Characterisation of protolimonoid biosynthesis in plants:
A key step towards engineering limonoids for crop protection.

A thesis submitted to the University of East Anglia in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Hannah Elizabeth Hodgson

December 2019

Abstract

The limonoids are a diverse class of plant natural product produced by the Meliaceae (Mahogany) and Rutaceae (Citrus) families. Limonoids are recognisable by their furan-containing tetracyclic triterpene structures, a number of which have human-usable biological activities. For instance, azadirachtin (from *Azadirachta indica* (neem)) is used to protect crops from insect pests. The current-use of limonoids is reliant on their extraction from plant materials, which has limited the extent of their use. Metabolic engineering could provide an alternative supply of limonoids. However, at the commencement of this project, the metabolic engineering of limonoids was not a possibility as the enzymes responsible for even the first step in their biosynthesis remained unknown. Here, the initiation of limonoid biosynthesis has been characterised in three diverse limonoid-producing species, *A. indica, Melia azedarach* (chinaberry) and *Citrus sinensis* (sweet orange). Three oxidosqualene cyclases (AiOSC1, MaOSC1 and CsOSC1) have been characterised as functional tirucalla-7,24-dien-3β-ol synthases which initiate limonoid biosynthesis. Further, two cytochrome P450s (MaCYP71CD2 and MaCYP71BQ5) from *M. azedarach* have been shown to collectively introduce three oxidations onto the side-chain of this scaffold to produce the protolimonoid melianol, a precursor to limonoids. The availability of Nagoya Protocol compliant *M. azedarach* plants has allowed the limonoid content of this species to be profiled and enabled the generation of a pseudochromosome-level genome assembly with accompanying RNA-Seq datasets. Utilising these resources a further four limonoid biosynthetic enzymes have be identified. Two of these enzymes (MaCYP88A108 and MaIsom1-I) are predicted to convert the internal scaffold of melianol to one of a mature limonoid and the other two enzymes (MaSDR1 and MaBAHD1) are thought to decorate the resultant structure. This biosynthetic knowledge will provide a basis for future metabolic engineering of limonoids for crop protection and has already enabled *Nicotiana benthamiana*-based investigations into the structure-activity relationships of limonoids.

Contents

List of Figures

List of Tables

Acknowledgements

I would firstly like to thanks to my supervisors Prof. Anne Osbourn and Dr. Jason Vincent. I am incredibly grateful to them both for creating a project I have truly enjoyed and providing me with so many opportunities throughout the course of my PhD. I have learnt a great deal from the expertise of them both. I would also like to thank my secondary supervisor Prof. Rob Field for his help and advice.

My PhD would not have been possible without the breadth of expertise of the Osbourn group and therefore I would like to thank all group members, past and present. In particular I would like to thank: Dr. Ramesha Thimmappa, for his supervision at the beginning of my PhD; Dr. James Reed, for his constant help; Dr. Michael Stephenson, for advice on everything chemistry related; and finally Dr. Zhenhua Liu, Charlotte Owen and Jenny Jo, for their friendship and support.

I am incredibly grateful for all of the JIC platforms, for all that they do, and would like to particularly thank: Dr. Lionel Hill and Dr. Paul Brett, of JIC metabolomics; Darrell Bean, Anna Jordon and Susie Gill of, JIC entomology; Dr. Matthew Hartley and Dr. Tjelvar Olsson, of JIC informatics; and Mike Ambrose, of JIC GRU.

I also have to thank all of my friends and family. I feel incredibly lucky to be surrounded by so many people that I love, all of whom have encouraged and supported me throughout this PhD. I owe an enormous thanks to my parents Liz and Mike for for their never-ending support throughout the last 22-years of my education. Their support has given me both the opportunity and confidence to pursue this PhD, which is a true privilege. I would also like to thank my brother Sam and friends Emily and Sophie for always loyally listening to me talk about my PhD- even if they are not entirely sure what I am going on about. Lastly, I have to thank my husband Will. Firstly for his endless patience with me and my PhD, but moreover because his presence in my life (with a little help from Arla the dog) has brought me so much happiness and contentment that I could not imagine completing this PhD with out him.

Abbreviations

ABSCH Access and Benefit Sharing Clearing House

BAHD Benzylalcohol acetyl-, anthocyanin-*O*-hydroxy-cinnamoyl-, anthranilate-*N*-hydroxy-cinnamoyl/benzoyl-, deacetylvindoline acetyltransferase

BGC Biosynthetic gene cluster

bp Base pairs

BVMO Baeyer-Villiger Monooxygenases

CTAB Cetyl trimethylammonium bromide

CYP Cytochrome P450

DSI Digital sequencing information

EI Earlham Institute

EIC Extracted ion chromatogram

ESI Electrospray ionization

GC-MS Gas chromatography mass spectrometry

HMW gDNA High molecular weight genomic DNA

ICH-VAST Institute of Chemistry- Vietnam Academy of Science and Technology

ICNaP Interdisciplinary Centre for Natural Products Research

IT-TOF Ion trap time of flight

JIC John Innes Centre

Kbp Kilo base pair

LC-MS Liquid chromatography mass spectrometry

MAT Mutually Agreed Terms

Mbp Mega base pair

MS Mass spectra

NCBS National Centre for Biological Sciences

NMR Nuclear magnetic resonance

NP Nagoya Protocol

OGD 2-Oxoglutarate Fe(II)-dependent oxygenases

OSC Oxidosqualene cyclase

PIC Prior Informed Consent

qRT-PCR Quantitative real-time reverse transcription PCR

SDR Short-chain dehydrogenase/reductase

SRA Sequence read archive

TIC Total ion chromatogram

TMS Trimethylsilyl

UHPLC Ultra high performance liquid chromatography

USTH University of Science and Technology, Vietnam

1

General introduction

Plants produce an abundance of diverse chemicals that enable them to survive and reproduce within their environment. Many of these chemicals (natural products) are considered 'secondary metabolites' as they do not appear to be essential for the physiological function of the plant, but can provide a competitive advantage in nature. Plant natural products have diverse functions including providing protection against pathogens and herbivores and attracting pollinators (1). In addition to their functions *in planta*, plant natural products have a long history of human use and the continued identification of novel, bioactive natural products suggests that many more new drugs and high-value compounds are likely to be discovered from plants (2). However, the complex chemical structures of many natural products make them recalcitrant to chemical synthesis. Furthermore, these compounds are often present at low levels in plant extracts as parts of complex mixtures, preventing commercially viable extraction (1, 3). This has restricted the use of many natural products with promising biological activities. Metabolic engineering has the potential to provide alternate sources of these natural products and so expand their usefulness to humans (1, 3).

The limonoids are a large and structurally diverse class of plant natural product. There are almost 2,000 limonoids isolated to date, with an array of associated biological activities (4–9). Limonoids are modified triterpenes and therefore members of the large terpene family which accounts for almost 40% of all plant natural products (10). Whilst triterpene structures are found across the plant kingdom (11), the occurrence of limonoids is species-specific and largely confined to the Meliaceae (Mahogany) and Rutaceae (Citrus) families (12–14). The limonoids have received scientific attention for over 75 years due to their bitter taste and anti-insect activity, which has been exploited for both traditional and commercial crop protection (12, 13, 15, 16). More recently the limonoids have been investigated for other uses, including as future pharmaceuticals (17). Whilst the diverse structures and biological activities of limonoids have been heavily investigated (4–6, 12–17), an understanding of their biosynthesis is currently lacking. This has

prevented metabolic engineering attempts thus far. The use of limonoids is therefore still reliant on extraction from plant material. Here, the structures, occurrence and biological activities of limonoids are introduced, along with the potential benefits their metabolic engineering could bring.

1.1 Structural diversity of limonoids

The prototypical limonoid is a tetracyclic C26 scaffold with a 6,6,6,5 ring system and an additional furan ring (Figure 1.1.1). Limonoids are considered tetranor-triterpenes, as their structure feasibly originates from a C30 triterpene which has lost four carbon atoms during furan ring formation. The structure of simple limonoids resembles tetracyclic triterpene skeletons such as euphol or tirucallol, which have been modified by the addition of a furan ring in place of the linear side chain and movement of the C30 methyl group (Figure 1.1.1).
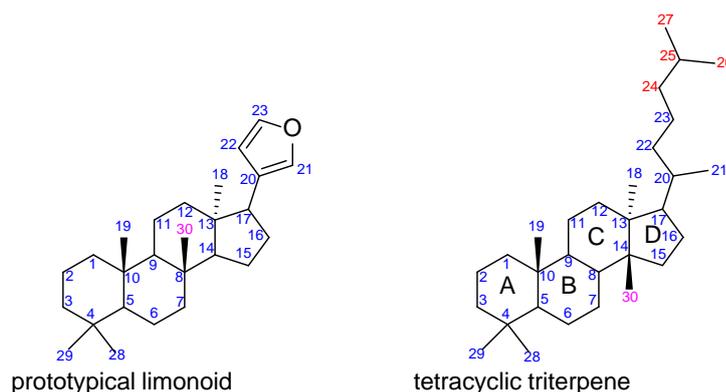
Figure 1.1.1: **Structures of a prototypical limonoid and a tetracyclic triterpene.** Carbon and ring numbering system for a prototypical limonoid structure and a tetracyclic triterpene (euphol or tirucallol type). The terminal four carbons, which are thought to be lost in furan ring formation, are indicated (red) along with the carbon involved in the methyl shift (pink).

Limonoids that contain this prototypical scaffold, such as 7-deacetylazadirone (Figure 1.1.2), are referred to as ring-intact limonoids. Ring-intact limonoids are considered the simplest class of limonoids. However their structures are still diverse. Ring-intact limonoids can be heavily oxygenated and their scaffolds modified, an example being the formation of ether bridges in sendanin (Figure 1.1.2).

The diversity of ring-intact limonoids is further expanded by the formation of *seco*-ring limonoid derivatives. The term *'seco'* refers to the cleavage of a ring with the addition of hydrogen atoms to the terminal groups. Within the limonoids *seco*-ring derivatives refer not only to demolished rings, but also those that have reformed
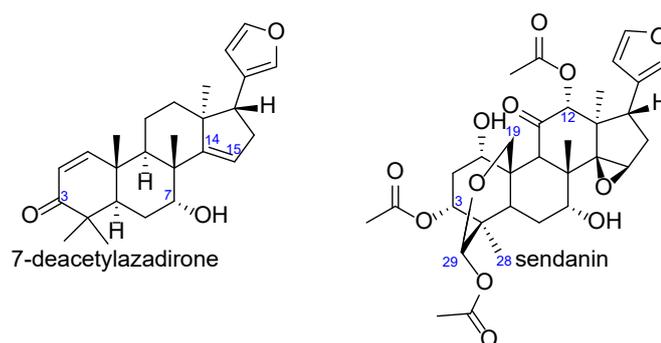
Figure 1.1.2: **Examples of ring-intact limonoids.**
7-Deacetylazadirone (isolated from both Meliaceae (5) and Rutaceae (18) species) and sendanin (isolated from two Meliaceae species, *Melia azedarach* and *Trichilia roka* (5)).

with a different configuration or that have undergone oxidative expansion, such as the A and D rings of limonin, respectively (5) (Figure 1.1.3). A recent review reported >300 *seco*-ring limonoids within the Meliaceae family, including *seco*-formation at each of the four of rings. Further, there are >100 Meliaceae limonoids structures that contain more than one *seco*-ring, resulting in highly rearranged structures such as methyl ivorensate (Figure 1.1.3), where only the B ring remains intact (5).

The ability of the limonoids to form *seco*-rings adds new layers of diversity. Firstly, the opening and expansion of the rings provides new positions that can be further modified. Secondly, significant changes to the structure and flexibility of the limonoid scaffold are introduced, which can lead to the 3D structures of *seco*-ring limonoids being distinctly different to their ring-intact precursors (19).

This formation of *seco* derivatives is considered a rarity in non-limonoid triterpenes. Non-limonoid *seco*-triterpenes do exist among the 19,461 triterpene currently known (20). However, ~40 of these structures are formed as part of scaffold generation (21) rather than by dedicated ring opening steps, as seen in the limonoids. The vast majority of *seco*-triterpenes are limonoids; however *seco*-triterpenes are also a feature of the quassinoids, a related class of nortriterpene produced by Simaroubaceae species within the Sapindales order (22).

In addition to *seco*-ring formation limonoids are also modified by oxidation, acylation and glycoslation, as commonly seen for other triterpene pathways (11). Limonoid glucosides are commonly found in Citrus species (12, 13) and decoration of the Meliaceae limonoids with various acyl groups is common. These acylations range from the addition of simple acetoxy groups such as those at C3, C12 and C29 in sendanin (Figure 1.1.2), to larger groups such as tiglate at C1 in salannin (Figure 1.1.3) (5).
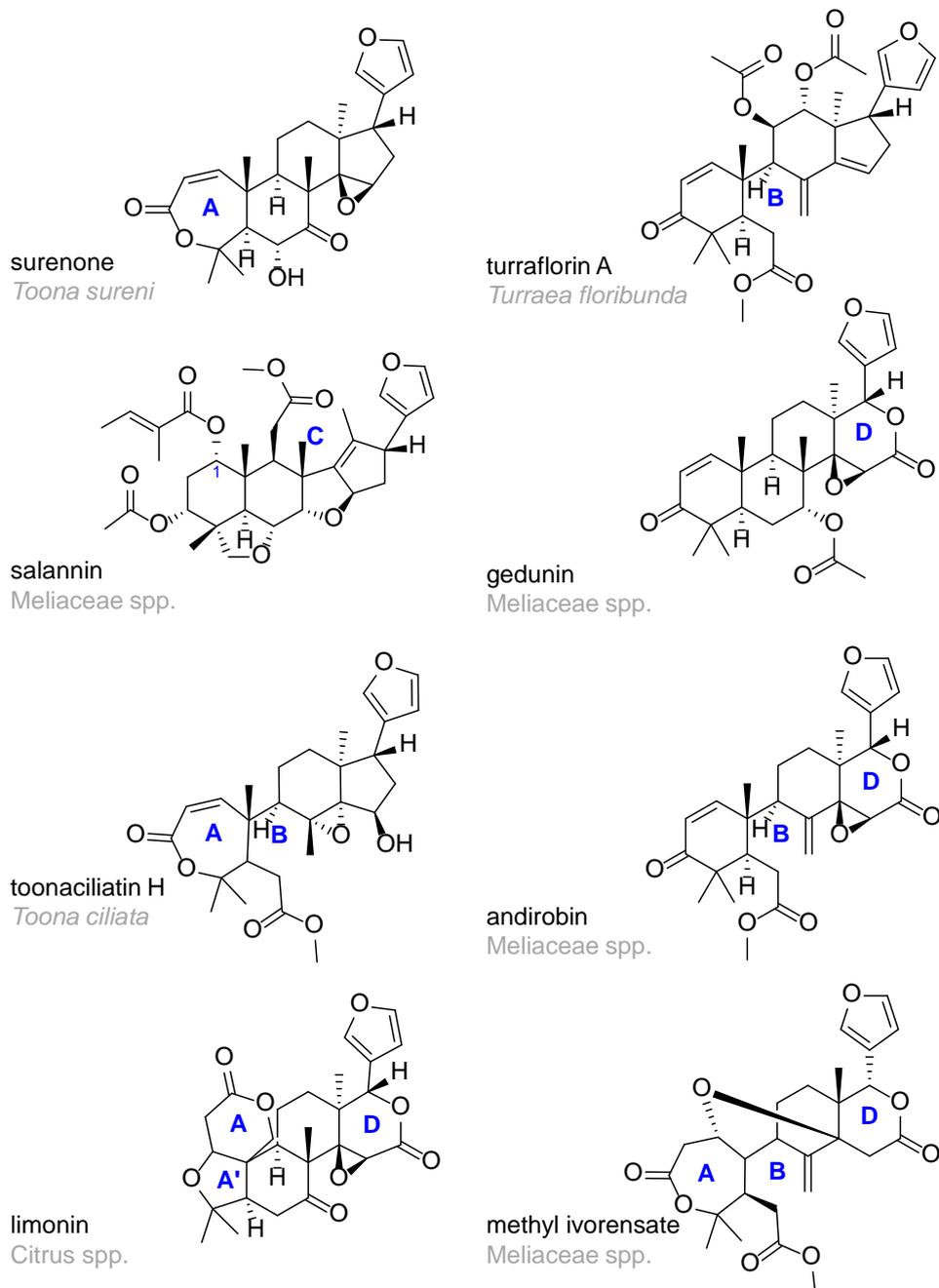
Figure 1.1.3: Examples of *seco*-ring limonoids.
Representative structures of *seco*-ring limonoids. The *seco*-rings are labelled in blue.
All limonoids shown are isolated from the Meliaceae family, with the exception of
limonin, which is Rutaceae-specific (4–6).

1.2   Occurrence of limonoids within the Sapindales order

The biosynthesis of limonoids is largely restricted to species within the Sapindales order of plants. Currently, the are two reported exceptions to this within the plant kingdom: the isolation of seven limonoids from two *Phyllanthus* species of the Euphorbiaceae family (Malpighiales order) (23, 24) and the isolation of five limonoids from *Kigela africana* of the Bignoniaceae family (Lamiales order) (25). Whether these represent isolated occurrences of convergent evolution or whether limonoids are produced more widely within these families is not yet known. Within the limonoid-producing families of the Sapindales order limonoid production appears to be conserved and has only been reported in the Meliaceae (Mahogany), Rutaceae (Citrus) and Simaroubaceae families. The vast majority of limonoids are specific to the Meliaceae (Mahogany) and Rutaceae (Citrus) families (Figure 1.2.2), only a small number of limonoids having been reported in the Simaroubaceae family. However the Simaroubaceae produce the separate but related class of nor-triterpenes, the quassinoids (26, 27) (Figure 1.2.2).

Within these limonoid-producing families there appears to be a degree of divergence towards production of family-specific limonoid types. The structures of limonoids from the Meliaceae and Rutaceae families are distinct from each other, with only a handful of structures reported from both families. One of the overlapping limonoids is 7-deacetylazadirone (Figure 1.2.2), a simple ring-intact limonoid that has been isolated from *Azadirachta indica* (neem) and *Walsura piscidia* of the Meliaceae (28) family, as well as the Rutaceae species *Teclea grandifolia*. The only *seco*-ring limonoids that are shared between these two families are members of the *seco*-A,D-ring obacunone class (Figure 1.2.1), which have been isolated from citrus species (29) and obscure Meliaceae genera (*Lovoa, Trichilia* and *Cedrela*) (5).

11β-acetoxyobacunyl acetate   7α-obacunol   obacunoic acid
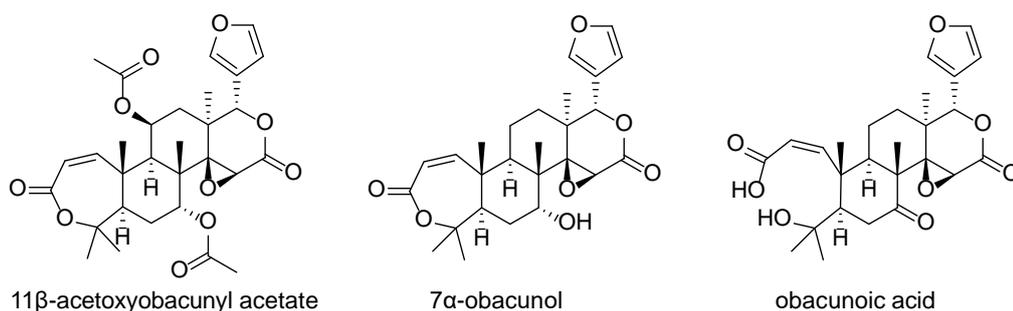
Figure 1.2.1: **Example limonoid structures belonging to the obacunone class.**
11β-Acetoxyobacunyl acetate from the Meliaceae species *Cedrela odorata*, 7α-obacunol from the Meliaceae species *Lovoa trichiliodes, Trichilia trifolia*, as well as a number of citrus species; and obacunoic acid from citrus species (4, 5).

Sapindales order

Biebersteiniaceae
Nitrariaceae
Sapindaceae
Kirkiaceae
Anacardiaceae
Burseraceae

Non-limonoid
producing
families

Rutaceae family
>70 limonoids
(including limonoid glycosides)

Citrus sinensis
(sweet orange)

limonin 17-ß-D-glucoside

7-deacetylazadirone

Simaroubaceae family
~20 limonoids and derivatives
~ 150 quassinoids

Ailanthus alitssima
(tree of heaven)

ailanthone

Meliaceae family
>1000 limonoids

Melioideae subfamily

Cedreloideae subfamily

Azadirachta & Melia Spp.
>300 limonoids

Azadirachta indica
(neem)

Khaya senegalensis
(African mahogany)

mexicanolide

azadirachtin

Figure 1.2.2: **Production of limonoids within the Sapindales.**
An overview of limonoid production within the Sapindales, including example
species and associated limonoid structures (5, 6, 12, 13, 26). An example quassinoid
structure (ailanthone), which is specific to the Simaroubaceae, is also included
(27, 30). The phylogeny of the Sapindales is based on phylogenetic studies (31, 32).
Images are from Wikimedia Commons (33–36).

