52	◄(*T_{trpC}-nodW-P_{trpC})*:pML2(+)WR	pKV63	*Aar*I	pKV64	pML3:*nptII*►:*nodM*►:◄*nodW*	13.0
3	pKV53	(*P_{trpC}-nodR-T_{trpC})*►:pML2(+)WF	pKV63	*Aar*I	pKV65	pML3:*nptII*►:*nodM*►:*nodR*►	13.0
3	pKV54	◄(*T_{trpC}-nodX-P_{trpC})*:pML2(+)WR	pKV63	*Aar*I	pKV66	pML3:*nptII*►:*nodM*►:◄*nodX*	13.2
3	pKV55	◄(*T_{trpC}-nodJ-P_{trpC})*:pML2(+)WR	pKV63	*Aar*I	pKV67	pML3:*nptII*►:*nodM*►:◄*nodJ*	12.9
3	pKV56	(*P_{trpC}-nodZ-T_{trpC})*►:pML2(+)WF	pKV63	*Aar*I	pKV68	pML3:*nptII*►:*nodM*►:*nodZ*►	13.0
1	pSK26	(P_{trpC}-*nptII*-T_{trpC})►:pML2(+)WF	pML3	*Aar*I	pSK33	pML3:*nptII*►	5.6
2	pSK21	◄(*T_{paxG}-paxG-P_{paxG})*:pML2(+)BR	pSK33	*Bsm*BI	pSK34	pML3:*nptII*►:◄*paxG*	8.2
3	pSK22	◄(*T_{paxM}-paxM-P_{paxM})*:pML2(+)WR	pSK34	*Aar*I	pSK36	pML3:*nptII*►:◄*paxG*:◄*paxM*	11.5
4	pSK29	◄(*T_{paxB}-nodB-P_{paxB})*:pML2(+)BR	pSK36	*Bsm*BI	pKV73	pML3:*nptII*►:◄*paxG*:◄*paxM:*◄*nodB*	14.1
5	pSK59	(*P_{paxC-2}-paxC-T_{paxC})*►:pML2(+)WF	pSK73	*Aar*I	pKV74	pML3:*nptII*►:◄*paxG*:◄*paxM:*◄*nodB*:*paxC*►	16.3
3	pSK28	(*P_{paxM}-nodM-T_{paxM})*►:pML2(+)WF	pSK34	*Aar*I	pSK35	pML3:*nptII*►:◄*paxG*:*nodM*►	11.5
4	pSK29	◄(*T_{paxB}-nodB-P_{paxB})*:pML2(+)BR	pSK35	*Bsm*BI	pSK38	pML3:*nptII*►:◄*paxG*:*nodM*►:◄*nodB*	14.1
5	pSK60	(*P_{paxC-2}-nodC-T_{paxC})*►:pML2(+)WF	pSK38	*Aar*I	pSK66	pML3:*nptII*►:◄*paxG*:*nodM*►:◄*nodB*:*nodC*►	16.3
6	pSK67	◄(*T_{trpC}-nodW-P_{trpC})*:pML2(+)BR	pSK66	*Bsm*BI	pSK68	pML3:*nptII*►:◄*paxG*:*nodM*►:◄*nodB*:*nodC*►:◄*nodW*	20.5
6	pSK80	◄(*T_{trpC}-nodW-P_{janD})*:pML2(+)BR	pSK66	*Bsm*BI	pSK81	pML3:*nptII*►:◄*paxG*:*nodM*►:◄*nodB*:*nodC*►:◄*nodW*	20.3

Table 9.6. Table of notable fungal strains generated in this thesis.

Newly generated *P. paxilli* strains	Description (plasmid/destination strain)	Notable indole diterpene phenotype
sKV50	pKV13/PN2290	none
sKV60	pKV25/PN2290	none
sKV44	pPN1783/PN2290	paxilline
sKV66	pKV27/PN2290	paxilline
sKV98	pKV34/PN2257	emindole SB (18)
sKV101	pPN1857/PN2257	paxilline
sKV136	pKV63/PN2257	emindole SB (18)
sKV88	pKV36/PN2458	none
sSK66-14	pSK66/PN2250	emindole SB (18)
sKV254	pKV74/PN2250	paspaline (11)
sKV170	pKV64/PN2257	nodulisporic acid F (82)
sKV157	pKV65/PN2257	emindole SB (18)
sKV160	pKV67/PN2257	emindole SB (18)
sKV185	pKV66/PN2257	emindole SB (18)
sKV193	pKV68/PN2257	emindole SB (18)
sSK68-11	pSK68/PN2250	nodulisporic acid F (82)
sSK81-4	pSK81/PN2250	nodulisporic acid F (82)

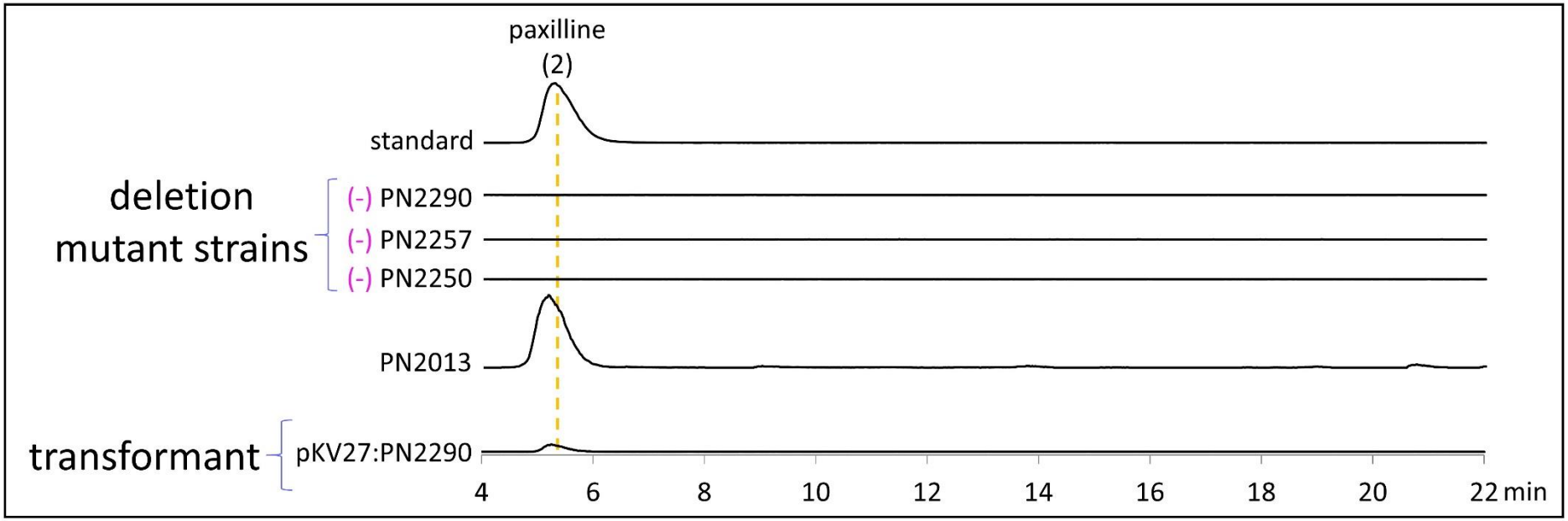


Figure 9.8. Extracted ion chromatograms (EICs) used to detect paxilline (2) (5.3 minutes, 436.248 ± 0.01 m/z). A pink negative sign (-) indicates traces where paxilline (2) was not detected. All EICs have been scaled to 300,000 AU.

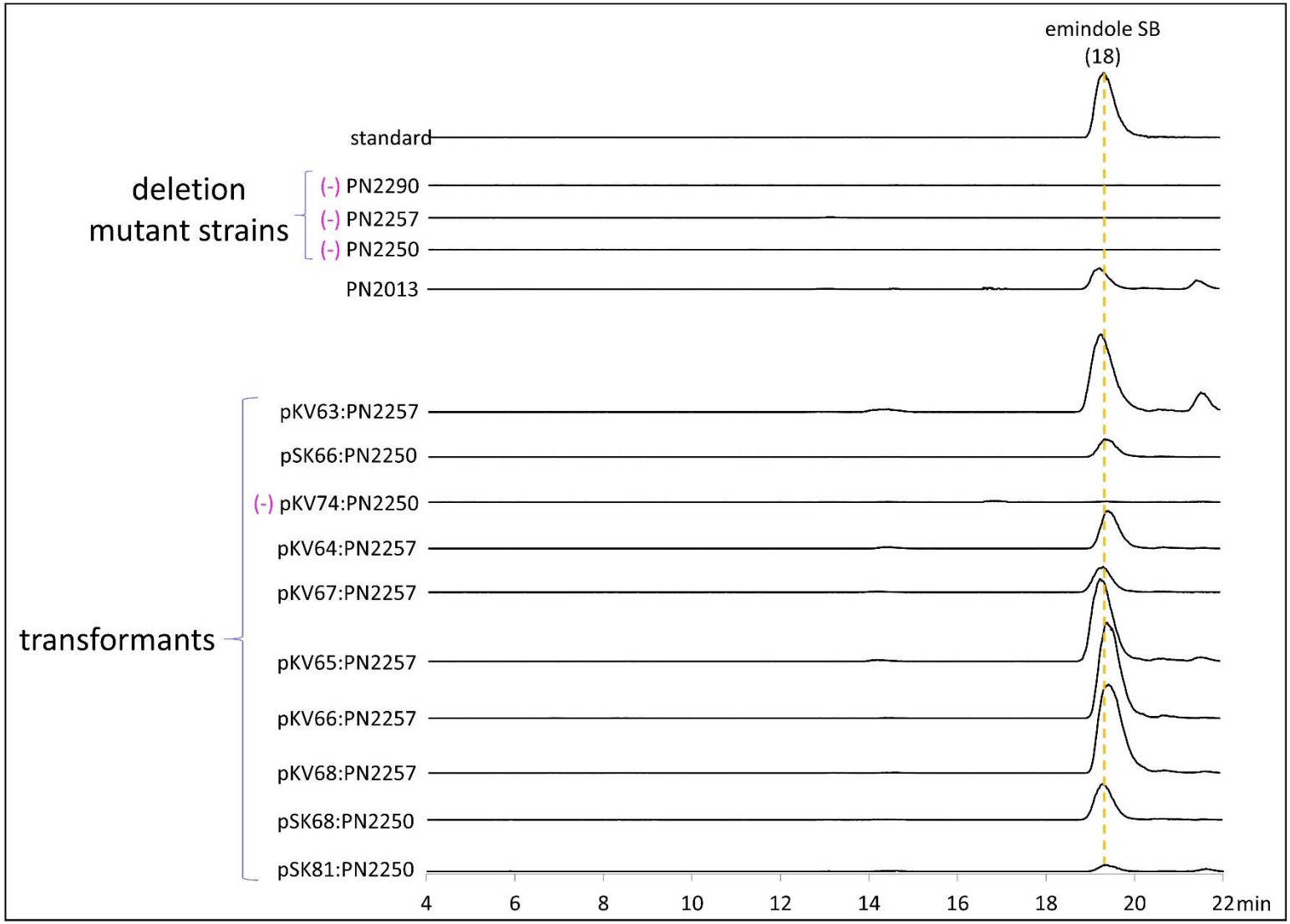


Figure 9.9. Extracted ion chromatograms (EICs) used to detect emindole SB (18) (19.3 min, 406.31 ± 0.01 m/z). A pink negative sign (-) indicates traces where emindole SB (18) was not detected. All EICs have been scaled to 300,000 AU.

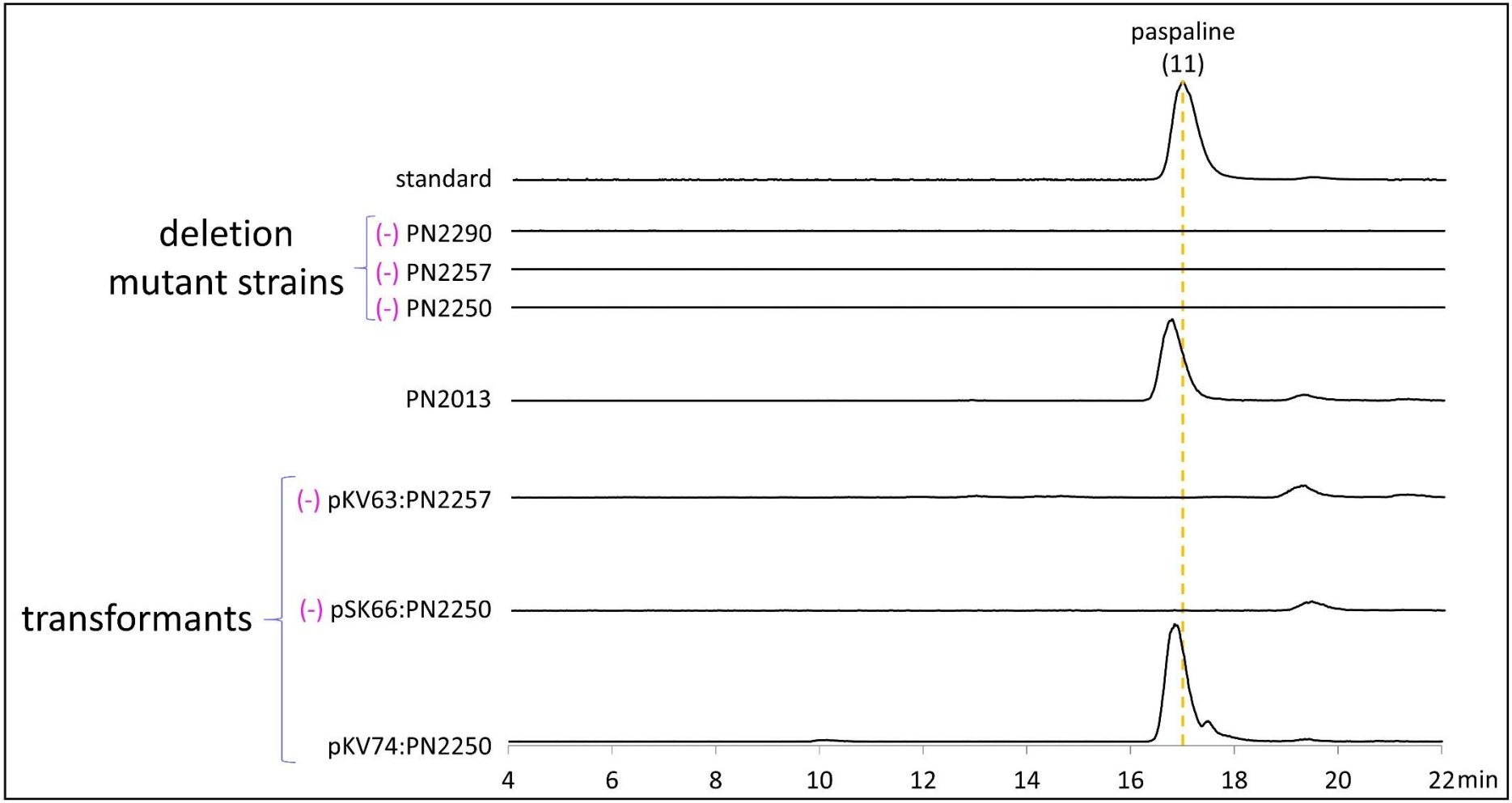


Figure 9.10. Extracted ion chromatograms (EICs) used to detect paspaline (11) (17.6 minutes, 422.305 ± 0.01 m/z). A pink negative sign (-) indicates traces where paspaline (11) was not detected. All EICs have been scaled to 300,000 AU.