1.2.1 The Rutaceae (Citrus) limonoids

The investigation of limonoids in commercial citrus species began in the 1940s and was driven by the need to identify the bitter components of their fruits and juices (4, 12, 13). The first citrus limonoid to be isolated was limonin, the most prevalent limonoid in citrus. Limonin was quickly determined to be one of two bitter components of citrus juices (the other being the flavonoid naringin) (4), however the structure of limonin was not determined until the 1960s (37).

Limonin has been intensively studied due to its role in delayed bitterness. Delayed bitterness, which is induced by cold temperatures, was a major factor in the 90 million USD freeze losses suffered by the juice industry of California between 1992 and 2006 (13). Bitterness is associated with limonoid aglycones structures that possess a closed, epoxide-containing D-ring, C-7 carbonyl and the characteristic double A-ring (38), such as limonin (Figure 1.2.3).
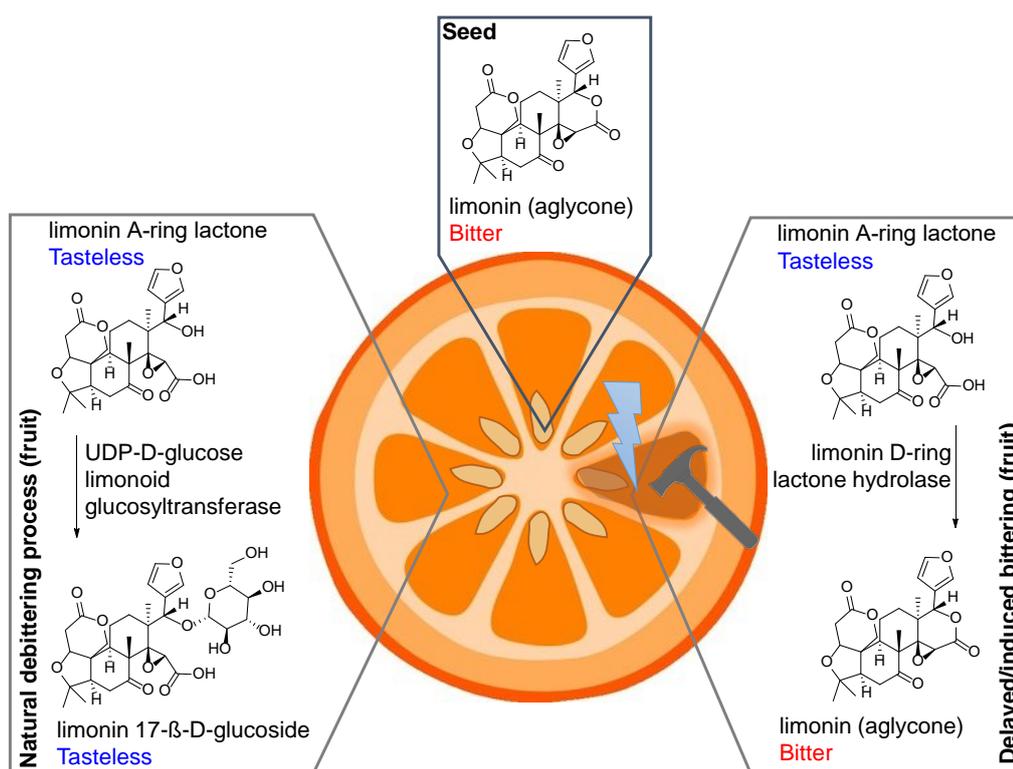
Figure 1.2.3: **Schematic of de-bittering process in citrus fruits.**
Location of accumulation of different types of limonoids in citrus fruits. The predominant limonoid (limonin) is shown in this example. The process of natural de-bittering (which occurs within the fruit) and delayed bittering (in damaged fruit) are illustrated. Bitterness (red) and tastelessness (blue) are indicated next to the relevant structures.

Most limonoid aglycones only accumulate to high levels in the seeds of citrus fruit, which confines their bitter taste. Limonin accumulates to levels of up to 10 mg/g DW in *Citrus sinensis* (sweet orange) seed kernels (29).   Contrastingly, the juice containing tissues of citrus fruit contain tasteless limonoid structures with an open D-ring, such as limonin-A-ring lactone (LARL) (Figure 1.2.3).   During the natural process of fruit maturation LARL is converted to limonin-17-β-D-glucoside by limonoid UDP-D-glucosyltransferase (39), currently the only limonoid-specific enzyme to be characterised by heterologous expression (40) (Figure 1.2.3).   In contrast to this natural process, when tissue containing juice is exposed to cold temperatures or mechanical damage, a separate enzyme converts LARL to limonin, causing delayed bitterness which impacts on commercial value (4, 12, 13) (Figure 1.2.3).

Early labelling studies within citrus species have shown that limonoid biosynthesis is not entirely fruit-specific.   In *Citrus limon* (lemon) stems, ^{14}C-labelled precursors were incorporated into the *seco*-A,D-ring limonoid nomilin (Figure 1.2.4) (41, 42). Further, ^{14}C-labelled nomilin was converted to limonin in both the leaves and fruits (41), and to obacunone in the stem (43) of *C. limon.* Therefore nomilin is considered an early *seco*-A,D-ring limonoid and the precursor to more complicated citrus limonoids, such as limonin.

Until recently it was thought that the Rutaceae family produce only 36 limonoid aglycones (4).   However, in recent years 32 new limonoid structures have been discovered from more obscure members of the family (6–9).   Despite these newly discovered structures, the limonoids of the Rutaceae show much less diversity than the Meliaceae, the vast majority of structures falling within the *seco*-A,D-ring class of limonoids (12).   The number of Rutaceae limonoids with a *seco*-C-ring are extremely limited and often highly rearranged, e.g.   the *seco*-B,C,D-ring limonoid 11β-hydroxyhortiolide C (Figure 1.2.4), which has been recently isolated from *Hortia oreadica* (44) and is an exception to this rule.

1.2.2   The Meliaceae (Mahogany) limonoids

The Meliaceae limonoids are quite distinct from those of the Rutaceae. In contrast to the relatively small numbers of Rutaceae limonoids there are ~1,879 Meliaceae structures (5–9). All of the classifications of *seco*-ring limonoids shown in Figure 1.1.3 have been found within the Meliaceae family.   This greater diversity can not simply be assigned to the larger number of plant species within this family.   The Rutaceae family contains 1,730 accepted species, compared with only 669 within the Meliaceae (45). It may be that the diversity of limonoids found within the Meliaceae
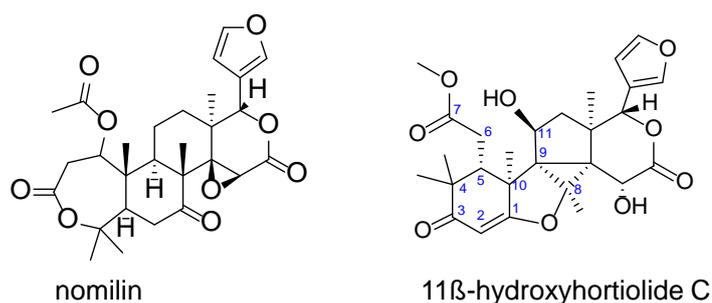
Figure 1.2.4: **Structures of nomilin and 11β-hydroxyhortiolide C.**
Nomilin has been isolated from various citrus species (4) and 11β-hydroxyhortiolide C has been isolated from *Hortia oreadica* (44).

family is a reflection of the hyperdiversity that has evolved within this family of tropical rainforest trees. The differentiation between Meliaceae species is thought to be due to recent rapid phenotypic radiation (31) and it seems feasible that the limonoid chemistry within these species simultaneously radiated.

Like the Rutaceae family, the Meliaceae family has a long history of investigation. It was traditionally known that the tropical tree *A. indica* (neem) was the only desert plant to be untouched by locusts (16). Further, in Asian countries, such as India, extracts of the seed kernels of *A. indica* are used in traditional crop protection. This phenomenon has attracted sustained scientific attention, with over 1,000 studies on *A. indica* alone (17). The first Meliaceae limonoid to be isolated from *A. indica* was nimbin, which was isolated from neem oil in 1942 (46), although the structure of nimbin remained unconfirmed for over 20 years (47). However, it was an investigation of the seeds of *A. indica* in 1968, by researchers at Keele University, that led to the isolation of the major anti-insect component produced by neem, azadirachtin (48, 49) (Figure 1.2.5).

The complex structure of azadirachtin took almost 20 years to solve (50–54) (Figure 1.2.5). Although azadirachtin is a potent insecticide, the potency of crude *A. indica* seed kernel extract is greater than azadirachtin alone, due to other anti-insect limonoids (such as salannin) acting synergistically (16) (Figure 1.2.5). A recent quantitative study suggested that azadirachtin accumulates to the highest levels in *A. indica* seed kernels (3.5 mg/g) followed by salannin (1.25 mg/g) and nimbin-type scaffolds (~0.2 mg/g) (55). This study revealed a variable occurrence of limonoids in different *A. indica* tissues (55) and raises the question of whether the whole biosynthetic pathway occurs in one tissue of the plant or in multiple tissues. Specialist triterpene-accumulating 'secretary cells' have been observed in all *A. indica* tissues assessed (56). However, with the exception of *A. indica* seeds, the major site of limonoid biosynthesis in most Meliaceae species has not been defined.
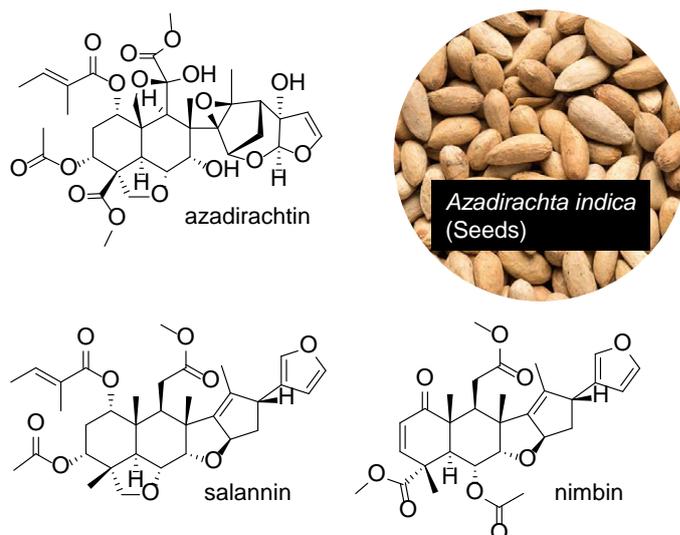
Figure 1.2.5: **Structure of selected limonoids isolated from *A. indica* seeds.**
Image credit Andrew Davis (JIC photography).

New limonoid structures are still being rapidly identified within the Meliaceae. A 2011 review of all Meliaceae limonoids reported 1,113 structures (5) from 256 different Meliaceae species and a later review (2014-2016) reported a further 339 limonoid structures (6). In addition to these limonoid-specific reviews, a further 427 Meliaceae limonoids were reported in triterpene-specific reviews (2012-2013) (7–9). The identification of most of these novel limonoids has come from more obscure members of the Meliaceae family which have only recently been investigated. In November 2019 six novel limonoids (munropins A-F) were reported from *Munronia pinnata*, including five *seco*-A,B-ring limonoids and one *seco*-C-ring limonoid (57). Additionally, novel limonoids are still being identified from well-known Meliaceae species. In 2019 a new limonoid (spirodione, Figure 1.2.6) was identified from *A. indica* seed kernels, one of only two spiro-type limonoid structures isolated to date (58).
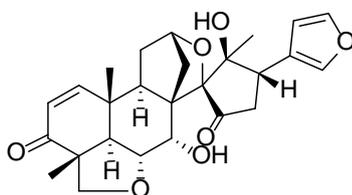
Figure 1.2.6: **Structure of spirodione.**

1.3   Biological activities of limonoids

Azadirachtin is often described as toxic to insects yet harmless to mammals (5). This has led to promotion of azadirachtin and related *seco*-C-ring limonoids as environmentally friendly, selective crop protection agents. *A. indica* seed kernel extracts have been commercially used for crop protection since 1985, with an even longer history of traditional use (15). However, the diversity of limonoid structures, particularly within the Meliaceae family, have potential beyond the anti-insect effect of *seco*-C-ring limonoids. A number of limonoids have been implicated as potential crop protection agents for microbial pathogens, and others are currently being investigated for useful pharmaceutical properties (5, 15). The majority of the primary literature reporting new limonoid structures have additionally used bio-assays to assign preliminary bioactivity to these structures. This has generated a large volume of limonoid bioactivity data which has been extensively reviewed (5, 6, 14–16, 59–63) and is introduced here.

1.3.1   Anti-insect activities of limonoids

The anti-insect properties of limonoids are wide-ranging with ~90 limonoids reported to be effective (5). Further, limonoids have been shown to affect a broad range of insects, including the Lepidoptera, Orthoptera, Coleoptera, Heteroptera and Diptera insect orders (14). Reviews of these properties have been complicated by the dual toxicity and antifeedant activity of limonoids such as azadirachtin. Antifeedancy may mask toxicity, and contrastingly toxicity may reduce feeding (14).

Azadirachtin has received the most attention because of its potent anti-insect effects (15, 16). These were first quantified for *Scistocera gregaria* (desert locust), where feeding reduction was observable at concentrations as low as 0.04 ppm (48). Cauterisation of the mouthpieces of *S. gregaria* permitted feeding (16), which subsequently revealed that azadirachtin was not merely a feeding deterrent but a potent toxin to insects (14). In comparison to its antifeedant activity, the toxicity of azadirachtin is more consistently observed between insect species (15) and further, toxicity is observed at lower concentrations (64). Azadirachtin has reportedly been tested on over 600 insect species, where the concentration required for activity ranged from 1-10 ppm (16).

The activity of azadirachtin is observed by a plethora of different responses, which early reviews summarise in three parts: i) antifeedancy, ii) growth regulatory effects and iii) reproductive effects (15). The observed activities within each group are complex. For instance, antifeedant effect can be be divided into primary

antifeedancy (associated with deterrents sensed by chemoreceptors) and a secondary antifeedancy (a reduction in food intake caused by ingestion). The growth regulatory and reproductive effects of azadirachtin could be symptoms of more general toxicity. The observed toxic effects occur across a wide axis ranging from a minor reduction in longevity to rapid mortality. Whilst moulting deformities are commonly observed and rationalised as a symptom of impaired growth regulation, other effects are more difficult to rationalise, such as the occurrence of 'black patches' on the cuticles of azadirachtin-treated insects (15).

Despite these complexities, the potency of azadirachtin is not disputed. The suitability of azadirachtin for crop protection extends beyond its potency, systemic uptake, rapid degradation in the environment and relatively low toxicity to beneficial insects all recommend azadirachtin as a suitable crop protection agent (15, 16). Subsequently, a number of commercial crop protection formulations have used *A. indica* seed kernel extracts due to their high azadirachtin content. One such formulation is NeemAZAL T/S (TrifoliaM), which has been developed in Germany for use in 'organic' agriculture based on the positive qualities of azadirachtin (Table 1.3.1). NeemAZAL T/S and other azadirachtin-based formations have been approved for use in the EU and legalised in 15 EU countries.

Table 1.3.1: **Summary of NeemAZAL T/S as advertised by Trifolio-M (65).**

Formulation:	*A. indica* oil and surfactants.
Active substance:	1% azadirachtin (10 g/L).
Targets:	Hemiptera, Homoptera, Lepidoptera, Coleoptera, Diptera, Thysanoptera and Acarina.
Resistance:	Unknown and deemed less likely due to complexity of mixture.
Uptake:	Systemic and translaminar.
Environmental:	Rapid degradation.
Effect on beneficial insect:	'Harmless' to most (including honey bees).
Effect on aquatic organism:	'Safe'.

The use of *A. indica* in pesticides is more extensive outside of Europe. In China there are three limonoid-based insecticides based on extracts of *A. indica* and closely related *Melia azedarach* and *Melia toosendan* (5). Further, there are now over 20 commercial *A. indica*-based pesticides in India, in addition to traditional use (16).

The traditional use of *A. indica* has been encouraged by many governments as an alternative to expensive commercial products (16). A recent study compared NeemAZAL T/S to a traditional Mali aqueous extract of *A. indica* seeds (200 mg/L azadirachtin) (66). The aqueous extract, even when diluted to recommended concentrations of NeemAzal T/S (25 mg/L), showed equivalent activity to the

commercial insecticide against tobacco whitefly (*Bemisia tabaci*) and cotton leafworm (*Spodoptera littoralis*) (66).

The anti-insect effects of limonoids are not limited to azadirachtin or *A. indica* alone. However, identifying which of the non-azadirachtin limonoids have potentially useful anti-insect activity, and the substructures associated with this, has historically been difficult due to the the vast amount of primary biological activity data which differs by species, larval stage and bioassay used (14).

The structure-activity relationships of limonoids are still not fully understood (14–16). However, some trends have been observed. *Seco*-C-ring limonoids, such as azadirachtin, salannin and nimbinene (Figure 1.2.5) are the most active class of limonoids (14). Further, the most active limonoids within this class often originate from *Azadirachta* and *Melia* species (5) (Figure 1.2.2). The second most active group are the ring-intact limonoids, within which a C14-C15 epoxide is associated with increased activity (14).

Comparative reviews have suggested that the *seco*-A,D-ring limonoids typical of the Rutaceae family are associated with comparatively low levels of anti-insect effect. This is in contrast to early reports that citrus limonoids possess useful levels of anti-insect activity (4, 13). However, it has been suggested that high levels of limonoid accumulation within citrus fruits may overcome this low potency to cause the observed anti-insect effects (12, 14).

Further structure-activity information has been provided by a large scale study utilising *Spodoptera litura* (tobacco cutworm) third instar larvae in a dual-choice antifeedant assay (67). This assay assessed 56 limonoid structures, which were a mixture of naturally isolated limonoids and photooxidation products (67). As expected, the azadirachtin-type limonoids possessed the most potent antifeedant activity followed by the ring-intact limonoids, such as azadirone.

A further observation was that photooxidation products containing an additional oxygen (in the form of an epoxide or hydroxy group) on their furan ring showed increased antifeedancy (67). This may explain the high potency of azadirachtin, which contains a hydroxy group at C23 of the furan ring. A further observation from this study was that free hydroxy groups in the A-ring of limonoid structures (at the C1 or C3 position) could further increase antifeedant activity, and indeed azadirachtol showed a greater antifeedant effect than azadirachtin itself (67) (Figure 1.3.1).
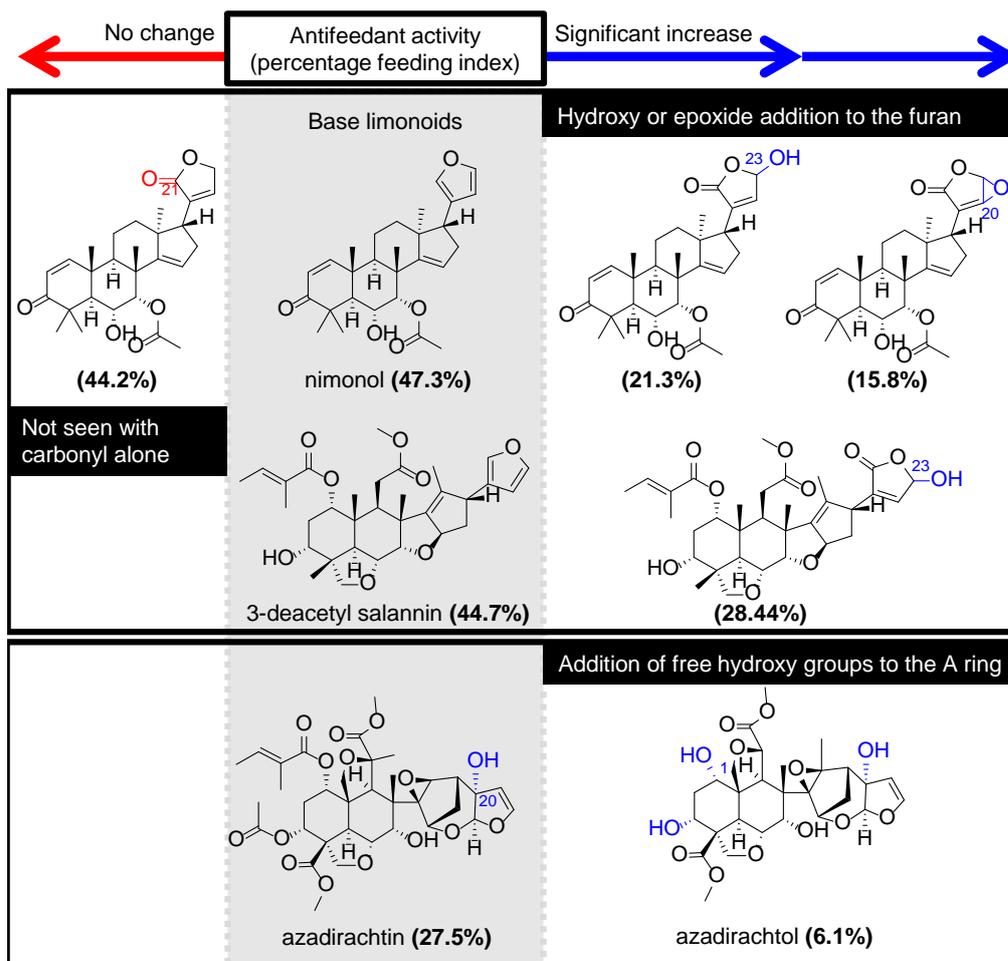
Figure 1.3.1: **Structural features associated with antifeedancy.**
Modifications associated with increased antifeedancy are indicated (blue) and for each structure the percentage feeding index, area fed/(area fed treated + area fed control) x 100, is provided (67). Limonoid structures are labelled. Where a name is not given, the structure is a photooxidation product.

1.3.2   Potential pharmaceutical activities of limonoids

The human use and associated bioactivity of limonoids extends beyond crop protection. A pseudonym for *A. indica*, the most prolific limonoid producing plant, is the 'village pharmacy' (15, 68), which anecdotally highlights the medicinal value of this tree and the limonoids it produces. *A. indica* has been used in India both traditionally and culturally for over 4,000 years (69). Its use in traditional medicine is extensive, ranging from antimicrobial to anti-inflammatory treatments. The other reported uses of *A. indica* are even broader, ranging from toothbrushes made of *A. indica* twigs to curries flavoured with *A. indica* flowers (69). Further, the historical and cultural use of limonoids is not confined to India. Toosendanin, a limonoid isolated from *M. toosendan*, has a long history of use in Chinese traditional medicine

and has been developed into a commercial treatment for parasitic worms in China (70). Some of these traditional uses are beginning to receive attention for use in modern western medicine. For instance, in Korea a clinical trial was performed based on the ring-intact limonoid 29-deacetylsendanin (often misnamed 28-deactylsendanin) which was identified as a dose-dependent inhibitor of herpes simplex virus-1 replication (IC50 of 1.46 mg/ml) (71).

Over recent years there has been increased interest in the *seco*-C-ring limonoid nimbolide (Figure 1.3.2) as a novel cancer treatment. Nimbolide has reported activity on 13 different cancer cell lines with an average IC50 of 3.8 μM (72) and lowest IC50 of 0.20 μM on Waldenstrom Macroglobulinemia cells (B lymphocytes) (72). Further, nimbolide has been shown to be active in certain animal models. Nimbolide (10 μg/kg body weight) was capable of reducing squamous cell carcinoma incidence to zero in 7,12-dimethylbenz[a]anthracene (DMBA)-induced hamster buccal pouch carcinogenesis experiments (73, 74).
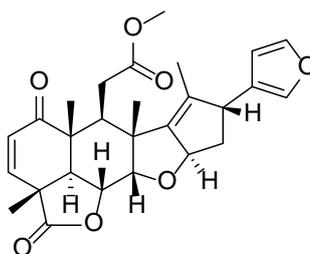
Figure 1.3.2: **Structure of nimbolide.**
Isolated from the Meliaceae species *A. indica* and *A. excelsa* (5).

Azadirachtin has also been cited as a potential anti-cancer agent, but has been surpassed by the superior activity of nimbolide (75–77). A plethora of alternative pharmaceutical uses have also been linked to azadirachtin in the literature, although many lack further supporting evidence (16). However, evidence is emerging for azadirachtin as a potential novel therapeutic agent for type 2 diabetes mellitus management. Azadirachtin has been shown to induce hypoglycemia in diabetic animal models (rats). Molecular investigations demonstrated that azadirachtin was able to inhibit and disaggregate amyrin fibrils (the hallmark of diabetic β-cell dysfunction) both *in vitro* and in pancreatic β-cell lines (INS-1E) (78).

In pharmaceutical-related bioassays, the citrus limonoids have often been observed to be less potent than Meliaceae limonoids. For instance, a similar DMBA-induced hamster buccal pouch study found that limonin was capable of reducing (60%), but not completely clearing tumour incidence (79). Additionally, there are historical reports that citrus limonoids have antioxidant properties, but these have subsequently been disproven (13). Despite this, investigations of the uses of citrus

limonoids as pharmaceuticals has continued, partially due to their existing occurrence in human diets and their availability as by-products of the juice industry (13). Mammalian studies have shown that orange juice consumption could reduce low-density lipoprotein cholesterol by up to 43% (80). Subsequent studies with human liver cell lines (HepG2) demonstrated that treatment with limonin could reduce levels of apolipoprotein B (one of the main constituent of low-density lipoprotein) by >70% (81). Limonin 17-O-β-D-glucopyranoside (300-400 μg/ml), rather than limonin itself, is the major limonoid constituent of orange juice and bioavailability studies have demonstrated that limonin is detectable in the blood after ingestion of limonin 17-O-β-D-glucopyranoside (13, 82), suggesting that limonin 17-O-β-D-glucopyranoside is converted to the active limonin form after consumption. Such studies were the basis of a 2013 clinical trial (NCT02011789) to investigate the effects of citrus limonoids on individuals with high-cholesterol. However the results of this trial have not yet been made available.

1.3.3   Molecular mechanism of action of limonoids

Despite the biological activity and physiological effects of limonoids being heavily investigated, their molecular mode of action remains unclear, particularity in terms of their toxicity. The antifeedant effect of limonoids such as azadirachtin is thought, as for other insect deterrents, to occur primarily though contact chemoreception by dedicated receptors on insect mouthparts. Early investigations of chemoreception of azadirachtin have demonstrated this, but reported high levels of variation both within and between insect species (15). For instance, a number of studies have shown that individual insects previously exposed to azadirachtin displayed a decreased chemoreception reaction and increased feeding when re-exposed, suggesting a degree of plasticity in the insect feeding response (15).

The mechanism of toxicity of azadirachtin to insects, both in terms of effects on growth and reproduction, is less clear. Early suggestions proposed that azadirachtin may interfere with insect moulting, because it has a similar structure to insect moulting hormones (ecdysones) (83). However, the correct structural assignment of azadirachtin proved this idea to be false, as the structures are distinct (16). An alternate suggestion was that azadirachtin could target and interfere with the final enzyme in the pathway for biosynthesis of active moulting hormones, ecdysone-20-monooxygenase (E-20-M). A concentration-dependent decrease in E-20-M activity was observed when tissue from three insect species, *Drosophila melanogaster* (fruit fly), *Aedes aegypti* (yellow fever mosquito) and *Manduca sexta* (tobacco hornworm), were exposed to azadirachtin (84). However, this was measured by assessing the incorporation of a labelled E-20-M substrate

([^{3}H]ecdysone) into 20-hydroxyecdysone, rather than by azadirachtin binding to E-20-M. Therefore the observed effect on E-20-M may not represent the direct target of azadirachtin.

The azadirachtin-sensitivity of insect cell lines, such as those of *Spodoptera frugiperda* (fall armyworm) (85), has aided further mechanistic studies (86–91). Such studies reached a consensus that ecdysone-pathway effects may be indirect secondary effects. The primary mode of action of azadirachtin is yet to be confirmed. However, there are two consistent observations of high-order effects of azadirachtin on insect cell lines that that may provide insights into modes of action: effects on microtubule formation (88, 89, 91); and on induction of cell cycle arrest and apoptosis (86, 87, 92). Whether these represent two independent modes of action or are intertwined remains unknown, as do the cellular targets of azadirachtin binding. The only reported binding of azadirachtin was in a *D. melanogaster* cell line (Kc167) where an azadirachtin analogue was shown to bind to a 591 kDa protein complex. The only protein to be identified from this complex was heat shock protein 60 (hsp60) subunit A (90), however how this binding links to predictions of the primary mode of action of azadirachtin remains unclear.

Further work has been undertaken on the effect of azadirachtin on cell cycle arrest and apoptosis. Studies in *S. litura* cell lines (SL-1) have linked azadirachtin induced-apoptosis to high levels of autophagy induced through the the insulin receptor and PI3K/AKT/TOR pathways (91). These studies have found azadirachtin treatment to be comparable to starvation activation of these pathways (91). Autophagy-induced apoptosis has previously been observed in insect cell lines. For instance, autophagy was activated in *Bobmyx mori* (silkworm) cell lines (Bm-12) by both starvation and 20-hydroxyecdysone treatment, the later being required for metamorphosis-related apoptosis within certain insect species (93). These findings have now been transferred to whole organism models, with azadirachtin-activated autophagy and apoptosis demonstrated in *S. litura in vivo* (92).

Despite this progress, and a number of *in silico* binding studies, knowledge of a target molecule for azadirachtin is still lacking. The only reported limonoid binding is for nimbolide, acting in its anti-cancer capacity. Nimbolide is known to impair cell proliferation in triple negative breast cancer cell lines (231MFP and HCC38). An activity-based protein profiling chemoproteomic platform was used to demonstrate that nimbolide binds covalently to RN114, an E3 ubiquitin-protein ligase (94). This binding has been pinpointed to a specific cysteine residue of RN114 which is involved in substrate recognition. The ability of nimbolide to bind to this residue is thought to be due to the cyclic enone in its A-ring, which is not seen in azadirachtin. Binding of nimbolide impairs the ability of RN114 to ubiquitinate and degrade the tumour suppressor CDKN1A (p21) (94). This leads to the stabilisation of p21 in cell

lines, with associated cell-cycle arrest and apoptosis under certain conditions (95). Although the discovery of RN114 as a target for nimbolide binding is a major advance in understanding the mechanisms of action of limonoids, it does not fully explain all of the observed activities of nimbolide and therefore multiple other sites are also suggested (94).

1.4   Availability of limonoids

The use of limonoids, both traditional and commercial, is currently entirely reliant on extraction from plant material. The by-products of the juice industry represent a potential source of limonoids as there is estimated to be ~0.5 kg of limonoid glucosides per ton of by-products (13). However, the citrus limonoids contained within these by-products are less potent, both in terms of anti-insect and pharmaceutical properties, than the *seco*-C-ring limonoids of the Meliaceae family.

Within the Meliaceae family there are currently no by-product sources of limonoids. Although azadirachtin-based insecticides are in commercial use, they are sourced by extraction from *A. indica* seeds, the cost of which is not in line with other natural product crop-protection agents extracted this way. In 2009, a kilogram of *A. indica* seed cost 2 USD compared to 1 USD for the same mass of dried *Chrysanthemum cinerariifolium* flowers from which the commercially used anti-insect compound pyrethrum is extracted. Further, there is 4-fold more pyrethrum in *C. cinerariifolium* flowers than there is azadirachtin in *A. indica* seeds (16). The relatively high cost of *A. indica* seeds has been cited as one of the major factors preventing wide-spread use of *A. indica*-based insecticides (16).

An alternative to extraction from plant material is chemical synthesis, which for some simple natural products has provided an effective alternative supply (3). Fragmented limonoids, such as fraxinellone (Figure 1.4.1), were successfully synthesised in the 1980s (96). However, the complex polycyclic nature of most *seco*-ring limonoids has largely prevented their synthesis (19). Despite this, the total chemical synthesis of azadirachtin has been achieved (97–99). Azadirachtin has 16 contiguous stereogenic centres and therefore completion of this synthesis took 22 years. In addition, total synthesis of limonin has also been accomplished (100). Although these chemical syntheses represent impressive achievements, the yields and processes are currently not commercially viable (Table 1.4.1). Simpler ring-intact limonoids such as azadiradione have been chemically synthesised at higher yields. However, such compounds are not priorities for commercial production due to their lower bioactivity and further conversion to bioactive *seco*-C-ring structures would not be trivial by chemical or enzymatic methods.

Table 1.4.1: **Examples of commercially relevant limonoid total chemical syntheses.**

Limonoid	Class	Precursor	Yield	Steps	Reference
azadirachtin	*seco*-C	multiple	0.00015%	71	(97–99)
limonin	*seco*-A,D	geraniol	0.0017%	35	(100)
azadiradione	intact	farnesol	1.35%	12	(101)

More recently the total synthesis of the *seco*-B,D-ring limonoid cipadonoid B (Figure 1.4.1) was achieved (102), with subsequent synthesis of the related limonoids khayasin, proceranolide and mexicanolide (103). The total syntheses of the *seco*-A,B,D-ring limonoids perforanoid A (Figure 1.4.1) and 10-*epi*-perforanoid A have also been achieved (104) (Figure 1.4.1). These total syntheses were more efficient than those of azadirachtin and limonin. However, these compounds are markedly different to the *seco*-C-ring limonoids that are the priority for crop protection and pharmaceutical uses.

fraxinellone   cipadonoid B   perforanoid A
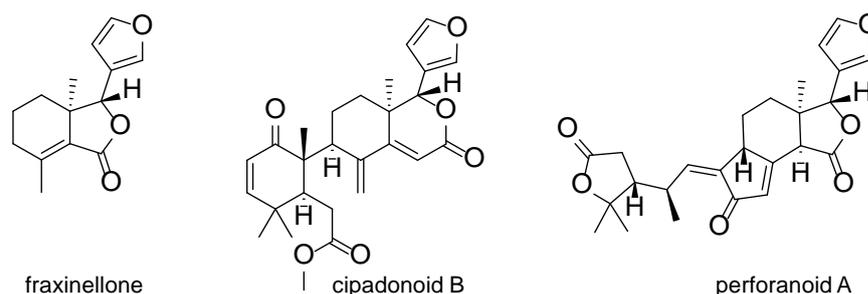
Figure 1.4.1: **Structure of fraxinellone, cipadonoid B and perforanoid A.**
Examples of limonoid structures for which total chemical synthesis has been achieved. The fragmented limonoid fraxinellone has been isolated from a number of Meliaceae and Rutaceae species. Cipadonoid B has been isolated from the Meliaceae species *Cipadessa cinerascen*, and perforanoid A from the Rutaceae species *Harrisonia perforata*.

In addition to total chemical syntheses, there have been a number of reported modifications to limonoid scaffolds by chemical and biological methods (5, 6, 19). Examples of biological transformations include microbial modification. This has been demonstrated by the fermentation of limonoids with a fungus capable of biotransformation of steroidal scaffolds *Cunninghamella echinulata* (105), which resulted in the addition of a C12 hydroxy group to simple ring-intact limonoid scaffolds (e.g. azadiradione) (106) and modifications to the furan ring of salannin (107). Currently, these modifications do not appear to be specific enough to aid the synthesis of commercially relevant limonoids.

1.5   Metabolic engineering of limonoids

Given the potent bioactivity of limonoids and their suitability for crop protection uses (16), it could be stated that limonoids are not being utilised to their full potential. The cost of extraction from plant material and challenge of chemical syntheses indicate that a novel approach to the production of limonoids is required in order for these compounds to be used more extensively. Metabolic engineering could represent such an alternative approach.

Metabolic engineering refers broadly to the manipulation or transfer of biosynthetic pathways. Metabolic engineering has led to the production of valuable natural products in alternative, heterologous hosts. For instance, pathways for fuels have been engineered into bacteria (108), and for opioids into yeast (109). Although production of commercially relevant yields was an initial challenge in this field, with time this is being resolved. For instance, steviol (a precursor to the plant-derived sweeteners, steviol glycosides) is now being fermented in baker's yeast in commercially viable quantities (110). Further, metabolic engineering within crop species has the potential to improve both the productivity and value of crops. For instance, the engineering of β-carotene production into edible parts of rice (Golden Rice) has the potential to combat vitamin A deficiency, a major cause of blindness in developing countries (111). Further to direct commercial uses, metabolic engineering can also aid scientific investigations. For example, a suite of differentially oxidised β-amyrin triterpene analogues were produced in milligram quantities by combinatorial biosynthesis in *Nicotiana benthamiana*, which enabled their high-throughput testing for biological activity (112).

Therefore, the potential opportunities of metabolic engineering for limonoid production are three-fold. Heterologous hosts could be used to produce higher titres of high-value limonoids, such as azadirachtin, for direct use by industry. Suites of limonoid analogues could be produced by combinatorial biosynthesis to enable rapid and thorough biological activity assessments and determination of structure-activity relationships. Finally, the pathway for production of anti-insect *seco*-C-ring limonoids could be engineered into crop plants to reduce insect herbivory and increase yields.

The current limitation of such approaches is the absence of any characterised limonoid biosynthetic enzymes, with the exception of two enzymes involved in the de-bittering of citrus limonoids (4, 13). Metabolic engineering frequently utilises enzymes from different species. For instance, the engineering of opioids into yeast required enzymes from the natural opioid-producing poppy, three additional plants, two bacteria and a mammal (109). However, in the case of limonoids, an

understanding of the route of biosynthesis is first required in order to enable such combinatorial biosynthesis for limonoid engineering.

Although there has been remarkably limited investigation of the biosynthesis of limonoids (16), over recent decades great progress has been made in the understanding of other triterpene biosynthetic pathways (11). Triterpene biosynthesis is initiated by the cyclisation of 2,3-oxidosqualene (a product of the mevalonate pathway). This step is performed by dedicated enzymes known as oxidosqualene cyclases (OSCs). Protonation of the C2-C3 epoxide of 2,3-oxidosqualene by OSCs produces a carbocation which, based on proximity specified by the structure of the OSC active site, initiates a series of electrophilic attacks that lead to ring formation (11, 113, 114). This reaction can lead to the formation of hundreds of different cyclic triterpene scaffolds (21, 114). Only two of these triterpene scaffolds (lanosterol and cycloartenol) are associated with primary metabolism (11) and the remainder are all postulated to be involved in secondary metabolism and therefore represent the first layer of diversity within triterpene biosynthesis. The OSC(s) and triterpene scaffold(s) which initiate limonoid biosynthesis have not yet been identified, although both tirucallol- and euphol- type triterpene scaffolds are speculated to be involved (5).

Cyclisation of triterpenes is commonly followed by oxidation, a process associated with increased diversity in many plant natural product pathways (1). Oxidation of triterpenes is most commonly performed by cytochrome P450 enzymes (CYPs), a class of monooxygenases which utilise O_2 to incorporate an oxygen atom into the target substrate (115). This reaction is catalysed by iron in the haem cofactor of CYPs and requires electrons from NADPH, which are donated via a partner reductase (116). CYPs have evolved to be highly diverse within plants. There are 244 CYP genes identified within *Arabidopsis thaliana* alone (117). In 2017 there were 80 known triterpene acting CYPs capable of performing diverse types of oxidations (118) and more triterpene acting CYPs have since been discovered (e.g. AsCYP72A475 from *Avena strigosa* (119) and a number of CYPs from the Brassicaeae (120, 121)).

Beyond oxidation, diversity is added to triterpene scaffolds by tailoring enzymes such as acyl and glycosyl transferases (11). Currently, the only functionally characterised limonoid acting enzyme is a glycosyltransferase from citrus species: limonoid UDP-D-glucosyltransferase (39, 40). Further, within the Meliaceae limonoids addition of acyl groups is common (5, 6) and therefore acyltransferases are likely to be heavily involved in limonoid biosynthesis. Although the predictions of these transferase functions appear obvious, there are key aspects of limonoid biosynthesis, such as furan ring formation, for which there is no knowledge of which type of enzymes are involved. Therefore studies of limonoid biosynthesis will inevitably reveal novel enzymes and biosynthetic processes.

1.6   PhD overview

The limonoids are a fascinatingly diverse class of plant natural products with both current and prospective human uses. Use of limonoids has been limited due to reliance on extraction from plant material, the cost of which is uncompetitive in comparison to other plant natural product based insecticides. Metabolic engineering approaches could provide a new, scalable and sustainable source of limonoids in the future. At present, a lack of understanding of limonoid biosynthesis has prevented this. Therefore the overarching aim of this project is to characterise limonoid biosynthesis, to enable future metabolic engineering of limonoids for crop protection. Initial gene discovery investigations drew on existing sequencing data to mine for OSCs and CYPs, which led to the discovery of three protolimonoid biosynthetic genes that together could produce melianol (Chapter 2). However, the available sequencing data within the Meliaceae family was limited and so to improve this access to limonoid producing Meliaceae material was required (Chapter 3). Once *M. azedarach* plants had been secured they were profiled both for limonoid content and expression of protolimonoid biosynthetic genes, which revealed both tissue and individual plant specific differences (Chapter 4). This knowledge was used to design RNA-Seq experiments for the discovery of downstream pathway candidates. Further, this material was used to generate a pseudochromosome-level genome sequence for *M. azedarach* (accession JPN11) (Chapter 5), the first to be generated for the Meliaceae family. These new *M. azedarach* resources were mined to identify further limonoid biosynthetic genes of diverse enzyme classes (Chapter 6). Finally, in parallel with the biosynthetic investigations of limonoids, an assay was developed to assess the anti-insect activity of transiently expressed limonoid biosynthetic genes within *N. benthamiana* (Chapter 7).