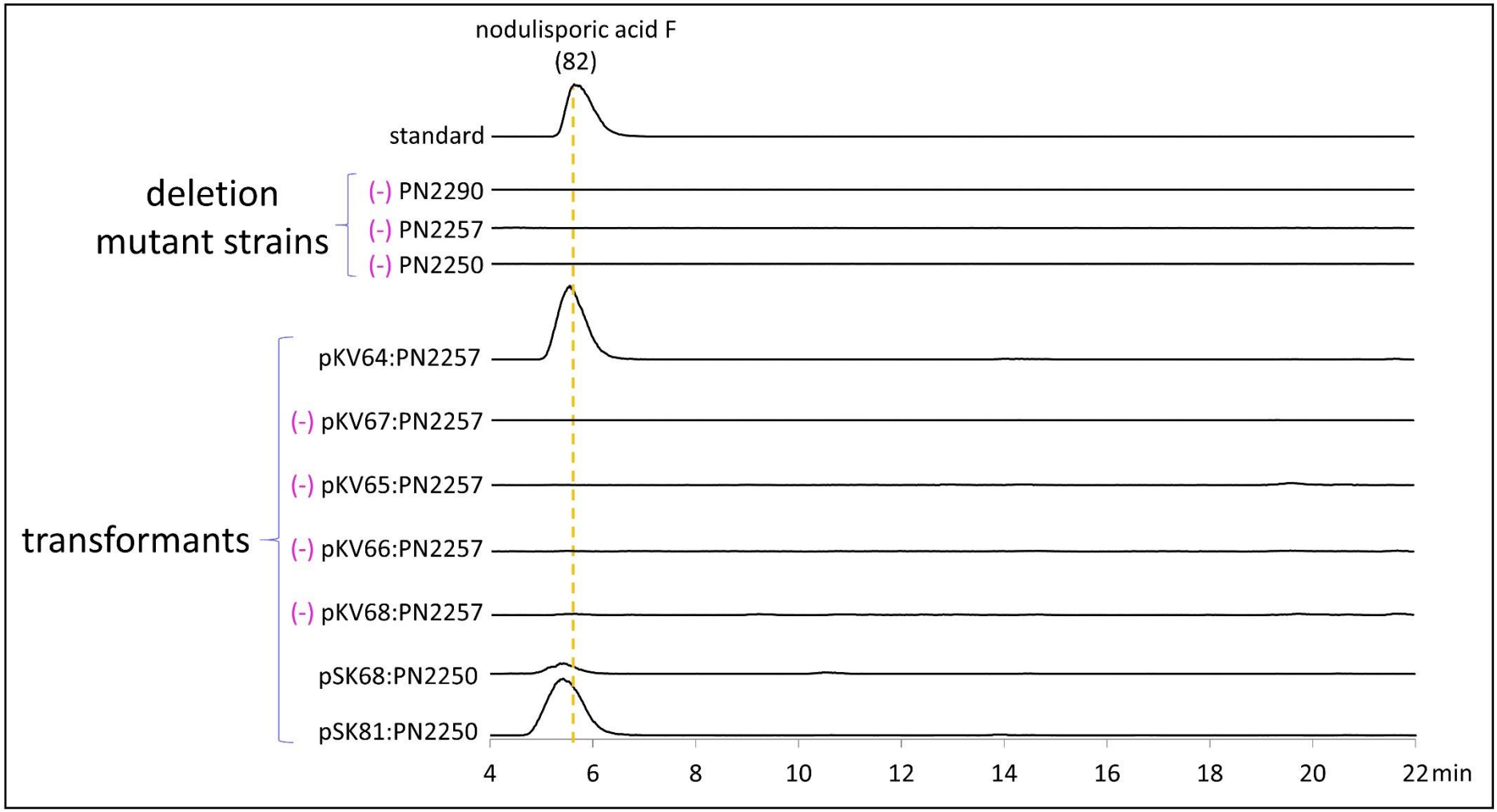
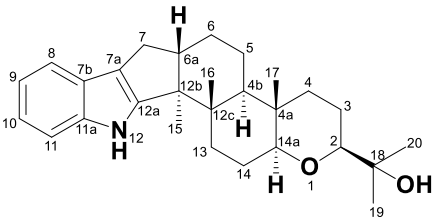


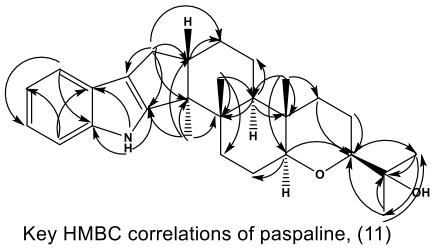
Figure 9.11. Extracted ion chromatograms (EICs) used to detect nodulisporic acid F (82) (6.2 minutes, 436.284 ± 0.01 m/z). A pink negative sign (-) indicates traces where nodulisporic acid F (82) was not detected. All EICs have been scaled to 300,000 AU.

Table 9.7. ^{1}H and ^{13}C NMR assignment of paspaline (11) in $CDCl_3$. Paspaline (11) spectra are depicted in Figure 9.12 to 9.16.

Position	Paspaline (11)		
	^{1}H	^{13}C	COSY
1	-	-	-
2	3.21, dd	84.64, CH	H3
3	1.35, m 1.67, m	21.93, CH_2	
4	1.12, m 1.83, m	37.62, CH_2	
4a		36.52, C	
4b	1.45, m	46.38, CH	H5
5	1.45, m 1.66, m	21.90, CH_2	
6	1.59, m 1.77, m	25.24, CH_2	H5
6a	2.76, m	48.72, CH	H6, H7
7	2.32, dd 2.66, dd	27.48, CH_2	
7a		118.22, C	
7b		125.09, C	
8	7.41, m	118.35, CH	H9
9	7.06, m	119.52, CH	
10	7.06, m	120.40, CH	
11	7.28, m	111.40, CH	H10
11a		139.94, C	
12	7.72, brs		
12a		150.82, C	
12b		52.97, C	
12c		39.98, C	
13	1.62, m 1.94, m	33.88, CH_2	
14	1.69, m	24.57, CH_2	
14a	3.01, dd	85.68, CH	
15	1.01, s	14.50, CH_3	
16	1.12, s	19.93, CH_3	
17	0.87, s	12.61, CH_3	
18		71.98, C	
19	1.18, s	26.02, CH_3	
20	1.16, s	23.64, CH_3	



paspaline (11)
$C_{28}H_{39}NO_2$
Exact Mass: 421.30
calc. [M + H]+ 422.3054 *m/z*
obs. [M + H]+ 422.3055 *m/z*



Key HMBC correlations of paspaline, (11)

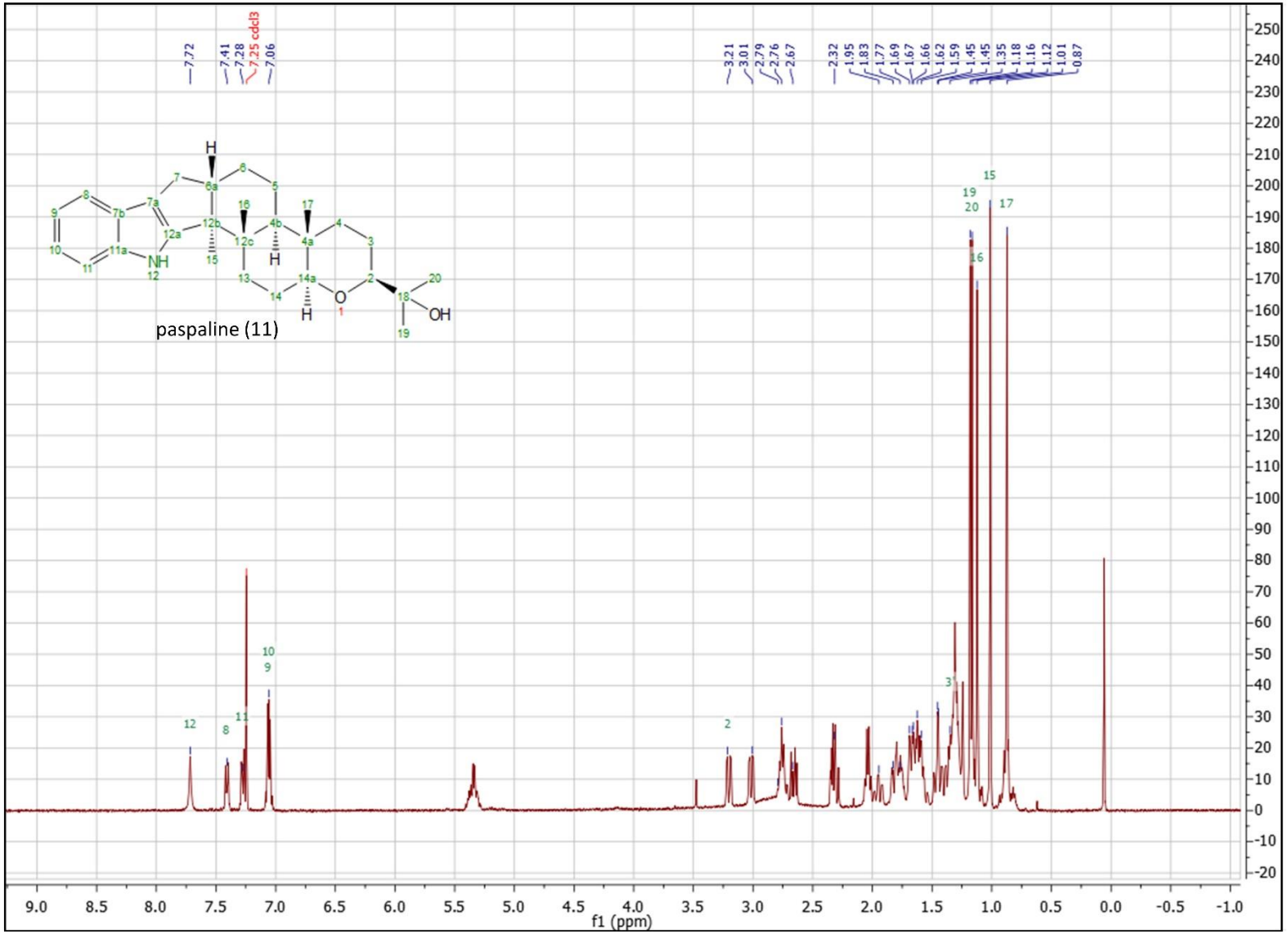


Figure 9.12. ^{1}H-NMR spectra for paspaline (11) standard in $CDCl_3$ at 400 MHz.

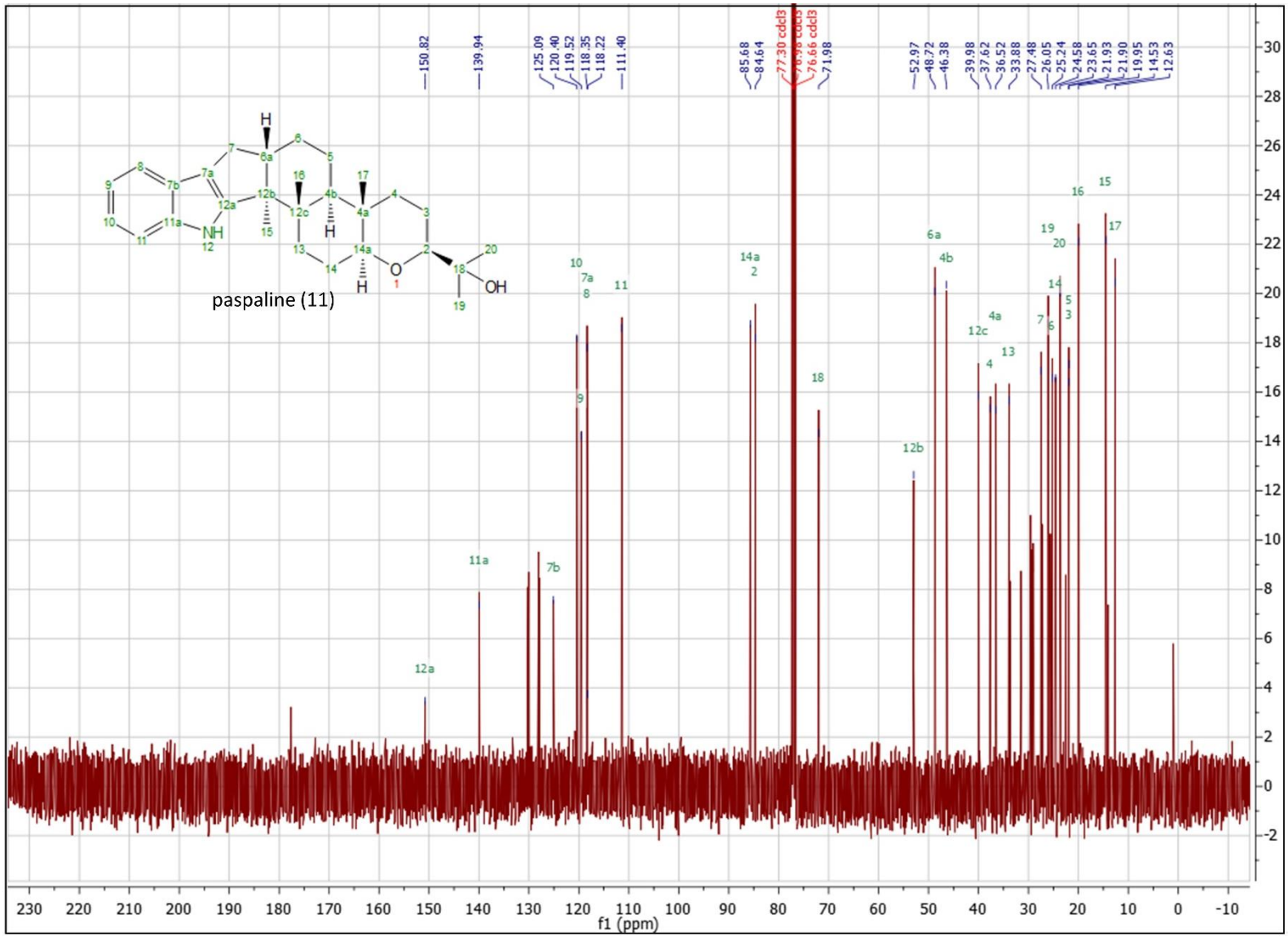


Figure 9.13. ^{13}C-NMR spectra for paspaline (11) standard in $CDCl_3$ at 100 MHz.

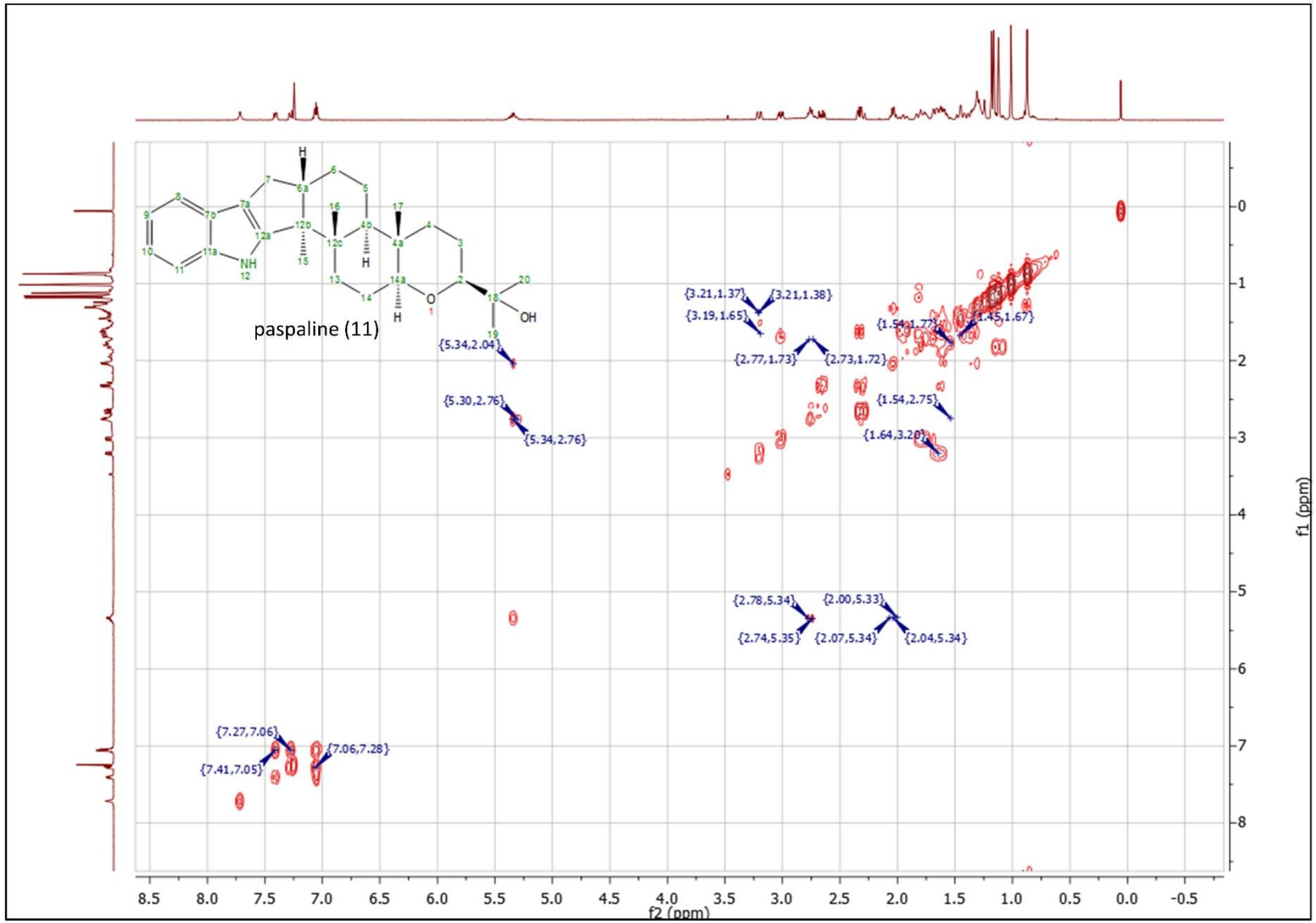


Figure 9.14. ^{1}H-^{1}H COSY-NMR spectra for paspaline (11) standard in $CDCl_3$ at 400 MHz.