2

Characterisation of initial protolimonoid biosynthesis

Synopsis

Upon commencing this project the initiation of limonoid biosynthesis remained speculative and there was an absence of characterised enzymes. In this chapter existing sequencing resources within the Meliaceae and Rutaceae have been mined in order to address this. Three OSCs (AiOSC1, MaOSC1 and CsOSC1) from three limonoid-producing species (*A. indica, M. azedarach* and *C. sinensis*) have been characterised as initiating limonoid biosynthesis by converting 2,3-oxidosqualene to tirucalla-7,24-dien-3β-ol. Two CYPs (MaCYP71CD2 and MaCYP71BQ5) from *M. azedarach* have also been characterised. Heterologous expression of these two CYPs in *N. benthamiana* has demonstrated that together they perform the three oxidations required for the spontaneous production of the protolimonoid melianol from a tirucalla-7,24-dien-3β-ol scaffold. The identification of homologs of these genes from *C. sinensis* and *A. indica* suggests that melianol biosynthesis may represents a shared start in limonoid biosynthesis.

2.1   Introduction

Although at the commencement of this project limonoid biosynthesis remained wholly uncharacterised, the understanding of the overarching process of the biosynthesis of triterpenes is well-understood (11). Labelling experiments have provided evidence that the substrate for limonoid biosynthesis, as for all other triterpenes, is 2,3-oxidosqualene, a product of the mevalonate (MVA) pathway. Incubation of ^{14}C-mevalonate with *A. indica* seed kernel extracts resulted in its incorporation into azadirachtin, salannin and nimbin (123). A separate study, in which ^{13}C-glucose was fed to *A. indica* cell cultures, revealed its incorporation into an additional six limonoids and provided further evidence that proved that the MVA pathway provides precursors to limonoid biosynthesis, by the use of pathway inhibitors (124).

A number of studies have utilised transcriptomic data to identify candidate MVA pathway genes from the limonoid-producing Meliaceae species *A. indica* and *M. azedarach* (55, 125–128). However, thus far the only genes characterised from these species are a functional farnesol diphosphate synthase (AiFDS), a squalene synthase (AiSQS) (55) and two 3-hydroxy-3-methylglutaryl coenzymeA reductases (AiHMGR1, AiHMGR2) (129).

Enzyme characterisation beyond the primary metabolism of the MVA pathway has not yet been achieved in limonoid-producing species. The initiation of triterpene biosynthesis involves the cyclisation of 2,3-oxidosqualene (Figure 2.1.1) to one of hundreds of different triterpene scaffolds (130), which is performed by OSCs. The lack of a characterised limonoid biosynthetic OSC means that the triterpene scaffold implicated in limonoid biosynthesis remains unconfirmed. Based on the tetracyclic structure of simple ring intact limonoids it has been proposed that the scaffold is likely to be either a euphane- (20*R*) or tirucallane- (20*S*) type scaffold (Figure 2.1.1).

The formation of the furan ring eliminates the stereochemistry of the C20 position. Therefore, based on mature limonoid structures alone, it is not possible to determine the stereochemistry of the precursor. However, immature limonoid structures (protolimonoids) are able to help discriminate this. For instance, the C20 stereochemistry of the protolimonoid melianol has been assigned as the *'S'* configuration (134), which would suggest a tirucallane rather than a euphane type scaffold (Figure 2.1.1). Further, although the C14 position of an alkene within mature limonoid scaffolds is not consistent with tirucallane or tirucalla-7,24-dien-3β-ol **(1)**, the C7 alkene seen in protolimonoids suggests that tirucalla-7,24-dien-3β-ol is the most likely limonoid initiating scaffold **(1)** (Figure 2.1.1).

Two OSCs have been identified from *A. indica* (55). However their functional
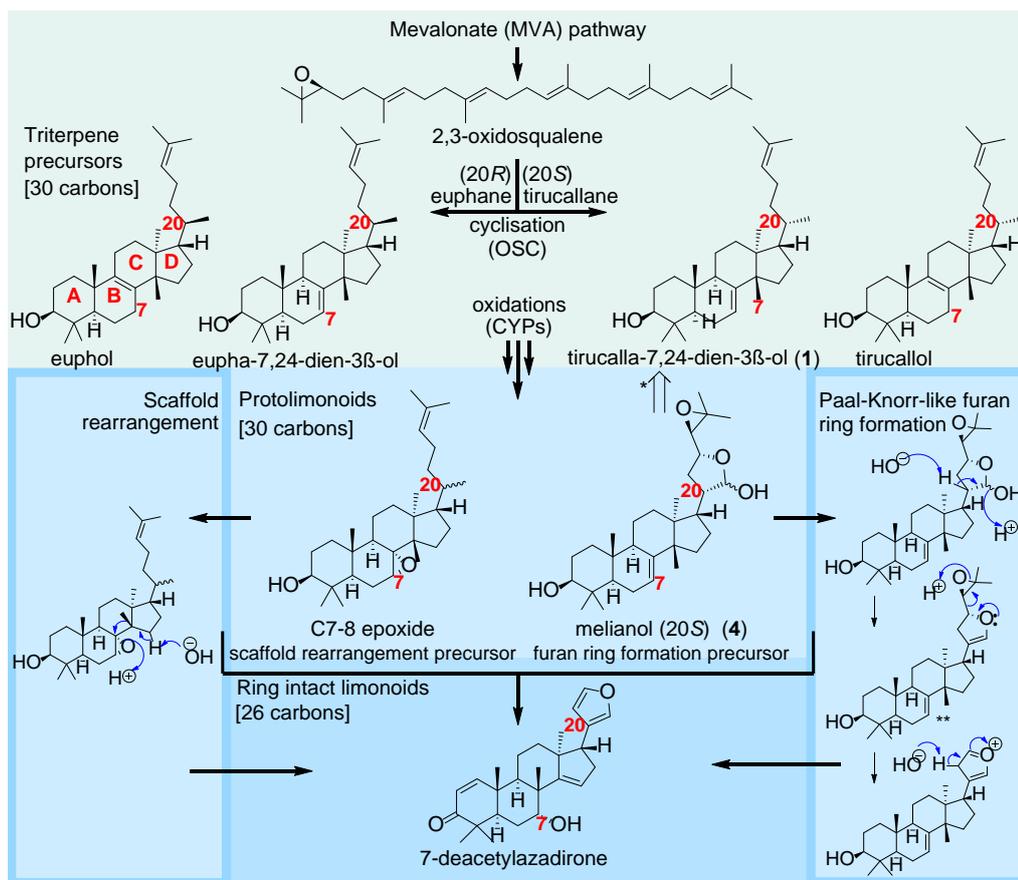
Figure 2.1.1: **Predicted initiation of protolimonoid biosynthesis.**
The triterpene precursor 2,3-oxidosqualene is proposed to be cyclised to an unconfirmed tetracyclic triterpene scaffold (predicted to be either a euphane (20*R*) or tirucallane (20*S*) type). The most likely triterpene precursor is tirucalla-7,24-dien-3β-ol **(1)**, as indicated by the retrosynthetic arrow (*). Subsequent biosynthesis of limonoids, such as the ring intact 7-deacetylazadirone (isolated from both Meliaceae (28) and Rutaceae (18) species), from a triterpene scaffold is predicted to occur through protolimonoid structures such as melianol **(4)** and require two major biosynthetic steps: scaffold rearrangements (predicted to be initated by oxidation), and furan ring formation accompanied by loss of four carbons (predicted to proceed by a Paal-Knorr-like mechanism (131, 132)). Isolation of nimbocinone, a feasible degradation product of this route (**), from *A. indica* (133) supports the conversion of protolimonoids to ring-intact limonoids through this mechanism. Figure adapted from authors previous publication (122).

characterisation has not yet been reported. A transcriptomic study of *Citrus grandis* (135) also identified an OSC candidate. Viral induced silencing of this OSC was reported to reduce levels of nomilin and limonin, however the method of quantification was not clear and the cyclisation product of this OSC remains unknown.

The conversion of a triterpene scaffold to a ring-intact limonoid, (e.g.

7-deacetylazadirone) is thought to occur through protolimonoid intermediates such as melianol **(4)** (Figure 2.1.1). Two major biosynthetic steps are implicated in this process: scaffold rearrangement, which is proposed to be initiated by epoxidation of the C7 alkene; and the cyclisation of the side chain to a mature furan ring with associated loss of four carbons. The isolation of hemiacetal-containing protolimonoids, such as melianol **(4)**, suggests that a Paal-Knorr-like route (131, 132) could be involved in furan ring formation (Figure 2.1.1). The diversity of protolimonoid structures isolated from Meliaceae and Rutaceae species makes it difficult to propose which of these arrangements may occur first (5, 136). A diverse array of oxidations are performed on triterpene scaffolds by CYPs (118) and therefore oxidation by CYPs could potentially initiate both of these rearrangements. A number of CYPs have been identified from transcriptomic studies of *A. indica* and *C. grandis* and suggested to be involved in limonoid biosynthesis (55, 125, 128, 135, 137, 138). However, none of these CYPs have been functionally characterised.

Based on these predictions (Figure 2.1.1), OSC and CYP enzyme classes are the primary targets for characterising the early steps in limonoid biosynthesis. Ample high-quality sequencing resources are available for the Citrus family (Citrus Genome Database (139)). However, it is not known whether the early steps in limonoid biosynthesis occur by a mechanism that is shared between the Rutaceae and Meliaceae families. Within the Meliaceae, availability of sequencing information is more limited. Although two draft genome sequences have been generated for *A. indica*, one of them was not of sufficient quality and the other was not available at the time that this work was carried out. Therefore, available RNA-Seq datasets for the Meliaceae (Table 2.1.1) (126, 128, 140) were used instead.

Table 2.1.1: **Summary of previous RNA-Seq experiments within the Meliaceae.**

Species	Tissues and replicates (n)	SRA ID	Reference
A. indica	5 (n=1)	SRP013453	(126, 140)
A. indica	13 (n=1)	SRP052002	(127)
A. indica	3 (n=1)	SRP061725	(55)
A. indica	2 (n=1)	NA	(125)
A. indica	2 (n=1)	SPR4299636	(141)
A. indica	1 (n=3)	SRP070635	(128)
M. azedarach	1 (n=3)	SRP070684	(128)
T. sinensis	1 (n=1)	SRP011871	NA

There have been six separate transcriptomic investigations of *A. indica* (55, 125–128, 140, 141), one of *M. azedarach* (128) and one of *Toona sinensis*. Although the assembled transcriptomes from these studies had not been released, the raw RNA-Seq reads for most of the studies were available on the NCBI's Sequence

Read Archive (SRA) (Table 2.1.1). In addition, although not available at the beginning of this project, the OneKP project have released assembled and annotated transcriptomes of *A. indica* (ID: UVDC) and *M. azedarach* (ID: CAIX, VCCF) leaf tissue (142, 143).

2.1.1   Aims

The availability of multiple RNA-Seq resources created an opportunity for the *de novo* assembly of transcriptomes from a number of Meliaceae species. Here, these transcriptomes, were used to mine for candidate limonoid biosynthetic genes. Based on knowledge of triterpene biosynthesis and predictions of limonoid biosynthetic pathways the enzyme classes prioritised for investigation were OSCs and CYPs. Candidates were then characterised by expression in either Yeast or *N. benthamiana*. Therefore the aims of this chapter are as follows:

1. Assembly of Meliaceae transcriptomes from existing RNA-Seq data.

2. Identification of candidate OSCs and characterisation by functional expression in Yeast.

3. Identification of candidate CYPs and characterisation by transient expression in *N. benthamiana*.

2.2   Results and Discussion

2.2.1   Assembly of Meliaceae transcriptomes from existing RNA-Seq data

Trinity *de novo* assembler (144) was used to assemble a number of transcriptomes based on existing RNA-Seq data within the Meliaceae family (Table 2.2.1). The *A. indica* transcriptome (Ai1) (126, 140) includes RNA-Seq reads from five tissues, enabling differential expression analysis to be performed. In contrast, the lack of different tissue types for the available RNA-Seq data from *M. azedarach* (Ma1) (128) prevented this. However, the inclusion of three biological repeats resulted in this transcriptome having the greatest coverage (110 Mbp) of those assembled (Table 2.2.1).

Table 2.2.1: **Details of Meliaceae assembled transcriptomes**

ID	Species	Transcripts	N50	Total bases (bp)	Reference
Ai1	*A. indica*	73,769	1,171	53,816,418	(126, 140)
Ai2	*A. indica*	70,275	1,526	62,521,671	(55)
Ma1	*M. azedarach*	154,767	1,226	110,216,786	(128)

Each transcriptome has been assigned an ID based on species and original data set: *A. indica* (Ai1, Ai2), *M. azedarach* (Ma1). Basic statistics from Trinity-generated (144) assemblies are included: total number of transcripts per assembly, N50 and total number of bases assembled (bp). References to the original datasets are provided. Specific BioSample IDs of the individuals used and SRA IDs for each tissue are available (Appendix B.1). Table adapted from the author's previous publication (122).

2.2.2   Identification of candidate OSCs and characterisation by functional expression in Yeast

The assembled transcriptomes (Ai1, Ai2 and Ma1) and also a *C. sinensis* protein annotation (GCF_000317415, (145)) were mined to identify candidate OSCs using BLAST. The protein sequences of functionally characterised OSCs were used as query sequences (11) and the results were filtered dependent on having the correct length and motif for a putative triterpene synthase OSCs.

A phylogenetic tree was constructed to compare the OSC candidates to those with known functions (Figure 2.2.1). From *C. sinensis* five phylogenetically distinct OSC candidates were identified, however only two unique candidate genes were identified from each of the Meliaceae species investigated (*A. indica* and *M. azedarach*). A number of OSC candidates were discounted from further analysis because they grouped with either the lanosterol or cycloartenol synthase clades: and so were likely to be involved in primary sterol biosynthesis (11). This resulted in one distinct candidate from each of *A. indica* (AiOSC1) and *M. azedarach* (MaOSC1), and three from *C. sinensis* (CsOSC1, CsOSC2 and CsOSC3). These were selected for further analysis (Figure 2.2.1, Appendix B.3).

One candidate from each of these three species (AiOSC1, MaOSC1 and CsOSC1) fell within a distinct subclade (subclade one). MaOSC1 and AiOSC1 share 97% amino acid sequence similarity with each other, and 86% with CsOSC1 (despite its occurrence within a separate family) (Appendix B.2). The shared nature of limonoid biosynthesis between these families implicates these as the strongest OSC candidates. The identification of only two candidate OSC genes in each of the two Meliaceae transcriptomes was unexpected. Most higher plants are known to contain more OSCs. For instance, the genome of *A. thaliana* has 13 OSC genes (148).

Figure 2.2.1: **Phylogenetic tree of candidate OSCs.**
Phylogenetic tree of candidate OSCs from *A. indica* (blue), *M. azedarach* (green) and *C. sinensis* (orange). Functionally characterised OSCs from other plant species (11) are included, with the two previously characterised tirucalla-7,24-dien-3β-ol synthases from *A. thaliana* (AtLUP5, AtPEN3) highlighted (yellow). Human and prokaryotic OSC sequences used as an outgroup are represented by the grey triangle. Candidate OSCs chosen for further analysis are indicated (circles) and their cloning described (Appendix B.3). The phylogenetic tree was constructed by FastTree V2.1.7 (146) and formatted using iTOL (147). Local support values from FastTree Shimodaira-Hasegawa (SH) test (between 0.6 and 1.0) are indicated at nodes. Scale bar, estimated number of amino acid substitutions per site. Figure adapted from the author's previous publication (122).

Therefore the identification of only two OSC genes in *A. indica* and *M. azedarach* is likely to be a limitation of utilising transcriptomic data. Candidate OSCs were initially characterised by heterologous expression in *Saccharomyces cerevisiae* strain GIL77. GC-MS analysis of triterpene extracts of yeast strains expressing candidate OSCs from subclade 1 (*AiOSC1, MaOSC1* and *CsOSC1*) revealed the production of a new peak **(1)** with the mass of 498.4, equivalent to a trimethylsilyl (TMS)-derivatised triterpene structure (Figure 2.2.2.A-B). Further, yeast expressing a known tirucalla-7,24-dien-3β-ol synthase from *A. thaliana* (AtLUP5) (149) produced a peak

with the same mass and retention time. Therefore AiOSC1, MaOSC1 and CsOSC1 were predicted to synthesise tirucalla-7,24-dien-3β-ol (Figure 2.2.2.C).
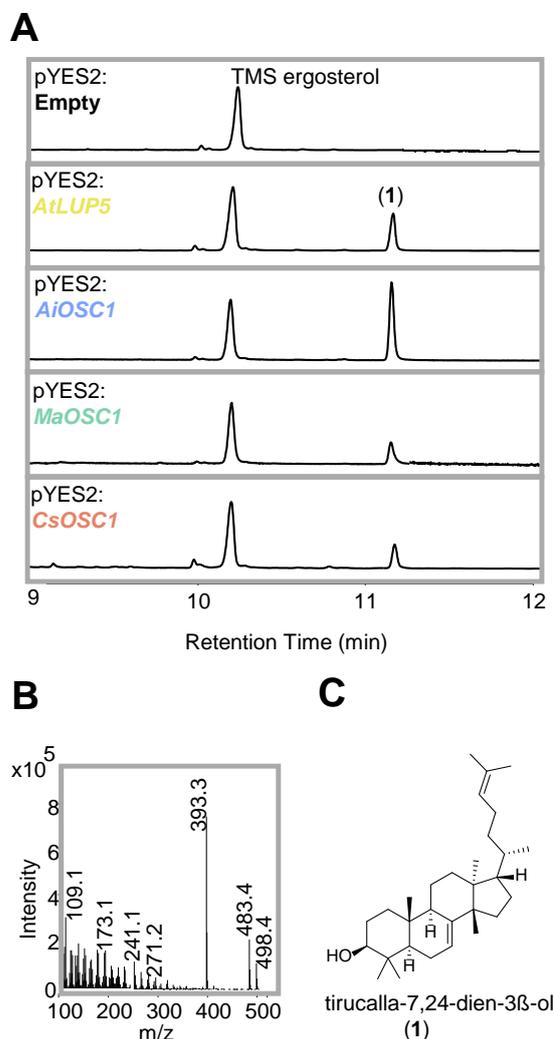
Figure 2.2.2: **Characterisation of candidate OSCs from subclade one.**
(A) GC-MS total ion chromatograms of derivatised extracts from yeast strains expressing candidate OSCs. Traces for the empty vector (pYES2) and strains expressing the candidates AiOSC1 (blue), MaOSC1 (green), CsOSC1 (orange) and the previously characterised AtLUP5 (yellow) are shown. (B) GC-MS mass spectra of TMS-tirucalla-7,24-dien-3β-ol **(1)**. (C) Confirmation of the structure of the cyclization product generated by AiOSC1 as tirucalla-7,24-dien-3β-ol **(1)** by NMR can be found in Appendix B.5 and functionally characterised OSCs are summarised in Appendix B.3. Figure adapted from the author's previous publication (122).

This prediction was confirmed by large-scale expression of *AiOSC1* in GIL77 to allow purification and structural characterisation by NMR (Appendix B.5). The two other OSC candidates from *C. sinensis*, CsOSC2 and CsOSC3, were characterised as a β-amyrin and a lupeol synthase, respectively (Appendix A.1, Appendix A.2). The three tirucalla-7,24-dien-3β-ol synthase OSCs characterised here are

phylogenetically distinct from the previously characterised AtLUP5 and AtPEN3 OSCs from *A. thaliana*, which are both capable of producing tirucalla-7,24-dien-3β-ol (148, 149). Further, the three OSC sequences previously identified in relation to limonoid biosynthesis (two from *A. indica* (55) and one from *C. grandia* (135)) also appear to be phylogenetically distinct from subclade one based on their reported phylogeny and closest BLAST hits.

The identification of homologous and functional tirucalla-7,24-dien-3β-ol producing OSCs from three diverse limonoid-producing species strongly suggests that this transformation represents the first step in limonoid biosynthesis. This is consistent with the prediction that tirucalla-7,24-dien-3β-ol is the most likely scaffold based on retro-synthetic analysis (Figure 2.1.1). However, this finding is in contrast to a previous feeding study in *A. indica*, which reported greater relative incorporation of ^{3}H-labelled euphol into the *seco*-C-ring limonoid nimbolide, compared to tirucallane precursors (150). However, this experiment suffered from several sources of experimental error, including inconsistent labelling of precursors and use of wet leaf weight in calculations, which may render the relative incorporation unreliable.