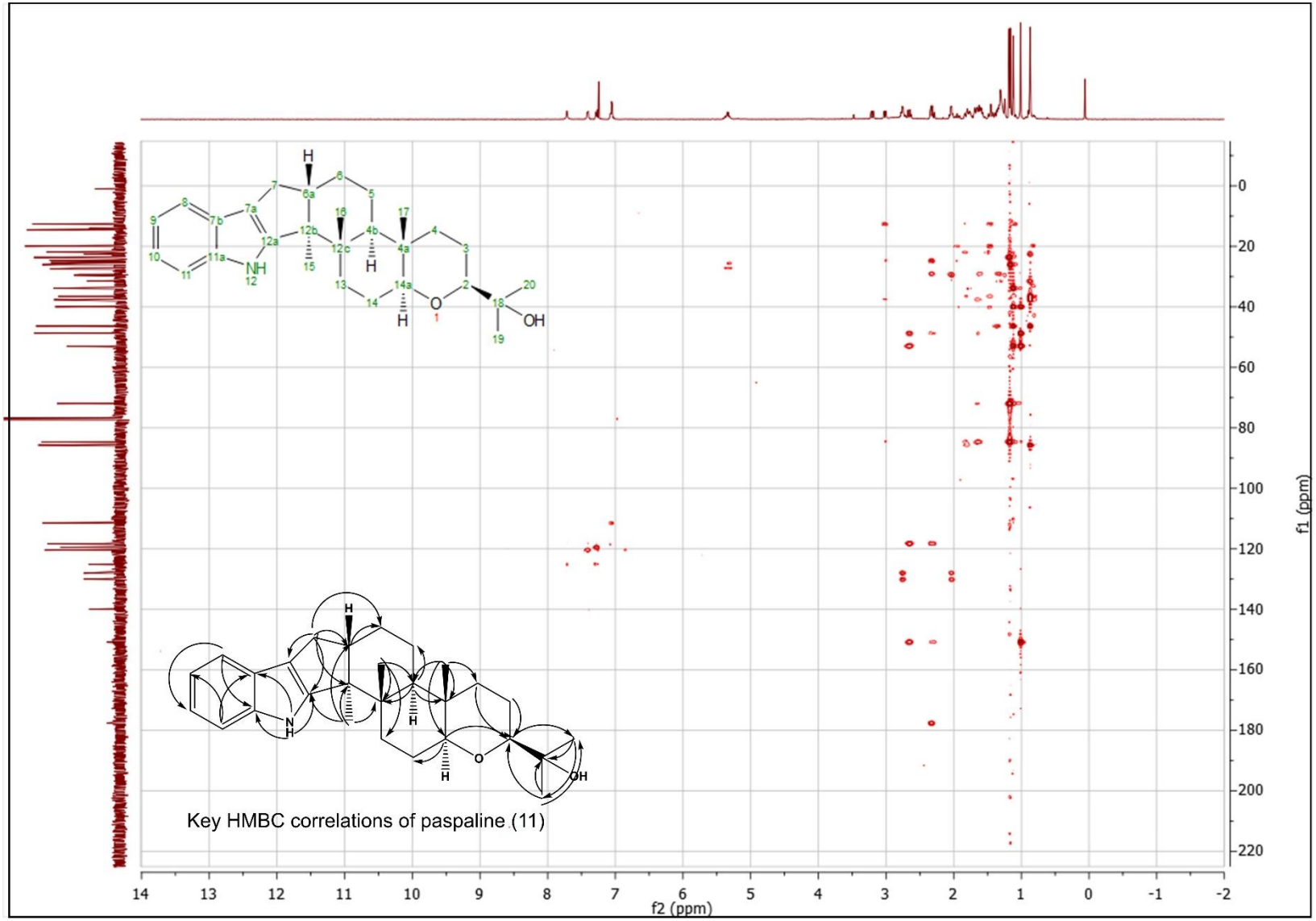


Figure 9.15. HMBC-NMR spectra for paspaline (11) standard in $CDCl_3$ at 400 MHz.

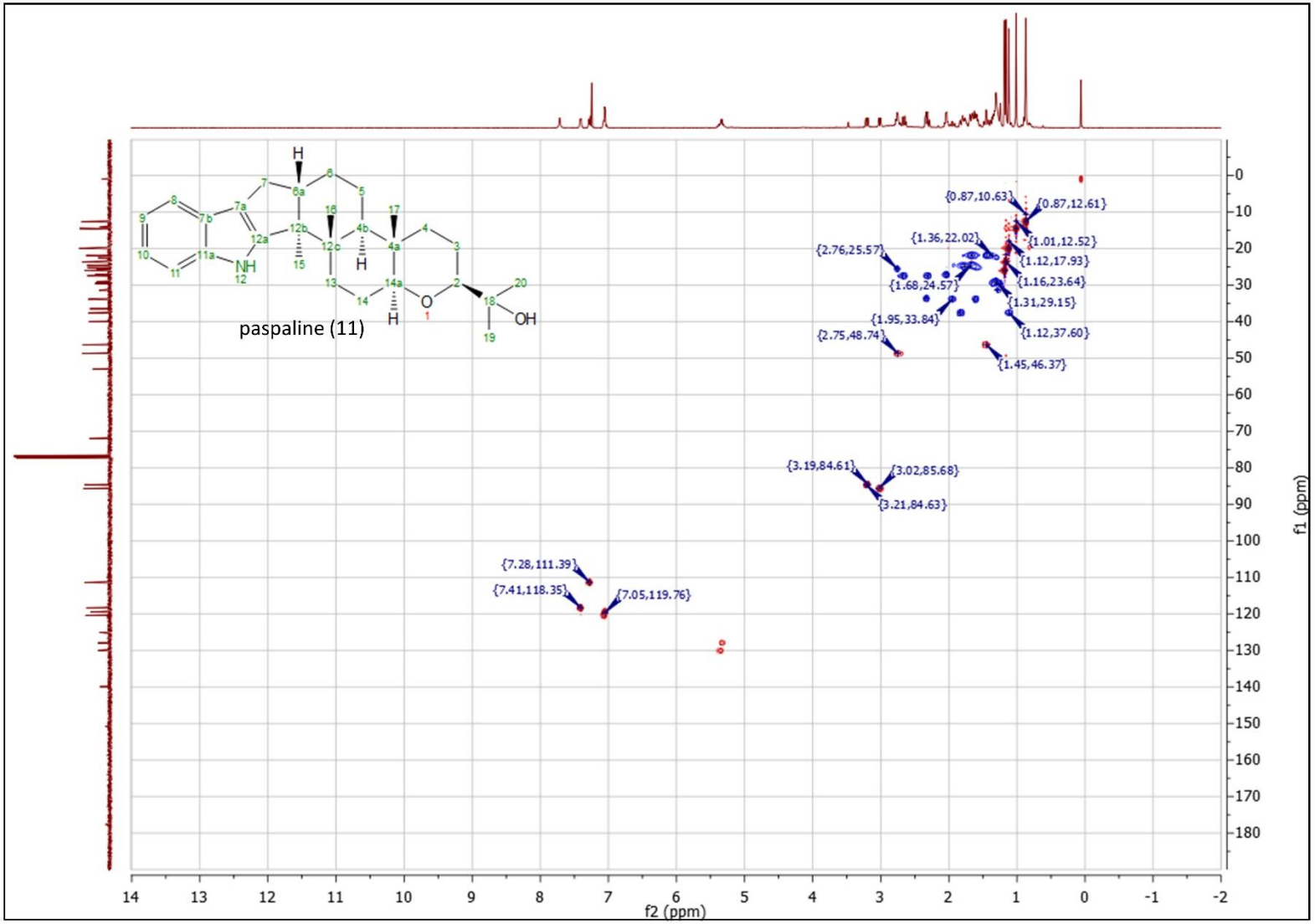
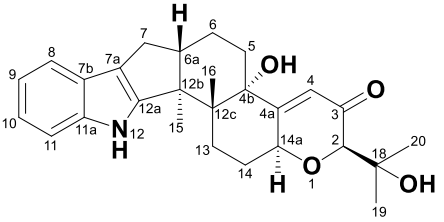


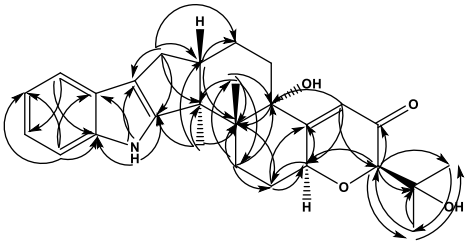
Figure 9.16. HSQC-NMR spectra for paspaline (11) standard in $CDCl_3$ at 400 MHz. Blue correlations represent secondary carbons and red correlations represent primary and tertiary carbons.

Table 9.8. ^{1}H and ^{13}C NMR assignment of paxilline (2) in $CDCl_3$. Paxilline (2) spectra are depicted in Figure 9.17 to 9.21.

Position	Paxilline (2)		
	^{1}H	^{13}C	COSY
1	-	-	-
2	3.69, s	83.16, CH	
3		199.41, C	
4	5.85, s	119.46, CH	
4a		168.64, C	
4b		77.20, C	
5	1.64 d, 2.03 m	34.13, CH_2	
6	1.76 m, 2.03 m,	20.87, CH_2	H5
6a	2.82, m	49.47, CH	H6, H7
7	2.44 dd, 2.75 dd	27.17, CH_2	
7a		117.45, C	
7b		125.03, C	
8	7.42, m	118.51, CH	H9
9	7.07, m	119.70, CH	
10	7.07, m	120.60, CH	
11	7.30, m	111.50, CH	H10
11a		139.73, C	
12	7.82, brs		
12a		151.64, C	
12b		50.80, C	
12c		43.12, C	
13	1.45, dd 2.80, m	27.92, CH_2	
14	2.32 m, 1.90 m	28.47, CH_2	
14a	4.84, t	72.67, CH	
15	1.32, s	16.24, CH_3	
16	1.01, s	19.53, CH_3	
17	-	-	-
18		72.66, C	
19	1.29, s	26.56, CH_3	
20	1.27, s	24.06, CH_3	



paxilline (2)
$C_{27}H_{33}NO_4$
Exact Mass: 435.24
calc. [M + H]+ 436.2482 *m/z*
obs. [M + H]+ 436.2485 *m/z*



Key HMBC correlations of paxilline, (2)

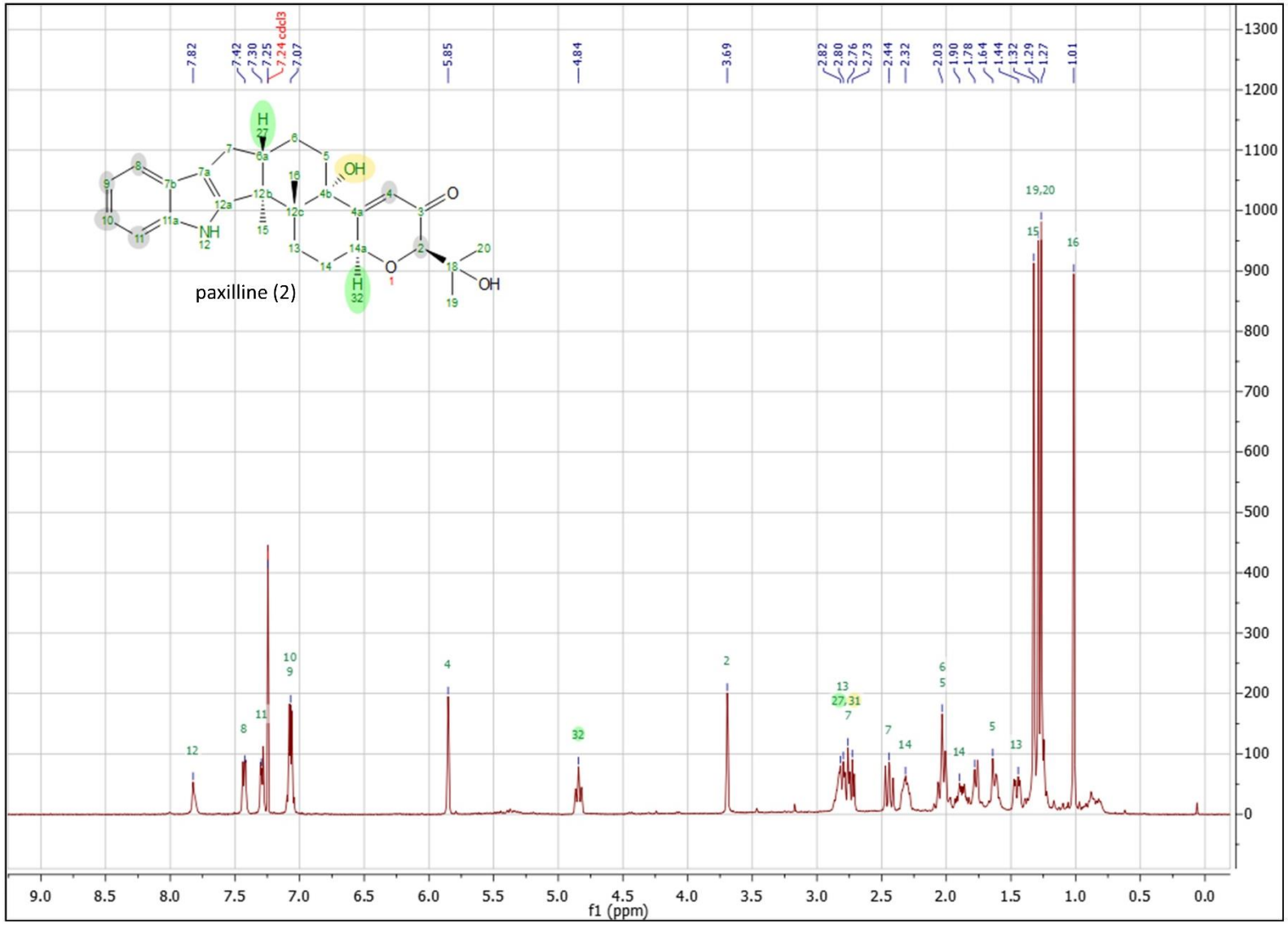


Figure 9.17. ^{1}H-NMR spectra for paxilline (2) standard in $CDCl_3$ at 400 MHz.

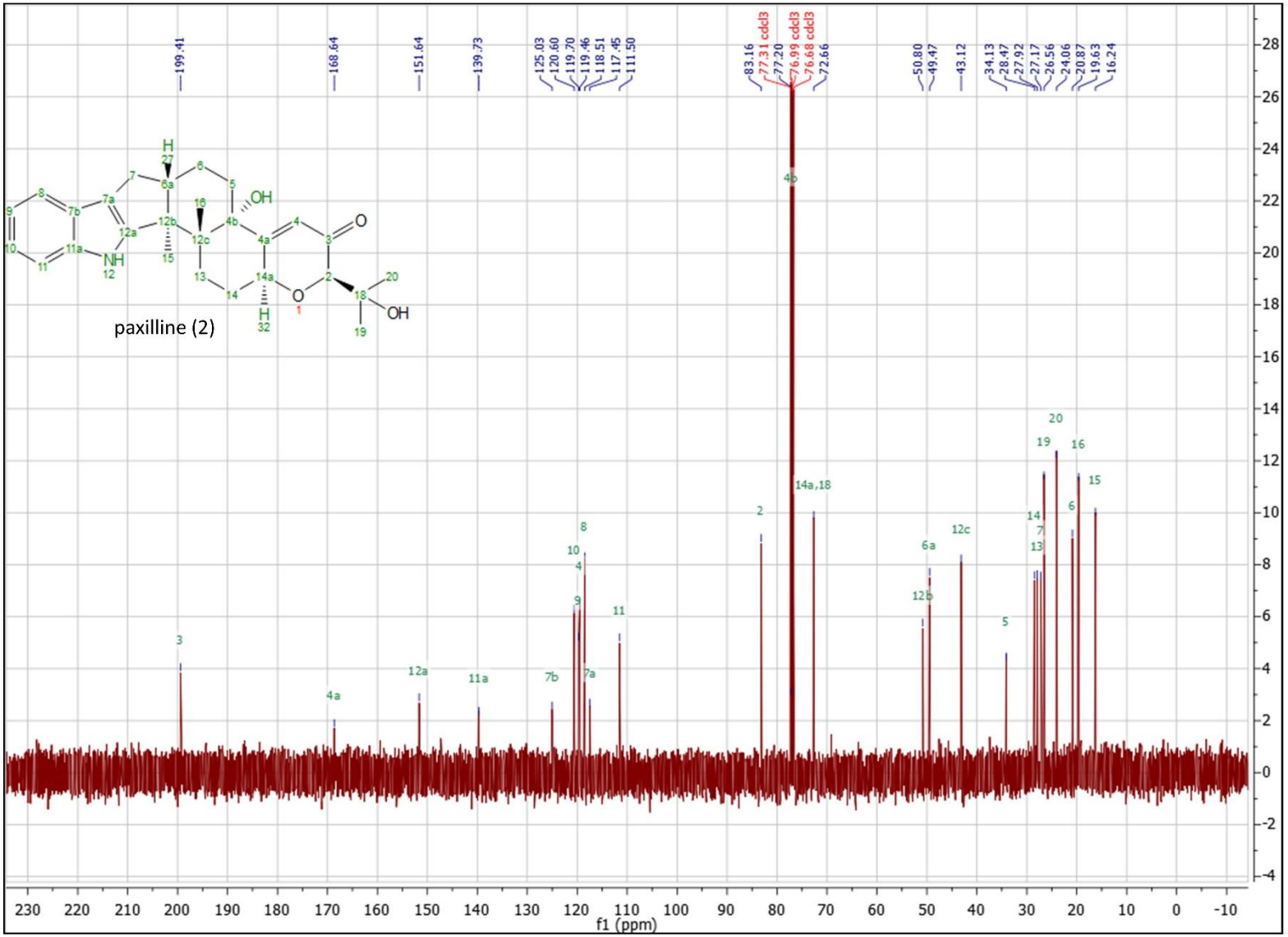


Figure 9.18. ^{13}C-NMR spectra for paxilline (2) standard in $CDCl_3$ at 100 MHz.

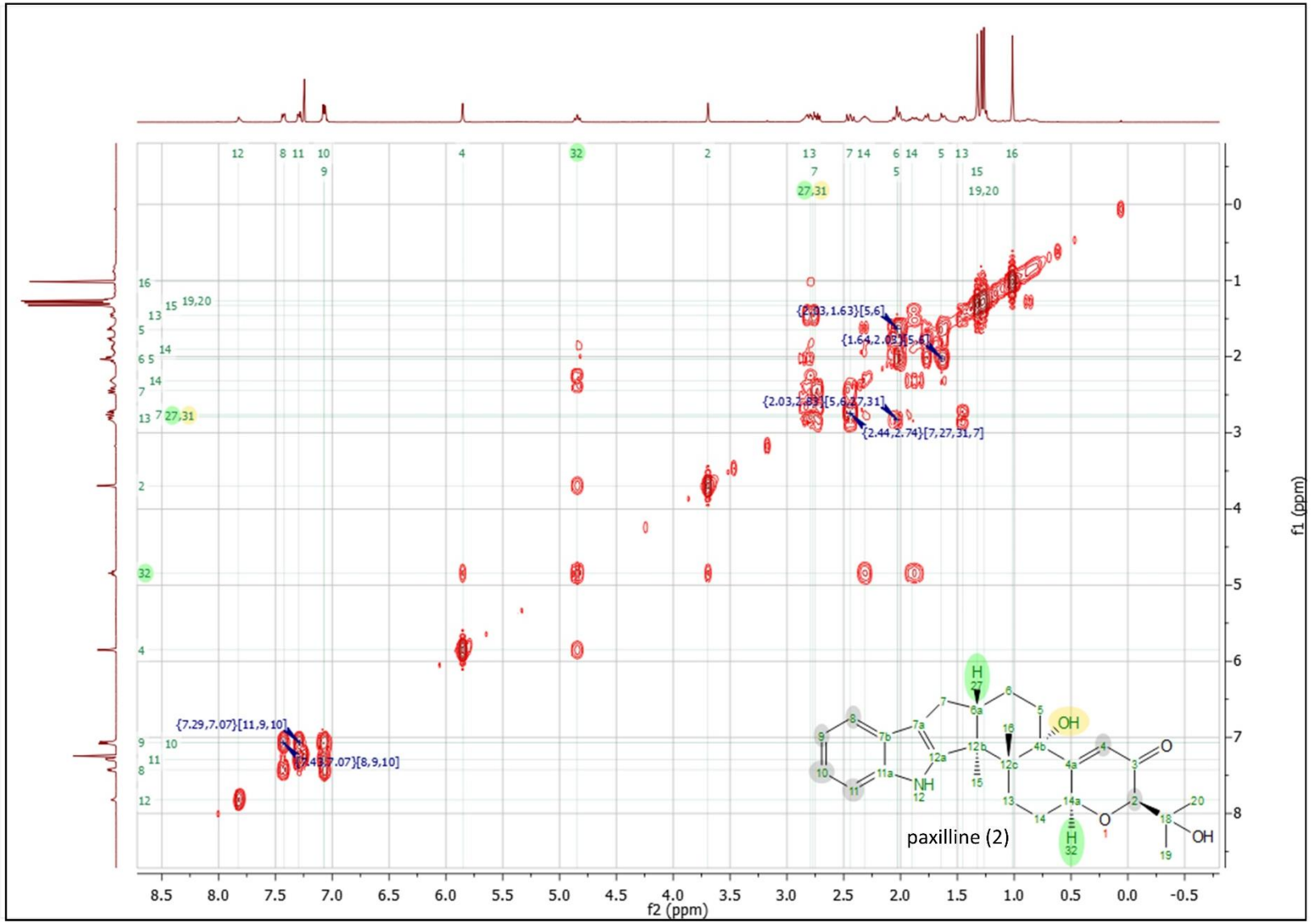


Figure 9 19. ^{1}H-^{1}H COSY-NMR spectra for paxilline (2) standard in $CDCl_3$ at 400 MHz.

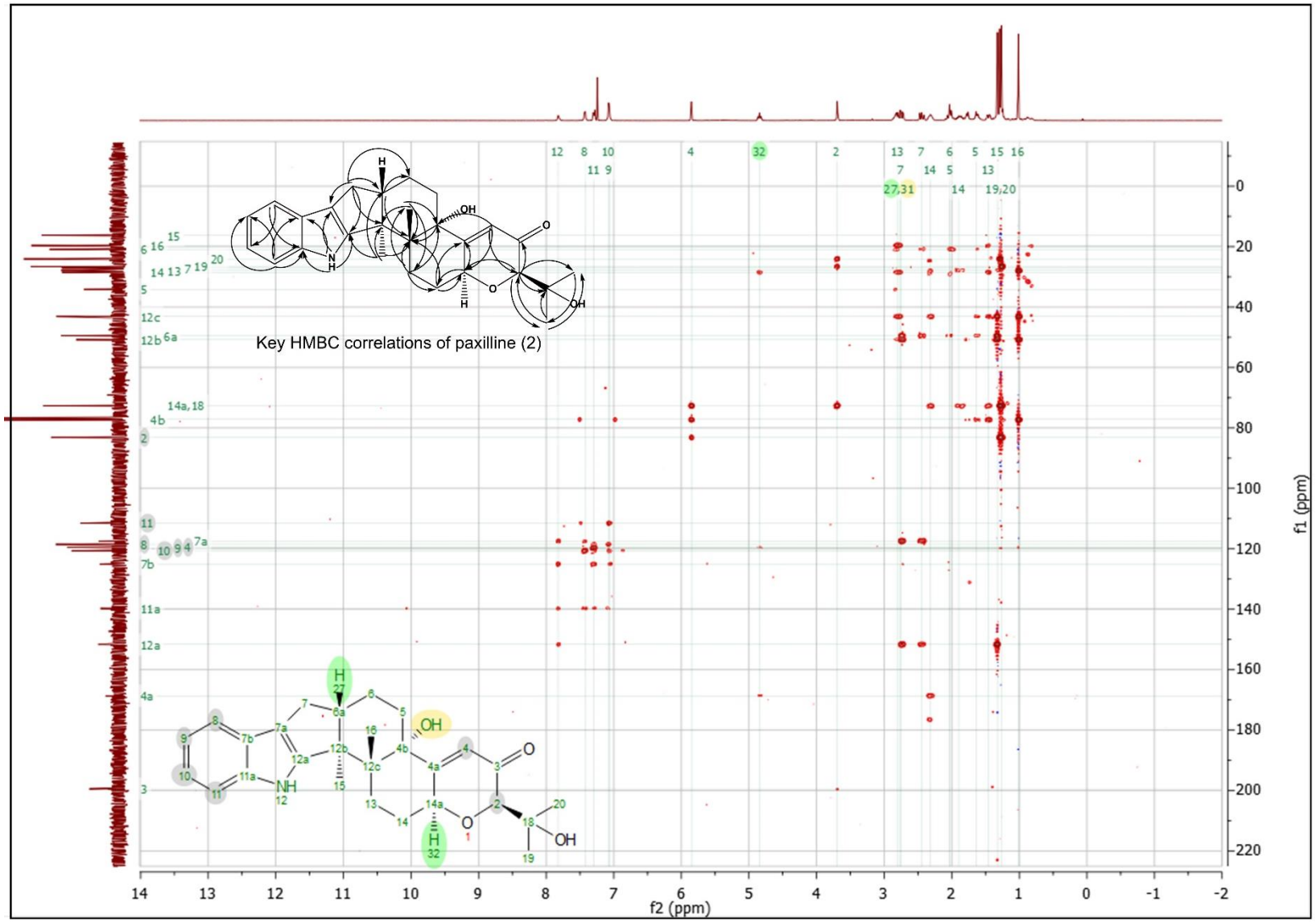


Figure 9.20. HMBC-NMR spectra for paxilline (2) standard in $CDCl_3$ at 400 MHz.

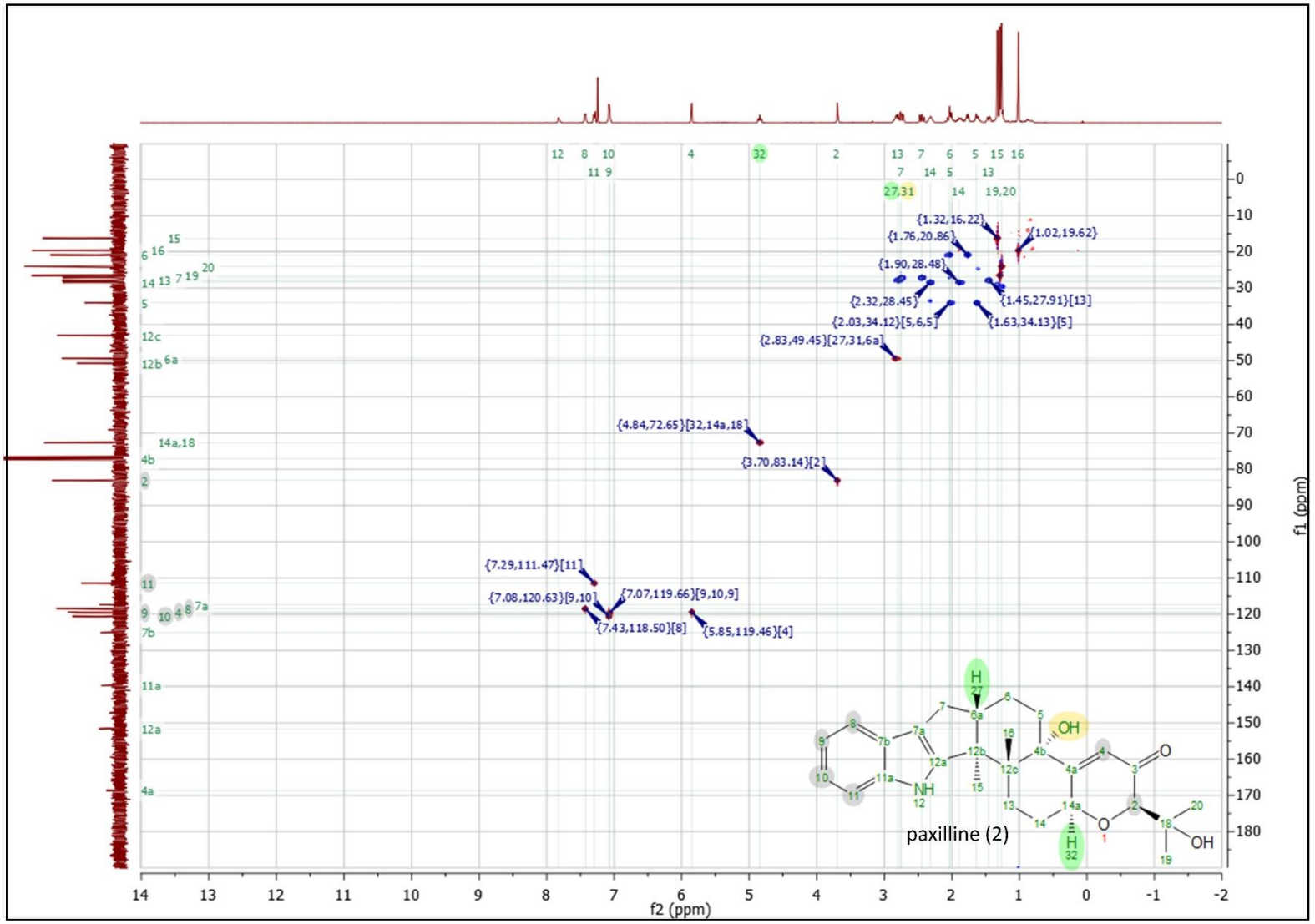
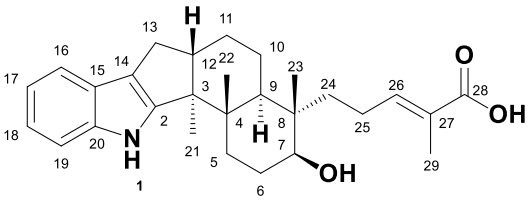


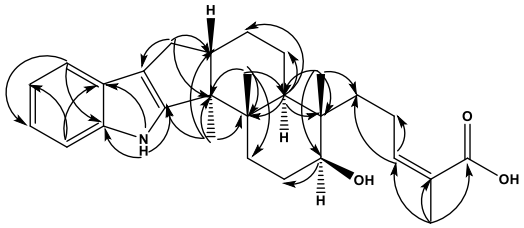
Figure 9.21. HSQC-NMR spectra for paxilline (2) standard in $CDCl_3$ at 400 MHz. Blue correlations represent secondary carbons and red correlations represent primary and tertiary carbons.

Table 9.9. ^{1}H and ^{13}C NMR assignment of nodulisporic acid F (NAF) (82) in CD_3OD. Spectra depicted in Figure 9.22 to 9.25.