2.2.3 Identification of candidate CYPs and characterisation by transient expression in *N. benthamiana*

To identify candidate CYPs from available *M. azedarach* transcriptomic data (Ma1) (128) a BLAST search was performed using annotated CYP amino acid sequences from *A. thaliana* (151) as a query. The resultant 1672 hits were filtered to give 103 full-length candidates CYPs. As the *M. azedarach* (Ma1) transcriptome was generated from leaf only tissue it was not possible to compare expression levels of genes in different tissues. It was however possible to perform such an analysis using the *A. indica* (Ai1) transcriptome (126, 140) (Figure 2.2.3).

Differential expression analysis for *A. indica* was aided by a better knowledge of tissue-specific limonoid occurrence in comparison with *M. azedarach*. Traditional knowledge and use of *A. indica* indicates that the highest abundance of limonoids is in the seed kernel (extracts of which are used for crop protection) and in the developing fruit. Metabolic profiling has proved this by demonstrating that azadiradione-type ring-intact limonoids were most abundant in the pericarp of the developing fruit, whilst the more highly modified *seco*-C-ring limonoids showed a more varied spatial occurrence, azadirachtin levels being highest in the developed seed kernel, nimocinol in the leaves and salannin in the stem and bark (55).

Differential expression analysis was performed by mapping raw RNA-Seq reads to
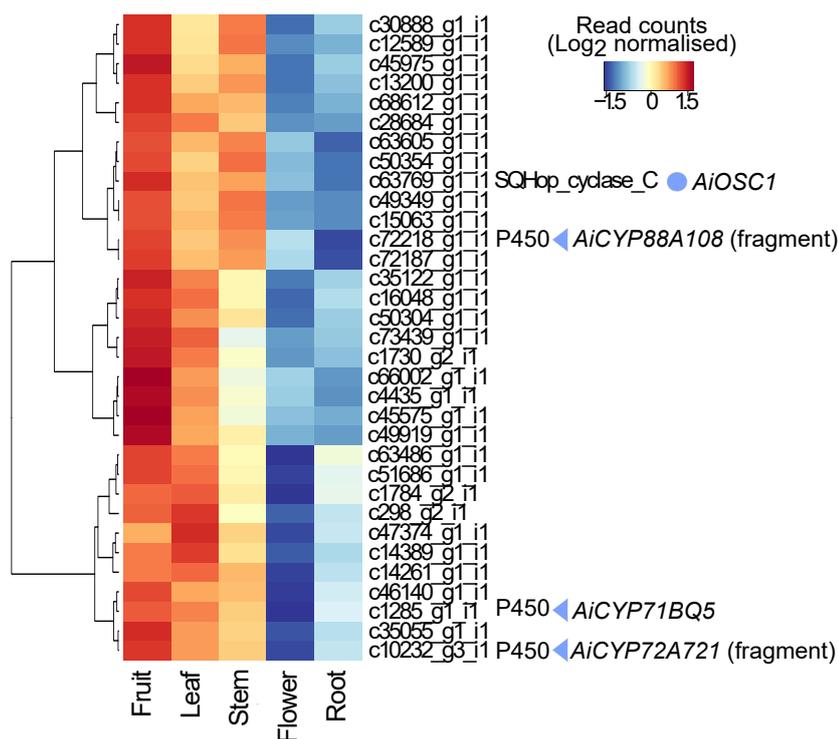
Figure 2.2.3: Hierarchical clustering of *A. indica* (Ai1) genes with *AiOSC1*.
A heatmap of a subset of differentially expressed (P-value < 0.05) genes with similar expression patterns to *AiOSC1* (blue circle). Expression profile across flower, root, fruit, stem and leaf tissues of *A. indica* are shown. Raw RNA-Seq reads (140, 152) were aligned to a Trinity-assembled transcriptome of the same dataset. Read counts were normalised to library size and $\log_2$-transformed. Values shown are scaled by row (gene) to emphasise differences across tissues. The Pfam descriptions for relevant predicted genes are indicated next to the contig number. Genes with no structural (Augustus) or functional (Pfam) annotations have not been included in this heatmap. The CYP candidates *AiCYP71BQ5, AiCYP72A721* and *AiCYP88A108* (blue triangles) are indicated, with the latter two being considered gene fragments (< 300 amino acids). Figure adapted from the author's previous publication (122).

the related *A. indica* transcriptome (Ai1) to identify a subset of genes differentially expressed across the available tissues (P-value < 0.05). Hierarchical clustering was used to extract a subset of differentially expressed genes with a similar expression pattern to the characterised tirucalla-7,24-dien-3β-ol synthase *AiOSC1*. Based on this analysis, *AiOSC1* and similarly expressed genes showed highest expression levels in the fruit (where simple ring-intact limonoids have been reported), followed by the leaf and stem (reported to contain high abundance of certain *seco*-C-ring limonoids) (Figure 2.2.3). Within this subset of differentially expressed genes were three CYP candidates (*AiCYP71BQ5, AiCYP72A721* and *AiCYP88A108*). However, because of the poor quality of this transcriptome (Ai1) these sequences were not full-length.

Full length homologs of *AiCYP71BQ5, AiCYP72A721* and *AiCYP88A108* were

identified from the 103 *M. azedarach* CYP sequences (*MaCYP71BQ5, MaCYP72A721* and *MaCYP88A108*). Phylogenetic analysis revealed that one of these, *MaCYP71BQ5*, fell within a discrete subclade of the CYP71 family (Figure 2.2.4). This subclade contained seven candidate sequences from *M. azedarach* but no sequences from the other non-limonoid-producing species used as comparisons (*A. thaliana* (151) and *Cucumis sativus* (153)). Therefore the homologs of the three genes identified as differentially expressed in *A. indica* (*MaCYP88A108, MaCYP71BQ5* and *MaCY72A720* (Figure 2.2.3)), along with the additional six genes that occurred within the distinct subclade (*MaCYP71CD2, MaCYP71BE124, MaCYP71D557, MaCYP71BQ6, MaCYP71BE123* and *MaCYP71BE125* (Figure 2.2.4)) were chosen as candidates for further analysis. A summary of candidate CYPs and their homologs from *A. indica* and *C. sinensis* is provided in Appendix B.4.

To investigate the function of these candidate CYPs, their coding sequences were cloned into pEAQ-HT-DEST1 expression vectors, along with *AiOSC1. Agrobacterium tumefaciens*-mediated transient expression in *N. benthamiana* was then used to determine their activity. GC-MS analysis of triterpene extracts from *N. benthamiana* leaves transiently expressing *AiOSC1* showed a clear peak **(1)** (Figure 2.2.5.A), which was identifiable as tirucalla-7,24-dien-3β-ol using previously purified standards. Thus demonstrating a consistent function of AiOSC1 in *N. benthamiana* and Yeast (Figure 2.2.2).

For the majority of candidate CYPs there was no detectable activity when they were co-expressed with *AiOSC1* (Appendix A.3). However a new peak was observed for two of the CYP candidates *MaCYP71CD2* and *MaCYP71BQ5*.

When *MaCYP71CD2* was co-expressed with *AiOSC1*, tirucalla-7,24-dien-3β-ol **(1)** was completely consumed and a new peak **(2)** was detected (Figure 2.2.5). This new peak **(2)** was determined by GC-MS to have a mass of 602.6 and LC-MS revealed a peak of 481.365 (Appendix A.4, Appendix A.5). These masses are consistent with the TMS derivative and sodium adduct, respectively, of tirucalla-7,24-dien-3β-ol **(1)** with the addition of one hydroxy group and one ketone (or the conversion of an alkene to a epoxide).

When *MaCYP71BQ5* was co-expressed with *AiOSC1*, tirucalla-7,24-dien-3β-ol **(1)** was partially consumed and a new peak **(3)** was detected (Figure 2.2.5.A). This new product was detectable only by GC-MS and had a mass of 498.4 (Appendix A.4) consistent with a TMS derivative of hydroxylated tirucalla-7,24-dien-3β-ol.

When all three genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*) were expressed together in *N. benthamiana* a new peak was formed **(4)** (Figure 2.2.5.B), which was only detectable by LC-MS. The absence of precursor peaks **(1-3)** suggests that
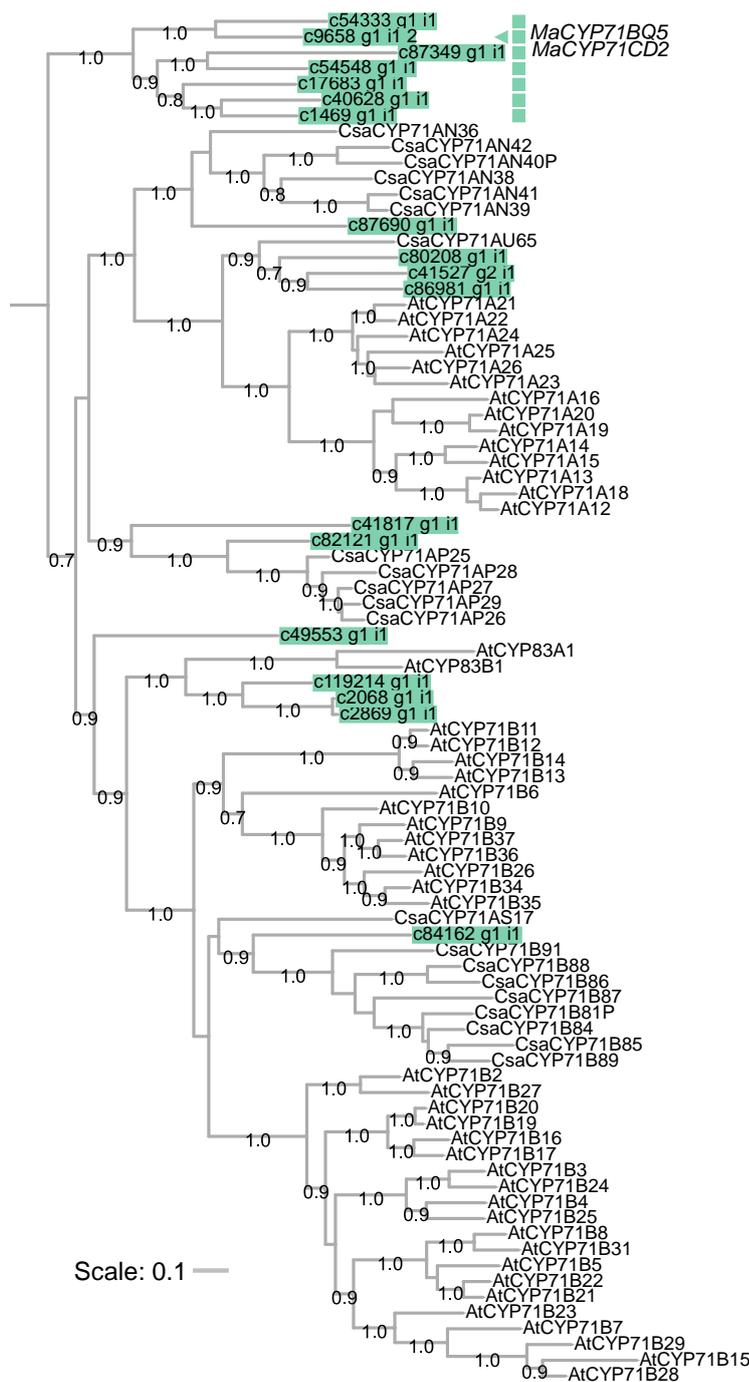
Figure 2.2.4: Phylogenetic tree of candidate CYPs from *M. azedarach* (Ma1).
A subset showing the CYP71 family of a larger phylogenetic tree (Appendix B.6).
Candidate CYPs from *M. azedarach* (green) and CYPs from *A. thaliana* (151) and
C. sativus (153) (black) are included. Candidates chosen for further analysis are
indicated based on their identification by: differential expression (triangle) or
phylogeny (square). The phylogenetic tree was constructed by FastTree V2.1.7 (146)
and formatted using iTOL (147). Local support values from FastTree Shimodaira-
Hasegawa (SH) test (between 0.6 and 1.0) are indicated at nodes. Scale bar, estimated
number of amino acid substitutions per site. Figure adapted from the author's
previous publication (122).

MaCYP71CD2 may act before MaCYP71BQ5, so explaining the minimal activity of MaCYP71BQ5 on tirucalla-7,24-dien-3β-ol **(1)** (Figure 2.2.5). This new peak **(4)** had a mass of 495.344 (Appendix A.5) which would be equivalent to the sodium adduct of tirucalla-7,24-dien-3β-ol with the three predicted oxidations performed by MaCYP71CD2 and MaCYP71BQ5.

To determine the structures of peaks **(2-4)** (Figure 2.2.5), large-scale *N. benthamiana* transient expression experiments were performed for each combination of genes. Structural elucidation of peak **(2)**, produced by co-expression of *AiOSC1* and *MaCYP71CD2*, revealed that MaCYP71CD2 simultaneously epoxidises the C24 alkene of tirucalla-7,24-dien-3β-ol **(1)** and hydroxylates the C23 position to produce a known protolimonoid, dihydroniloticin **(2)** (Figure 2.2.6). A literature review of dihydroniloticin-type protolimonoids (Appendix B.10) revealed that dihydroniloticin has been isolated from 16 different limonoid-producing species across the Meliaceae, Rutaceae and Simaroubaceae families and the C3 ketone of this structure, niloticin, from 23 species (Appendix B.10).

Structural elucidation of peak **(3)**, produced by co-expression of *AiOSC1* and *MaCYP71BQ5*, revealed that MaCYP71BQ5 is capable of oxidising the C21 position of tirucalla-7,24-dien-3β-ol **(1)** to produce tirucalla-7,24-dien-3,21β-diol **(3)** (Figure 2.2.6), another protolimonoid. This structure has been previously isolated, although only on one occasion, from an obscure member of the Simaroubaceae family (*Picrasma quassioides*) (154). The less extensive occurrence of tirucalla-7,24-dien-3,21β-diol **(3)** in limonoid-producing species, compared to dihydroniloticin **(2)**, further suggests that dihydroniloticin is the correct substrate of MaCYP71BQ5.

Structural elucidation of peak **(4)**, produced by co-expression of *AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*, did not reveal the expected tirucalla-7-ene-24,25-epoxy-dien-3β,21,23-triol, but instead a protolimonoid with a hemiacetal ring (melianol) **(4)** (Figure 2.2.6). This instead suggests that the oxidation of the C21 position, performed by MaCYP71BQ5, does not introduce a hydroxyl group but an aldehyde group. The combination of these three oxidations on the tail of tirucalla-7,24-dien-3β-ol is thought to cause spontaneous hemiacetal formation by nucleophilic attack of the C21 position (Figure 2.2.6). In solution melianol exists as an epimeric mixture, the C21 stereochemistry changing as the hemiacetal ring opens and reforms (155). This has also been observed in other similar hemiacetal containing protolimonoids such as turreanthin (156) and melianone (157). Melianol **(4)** and its C3 ketone derivative melianone, have been isolated a total of 8 and 18 times, respectively, from limonoid-producing species of the Meliaceae, Rutaceae and Simaroubaceae families (Appendix B.10).
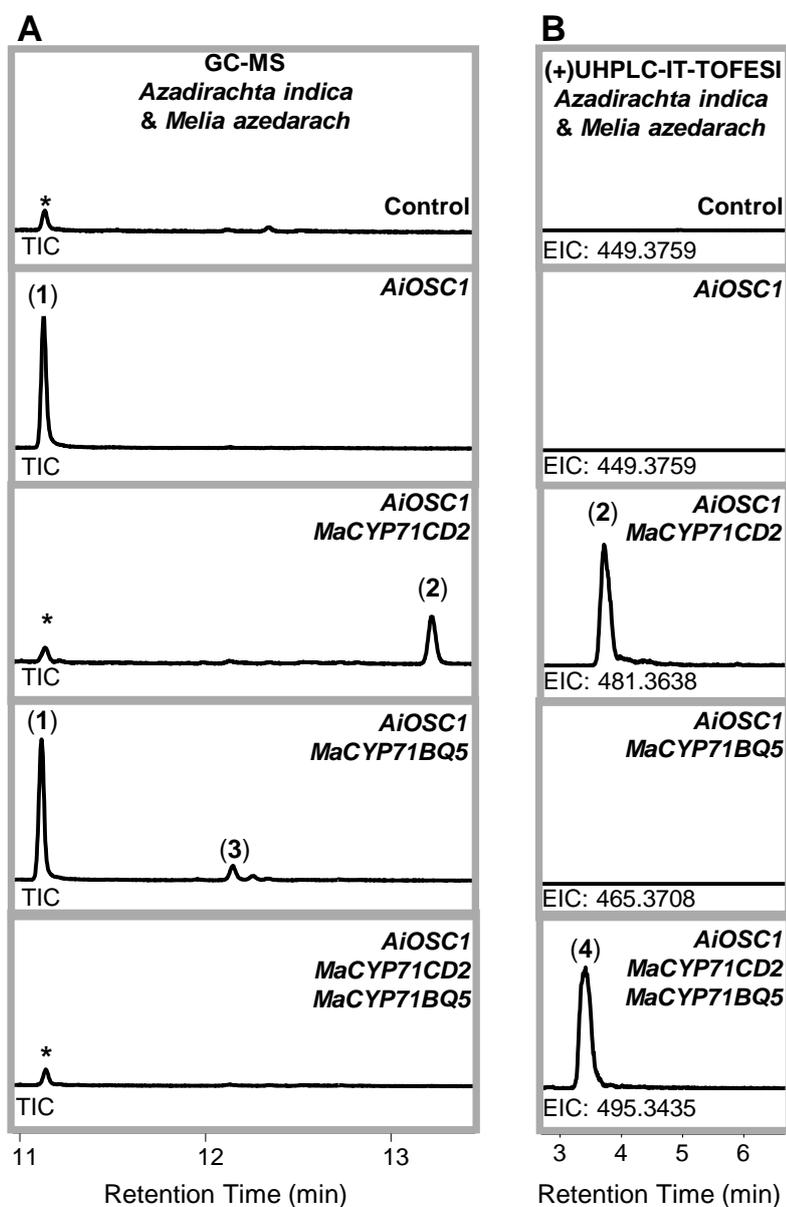
Figure 2.2.5: **Heterologous expression of candidate CYPs from *M. azedarach* (Ma1).** GC-MS total ion chromatograms (A) and UHPLC-IT-TOF ESI extracted ion chromatograms (B) of triterpene extracts from agro-infiltrated *N. benthamiana* leaves expressing *A. indica* and *M. azedarach* candidate genes in the pEAQ-HT-DEST1 vectors. Mass spectra for labelled peaks **(1-4)** are provided (Appendix A.4, Appendix A.5). The peak marked with an asterisk is an endogenous *N. benthamiana* peak, not tirucalla-7,24-dien-3β-ol **(1)** (mass spectra provided in Appendix A.6). UHPLC-IT-TOF analysis was performed using the 'protolimonoid' method. Figure adapted from the author's previous publication (122).

The identification of three biosynthetic genes from *M. azedarach* that are capable of full conversion of 2,3-oxidosqualene to the protolimonoid melianol **(4)** suggests that this pathway represents the initiation of limonoid biosynthesis (Figure 2.2.6). This is supported by the efficient conversion carried out by these enzymes in transient plant expression experiments (Figure 2.2.5), and by reports of melianol and dihydroniloticin in limonoid-producing species (Appendix B.10). Furthermore, homologs of MaCYP71CD2 and MaCYP71BQ5 were identifiable in *A. indica* (Ai1) with > 96% amino acid sequence identity and *C. sinensis* (Cs1) with > 86% amino acid sequence identity (Appendix B.2). Collaborators at Stanford University have identified (by co-expression) and cloned the two *C. sinensis* homologs of the CYPs characterised here and confirmed that these enzymes (CsCYP71CD1 and CsCYP71BQ5) are functionally equivalent to MaCYP71CD2 and MaCYP71BQ5 (122). This suggests that the three-step pathway to melianol characterised in *M. azedarach* is shared with *C. sinensis*, which implicates melianol biosynthesis as the initiation of limonoid biosynthesis across both the Meliaceae and Rutaceae families.

| 2,3-oxidosqualene | tirucalla-7,24-dien-3ß-ol (**1**) | dihydroniloticin (**2**) | intermediate | melianol (**4**) |
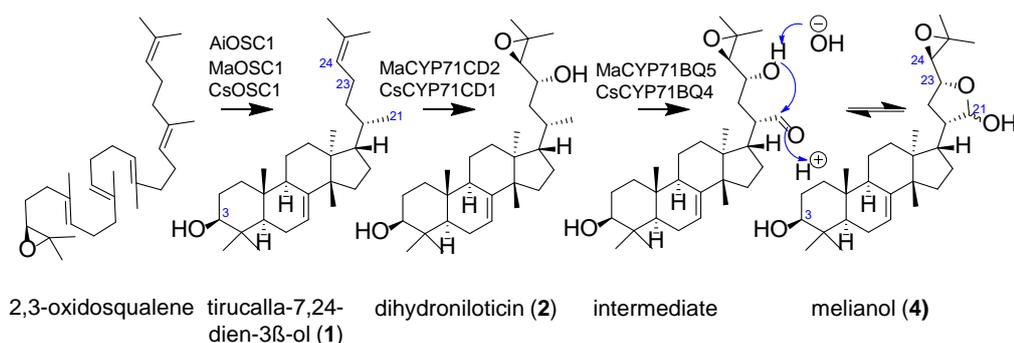
Figure 2.2.6: **Proposed melianol biosynthetic pathway.**
NMR confirmation of all structures is available (Appendix B.5, Appendix B.8, Appendix B.7, Appendix B.9). Figure reproduced from the author's previous publication (122).