Position	nodulisporic acid F (82)	
	^{1}H	^{13}C
1	-	-
2		150.78, C
3		52.90, C
4		39.00, C
5	1.84 m, 1.86 m	32.31, CH_2
6	1.77 m, 1.86 m	26.93, CH_2
7	3.53, dd	72.51, CH
8		41.13, C
9	1.76, t	39.94, CH
10	1.49 d, 1.65 m	22.62, CH_2
11	1.65 m, 1.77 m	24.90, CH_2
12	2.75, m	48.84, CH
13	2.29 dd, 2.61 dd	26.85, CH_2
14		116.55, C
15		124.82, C
16	7.27, brd	117.27, CH
17	6.94, dt	119.27, CH
18	6.90, dt	118.26, CH
19	7.27, brd	111.20, CH
20		140.64, C
21	1.01, s	13.46, CH_3
22	1.12, s	18.02, CH_3
23	0.84, s	15.84, CH_3
24	1.70 m, 1.36 m	35.61, CH_2
25	2.13 m, 2.20 m	22.02, CH_2
26	6.79, t	142.60, CH
27		127.45, C
28		170.73, C
29	1.85, s	11.09, CH_3



nodulisporic acid F (82)
$C_{28}H_{37}NO_3$
Exact Mass: 435.28
calc. [M + H]+ 436.2846 *m/z*
obs. [M + H]+ 436.2870 *m/z*



Key HMBC correlations of nodulisporic acid F (82)

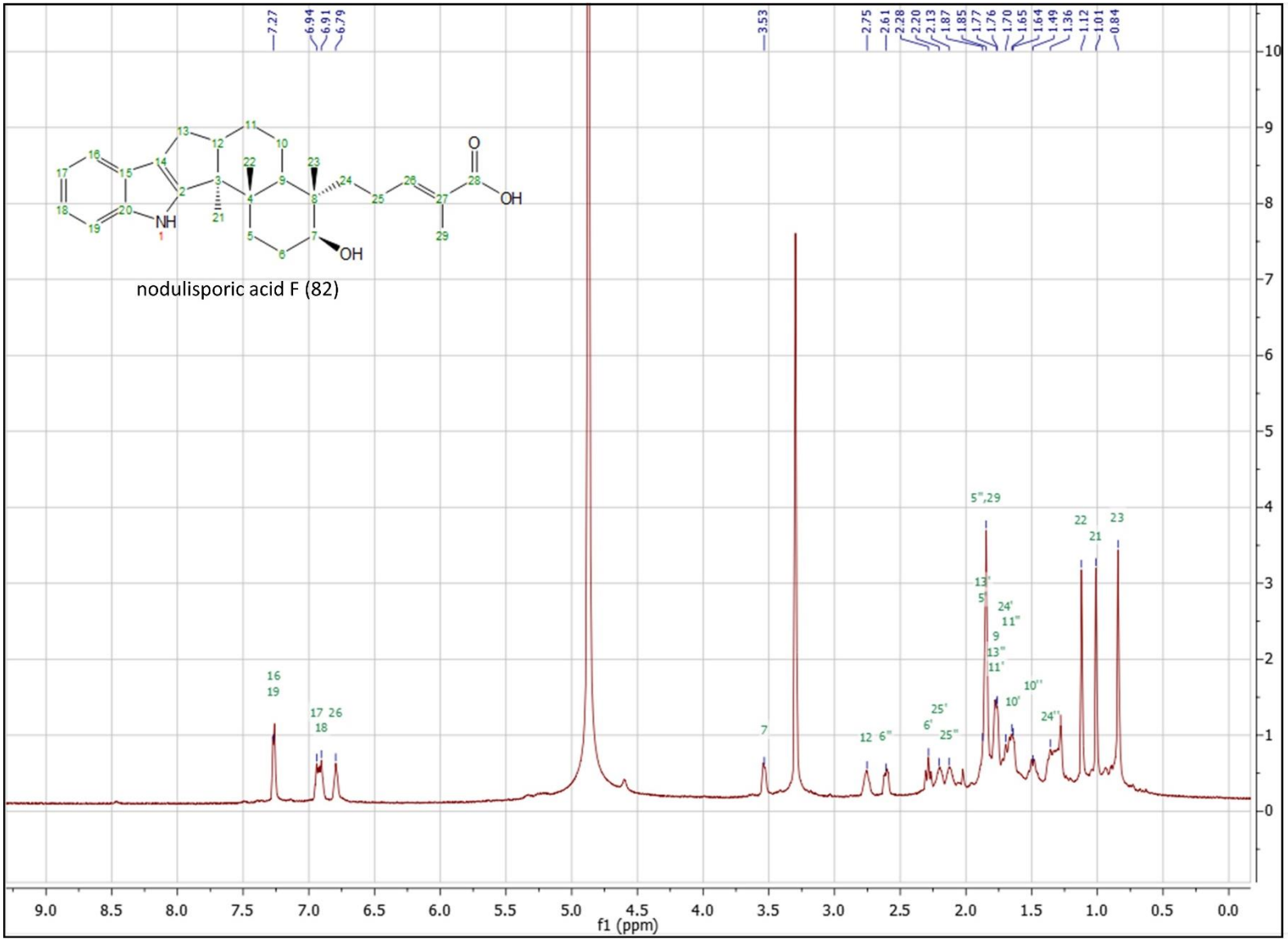


Figure 9.22. ^{1}H-NMR spectra for nodulisporic acid F (82) standard in CD_3OD at 400 MHz.

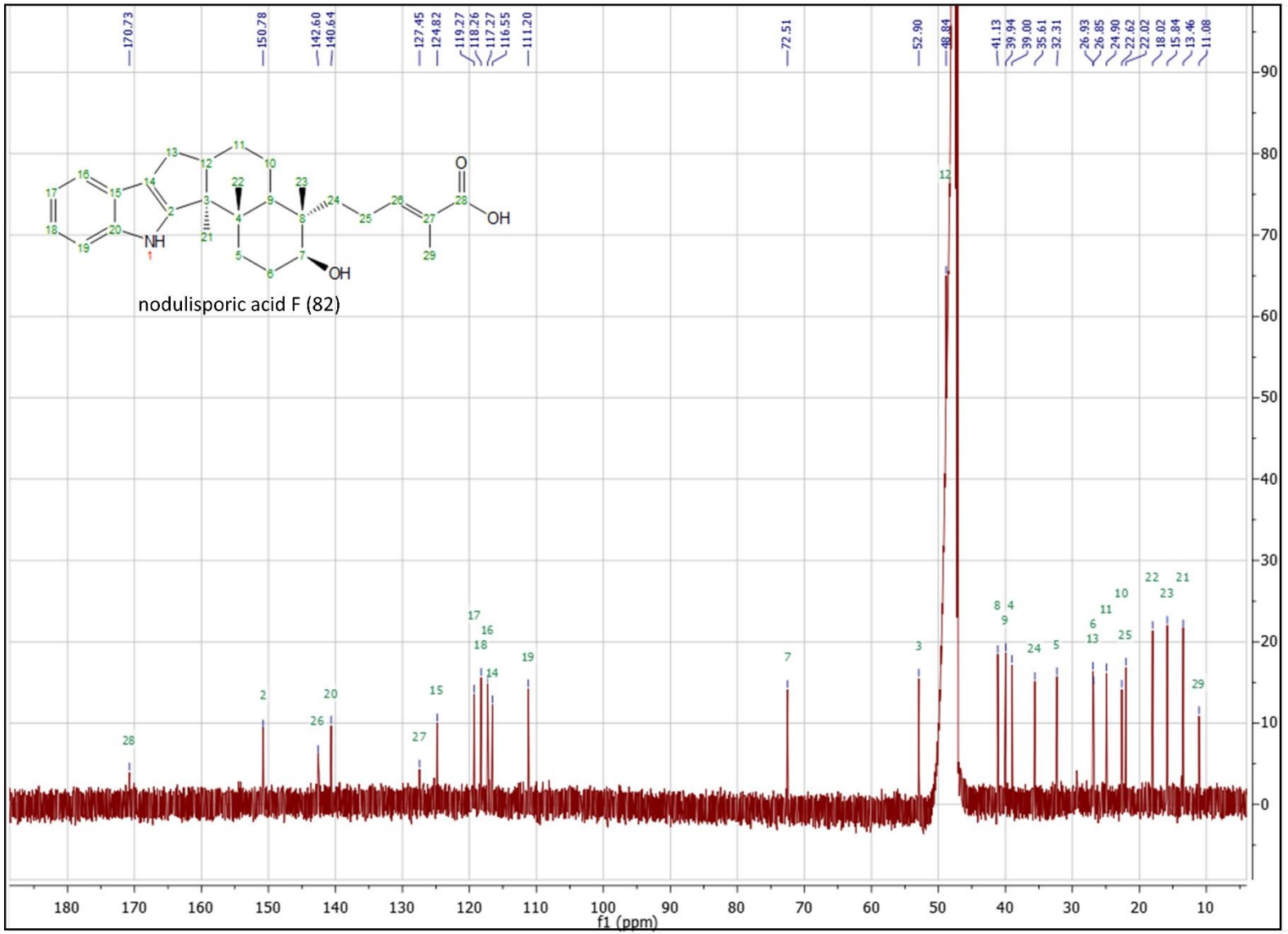


Figure 9.23. ^{13}C-NMR spectra for nodulisporic acid F (82) standard in CD_3OD at 100 MHz.

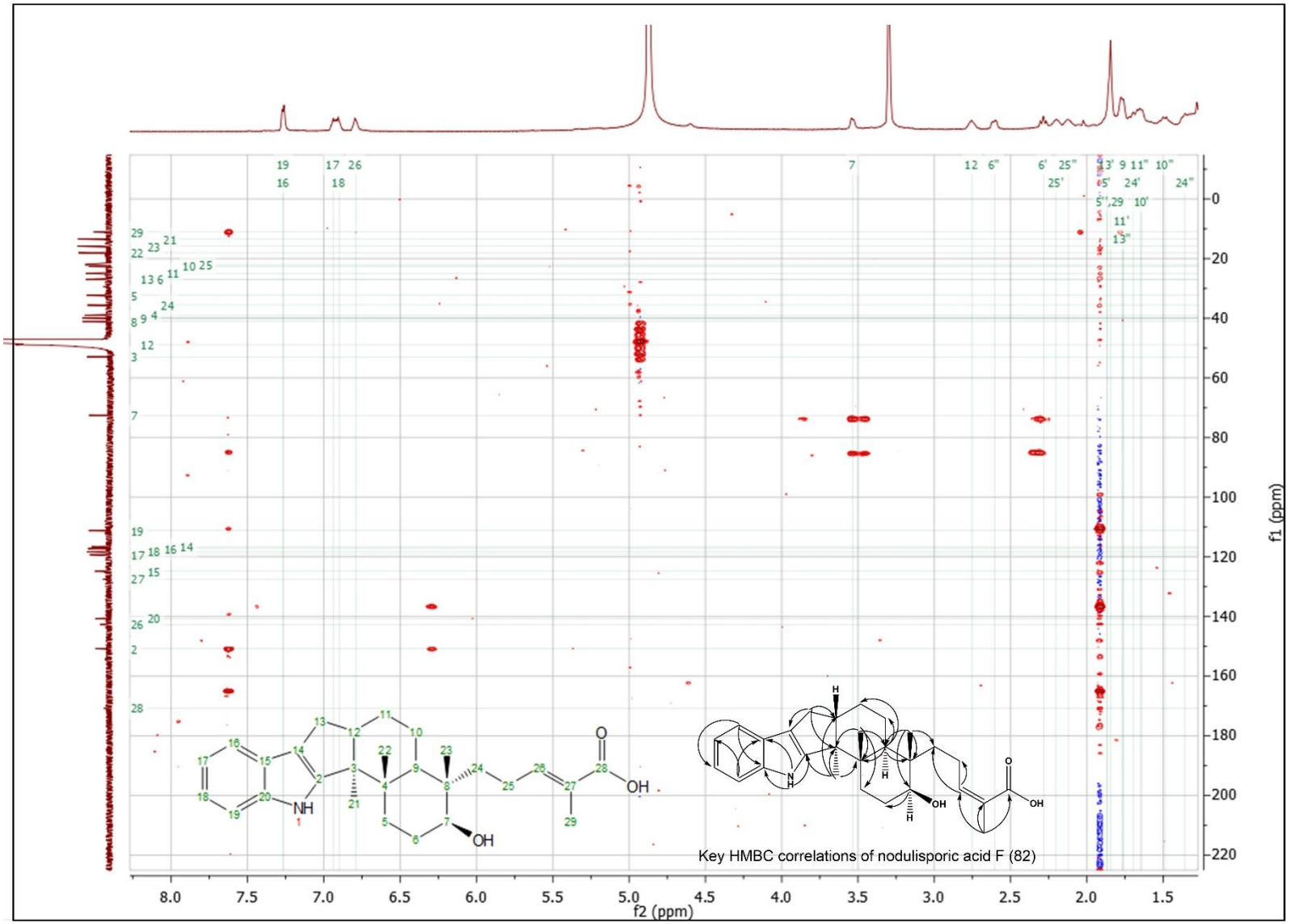


Figure 9.24. HMBC-NMR spectra for nodulisporic acid F (82) standard in CD_3OD at 600 MHz.

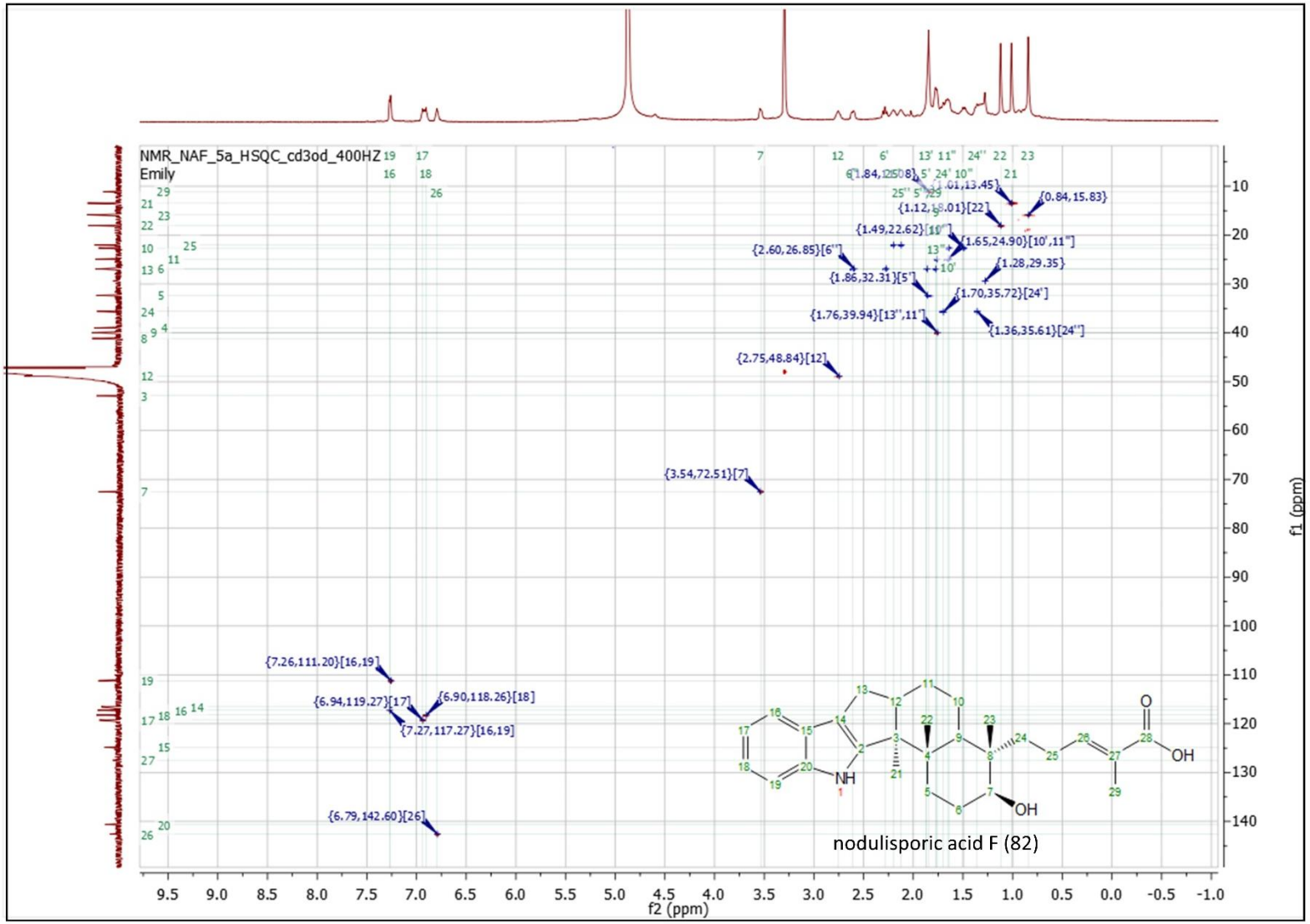
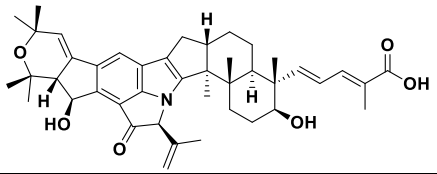
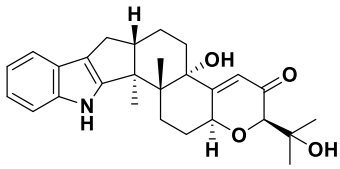
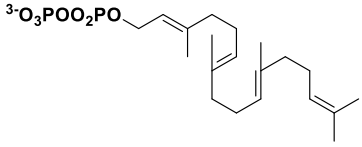
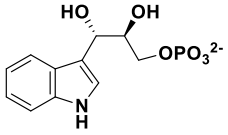
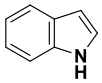
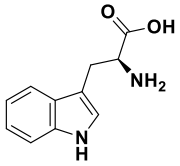
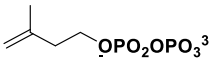
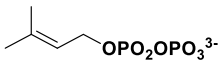
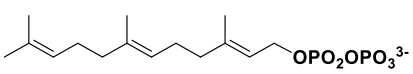
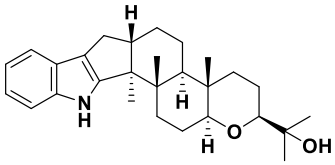
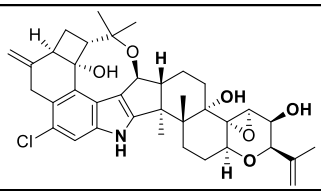
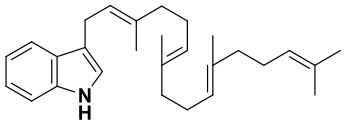
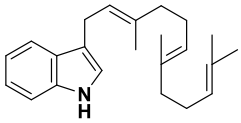
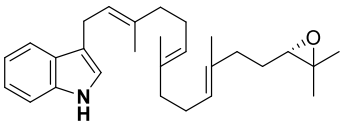
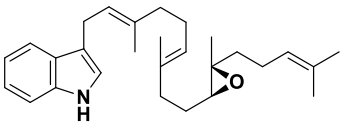
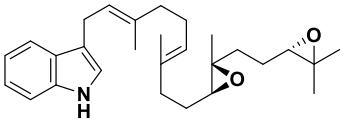
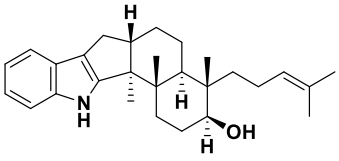
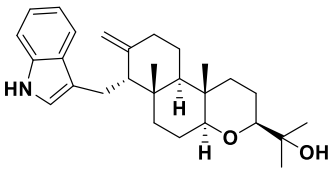
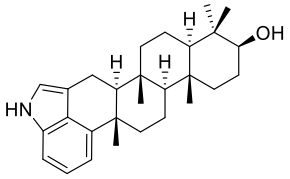
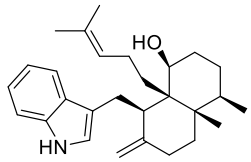
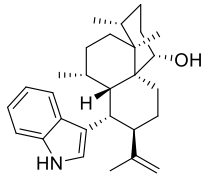
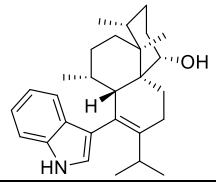
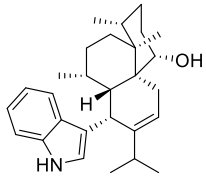
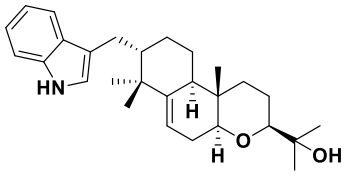
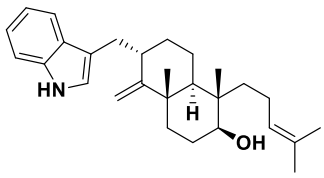
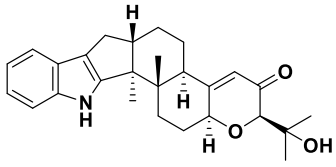
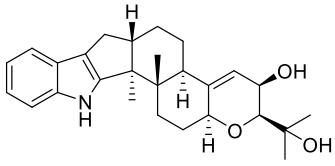
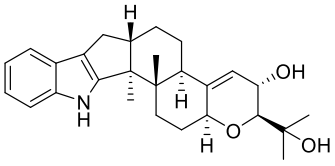
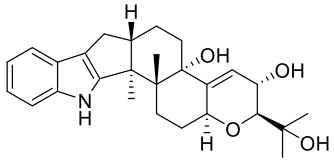


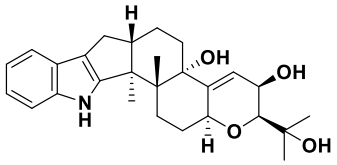
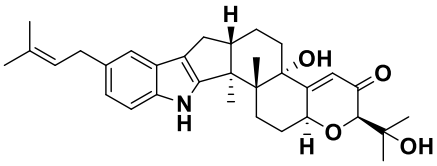
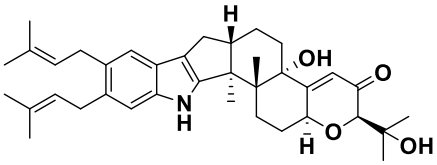
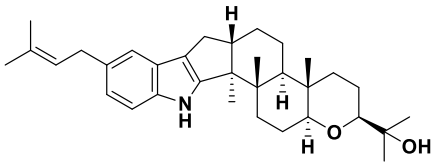
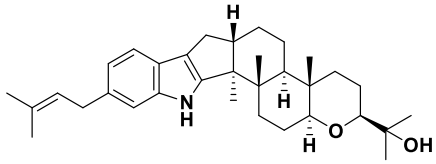
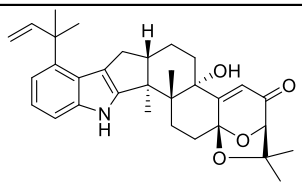
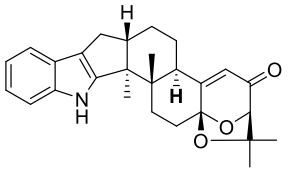
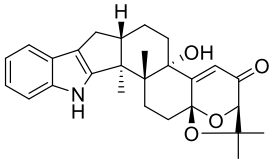
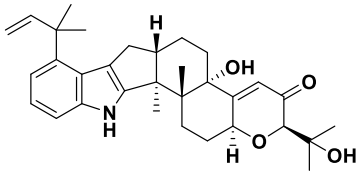
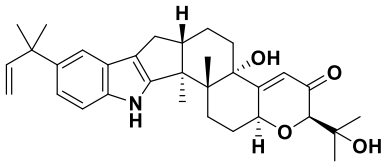
Figure 9.25. HSQC-NMR spectra for nodulisporic acid F (82) standard in CD_3OD at 400 MHz. Blue correlations represent secondary carbons and red correlations represent primary and tertiary carbons.

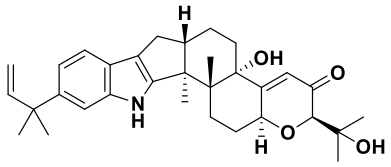
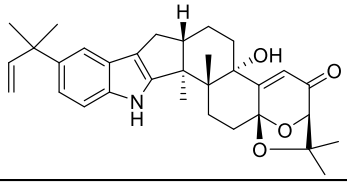
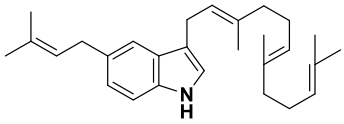
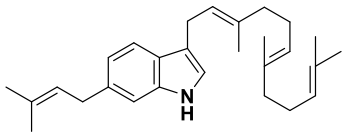
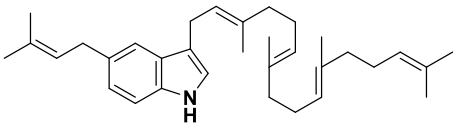
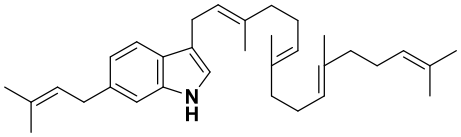
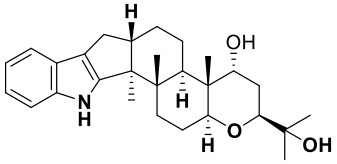
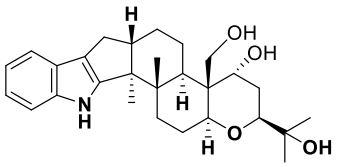
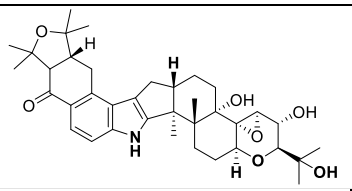
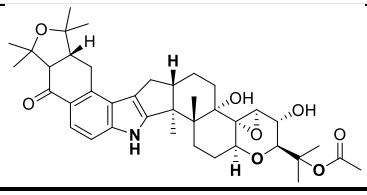
Table 9.10. Table of numbered compounds in this thesis.

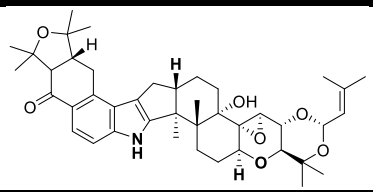
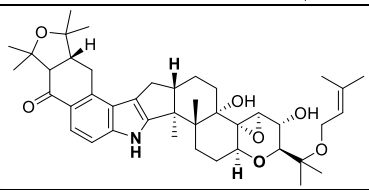
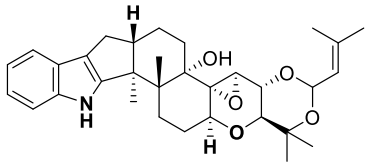
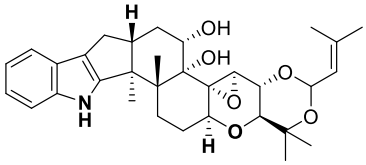
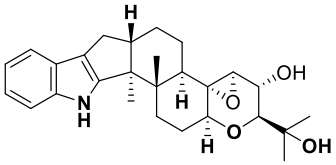
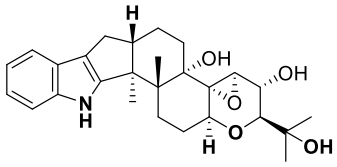
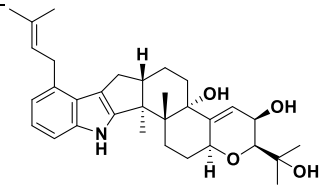
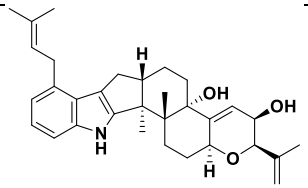
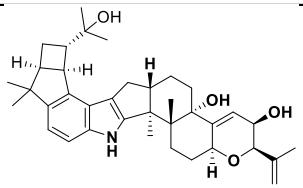
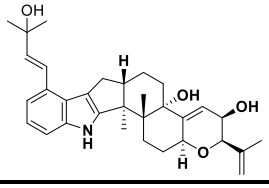
Number	Compound	Structure
1	nodulisporic acid A	
2	paxilline	
3	geranylgeranyl pyrophosphate	
4	indole-3-glycerol phosphate	
5	indole	
6	tryptophan	
7	isopentyl pyrophosphate	
8	dimethylallyl pyrophosphate	
9	acetate anion	
10	farnesyl pyrophosphate	

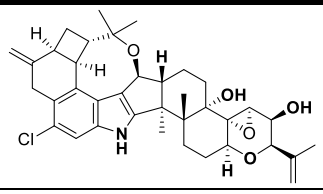
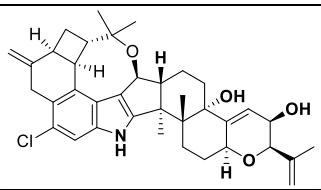
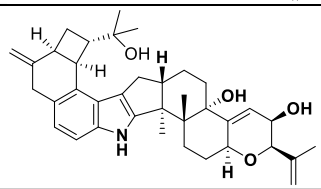
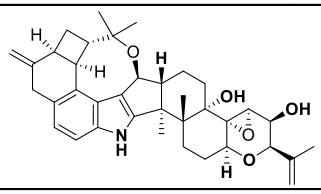
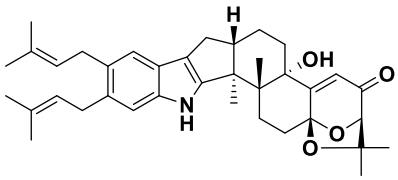
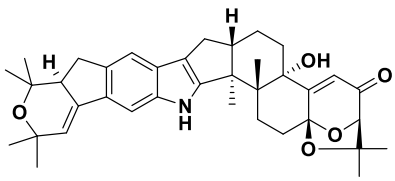
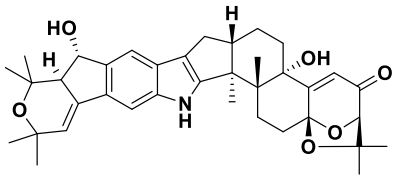
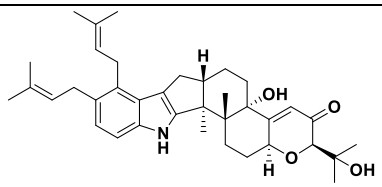
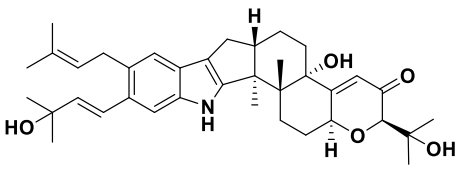
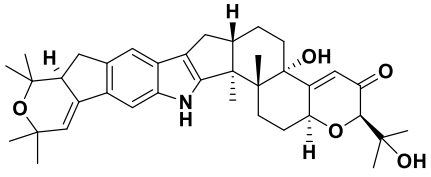
Number	Compound	Structure
11	paspaline	
12	penitrem A	
13	3-geranylgeranylindole	
14	farnesylindole	
15	17,18-epoxy-3-geranylgeranylindole	
16	13,14-epoxy-3-geranylgeranylindole	
17	bisepoxy-3-geranylgeranylindole	
18	emindole SB	
19	emindole DB	
20	petromindole	

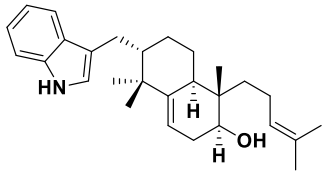
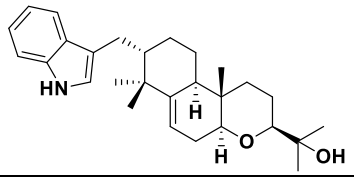
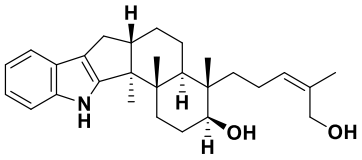
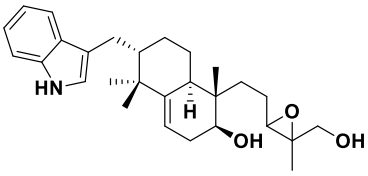
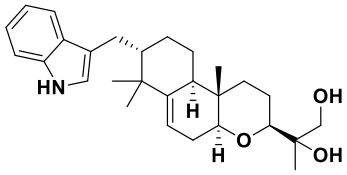
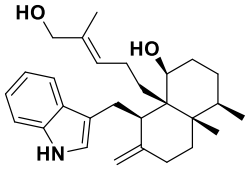
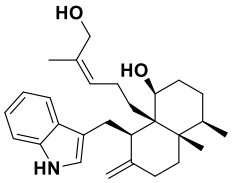
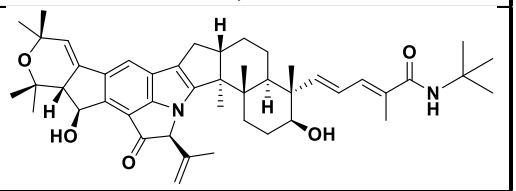
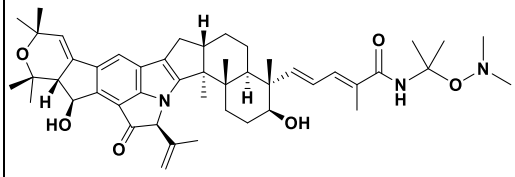
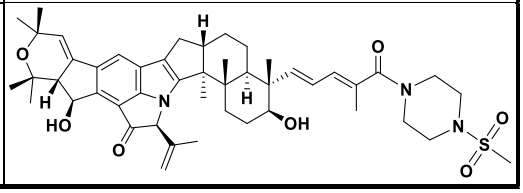
Number	Compound	Structure
21	anominine	
22	10,23-dihydro-24,25-dehydroaflavinine	
23	aflavinine	
24	aflavinine congener	
25	emindole PA precursor	
26	emeniveol	
27	9-desoxypaxilline	
28	β-PC-M6	
29	α-PC-M6	
30	α-paxitriol	