2.3   Conclusion

Here, the first characterisation of protolimonoid biosynthetic enzymes is reported. Pre-existing sequencing resources from three diverse limonoid-producing species (*A. indica, M. azedarach* and *C. sinensis*) have been utilised to identify and characterise one OSC from each species (AiOSC1, MaOSC1 and CsOSC1) capable of converting 2,3-oxidosqualene to the triterpene scaffold tirucalla-7,24-dien-3β-ol, thereby initiating limonoid biosynthesis. Further, two CYPs (MaCYP71CD2 and MaCYP71BQ5) from *M. azedarach* have been shown to collectively carry out three oxidations on the side-chain of this scaffold, which together cause spontaneous hemiacetal ring formation and production of the protolimonoid melianol.

The relatively widespread occurrence of melianol, along with its intermediate dihydroniloticin, across the Meliaceae, Rutaceae and Simaroubaceae families (Appendix B.10), suggests that the biosynthesis of melianol represents a conserved pathway for initiation of limonoid biosynthesis in these different plant genera. In further support of this, homologs of melianol biosynthetic enzymes in *A. indica* and *C. sinensis* share > 85% amino acid sequence identity to the *M. azedarach* enzymes characterised here. Characterisation of the *C. sinensis* homologous CYPs by our Stanford collaborators has provided further confirmation that this pathway is indeed shared between *M. azedarach* and *C. sinensis* (122).

The characterisation of these three protolimonoid biosynthetic genes has provided a foundation for the identification of later steps in limonoid biosynthesis (Chapter 6) and investigation of the structure-activity relationships of protolimonoids (Chapter 7).

2.4   Chapter-specific materials and methods

Methods used in this chapter have been described in the author's previous publication (122) and are reproduced here. All NMR experiments and analyses were performed by Dr. Michael Stephenson.

2.4.1   Identification of candidate genes

BLAST+ V2.7.1 (158) protein searches were used to identify candidate genes. To identify candidate OSCs protein sequences of 83 functionally characterised OSCs (11) were used as a query. Candidates were filtered based on the presence of the conserved triterpene synthase OSC motif (DCTAE) (159) and length (between 700 and 1000 amino acids). Candidate CYPs were identified by using *A. thaliana* CYP protein sequences (153) as a query and filtered by length (300-700 amino acids) and presence of protein coding region (Augustus V3.2.2 (160)) and 'cytochrome P450' pFAM annotations (HMMSCAN (EMBL-EBI)). CYPs were named following convention by the Cytochrome P450 Nomenclature Committee (161). Once identified protein sequences of candidates were aligned to query sequences with MUSCLE V3.8.31 (162). FastTree V2.1.7 (146) was used to construct phylogenetic trees which were formatted using iTOL (147).

2.4.2 *A. indica* (Ai1) differential gene expression analysis

Differential analysis was performed using Trinity-assembled *A. indica* transcriptome (Ai1) as reference sequence to align raw RNA-Seq reads from fruit, root, leaf, stem and flower tissues (126, 140). Transcript abundance estimation was performed using a script provided within the Trinity *de novo* assembler package 'align and estimate abundance' (144). Briefly, raw RNA-Seq reads for each tissue were aligned to the transcriptome (BowTie V1.0.1 (163)) and abundance per gene was estimated (RSEM V1.3.0 (164)) using Trinity transcripts as a proxy for genes. The resultant estimated counts per gene were converted to integers and genes scoring less than one count per million in two or more tissues were excluded from the analysis.

Data were normalized to account for differences in library size by using a trimmed mean of M-values (TMM) method (EdgeR V3.22.5 (165)). Due to a lack of replicates in the published dataset, a dispersion value could not be calculated and was therefore manually estimated at 0.05. A genewise negative binomial generalized linear model (EdgeR V3.22.5 (165)) was used to identify differentially expressed genes (likelihood ratio test = 1 and P-value < 0.05). Hierarchical clustering analysis was subsequently used to identify a cluster of genes with similar expression patterns to *AiOSC1*. HMMSCAN (EMBL-EBI) was used to assign pFAM domains to this subset of genes (E-value = 1).

2.4.3 Purification of protolimonoids and structural elucidation

Purification of tirucalla-7,24-dien-3ß-ol (AiOSC1 product)

Two L of GIL77 cells expressing *AiOSC1* (pYES2) were cultured and pelleted, yielding 15.28 g of material. Saponification was performed in 100 ml of reagent and triterpenes were extracted by addition of an equal volume of hexane in triplicate yielding 220 mg of dried crude extract. Fractionation using Isolera Prime (Biotage) (Table 2.4.1) yielded 1 mg of purified tirucalla-7,24-dien-3β-ol enabling structural confirmation by NMR (Appendix B.5).

Purification of dihydroniloticin (the product of AiOSC1 and MaCYP71CD2)

Using the previously described vacuum infiltration method (112, 166) 144 *Nicotiana benthamiana* plants were infiltrated with equal volumes of *A. tumefaciens* strains containing pEAQ-HT-DEST1 expression construct for *AstHMGR, AiOSC1* and *MaCYP71CD2.* Initial extraction was performed on dried leaf material (68.64 g)

Table 2.4.1: **Isolera Prime fractionation conditions.**

Target	Column	Solvents	Gradient	Yield
AiOSC1 product	SNAP Ultra 10g	A: Hexane B: EtOAc	0-2%(1CV), 2-25% (10CV), 25-100% (3CV)	1 mg
AiOSC1 MaCYP71BQ5 product	KP-Sil 50g	A:Hexane B: EtOAc	0-100%(45CV), 100%(11CV)	50 mg
	SNAP Ultra 10g	A:Hexane B: EtOAc	20-50% (176CV),50%(5CV)	10 mg
	SNAP Ultra 10g	A:Hexane B: EtOAc	20%(176CV)	4 mg
AiOSC1 MaCYP71CD2 product	SNAP KP-Sil 25g	A:Hexane B: EtOAc	6-100% (10CV)	190 mg
	SNAP Ultra 10g	A:Hexane B: EtOAc	10%(1CV), 10-60%(10CV), 60%(5CV)	170 mg
	SNAP Ultra 10g	A:Hexane B: EtOAc	10% (3CV), 10-34% (143CV)	130 mg
	SNAP Ultra 10g	A:Hexane B: EtOAc	0-30%(181CV)	120 mg
AiOSC1 MaCYP71CD2 MaCYP71BQ5 product	SNAP KP-Sil 100g	A:Hexane B: EtOAc	6-100% (13CV)	410 mg
	SNAP Ultra 10g	A:Hexane B: EtOAc	20-53% (116CV)	220 mg
	SNAP Ultra 10g	A: Hexane B:DCM	10-41% (140CV)	34 mg
	SNAP Ultra 10g	A: Hexane B:DCM	10-30% (87CV), 30% (79CV)	4 mg

Details of conditions used for Isolera Prime fractionation including: column, solvent system, percentage gradient of solvent B, column volume and dry weight of resulting extract (yield). All samples were dry-loaded onto Isolera Prime (Biotage) using Celite (Sigma-Aldrich). Table reproduced from the author's previous publication (122).

following the large-scale triterpene extraction protocol previously described (166). Successive rounds of fractionation were performed using Isolera Prime (Biotage) (Table 2.4.1). To achieve final purification, the sample was dissolved in a minimal amount of ethanol and agitated (15 min) with activated charcoal (Sigma-Aldrich). This yielded 86 mg of dihydroniloticin, enabling structural confirmation by NMR (Appendix B.7).

Purification of tirucalla-7,24-dien-3β,21-diol (the product of AiOSC1 and MaCYP71BQ5)

Five L of *S. cerevisiae* Y21900 cells expressing *AiOSC1* (pYES2), *AtATR2* (pAG425gal) and *MaCYP71BQ5* (pAG423gal) were cultured and pelleted. Saponification was carried out in 500 ml of reagent. Extraction was performed in 1.5 L of hexane, yielding 1.13 g of dried crude extract. Successive fractionation using Isolera Prime (Biotage) (Table 2.4.1) yielded 4 mg of tirucalla-7,24-dien-3β,21-diol, enabling structural confirmation by NMR (Appendix B.8).

Purification of melianol (the product of AiOSC1, MaCYP71CD2 and MaCYP71BQ5)

Using vacuum infiltration (112, 166) 160 *N. benthamiana* plants were agroinfiltrated with equal volumes of *A. tumefaciens* strains containing pEAQ-HT-DEST1 expression constructs of *AstHMGR, AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*. Initial extraction was performed on dried leaf material (198.5 g) following the large-scale triterpene extraction protocol previously described (166). Successive rounds of fractionation were performed using Isolera Prime (Biotage) (Table 2.4.1). To achieve final purification, re-crystallisation was performed by dissolving the sample in a minimal volume of methanol (90°C), covering and allowing crystals to form at room temperature. Crystals were washed in ice-cold methanol. The re-crystallisation process was repeated, yielding ~6 mg of melianol. Structural confirmation was carried out by NMR (Appendix B.9).

3

Securing material for research under the Nagoya Protocol

Synopsis

The trade and use of plants has been a crucial part of human history and scientific advancement. However, certain modern bioprospecting cases have sparked controversy when commercial gain has not been shared with the communities of origin. This concern has led to the development of a legal framework for access and benefit sharing, the Nagoya Protocol (NP), a supplementary agreement to the Convention on Biological Diversity, which gives countries sovereign rights over their biodiversity. The NP does not aim to create a barrier to research. However, within this project attempts to access material through the NP were fraught with difficulties and eventually unsuccessful. Therefore material which was out of scope of the NP had to be secured in order to enable research to proceed. In this chapter the experience of trying to access material for use in this project is used as the basis for discussion of the efficacy of the NP and its potential impact on research.

3.1　Introduction

The Nagoya Protocol (NP) is an international agreement on access and benefit sharing. Its aim is to facilitate and provide a legal frame work for the implementation of objective 3 of the Convention on Biological Diversity-

> 'The fair and equitable sharing of the benefits arising out of the utilization of genetic resources, including by appropriate access to genetic resources and by appropriate transfer of relevant technologies, taking into account all rights over those resources and to technologies, and by appropriate funding.'

The 25-page text of the NP is the result of 18 years of discussion and negotiation, and was finally approved and endorsed at the Convention on Biological Diversity's 10th Conference of the Parties in Nagoya, Japan 2010 (Figure 3.1.1).
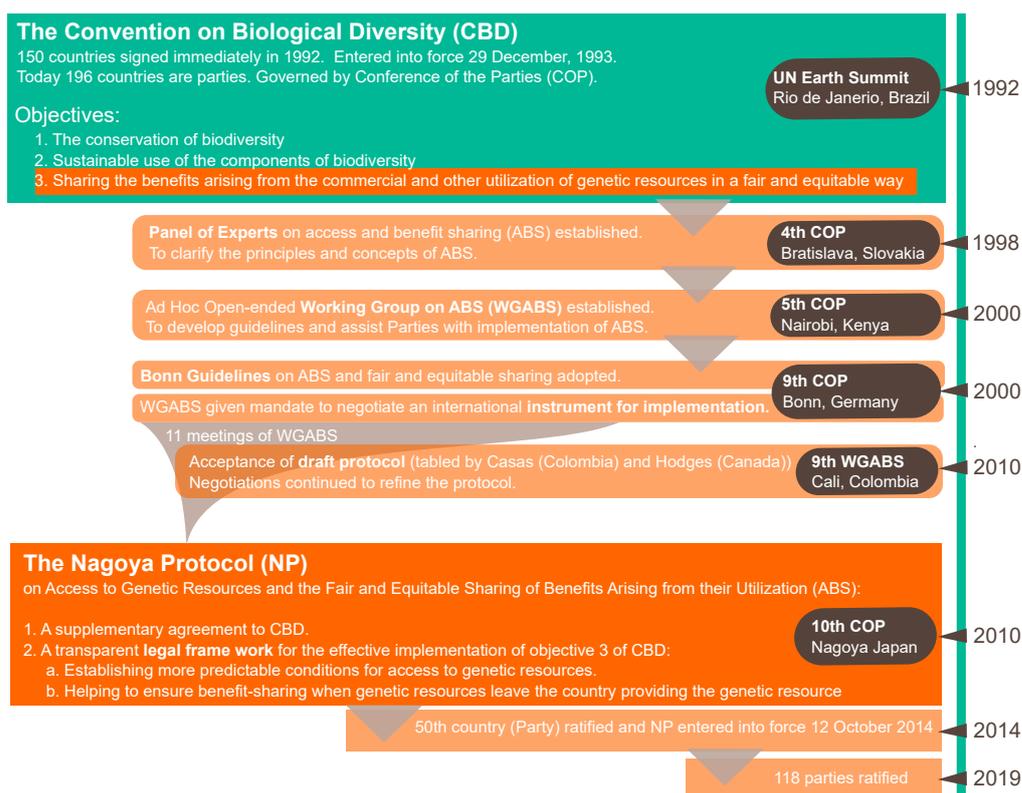
Figure 3.1.1: **History of the Nagoya Protocol.**
Time-line showing the development and entry into force of the NP, a supplementary agreement to the Convention on Biological Diversity. Information provided by the Convention on Biological Diversity (167).

The NP negotiations were fraught with complexities, as has been discussed in detail (168–171), due to a number of opposing views between different groups of countries.

The group of Like-Minded Megadiverse Countries consists of 17 countries (Bolivia, Brazil, China, Colombia, Costa Rica, Democratic Republic of the Congo, Ecuador, India, Indonesia, Kenya, Madagascar, Malaysia, Mexico, Peru, Philippines, South Africa and Venezuela) that together contain the majority of the earths biodiversity (60-70%) (172). This group, with the motivation to protect their genetic resources, advocated for binding international rules that were broad in nature and would establish compliance mechanisms and minimum requirements of countries (169). A group of developed countries with less biodiversity, but large biology-based research sectors (e.g. European Union and Canada) had contrasting views. These countries, to protect their research industries, advocated for a balance between ensuring sharing of benefits whilst still allowing access to genetic resources and more flexible rules to suit country-specific requirements (169). Because such starkly conflicting opinions had to be overcome, establishment of the NP was heralded as a success in itself (168, 169). The NP came into force four years later after 50 countries had ratified (Figure 3.1.1).

The NP summarises its aims as: (i) providing internationally predictable conditions for *'users'* to access genetic resources and (ii) helping to ensure benefits arising from the use of such genetic resources are shared with the *'providers'*. The basic mechanism involves giving Parties to the Convention on Biological Diversity sovereign rights over their biodiversity, which is defined as genetic resources and the traditional knowledge associated with these genetic resources. Importantly, *'derivatives'* of genetic resources such as natural products are also included in this definition. The principle of the NP is that, in order for users to access genetic resources for use in research, two things need to be negotiated and documented between the providers and users, and subsequently approved by the provider's National Competent Authority. These are: Prior Informed Consent (PIC), which serves as permission to sample resources; and Mutually Agreed Terms (MAT), which details the planned use of the resource and the benefits to be shared with the providing country. These benefits (which can be monetary or non-monetary) act to incentivise the protection and preservation of biodiversity in line with the overarching goals of the Convention on Biological Diversity. Therefore the process of the NP involves sharing benefits in exchange for resources with which to conduct research on (Figure 3.1.2).

An additional specification of the NP is that country-specific information will be available on an official international website (The Access and Benefit Sharing Clearing House; ABSCH, (173)). This information includes contact details for the designated National Focal Point and documentation of Internationally Recognised Certificates of Compliance (which confirm that PIC and MAT have been approved by the provider's National Competent Authority).
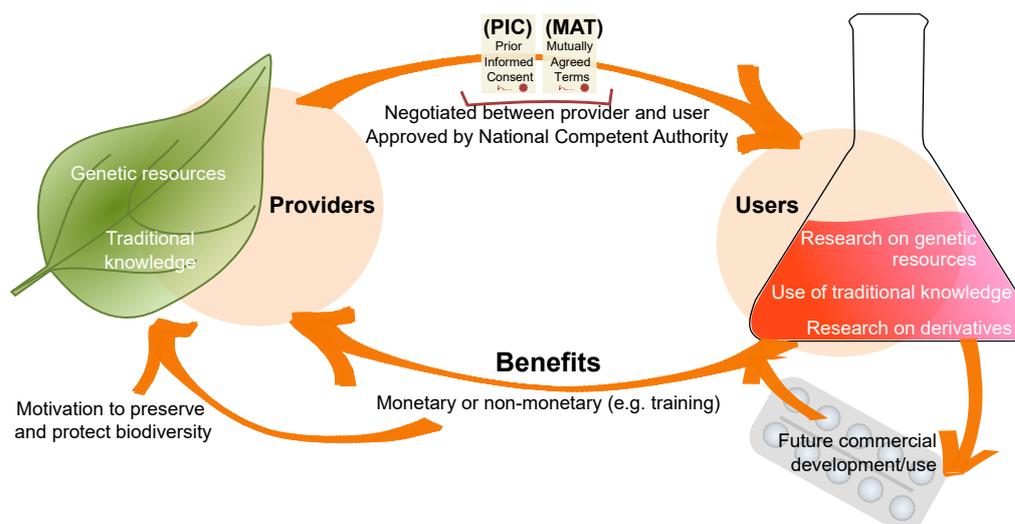
Figure 3.1.2: **Scheme illustrating the process of accessing genetic resources through the NP.**
In brief, the requirement of a PIC and MAT for the transfer of genetic resources from provider to user and for the return of benefits to the provider.

3.1.1 Aims

The purpose of this chapter is to discuss the potential impacts of the NP on research from the perspectives of a researcher in the field of plant natural product biosynthesis. The aims are to:

1. Introduce and discuss areas of clarity and contentious issues of the NP.

2. Use experiences of attempting to secure Meliaceae genetic resources for this project as case studies for accessing genetic resources in and out of scope of the NP.

3. Discuss the differing perspectives of scientists and the public based on engagement through cross disciplinary Global Garden workshops.

3.2 The NP- clarity and contentious issues

3.2.1 Historical case-studies of 'unfair' bioprospecting

Although the major motivation of the NP is to implement the Convention on Biological Diversity's goal of protecting biodiversity, the NP has also been shaped by cases of perceived unfair use of genetic resources, i.e. where a commercial gain has been made without the benefits of this being shared with the country or community

of origin.   Such cases were the fuel to the arguments of groups such as the Like-Minded Megadiverse Countries who were concerned that their biodiversity would be utilised without ensuing benefits being shared (174).

There are numerous examples of plant species being involved in such cases (175). For instance, the fruits of the Amazonian shrub *Myrciaria dubia* (camu-camu) (Figure 3.2.1.A), have been reported to produce up to 30 mg/g absorbic acid (vitamin C) (176), 30-fold greater than the levels reported in orange (177). *M. dubia* originates from the flood-plain environment of the Peruvian Amazon and has been traditionally utilised as a flavouring in Peruvian jams and juices, but has also been encouraged by the government as a cash crop (174, 177, 178).   Use of camu-camu extracts in the absence of benefit sharing agreements with Peru appears to be widespread with the National Anti-Piracy commission of Peru identified more than 50 claims within Japanese patents, and one within a European Union patent, which use camu-camu extract (174).

One of the most infamous cases of perceived unfair bioprospecting, is the now widespread commercial use of the natural sweeteners, steviol glycosides, which together give *Stevia rebaudiana* extracts (Figure 3.2.1.B) up to 300 times the sweetness of sucrose (179). This plant species, called kaa he-he (sweet herb), was traditionally used by the Guarani people (South America), as a sweetener for their herbal green tea (mate) and a general flavour enhancer (180, 181). By 2015 steviol glycosides were being used in both Coco-Cola and Pepsi-Co drinks with China performing 80% of all *S. rebaudiana* cultivation (182). Further, Cargill and Evolva have developed Yeast strains capable of producing, by fermentation, rebaudiosides D and M which, although are only 1% of *S. rebaudiana* leaf extracts, are the sweetest steviol glycosides. They claim this development will save resources such as land-use and energy requirements to produce the sweetener (183). However, benefits from the commercialisation of these natural products, whether fermented or extracted from the plant, have not been shared with the Guarani people or the countries of Paraguay or Brazil where this species is native (182). However, this may be subject to change in the future. For instance in a similar case of perceived unfair bioprospecting, the worldwide sale of *Aspalathus linearis* (rooibos) (Figure 3.2.1.C) based herbal teas, an agreement has recently (November 2019) been signed that will see 1.5% of the value of unprocessed rooibos shared with the San and Khoi communities of the Cederberg region of South Africa, where this plant is endemic (184).

Finally, *A. indica* (neem) (Figure 3.2.1.D) is a tree of vast cultural and historical importance in India, as evidenced by publications such as 'Neem: India's Miraculous Healing Plant' (185), 'Neem: The Divine Tree *Azadirachta Indica*' (186) and 'Neem: The Ultimate Herb' (187). Often known as the 'village pharmacy', neem extracts from different tissues are reported as traditional remedies in India for a plethora of

different aliments (68), with limonoids being the abundant natural product thought to be responsible for these uses (16). The use of neem extracts has been the subject of a 10 year patent dispute which ended in the overturning of an American patent when evidence was provided for comparable historical use in India. The overturned patent was for fungicidal use, but this ruling has raised questions about the validity of the nine other patents involving Neem (188).

Figure 3.2.1: **Plant species involved in access and benefit sharing disputes**
(A) *M. dubia* (camu-camu) (B); *S. rebaudiana* (sweetleaf); (C) *A. linearis* (rooibos) and (D) *A. indica* (neem). Image credit (A) Flickr (189) (B and C) Wikimedia Commons (190, 191), (D) Andrew Davis (JIC photography).

There is a clear sense of injustice associated with high profile and high profit bioprospecting examples. However, historic scientific advances, particularly in the field of natural products, have flourished through the freedom to investigate genetic resources without regulatory barriers, based on the pre-NP assumption that genetic resources were the 'common heritage of humankind' (192). When considering that ~70% of drugs and pesticides are natural products or natural product-inspired (193, 194) it is clear this freedom has had a significant impact on modern life. This is the basis of the counter-argument to the concerns of unfair bioprospecting. If regulatory frameworks, such as the NP, become a barrier to research, then this could significantly impede the rate of scientific progress and the discovery and development of products that could help humanity as a whole. A recent analysis suggested that less than 1% of the estimated number of species on earth are mentioned in patents (175). Although the patenting of biology has its controversies (192), this highlights the huge expanse of biodiversity yet to be investigated that could potentially benefit humankind.

3.2.2   Issues of clarity within the NP

The comment that the NP is '*a masterpiece in creative ambiguity*' has been reiterated and discussed a number of times (168, 169, 171). Due to the vastly different opinions and interests of the countries involved in the negotiation of the NP, there were a number of issues for which an agreement could not be reached. Therefore, by

remaining silent or being ambiguous on certain contentious issues the NP was able to reach acceptance (170). This has gained praise from some commentators as a creative solution to the huge conflict of interest between such countries (168, 169) (Section 3.2.1). However, to others, the lack of clarity that comes with the ambiguity does not provide a clear legal framework (195), which undermines the mission of the NP.

Regardless, the '*creative ambiguity*' means that critical decisions will be made during country-specific implementation of NP legislation (169). Subsequently how the NP is applied and policed could be drastically different across the world. Therefore the adoption of the NP is only the starting point and it is the implementation of both user-specific and provider-specific legislation that will determine the impact of the NP (169). Further compounding this, there is an argument that the guidelines of the NP largely represent only *'floors'* or minimum standards required for implementation and lack *'ceilings'*, or upper limits of what a Party may impose in its legislation (171).

There are obvious hypothetical questions that arise when giving a country sovereign rights over biodiversity. For instance, who owns the genetic resources residing in the 45% of the earth's surface covered by international waters (the high seas), to which no country has sovereign rights? Who would receive benefits from research on a species that is endemic to multiple countries (transboundary)? How long does an invasive species have to be established before a country can claim sovereign rights over it? These issues are not clearly resolved by the NP. Within the NP the statements surrounding the issue of transboundary genetic resources and instances where obtaining a PIC would not be possible include the following language: *'Recognizing that an innovative solution is required', '...those Parties shall endeavour to cooperate.'* and *''...consider the need for and modalities of a global multilateral benefit sharing mechanism'*. The language used is of recognition and consideration, a feature of the NP (171), but how such language will be translated into legislation is not known.

Further, although the distinction between commercial and non-commercial research was a feature of NP discussions, this subsequently became a point the NP was silent on. The only provision on this issue is Article 8, *'Special considerations'*, which encourages simplified provisions for non-commercial research and expeditious access to resources in the case of emergencies. The degree of this simplification and mechanisms to implement it are not specified and will be yet another element of the NP to be determined by the legislation of individual countries.

Arguably the most significant areas of the NP that lack clarity can be summarised as the temporal scope of the NP (retroactivity) and the breadth of Convention on Biological Diversity and NP coverage, both of which have both been discussed in

detail (170). Temporal scope becomes significant when considering whether the NP has been triggered at the time a genetic resource is sampled (and crosses borders from provider to user country) or at the time a genetic resource is used in research. For instance, would using plant material that had been growing in a UK botanical garden for decades still trigger the NP? Despite being a contentious issue in the 11 meetings that produced the NP (Figure 3.1.1), the text remains silent on this issue and therefore is another example of where legislation will determine impact. However, in this case, this relates to the legislation of the user country rather than the provider.

An example of how drastically legislation will determine the impacts of the NP can be taken from the EU, which has provided legislation for user-compliance and left its member states free to choose whether or not to draft and implement their own provider legislation. The EU legislation clearly state that the NP does not apply retroactively:

> 'The EU ABS Regulation applies from 12 October 2014, which is the date when the Nagoya Protocol entered into force for the Union. Genetic resources accessed prior to that date fall outside the scope of the Regulation even if utilisation of those resources occurs after 12 October 2014 (see Article 2(1) of the Regulation).'(196)

The legislation further states that accessing genetic resources from countries that are Party to the NP, but choose not to exercise their sovereign rights over biodiversity, such as the UK, are also out of scope of the NP.

> 'A state must exercise sovereign rights over genetic resources for them to be in the scope of the Regulation.' (196)

This legislation therefore creates a new pool of out of scope genetic resources. This would include resources that have been growing in botanical gardens, nurseries or even as invasive species as well as genetic resources from countries which have chosen not to exercise their sovereign right over biodiversity.

The definition of *'utilization of genetic resources'* is well defined by the NP, with broad definitions making it clear that the NP is relevant to all biology-based industries, including pharmaceutical companies, regardless of whether they are working on an actual organism or a natural product. However, the scope of what is classed as a derivative of a genetic resource remains a matter for debate (169, 170, 197). The crux of this debate is whether Digital Sequencing Information (DSI) should be classed as derivatives of genetic resources and therefore be in scope

of the NP. The advent of high-throughput sequencing, DNA synthesis and synthetic biology methods mean that it is possible for a genetic resource to be sequenced in its country of origin and for this sequence information to be used to synthesise genes for use in research elsewhere in the world. The use of DSI is currently out of scope of the NP, but its future inclusion is still being debated. Many biodiversity rich countries such as Brazil, Peru and the Philippines are arguing for its inclusion and using language in their legislation that stretches genetic resources to cover DSI (198). This debate is still ongoing (197). Briefly, the 13th Conference of the Parties called for a fact-finding study to clarify terminology and concepts around this issue by establishing an *'Ad Hoc Technical Expert Group'*. Their meeting in Montreal (February 2018) was unsuccessful in reaching an agreement between the vastly different viewpoints of the biodiversity rich countries and countries where biology-based research takes place. Subsequently, the group failed to agree on terminology and certain individuals dissociated themselves from the final report (197). At the 14th Conference of the Parties, in Sharm El-Sheik (November 2018), a new *'Ad Hoc Technical Expert Group'* was called to again consider the terminology and use of DSI, the results of which are expected in 2020. Until these results are released DSI is not considered part of the NP

3.2.3   The impacts of the infancy of the NP

In the five years since the NP came into force the lack of clarity on how to implement and abide by the policy has been further complicated by its infancy. After the 18-year wait for the NP (Figure 3.1.1), there will potentially be an even longer delay before individual provider countries decide how they will devise and implement their own legislation (171). The ABSCH website currently contains 233 legislative, administrative and policy documents. However, these 233 items have been uploaded by only 60 governments. Thus almost half of the Parties to the NP are yet to upload policy information over five years after the NP came into force (Table 3.2.1). Even considering that some Parties are newly joined and others will choose not to exercise their sovereign rights over biodiversity, this is not a reassuring statistic. Further, as was the case in 2016 (171), the majority of these documents contain information regarding pre-NP legislation. Although there is only one Party (recently joined) lacking a National Focal Point, nearly half of the countries party to the protocol have not as yet assigned a National Competent Authority (Table 3.2.1). Further, although not a necessity of the NP, the development of model contractual clauses are strongly encouraged, in Article 19 of the NP, and have been recognised as a crucial way to prevent the regulatory barrier of the NP becoming unnavigable. However, at present there is only one example of a model contractual clause on the Access and Benefit Sharing Clearing House (ABSCH) (Table 3.2.1).

Table 3.2.1: **Summary of information from the ABSCH website.**

ABSCH Information	Total	No. of Parties
National Focal Point	117	117
Competent National Authority	105	62
Legislative, Administrative or Policy Information	233	60
ABS procedure	6	5
National Model Contractual Clause	1	1
Internationally Recognised Certificates of Compliance	1107	17

Summary of available information on ABSCH. Data recovered September 2019 for the 118 Parties to NP. Including total counts for each piece of information and the numbers of Parties with information uploaded.

Despite the hurdles involved in abiding by the NP, the first Internationally Recognised Certificate of Compliance, confirming PIC and MAT, was granted by India only a year after the NP came into force in October 2015. The ABSCH website reveals that over a thousand Internationally Recognised Certificates of Compliances have been issued in the four years since, which is an achievement. However closer analysis revealed that ~80% of the Internationally Recognised Certificate are from India and France (Figure 3.2.2) and only 17 of the 118 Parties who have signed up to the NP have issued Internationally Recognised Certificates of Compliance thus far.
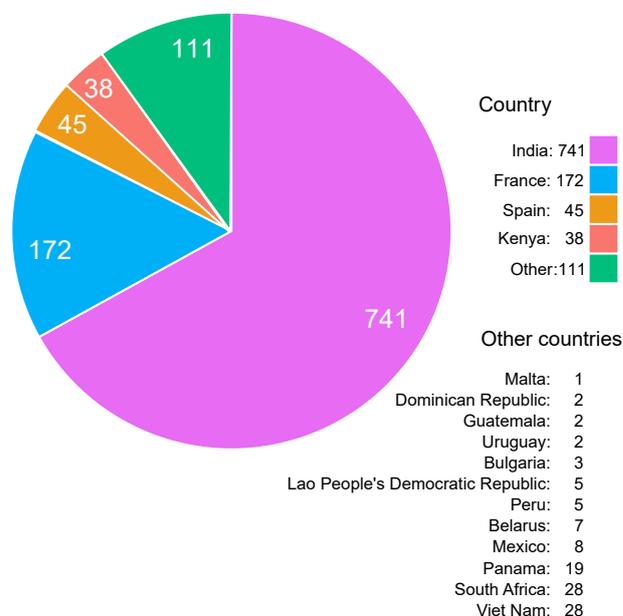
Figure 3.2.2: **Issuance of Internationally Recognised Certificates of Compliance**
The relative contributions by the 17 Parties who have successfully issued Internationally Recognised Certificates of Compliance (to a total of 1107). Data taken from ABSCH website September 2019.

3.3   Accessing NP-competent Meliaceae material

To progress this project, access to limonoid-producing Meliaceae material was required to inform and enable the generation of RNA-Seq data, which would greatly aid the identification of limonoid biosynthetic genes. Therefore a key aim of this project, in parallel to biosynthetic investigations, became to secure Meliaceae plant material for research. A number of different routes have been explored to access such plant material and are described in this section as a case study of accessing genetic resources for use in preliminary research.

3.3.1   Attempts to access material in scope of the NP

One potential route of accessing Meliaceae material was via an existing collaboration with researchers in Vietnam. A Royal Society International Partnership award was secured for a collaboration, between the John Innes Centre (JIC) and the Institute of Chemistry- Vietnam Academy of Science and Technology (ICH-VAST) (Hanoi, Vietnam), on a project entitled *'Analysis of triterpene biosynthesis and diversity in neem'*. Throughout the course of this project six trips between Hanoi and Norwich were arranged along with regular correspondence by email, Skype and FaceTime. As part of this process collaborators sourced *A. indica* and *A. excelsa* material from contacts in Southern Vietnam and a sampling trip to collect *M. azedarach* seeds was undertaken by JIC and ICH-VAST staff (including the author) in the Vĩnh Phúc province of Vietnam (Figure 3.3.1). One of the objectives of the project was to apply for a PIC and MAT to secure permission for the transfer these genetic resources to JIC.

To enable the transfer of these genetic resources a Material Transfer Agreement and Memorandum of Understanding were negotiated. Vietnam's pre-NP legislation forms for plant export were completed and a drafted MAT that both JIC and collaborators at ICH-VAST were willing to sign was negotiated. The process of negotiation and signing of these documents was not trivial and required a great deal of time and effort from both parties as summarised (Table 3.3.1). Unfortunately, although the collaborators at ICH-VAST were satisfied with the MAT, the director of ICH-VAST was unwilling to sign the MAT. The National Competent Authority of Vietnam specified that the signature of the director was required to proceed, and further that only certain institutes and organisations are capable of certifying such documents. The end result of this was that the MAT was not signed and a solution to this issue has not so far been found.

A number of recurring issues contributed to the difficulties of the negotiation process (Table 3.3.1). Firstly, the language barrier between the Vietnamese and UK

Figure 3.3.1: **Sampling trip in Vĩnh Phúc province, Vietnam.**
Fruits of the *M. azedarach* tree (in the background) were collected and processed to seed. Left to right: Dang (local guide), Hoang (ICH-VAST student), Dr. Le (University of Science and Technology, Vietnam (USTH)), Dr. Duoing (ICH-VAST), Dr. Oanh (USTH), Dr. Ramesha Thimmappa (JIC). Image credit: author.

participants greatly slowed progress in negotiations as translations of documents were required (in both directions). Moreover the differences between these two languages were accentuated in legal documentation. This resulted in ICH-VAST requesting changes in wording that it was difficult for the UK legal contract team to understand. Examples of this were requests to change *'agreement'* to *'contract'* and for the inclusion of a reference to the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITIES) in the Material Transfer Agreement, despite the species it related to not being protected by CITIES (Table 3.3.1). Undoubtedly, some of JIC's requests for changes to language would have seemed equally unnecessary to ICH-VAST.

Table 3.3.1: **Summary of collaboration attempts to secure Meliaceae material in scope of the NP.**

Date	Progress & complications	Outputs & benefits shared
2016		
Apr	Royal Society funding for this project commenced and initial discussions between JIC and ICH-VAST regarding this project began. Preliminary sampling was undertaken by ICH-VAST of *A. indica, A. excelsa* and *M. azedarach.*	Analytical limonoid standards sent to ICH-VAST for preliminary profiling of samples.
Jul	NP-related discussions started between JIC and ICH-VAST. The first items to negotiate were the MTA and MoU between JIC and ICH-VAST. Additionally, the pre-NP forms required for plant export from Vietnam were identified by Mike Ambrose (Mẫu số 07/TT & 08/TT).	Two scientists from ICH-VAST visited JIC for a two week period where they attended an OpenPlant conference and received scientific training in molecular biology techniques.
Aug	Negotiations of an MTA and MoU between JIC and ICH-VAST were continued. Collaborators reported that *A. excelsa* was unavailable as seed and therefore this species was excluded from the project.	
Sept	ICH-VAST and JIC staff undertook sampling of *M. azedarach* fruits in the Vĩnh Phúc Provence (Vietnam). Seeds were extracted and dried in preparation for shipping. Translations of MTA and MoU were required to proceed with negotiations, which highlighted differing expectations of these documents between JIC and ICH-VAST. These included ICH-VAST requesting to include: a reference to CITIES within the MTA, a signature from the JIC director (rather than Head of Contracts), an official JIC stamp in addition to the signature and colour copies of all documentation.	Two JIC scientists visited ICH-VAST for one week period and during this time a MTA and MoU were negotiated and signed between JIC and ICH-VAST. In addition the plant export forms for Vietnam (Mẫu số 07/TT & 08/TT) were completed, signed and sent to Ministry of Forestry.
Nov	Sending of samples had been postponed while awaiting a response from the Ministry of Forestry regarding the plant export forms. In anticipation of the response, *A. indica* seeds (from the Ninh Thuãn Provence) and *M. azedarach* seeds (from the Vĩnh Phúc Provence) were sent to JIC (with signed MTA and MoU between JIC and ICH-VAST). In absence of confirmation from the Vietnamese government, these seeds were stored and not utilised for research.	
Dec	Due to the continued delay in response from the Ministry of Forestry, JIC suggested that ICH-VAST should contact the National Focal Point for Vietnam (at the Vietnam Biodiversity Conservation Agency).	

Table 3.3.1: (continued)

Date	Progress & complications	Outputs & benefits shared
2017		
Feb	ICH-VAST and JIC both contacted the National Focal Point, who advised that plant export forms had been submitted to the wrong government agency. One ICH-VAST researcher had to leave the project.	One ICH-VAST scientist visited JIC. The plant export forms were resubmitted to the National Focal Point (Vietnam Biodiversity Conservation Agency).
May	Feedback regarding the plant export forms was received from the National Focal Point of Vietnam. A Vietnamese-English translation of the feedback was required before it could be actioned. The feedback confirmed that a PIC and MAT were required in addition to the plant export forms. Therefore, ICH-VAST began attempts to secure a retrospective PIC from the Vĩnh Phúc provence where *M. azedarach* had been sampled. In the absence of an example document on the ABSCH website, the author drafted a MAT (with guidance from Mike Ambrose), based on existing collaboration documents between JIC and ICH-VAST (MTA, MoU, 07/TT and 08/TT), which was designed to satisfy the feedback from the National Focal Point. A draft MAT was sent to ICH-VAST to begin negotiations of this document.	
Jun	JIC reiterated to ICH-VAST the importance of the MAT document and requested their feedback.	
Jul	Post-NP access and benefit sharing policy of Vietnam was uploaded onto the ABSCH website. ICH-VAST confirmed that they were unable to secure a retrospective PIC for the *M. azedarach* seeds sampled from the Vĩnh Phúc Provence, confirming that these seeds were unable to be used in research. To remedy this, ICH-VAST planned a resampling of *M. azedarach* in the Bắc Giang Provence. ICH-VAST began working towards gaining PIC for this location and JIC began changing documentation to accommodate this.	
Sept	A new ICH-VAST student was assigned to project.	
Nov	The ICH-VAST student had to leave the project. However, a second ICH-VAST student was assigned to project. Planned re-sampling location was changed to a location on the site of ICH-VAST (Hanoi). ICH-VAST began working towards gaining PIC from this new location and JIC began changing the relevant documentation. ICH-VAST accepted conditions of draft MAT negotiated between JIC an ICH-VAST.	The protocol for an insect assay (utilising diamond back moths) developed by the JIC insectary was shared with ICH-VAST.

Table 3.3.1: (continued)

Date	Progress & complications	Outputs & benefits shared
2018		
Jan	The new ICH-VAST student meet with staff at the Vietnam Biodiversity Conservation Agency to discuss paper work associated with this project. They confirmed that a specific document 'Form03' was required as MAT rather than document negotiated between JIC and ICH-VAST. Therefore, terms from the draft MAT were transferred into Form03 and sent to JIC contracts to review whether it would be appropriate for JIC to sign.	The pre-NP plant export forms were updated (with new sampling location and dates) and resubmitted to the Vietnam Biodiversity Conservation Agency.
Feb	Acceptable draft terms of Form03 were negotiated between JIC and ICH-VAST and subsequently sent to JIC contracts for final acceptance. An additional ICH-VAST researcher was assigned to project.	
Mar	ICH-VAST student had to leave the project and had to cancel scheduled trip to JIC. Negotiations of Form03 continued under guidance of the JIC contracts office.	An ICH-VAST scientist visited JIC for two weeks, where they received scientific training in GCMS, LCMS and transient expression.
Apr	End of two year funding from the Royal Society.	
May	A finalised and acceptable Form03 was produced by JIC contracts and sent to ICH-VAST.	
Jun	ICH-VAST requested language-based changes to Form03 e.g. 'contract' to 'agreement'. Form03 was modified to accommodate these changes (with acceptance of JIC contracts) and returned to ICH-VAST.	
Sept	ICH-VAST & JIC both accepted the terms of Form03.	
Oct	Despite acceptance by ICH-VAST collaborators, the Director of their institute was unwilling to sign Form03. A solution was suggested to sign Form03 with collaborator in their role as Director of ICNaP. JIC proceeded to change all documentation to accommodate this.	
Dec	It was determined that ICNaP had not yet secured the correct stamp and therefore signature from ICNaP would not be acceptable to the Vietnam Biodiversity Conservation Agency.	
2019	To resolve this ICNaP are currently attempting to secure a stamp.	