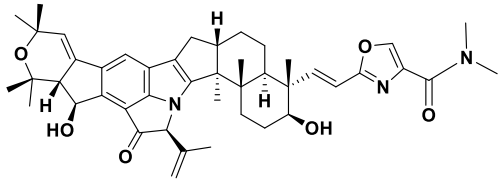
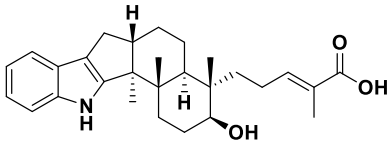
Number	Compound	Structure
31	β-paxitriol	
32	21-prenylpaxilline	
33	21,22-bisprenylpaxilline	
34	21-prenylpaspaline	
35	22-prenylpaspaline	
36	aflatrem	
37	paspalicine	
38	paspalinine	
39	reversely 20-prenylated paxilline	
40	reversely 21-prenylated paxilline	

Number	Compound	Structure
41	reversely 22-prenylated paxilline	
42	β-aflatrem	
43	5-prenyl-farnesylindole	
44	6-prenyl-farnesylindole	
45	5-prenyl-3-geranylgeranylindole	
46	6-prenyl-3-geranylgeranylindole	
47	terpendole E	
48	terpendole F	
49	lolitriol	
50	lolitrem J	

Number	Compound	Structure
51	lolitrem B	
52	lolitrem E	
53	terpendole C	
54	terpendole M	
55	9-desoxyterpendole I	
56	terpendole I	
57	20-prenyl-β-paxitriol	
58	20-prenylpenijanthine A	
59	PC-M4	
60	PC-M5	

Number	Compound	Structure
61	penitrem F	
62	penitrem C	
63	secopenitrem D	
64	penitrem B	
65	shearinine K	
66	shearinine A	
67	shearinine D	
68	20,21-bisprenylpaxilline	
69	22-protoshearinine	
70	shearinine B	

Number	Compound	Structure
71	AtS5B1 compound 1	
72	AtS5B1 compound 2	
73	AtS5-P450 compound 3	
74	AtS5-P450 compound 4	
75	AtS5-P450 compound 5	
76	AtS5-P450 compound 1	
77	AtS5-P450 compound 2	
78	*N* -tert-butyl nodulisporamide	
79	$NHC(Me)_2C(O)NMe_2$-nodulisporic acid A	
80	*N* -1-[(4-SO_2Me)Piperazinyl]-nodulisporic acid A	

Number	Compound	Structure
81	dimethylamino oxazole-nodulisporic acid A	
82	nodulisporic acid F	

10. REFERENCES

1. Van de Bittner, K. C.; Nicholson, M. J.; Bustamante, L. Y.; Kessans, S. A.; Ram, A.; van Dolleweerd, C. J.; Scott, B.; Parker, E. J., Heterologous biosynthesis of nodulisporic acid F. *J. Am. Chem. Soc.* **2018,** *140* (2), 582-585.
2. Nicholson, M. J.; Van de Bittner, K. C.; Ram, A.; Bustamante, L. Y.; Parker, E. J., Draft Genome Sequence of the Filamentous Fungus *Hypoxylon pulicicidum* ATCC 74245. *Genome A.* **2018,** *6* (2).
3. van Dolleweerd, C. J.; Kessans, S. A.; Van de Bittner, K. C.; Bustamante, L. Y.; Bundela, R.; Scott, B.; Nicholson, M. J.; Parker, E. J., MIDAS: A modular DNA assembly system for synthetic biology. *ACS Synth. Biol.* **2018,** *7* (4), 1018-1029.
4. Raistrick, H., Bakerian lecture: a region of biosynthesis. *P. Roy. Soc, Lond. B. Bio.* **1950,** *136* (885), 481-508.
5. Keller, N. P.; Turner, G.; Bennett, J. W., Fungal secondary metabolism—from biochemistry to genomics. *Nat. Rev. Microbiol.* **2005,** *3* (12), 937.
6. Peldáez, F., Biological activities of fungal metabolites. In *Handbook of industrial mycology*, CRC Press: 2004; pp 71-121.
7. Gloer, J. B.; TePaske, M. R.; Sima, J. S.; Wicklow, D. T.; Dowd, P. F., Antiinsectan aflavinine derivatives from the sclerotia of *Aspergillus flavus*. *J. Org. Chem.* **1988,** *53* (23), 5457-5460.
8. Wicklow, D. T.; Dowd, P. F.; Tepaske, M. R.; Gloer, J. B., Sclerotial metabolites of *Aspergillus flavus* toxic to a detritivorous maize insect (*Carpophilus hemipterus*, Nitidulidae). *T. Brit. Mycol. Soc.* **1988,** *91* (3), 433-438.
9. Laakso, J. A.; Dowd, P. F.; Gloer, J. B.; Wicklow, D. T. Radarin antiinsectan metabolites. WO1993002049 A1, 1993/02/04/, 1993.
10. Li, C.; Gloer, J. B.; Wicklow, D. T.; Dowd, P. F., Thiersinines A and B: novel antiinsectan indole diterpenoids from a new fungicolous *Penicillium* species (NRRL 28147). *Org. Lett.* **2002,** *4* (18), 3095-3098.
11. Meinke, P. T.; Smith, M. M.; Shoop, W. L., Nodulisporic acid: its chemistry and biology. *Curr. Top. Med. Chem.* **2002,** *2* (7), 655-674.
12. Nakazawa, J.; Yajima, J.; Usui, T.; Ueki, M.; Takatsuki, A.; Imoto, M.; Toyoshima, Y. Y.; Osada, H., A Novel Action of Terpendole E on the Motor Activity of Mitotic Kinesin Eg5. *Chem. Biol.* **2003,** *10* (2), 131-137.
13. Sallam, A. A.; Ayoub, N. M.; Foudah, A. I.; Gissendanner, C. R.; Meyer, S. A.; El Sayed, K. A., Indole diterpene alkaloids as novel inhibitors of the Wnt/β-catenin pathway in breast cancer cells. *Eur. J. Med. Chem.* **2013,** *70*, 594-606.

14. Ogata, M.; Ueda, J.-y.; Hoshi, M.; Hashimoto, J.; Nakashima, T.; Anzai, K.; Takagi, M.; Shin-ya, K., A Novel Indole-diterpenoid, JBIR-03 with Anti-MRSA Activity from *Dichotomomyces cejpii* var. *cejpii* NBRC 103559. *J. Antibiot.* **2007,** *60* (10), 645-648.
15. Fan, Y.; Wang, Y.; Liu, P.; Fu, P.; Zhu, T.; Wang, W.; Zhu, W., Indole-diterpenoids with anti-H1N1 activity from the aciduric fungus *Penicillium camemberti* OUCMDZ-1492. *J. Nat. Prod.* **2013,** *76* (7), 1328-1336.
16. Knaus, H.-G.; McManus, O. B.; Lee, S. H.; Schmalhofer, W. A.; Garcia-Calvo, M.; Helms, L. M. H.; Sanchez, M.; Giangiacomo, K.; Reuben, J. P., Tremorgenic Indole Alkaloids Potently Inhibit Smooth Muscle High-Conductance Calcium-Activated Potassium Channels. *Biochemistry* **1994,** *33* (19), 5819-5828.
17. Bills, G. F.; González-Menéndez, V.; Martín, J.; Platas, G.; Fournier, J.; Peršoh, D.; Stadler, M., *Hypoxylon pulicicidum* sp. nov. (Ascomycota, Xylariales), a Pantropical Insecticide-Producing Endophyte. *PLOS ONE* **2012,** *7* (10), e46687.
18. Acklin, W.; Weibel, F.; Arigoni, D., On the biosynthesis of paspalin and related metabolites from *Claviceps paspali*. *Chimia* **1977,** *31* (2), 63.
19. De Jesus, A. E.; Steyn, P. S.; Van Heerden, F. R.; Vleggaar, R.; Wessels, P. L.; Hull, W. E., Structure and biosynthesis of the penitrems A-F, six novel tremorgenic mycotoxins from *Penicillium crustosum*. *J. Am. Chem. Soc.* **1981,** (6), 289-291.
20. de Jesus, A. E.; Hull, W. E.; Steyn, P. S.; van Heerden, F. R.; Vleggaar, R.; Wessels, P. L., High-field ^{13}C n.m.r. evidence for the formation of [1,2-^{13}C]acetate from [2-^{13}C]acetate during the biosynthesis of penitrem A by *Penicillium crustosum*. *J. Chem. Soc. Chem. Comm.* **1982,** *0*, 837-838.
21. de Jesus, A. E.; Gorst-Allman, C. P.; Steyn, P. S.; van Heerden, F. R.; Vleggaar, R.; Wessels, P. L.; Hull, W. E., Tremorgenic mycotoxins from *Penicillium crustosum*. Biosynthesis of penitrem A. *J. Am. Chem. Soc.* **1983**, 1863-1868.
22. Mantle, P. G.; Weedon, C. M., Biosynthesis and transformation of tremorgenic indolediterpenoids by *Penicillium paxilli* and *Acremonium lolii*. *Phytochemistry* **1994,** *36* (5), 1209-1217.
23. Young, C.; Itoh, Y.; Johnson, R.; Garthwaite, I.; Miles, C. O.; Munday-Finch, S. C.; Scott, B., Paxilline-negative mutants of *Penicillium paxilli* generated by heterologous and homologous plasmid integration. *Curr. Genet.* **1998,** *33* (5), 368-377.
24. Young, C.; McMillan, L.; Telfer, E.; Scott, B., Molecular cloning and genetic analysis of an indole-diterpene gene cluster from *Penicillium paxilli*. *Mol. Microbiol.* **2001,** *39* (3), 754-764.
25. Zhang, S.; Monahan, B. J.; Tkacz, J. S.; Scott, B., Indole-Diterpene Gene Cluster from *Aspergillus flavus*. *Appl. Environ. Microb.* **2004,** *70* (11), 6875-6883.
26. Young, C. A.; Bryant, M. K.; Christensen, M. J.; Tapper, B. A.; Bryan, G. T.; Scott, B., Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass. *Mol. Genet. Genomics* **2005,** *274* (1), 13-29.
27. Young, C. A.; Felitti, S.; Shields, K.; Spangenberg, G.; Johnson, R. D.; Bryan, G. T.; Saikia, S.; Scott, B., A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte *Neotyphodium lolii*. *Fungal Genet. Biol.* **2006,** *43* (10), 679-693.
28. Nicholson, M. J.; Koulman, A.; Monahan, B. J.; Pritchard, B. L.; Payne, G. A.; Scott, B., Identification of Two Aflatrem Biosynthesis Gene Loci in *Aspergillus flavus* and Metabolic Engineering of *Penicillium paxilli* To Elucidate Their Function. *Appl. Environ. Microb.* **2009,** *75* (23), 7469-7481.
29. Saikia, S.; Takemoto, D.; Tapper, B. A.; Lane, G. A.; Fraser, K.; Scott, B., Functional analysis of an indole-diterpene gene cluster for lolitrem B biosynthesis in the grass endosymbiont *Epichloë festucae*. *FEBS Lett.* **2012,** *586* (16), 2563-2569.
30. Motoyama, T.; Hayashi, T.; Hirota, H.; Ueki, M.; Osada, H., Terpendole E, a Kinesin Eg5 Inhibitor, Is a Key Biosynthetic Intermediate of Indole-Diterpenes in the Producing Fungus *Chaunopycnis alba*. *Chem. Biol.* **2012,** *19* (12), 1611-1619.
31. Liu, C.; Tagami, K.; Minami, A.; Matsumoto, T.; Frisvad, J. C.; Suzuki, H.; Ishikawa, J.; Gomi, K.; Oikawa, H., Reconstitution of Biosynthetic Machinery for the Synthesis of the Highly Elaborated Indole Diterpene Penitrem. *Angew. Chem. Int. Edit.* **2015,** *54* (19), 5748-5752.

32. Nicholson, M. J.; Eaton, C. J.; Stärkel, C.; Tapper, B. A.; Cox, M. P.; Scott, B., Molecular Cloning and Functional Analysis of Gene Clusters for the Biosynthesis of Indole-Diterpenes in *Penicillium crustosum* and *P. janthinellum*. *Toxins* **2015,** *7* (8), 2701-2722.
33. Tang, M.-C.; Lin, H.-C.; Li, D.; Zou, Y.; Li, J.; Xu, W.; Cacho, R. A.; Hillenmeyer, M. E.; Garg, N. K.; Tang, Y., Discovery of Unclustered Fungal Indole Diterpene Biosynthetic Pathways through Combinatorial Pathway Reassembly in Engineered Yeast. *J. Am. Chem. Soc.* **2015,** *137* (43), 13724-13727.
34. McMillan, L. K.; Carr, R. L.; Young, C. A.; Astin, J. W.; Lowe, R. G. T.; Parker, E. J.; Jameson, G. B.; Finch, S. C.; Miles, C. O.; McManus, O. B.; Schmalhofer, W. A.; Garcia, M. L.; Kaczorowski, G. J.; Goetz, M.; Tkacz, J. S.; Scott, B., Molecular analysis of two cytochrome P450 monooxygenase genes required for paxilline biosynthesis in *Penicillium paxilli*, and effects of paxilline intermediates on mammalian maxi-K ion channels. *Mol. Genet. Genomics* **2003,** *270* (1), 9-23.
35. Saikia, S.; Parker, E. J.; Koulman, A.; Scott, B., Four gene products are required for the fungal synthesis of the indole-diterpene, paspaline. *FEBS Lett.* **2006,** *580* (6), 1625-1630.
36. Saikia, S.; Parker, E. J.; Koulman, A.; Scott, B., Defining Paxilline Biosynthesis in *Penicillium paxilli* FUNCTIONAL CHARACTERIZATION OF TWO CYTOCHROME P450 MONOOXYGENASES. *J. Biol. Chem* **2007,** *282* (23), 16829-16837.
37. Saikia, S.; Scott, B., Functional analysis and subcellular localization of two geranylgeranyl diphosphate synthases from *Penicillium paxilli*. *Mol. Genet. Genomics* **2009,** *282* (3), 257-271.
38. Scott, B.; Young, C. A.; Saikia, S.; McMillan, L. K.; Monahan, B. J.; Koulman, A.; Astin, J.; Eaton, C. J.; Bryant, A.; Wrenn, R. E.; Finch, S. C.; Tapper, B. A.; Parker, E. J.; Jameson, G. B., Deletion and Gene Expression Analyses Define the Paxilline Biosynthetic Gene Cluster in *Penicillium paxilli*. *Toxins* **2013,** *5* (8), 1422-1446.
39. Liu, C.; Minami, A.; Noike, M.; Toshima, H.; Oikawa, H.; Dairi, T., Regiospecificities and prenylation mode specificities of the fungal indole diterpene prenyltransferases AtmD and PaxD. *Appl. Environ. Microb.* **2013,** *79* (23), 7298-7304.
40. Tagami, K.; Liu, C.; Minami, A.; Noike, M.; Isaka, T.; Fueki, S.; Shichijo, Y.; Toshima, H.; Gomi, K.; Dairi, T.; Oikawa, H., Reconstitution of Biosynthetic Machinery for Indole-Diterpene Paxilline in *Aspergillus oryzae*. *J. Am. Chem. Soc.* **2013,** *135* (4), 1260-1263.
41. Tagami, K.; Minami, A.; Fujii, R.; Liu, C.; Tanaka, M.; Gomi, K.; Dairi, T.; Oikawa, H., Rapid Reconstitution of Biosynthetic Machinery for Fungal Metabolites in *Aspergillus oryzae*: Total Biosynthesis of Aflatrem. *ChemBioChem* **2014,** *15* (14), 2076-2080.
42. Liu, C.; Noike, M.; Minami, A.; Oikawa, H.; Dairi, T., Functional analysis of a prenyltransferase gene (paxD) in the paxilline biosynthetic gene cluster. *Appl. Microbiol. Biot.* **2014,** *98* (1), 199-206.
43. Liu, C.; Minami, A.; Dairi, T.; Gomi, K.; Scott, B.; Oikawa, H., Biosynthesis of Shearinine: Diversification of a Tandem Prenyl Moiety of Fungal Indole Diterpenes. *Org. Lett.* **2016,** *18* (19), 5026-5029.
44. Byrne, K. M.; Smith, S. K.; Ondeyka, J. G., Biosynthesis of Nodulisporic Acid A: Precursor Studies. *J. Am. Chem. Soc.* **2002,** *124* (24), 7055-7060.
45. Fueki, S.; Tokiwano, T.; Toshima, H.; Oikawa, H., Biosynthesis of Indole Diterpenes, Emindole, and Paxilline: Involvement of a Common Intermediate. *Org. Lett.* **2004,** *6* (16), 2697-2700.
46. Uchida, R.; Tomoda, H.; Ōmura, S., Biosynthesis of sespendole. *J. Antibiot.* **2006,** *59* (5), 298-302.
47. Zhao, L.; Chang, W.-c.; Xiao, Y.; Liu, H.-w.; Liu, P., Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annual review of biochemistry* **2013,** *82*, 497-530.
48. Nozawa, K.; Yuyama, M.; Nakajima, S.; Kawai, K.-i.; Udagawa, S.-i., Studies on fungal products. Part 19. Isolation and structure of a novel indoloditerpene, emindole SA, from *Emericella striata*. *J. Am. Chem. Soc.* **1988,** (8), 2155-2160.
49. Xiong, Q.; Zhu, X.; Wilson, W. K.; Ganesan, A.; Matsuda, S. P. T., Enzymatic Synthesis of an Indole Diterpene by an Oxidosqualene Cyclase: Mechanistic, Biosynthetic, and Phylogenetic Implications. *J. Am. Chem. Soc.* **2003,** *125* (30), 9002-9003.

50. Singh, S. B.; Ondeyka, J. G.; Jayasuriya, H.; Zink, D. L.; Ha, S. N.; Dahl-Roshak, A.; Greene, J.; Kim, J. A.; Smith, M. M.; Shoop, W.; Tkacz, J. S., Nodulisporic acids DF: structure, biological activities, and biogenetic relationships. *J. Nat. Prod.* **2004,** *67* (9), 1496-1506.
51. Ondeyka, J. G.; Byrne, K.; Vesey, D.; Zink, D. L.; Shoop, W. L.; Goetz, M. A.; Singh, S. B., Nodulisporic Acids C, C1, and C2: A Series of D-Ring-Opened Nodulisporic Acids from the Fungus *Nodulisporium* sp. *J. Nat. Prod.* **2003,** *66* (1), 121-124.
52. Ondeyka, J. G.; Dahl-Roshak, A. M.; Tkacz, J. S.; Zink, D. L.; Zakson-Aiken, M.; Shoop, W. L.; Goetz, M. A.; Singh, S. B., Nodulisporic acid B, B1, and B2: a series of 1'-deoxy-nodulisporic acids from *Nodulisporium* sp. *Bioorg. Med. Chem. Lett.* **2002,** *12* (20), 2941-2944.
53. Hensens, O. D.; Ondeyka, J. G.; Dombrowski, A. W.; Ostlind, D. A.; Zink, D. L., Isolation and structure of nodulisporic acid A1 and A2, novel insecticides from a *Nodulisporium* sp. *Tetrahedron Lett.* **1999,** *40* (30), 5455-5458.
54. Meinke, P. T.; Ayer, M. B.; Colletti, S. L.; Li, C.; Lim, J.; Ok, D.; Salva, S.; Schmatz, D. M.; Shih, T. L.; Shoop, W. L.; Warmke, L. M.; Wyvratt, M. J.; Zakson-Aiken, M.; Fisher, M. H., Chemical modification of nodulisporic acid A: preliminary structure–activity relationships. *Bioorg. Med. Chem. Lett.* **2000,** *10* (20), 2371-2374.
55. Meinke, P. T.; Colletti, S. L.; Fisher, M. H.; Wyvratt, M. J.; Shih, T. L.; Ayer, M. B.; Li, C.; Lim, J.; Ok, D.; Salva, S.; Warmke, L. M.; Zakson, M.; Michael, B. F.; Demontigny, P.; Ostlind, D. A.; Fink, D.; Drag, M.; Schmatz, D. M.; Shoop, W. L., Discovery of the development candidate N-tert-butyl nodulisporamide: a safe and efficacious once monthly oral agent for the control of fleas and ticks on companion animals. *J. Med. Chem.* **2009,** *52* (11), 3505-3515.
56. Ostlind, D. A.; Felcetto, T.; Misura, A.; Ondeyka, J.; Smith, S.; Goetz, M.; Shoop, W.; Mickle, W., Discovery of a novel indole diterpene insecticide using first instars of *Lucilia sericata*. *Med. Vet. Entomol.* **1997,** *11* (4), 407-408.
57. Ostlind, D. A.; Cifelli, S.; Conroy, J. A.; Mickle, W. G.; Ewanciw, D. V.; Andriuli, F. J.; Ho, P., *A novel Cimex lectularius - Rodent assay for the detection of systemic ectoparasiticide activity*. 2001; Vol. 26, p 181-186.
58. Ondeyka, J. G.; Helms, G. L.; Hensens, O. D.; Goetz, M. A.; Zink, D. L.; Tsipouras, A.; Shoop, W. L.; Slayton, L.; Dombrowski, A. W.; Polishook, J. D.; others, Nodulisporic acid A, a novel and potent insecticide from a *Nodulisporium* sp. Isolation, structure determination, and chemical transformations. *J. Am. Chem. Soc.* **1997,** *119* (38), 8809-8816.
59. Shoop, W. L.; Gregory, L. M.; Zakson-Aiken, M.; Michael, B. F.; Haines, H. W.; Ondeyka, J. G.; Meinke, P. T.; Schmatz, D. M., Systemic efficacy of nodulisporic acid against fleas on dogs. *J. Parasitol.* **2001,** *87* (2), 419-423.
60. Berger, R.; Shoop, W. L.; Pivnichny, J. V.; Warmke, L. M.; Zakson-Aiken, M.; Owens, K. A.; deMontigny, P.; Schmatz, D. M.; Wyvratt, M. J.; Fisher, M. H., Synthesis of Nodulisporic Acid 2 ''-Oxazoles and 2 ''-Thiazoles. *Org. Lett.* **2001,** *3* (23), 3715-3718.
61. Ludmerer, S. W.; Warren, V. A.; Williams, B. S.; Zheng, Y.; Hunt, D. C.; Ayer, M. B.; Wallace, M. A.; Chaudhary, A. G.; Egan, M. A.; Meinke, P. T., Ivermectin and nodulisporic acid receptors in *Drosophila melanogaster* contain both γ-aminobutyric acid-gated Rdl and glutamate-gated GluClα chloride channel subunits. *Biochemistry* **2002,** *41* (20), 6548-6560.
62. Smith, M. M.; Warren, V. A.; Thomas, B. S.; Brochu, R. M.; Ertel, E. A.; Rohrer, S.; Schaeffer, J.; Schmatz, D.; Petuch, B. R.; Tang, Y. S.; Meinke, P. T.; Kaczorowski, G. J.; Cohen, C. J., Nodulisporic Acid Opens Insect Glutamate-Gated Chloride Channels: Identification of a New High Affinity Modulator. *Biochemistry* **2000,** *39* (18), 5543-5554.
63. Davies, T.; Field, L.; Usherwood, P.; Williamson, M., DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB life* **2007,** *59* (3), 151-162.
64. Bass, C.; Schroeder, I.; Turberg, A.; Field, L. M.; Williamson, M. S., Identification of mutations associated with pyrethroid resistance in the para-type sodium channel of the cat flea, *Ctenocephalides felis*. *Insect Biochem. Molec.* **2004,** *34* (12), 1305-1313.

65. Saleem, M. A.; Ahmad, M.; Ahmad, M.; Aslam, M.; Sayyed, A. H., Resistance to selected organochlorin, organophosphate, carbamate and pyrethroid, in *Spodoptera litura* (*Lepidoptera: Noctuidae*) from Pakistan. *J. Econ. Entomol.* **2008,** *101* (5), 1667-1675.
66. Shad, S. A.; Sayyed, A. H.; Fazal, S.; Saleem, M. A.; Zaka, S. M.; Ali, M., Field evolved resistance to carbamates, organophosphates, pyrethroids, and new chemistry insecticides in *Spodoptera litura* Fab.(*Lepidoptera: Noctuidae*). *J. Pest Sci.* **2012,** *85* (1), 153-162.
67. Tong, H.; Su, Q.; Zhou, X.; Bai, L., Field resistance of *Spodoptera litura* (*Lepidoptera: Noctuidae*) to organophosphates, pyrethroids, carbamates and four newer chemistry insecticides in Hunan, China. *J. Pest Sci.* **2013,** *86* (3), 599-609.
68. Smith, A. B.; Davulcu, A. H.; Cho, Y. S.; Ohmoto, K.; Kürti, L.; Ishiyama, H., Indole Diterpene Synthetic Studies. Total Synthesis of (+)-Nodulisporic Acid F and Construction of the Heptacyclic Cores of (+)-Nodulisporic Acids A and B and (−)-Nodulisporic Acid D. *J. Org. Chem.* **2007,** *72* (13), 4596-4610.
69. Zou, Y.; Melvin, J. E.; Gonzales, S. S.; Spafford, M. J.; Smith, A. B., Total Synthesis of (−)-Nodulisporic Acid D. *J. Am. Chem. Soc.* **2015,** *137* (22), 7095-7098.
70. Melvin, J. E., Studies toward the synthesis of nodulisporic acids A and D. *University of Pennsylvania* **2014,** *PhD* (PN, USA), 1-273.
71. Wu, W.; Davis, R. W.; Tran-Gyamfi, M. B.; Kuo, A.; LaButti, K.; Mihaltcheva, S.; Hundley, H.; Chovatia, M.; Lindquist, E.; Barry, K., Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels. *Appl. Microbiol. Biot.* **2017,** *101* (6), 2603-2618.
72. Itoh, Y.; Johnson, R.; Scott, B., Integrative transformation of the mycotoxin-producing fungus, *Penicillium paxilli*. *Curr. Genet.* **1994,** *25* (6), 508-513.
73. Terpe, K., Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biot.* **2006,** *72* (2), 211.
74. Belofsky, G. N.; Gloer, J. B.; Wicklow, D. T.; Dowd, P. F., Antiinsectan alkaloids: Shearinines A-C and a new paxilline derivative from the ascostromata of *Eupenicillium shearii*. *Tetrahedron* **1995,** *51* (14), 3959-3968.
75. Belofsky, G. N.; Gloer, J. B.; Wicklow, D. T.; Dowd, P. F. Indole antiinsectan metabolites from the ascostromata of *Eupenicillium shearii*. US5492902 A, 1996/02/20/, 1996.
76. Gloer, J. B.; Rinderknecht, B. L.; Wicklow, D. T.; Dowd, P. F., Nominine: a new insecticidal indole diterpene from the sclerotia of *Aspergillus nomius*. *J. Org. Chem.* **1989,** *54* (11), 2530-2532.
77. Laakso, J. A.; Gloer, J. B.; Wicklow, D. T.; Dowd, P. F., Sulpinines A-C and secopenitrem B: new antiinsectan metabolites from the sclerotia of *Aspergillus sulphureus*. *J. Org. Chem.* **1992,** *57* (7), 2066-2071.
78. Laakso, J. A.; TePaske, M. R.; Dowd, P. F.; Gloer, J. B.; Wicklow, D. T.; Staub, G. M. Indole antiinsectan metabolites. US5227396 A, 1993/07/13/, 1993.
79. Sallam, A. A.; Houssen, W. E.; Gissendanner, C. R.; Orabi, K. Y.; Foudah, A. I.; El Sayed, K. A., Bioguided discovery and pharmacophore modeling of the mycotoxic indole diterpene alkaloids penitrems as breast cancer proliferation, migration, and invasion inhibitors. *Med. Chem. Comm.* **2013,** *4* (10), 1360.
80. Gao, S.-S.; Li, X.-M.; Williams, K.; Proksch, P.; Ji, N.-Y.; Wang, B.-G., Rhizovarins A–F, Indole-Diterpenes from the Mangrove-Derived Endophytic Fungus *Mucor irregularis* QEN-189. *J. Nat. Prod.* **2016,** *79* (8), 2066-2074.
81. Nagumo, Y.; Motoyama, T.; Hayashi, T.; Hirota, H.; Aono, H.; Kawatani, M.; Osada, H.; Usui, T., Structure-Activity Relationships of Terpendole E and Its Natural Derivatives. *ChemistrySelect* **2017,** *2* (4), 1533-1536.
82. Tarui, Y.; Chinen, T.; Nagumo, Y.; Motoyama, T.; Hayashi, T.; Hirota, H.; Muroi, M.; Ishii, Y.; Kondo, H.; Osada, H., Terpendole E and its Derivative Inhibit STLC-and GSK-1-Resistant Eg5. *ChemBioChem* **2014,** *15* (7), 934-938.
83. Sun, X.; Shen, X.; Jain, R.; Lin, Y.; Wang, J.; Sun, J.; Wang, J.; Yan, Y.; Yuan, Q., Synthesis of chemicals by metabolic engineering of microbes. *Chem. Soc. Rev.* **2015,** *44* (11), 3760-3785.

84. Dombrowski, A. W.; Endris, R. G.; Helms, G. L.; Hensens, O. D.; Ondeyka, J. G.; Ostlind, D. A.; Polishook, J. D.; Zink, D. L. United States Patent: 5399582 - Antiparasitic agents. 5399582, 1995/03/21/, 1995.
85. Byrd, A. D.; Schardl, C. L.; Songlin, P. J.; Mogen, K. L.; Siegel, M. R., The B-tubulin gene of *Epichloe typhina* from perennial ryegrass (*Lolium perenne*). *Curr. Genet.* **1990,** *18* (4), 347-354.
86. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D., Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **2006,** *7* (1), S10.
87. de Marco, A.; Deuerling, E.; Mogk, A.; Tomoyasu, T.; Bukau, B., Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC biotechnology* **2007,** *7*, 32.
88. Miyazaki, K., Molecular engineering of a PheS counterselection marker for improved operating efficiency in *Escherichia coli*. *BioTechniques* **2015,** *58* (2), 86-8.
89. Patron, N. J.; Orzaez, D.; Marillonnet, S.; Warzecha, H.; Matthewman, C.; Youles, M.; Raitskin, O.; Leveau, A.; Farre, G.; Rogers, C.; Smith, A.; Hibberd, J.; Webb, A. A.; Locke, J.; Schornack, S.; Ajioka, J.; Baulcombe, D. C.; Zipfel, C.; Kamoun, S.; Jones, J. D.; Kuhn, H.; Robatzek, S.; Van Esse, H. P.; Sanders, D.; Oldroyd, G.; Martin, C.; Field, R.; O'Connor, S.; Fox, S.; Wulff, B.; Miller, B.; Breakspear, A.; Radhakrishnan, G.; Delaux, P. M.; Loque, D.; Granell, A.; Tissier, A.; Shih, P.; Brutnell, T. P.; Quick, W. P.; Rischer, H.; Fraser, P. D.; Aharoni, A.; Raines, C.; South, P. F.; Ane, J. M.; Hamberger, B. R.; Langdale, J.; Stougaard, J.; Bouwmeester, H.; Udvardi, M.; Murray, J. A.; Ntoukakis, V.; Schafer, P.; Denby, K.; Edwards, K. J.; Osbourn, A.; Haseloff, J., Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *The New phytologist* **2015,** *208* (1), 13-9.
90. Yelton, M. M.; Hamer, J. E.; Timberlake, W. E., Transformation of *Aspergillus nidulans* by using a *trpC* plasmid. *PNAS* **1984,** *81* (5), 1470-1474.
91. Vollmer, S. J.; Yanofsky, C., Efficient cloning of genes of *Neurospora crassa*. *PNAS* **1986,** *83* (13), 4869-4873.
92. Oliver, R. P.; Roberts, I. N.; Harling, R.; Kenyon, L.; Punt, P. J.; Dingemanse, M. A.; van den Hondel, C. A. M. J. J., Transformation of *Fulvia fulva*, a fungal pathogen of tomato, to hygromycin B resistance. *Curr. Genet.* **1987,** *12* (3), 231-233.
93. Penn, J.; Mantle, P. G., Biosynthetic intermediates of indole-diterpenoid mycotoxins from selected transformations at C-10 of paxilline. *Phytochemistry* **1994,** *35* (4), 921-926.
94. Penn, J.; Swift, R.; Wigley, L. J.; Mantle, P. G.; Bilton, J. N.; Sheppard, R. N., Janthitrems B and C, two principal indole-diterpenoids produced by *Penicillium janthinellum*. *Phytochemistry* **1993,** *32* (6), 1431-1434.
95. Namiki, F.; Matsunaga, M.; Okuda, M.; Inoue, I.; Nishi, K.; Fujita, Y.; Tsuge, T., Mutation of an Arginine Biosynthesis Gene Causes Reduced Pathogenicity in *Fusarium oxysporum* f. sp. melonis. *Mol. Plant Microbe In.* **2001,** *14* (4), 580-584.