Collaboration between JIC (UK) and ICH-VAST (Vietnam). Additionally, in October 2018 a collaboration between an individual from ICH-VAST in their capacity as director as Interdisciplinary Centre for Natural Products Research (ICNaP) was formed. Both progress (black) and complications (red) are detailed along with outputs that could be considered benefits shared to providers (green). Updates from the Vietnam national focal point (blue) are also included.

Secondly, although it may seem trivial, the issue that eventually prevented progress was the difference in what is considered best practice for signing official documents. For instance, in the UK the signature of an official document is normally performed by the Head of Contracts. However, this was not sufficient in Vietnam and the signature of official documents had to be from the Director of the institute. Further, while in the UK a signature is normally sufficient, in Vietnam the signature must be accompanied by an official stamp and any copies of the document must be in full colour (Table 3.3.1). Although these issues are all minor and in most cases resolvable, their resolution takes thought, time and often a face-to-face discussion. Further, such issues are likely to be not only country-specific, but individual provider-specific. This means that generic guidelines intended to help the NP process may not be useful in resolving these individual cases. In a time-limited research project where material is needed quickly, dedicating the time to work through these issues without the certainty of receiving material is a high-risk strategy.

The infancy of NP implementation (Section 3.2.3) certainly compounded these issues. Vietnam is one of 17 countries to have issued Internationally Recognised Certificates of Compliances, although the earliest issue from Vietnam was in 2018, when the funding for this collaboration was coming to an end. Nonetheless, the Internationally Recognised Certificates of Compliances from Vietnam demonstrate that it is a relatively proactive Party of the NP and that securing in scope genetic resources from this country is possible. The Vietnam-specific NP guidelines were not available on the ABSCH website until 2017 and therefore time was spent drafting an MAT before the template MAT document (Form03) became available. The lack of information on the ABSCH (specifically National Competent Authority and National Focal Point) also caused further confusion regarding which Vietnam authority documentation should be submitted to (Table 3.3.1).

However, the major issue encountered was the lack of understanding of the NP at ICH-VAST (perhaps caused by inadequate efforts to raise awareness at this level). This lack of understanding meant that the motivation of how the NP could secure benefits and protect biodiversity (Figure 3.1.2) was not appreciated. Subsequently, the initiative and motivation to abide by the NP came from JIC and therefore at each step of the process (the majority of which had to be performed in Vietnam by ICH-VAST (Table 3.3.1)) JIC had to research procedure and encourage its actioning. Frequent staffing changes at ICH-VAST meant that explanations of the NP and its importance had to be reiterated. Further, the lack of awareness of the good intentions of the NP may be the reason why the director of ICH-VAST was eventually unwilling to sign the MAT.

Although in this example accessing genetic resources through the NP was

unsuccessful, not all aspects of the NP failed. As highlighted in Table 3.3.1, a number of non-monetary benefits have been shared with ICH-VAST. Therefore, this example could be summarised by the schematic shown in Figure 3.3.2, whereby no genetic resources were received for research because the negotiation of PIC and MAT created a barrier. However, as part of the collaboration, benefits flowed back to the providers. It is worth noting that similar benefits would have probably been shared in such a collaboration even in absence of the NP.
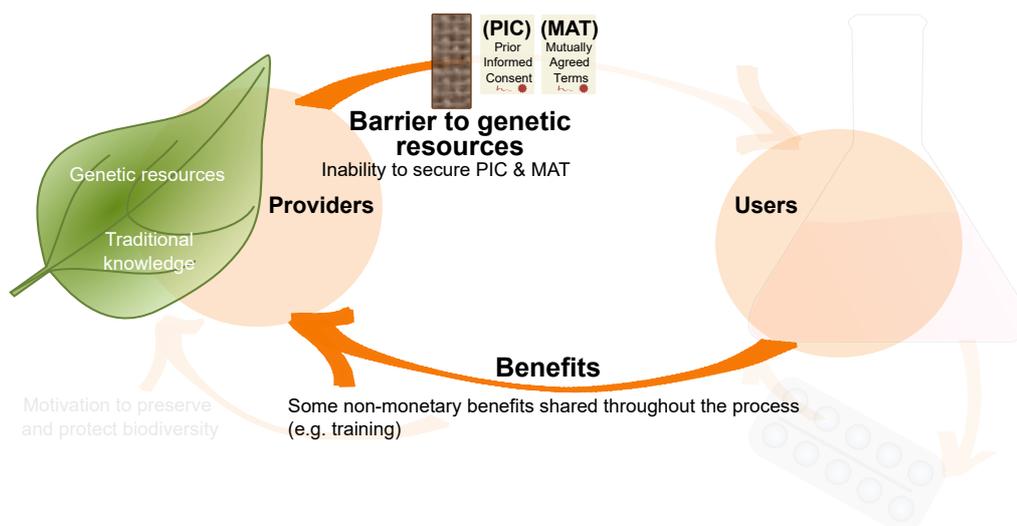
Figure 3.3.2: **Scheme illustrating when negotiation of PIC and MAT becomes a barrier.**

3.3.2 Accessing material that is out of scope of the NP

Although committed to attempting to access material through the NP, in July 2017 when it became apparent that resampling would be required (Table 3.3.1), there was a realisation that securing material in scope of the NP may not be successful within the time-frame of this project. It was therefore decided that material would have to be sourced that was out of scope of the NP for the continuation of this project.

As discussed (Section 3.2.2) there are a number of situations where, based on EU user legislation, non-native genetic resources in the UK are out of scope of the NP. Therefore a straight forward approach to accessing genetic resources was to contact UK nurseries to determine whether they had accessed genetic resources before the NP came into force. The Royal Horticultural Society website listed no UK supplier of *A. indica* at the time of searching. Therefore, the close relative of *A. indica* and second most prolific limonoid-producing Meliaceae species, *M. azedarach*, became the focus of attempts to access genetic resources. In 2017 there were nine UK based suppliers of *M. azedarach*. A template email was drafted (Appendix B.11) requesting

information about provenance and sent to these suppliers. Any suppliers detailed in the responses to these email were also contacted. The responses of the total of 12 businesses contacted are summarised in Table 3.3.2.

Table 3.3.2: **Summary of responses of out of scope suppliers.**

Supplier	Reply	Provenance	Date
Provender	Y	Purchased: unnamed French supplier	NA
Planfor	Y	Purchased: unnamed French supplier	NA
Brighton-Plants	Y	Purchased: Chiltern Seeds	NA
Plant Base	Y	Purchased: Sandeman Seeds	Pre Oct14
Sandeman Seeds	Y (OS)	Collected: India, Spain & Morocco	NA
Crûg Farm	Y	Collected: Japan	Post Oct14
Cross-Common	Y (OS)	NA	NA
Van de Berk	Y (OS)	NA	NA
Burncoose	Y	Purchased:Rein and Mark Bulk	Pre Oct14
Rein and Mark Bulk	Y (OS)	NA	NA
Botanica	N	NA	NA
Chiltern Seeds	N	NA	NA

Information includes whether they replied (Y) or not (N) and whether the species was out of stock (OS), the provenance they supplied and a date of access pre- or post-October 2014 (if supplied). Where response lacked enough detail 'NA' is specified. Information that could potentially be out of scope of the NP is highlighted (green).

Considering the only benefit to the suppliers contacted was the potential sale of a plant, and most customers would not ask questions of provenance, a surprising number of suppliers responded to this email and attempted to detail the provenance of their *M. azedarach* plants. A number of replies were too brief to be confident of the provenance of the plants (e.g. *'the seeds were from France'*). However, two suppliers had accessed *M. azedarach* before October 2014 (Table 3.3.2) and therefore according to EU legislation were out of scope of the NP.

A separate supplier (Crûg Farm Plants (199)) had collected seeds of *M. azedarach* themselves and were aware of the NP. This meant that they were willing and able to provide ample evidence about provenance for the purpose of NP due diligence. Although their collection of *M. azedarach* took place after the NP came into force, the collection was from Japan, a country not exercising its sovereign rights over genetic resources, and was therefore out of scope of the NP. Contact with the National Focal Point for Japan by email (Appendix B.12) confirmed that these accessions were suitable for use in research and that the Crûg Farm Plants sampling site (Figure 3.3.3) was not subject to sampling restrictions or alternate access and benefit sharing legislation.
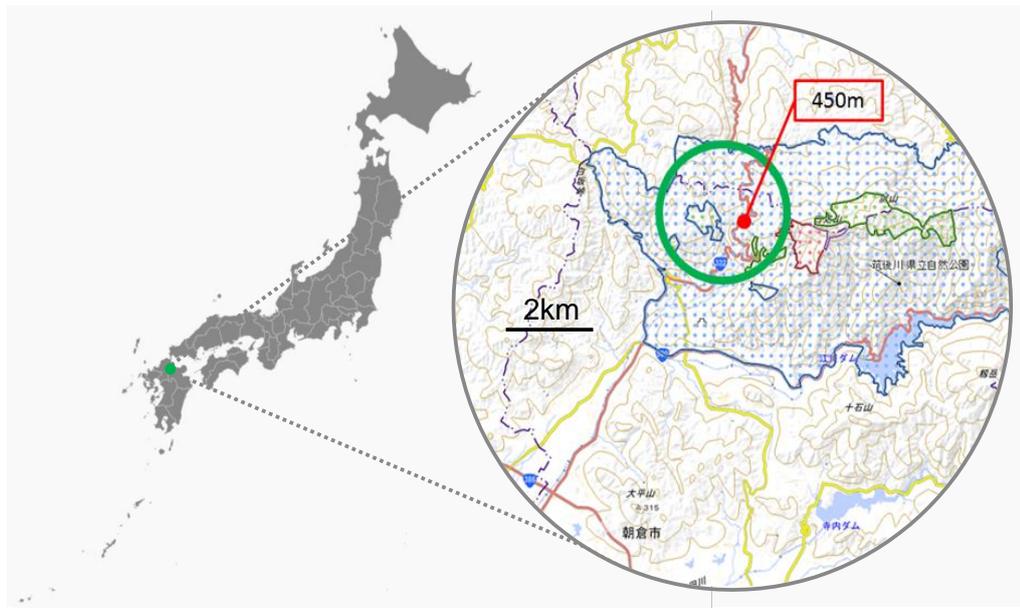
Figure 3.3.3: **Sampling site of *M. azedarach* seeds by Crûg Farm Plants (199) in Japan.**
The green circle shows the area of Chikugogawa Prefectural Natural Park where seeds were collected (indicated by Bleddyn Wynn-Jones of Crûg Farm Plants). The blue area is the 'Ordinary Zone' of the natural park which the National Focal Point of Japan confirmed is free from sampling restrictions (Appendix B.12). Map of Japan, Free Vector Maps; detailed map of Chikugogawa Prefectural Natural Park, Seishu Okuda National Focal Point of Japan.

The verification of the provenance by Crûg Farm provided the most convincing due diligence of all the suppliers. Therefore 16 individual *M. azedarach* plants (~6 months old) were ordered online and delivered to JIC, free of research restrictions, within two weeks of the initial search for genetic resources that were out of scope of the NP.

UK Botanical gardens were also contacted in an attempt to access Meliaceae genetic resources. Botanical gardens are estimated to contain ~30% of all plant species diversity (200) and are therefore in an interesting position with regard to the NP, although they are still obliged to abide by existing non-NP contracts that they have in place with providing countries.

Certain botanical gardens, such as Cambridge University Botanical Gardens (201), were quick to set up NP-compliant Material Transfer Agreements to facilitate transfer and use of out of scope material in non-commercial research. JIC signed such an agreement, as part of this project, to allow genetic resources from two Meliaceae species to be sourced from the garden in November 2016. A similar agreement was later signed between JIC and Kew Royal Botanical Gardens (202) to facilitate the sending of dried leaf material from 13 species in September 2017. Despite this, the *M. azedarach* plants sourced from Crûg Farm Plants were selected

as the focus for research on this project. Firstly, due to the high number of limonoids they produce and their close relation to *A. indica* and secondly, due to the availability of whole plants free from any restrictions of a Material Transfer Agreement.

The out of scope attempts to access genetic resources within this project were significantly simpler, quicker and more effective than attempts to access genetic resources in scope of the NP. Although the success of this out of scope approach is dependant on the species of interest being available, the large numbers of nurseries and botanical gardens in the UK mean it is highly likely that a plant of interest, or alternatively a close relative, will be accessible via this route. For a time-limited research project this approach represents a low-risk strategy of accessing genetic resources. However, the NP process has effectively been bypassed and there have been no negotiations, capacity building or sharing of benefits delivered back to the original providers as depicted in Figure 3.3.4.
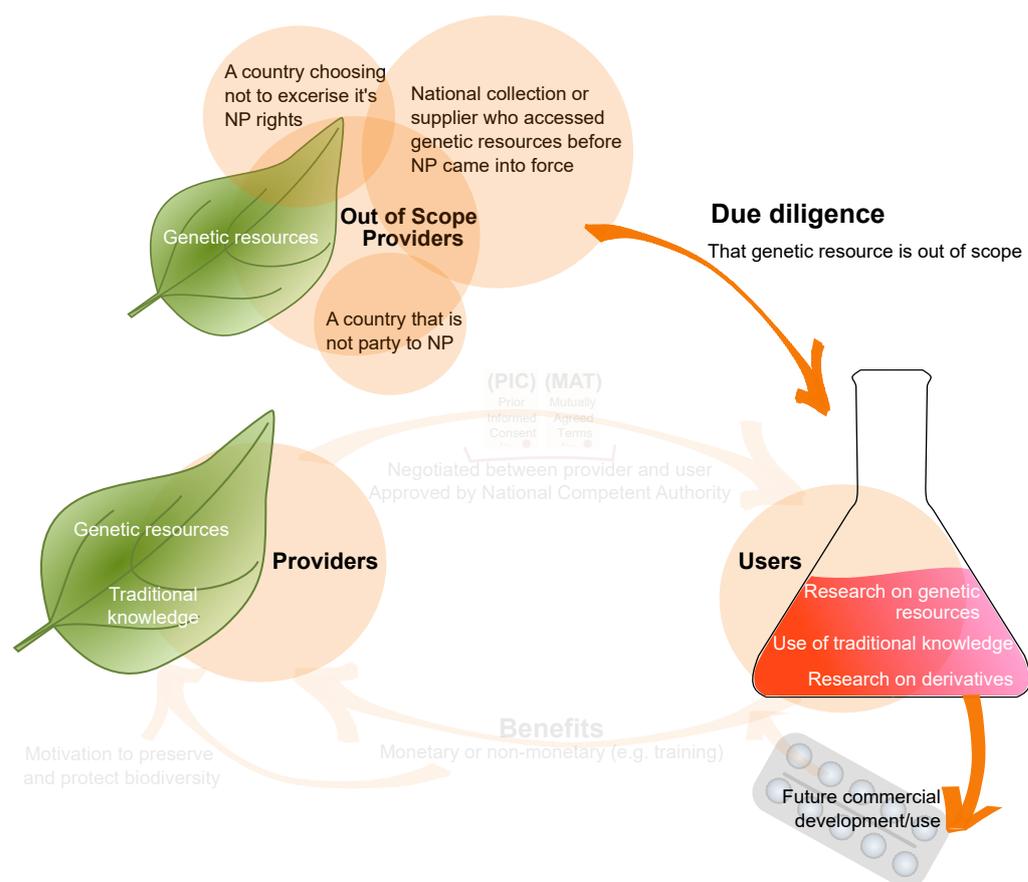
Figure 3.3.4: **Scheme illustrating accessing plant genetic resources by means out of scope of the NP.**

3.4 Different perspectives on the NP

The Global Garden workshops are a series of day-long workshops based on the established idea of Science, Arts, Writing (SAW) workshops (203, 204). SAW workshops aim to engage children with a particularly scientific topic by the design and delivery of a science, an art and a writing activity by an scientist, writer and artist, respectively. Scientific images from the field are used by the writer and artist to design creative activities which complement the scientific concept. SAW workshops have been demonstrated to be an effective way of engaging with the public on scientific subjects, including complex and potentially contentious issues (e.g. synthetic biology (205)).

The Global Garden workshops are day-long events which introduce case studies of the humans use of plants to participants from a range backgrounds, with the aim of sparking discussion. Case studies include the human use of: steviol glycosides, from *S. rebaudiana* (sweetleaf), as sweeteners (Figure 3.4.1); vanillin flavouring, from *Vanilla planifolia* (vanilla); the chemotherapeutic agents vinblastine and vincristine, from *Catharanthus roseus* (Madagascan periwinkle); the anti-malarial agent artemisinin, from *Artemisisa annua* (sweet wormwood); and the vaccine adjuvant QS-21, from *Quilliaja saponaria* (soapbark). The participants, as well as being invited to discuss these case studies, take part in laboratory practicals, such as extracting colourful pigments from plant material, as well as creating poetry and paintings.

During the Global Garden workshops some interesting anecdotal differences of opinion regarding the NP were raised. At a Global Garden workshop for members of the public without a scientific background, a strong theme throughout discussions was the unfairness of bioprospecting cases such as the widespread use of steviol glycosides (Figure 3.4.1) without benefits being shared. Therefore, although not discussed in detail, the NP was positively received as a potential way to stop the perceived threat of benefits not being shared with indigenous people.

Contrastingly, during a workshop with a group of scientists, the perceived risk was very different. The recurring discussion of the scientists was that the layer of regulation the NP represented was a clear barrier to the progress of research. Therefore the perceived threat was the barrier to the discovery and development of new natural products for human-use, that could have a positive impact on a wider population. At a different Global Garden workshop attended by a mixture of scientists and non-scientists, the discussion became largely side-tracked by the issue of digital sequencing information, after a laboratory practical where DNA was extracted from a strawberry.

Sweetleaf (*Stevia rebaudiana*)

Stevia rebaudiana (*S. rebaudiana*) is native to parts of Brazil and Paraguay. The leaves of the plant have been used as a sweetener (known as *ka'a he'e*) for thousands of years by the Guaraní peoples of South America.

In 1931, two French chemists successfully isolated chemicals known as steviosides, which are responsible for the sweetness of *S. rebaudiana* leaves. These chemicals are extremely potent—150 to 300 times sweeter than sugar. Use of steviosides as 'natural' sweeteners is now widespread.

In 2013, Coca Cola began using steviosides in a new product 'Coca-Cola Life,' promoting its 'natural' and 'calorie-free' properties.

Cultivation of *S. rebaudiana* has expanded to meet this demand. In 2011, 80% of all *S. rebaudiana* extract came from China.

Sweetleaf plant

The most commonly isolated steviosides have a bitter aftertaste that can impair flavouring. The Swiss company Evolva has recently produced 'better-tasting' steviosides in yeast and patented this process. In partnership with Cargill they plan to launch the new sweetener 'EverSweet' in 2018.

In recent years, the wide-spread use of steviosides has highlighted the issue of biopiracy, since despite the considerable revenue these chemicals bring to the food industry no benefits or royalties flow back to the Guaraní people. The representatives of the Guaraní, supported by international organisations such as Public Eye, continue to petition for some form of compensation.

Scanning electron micrograph of the surface of a human tongue

Stevioside molecular model

Figure 3.4.1: **Example Global Garden case study on human use of *S. rebaudiana*.**

The Global Garden case studies were also presented to a group of Key Stage 4 students taking part in a Brilliant Club (206) project *'Plants: The green chemists'*. The case studies were introduced at the end of session about the vast variety of plant natural products and their usefulness to humans. Interestingly, this group showed the most balanced discussion with most participants recognising the sensitivities associated with the origins of plants, but also the need for research on plant natural products not to be limited. These examples demonstrate how the introduction of a complex multifaceted issue, such as the human use of plants and its regulation by the NP, can easily be influenced by prior knowledge, emotion, and the nature of the discussion.

Perhaps, in terms of different perspectives on the NP, the most important consideration is not the opinion of different stakeholders, but the knowledge of the NP's existence. Although it is difficult to determine the true prevalence of this issue, based on anecdotal comments and conversations with other scientists, there is certainly a population of scientists who are unaware of the NP's existence. Further, there are certain sectors of biology-based industry, such as cosmetics, where a lack of awareness of the NP and access and benefit sharing has been cited (169, 207). This issue is exacerbated by the fact that the providers of the genetic resources may also lack knowledge of the NP, as demonstrated (Section 3.3.1).

3.5 Conclusion

Based on the complexities of the NP discussed here and on first-hand experience of attempting to access material, it appears at present that the NP is not achieving its aims. In this project, when attempting to access genetic resources in a time- and financially- conscious manner, the strategy with the lowest risk and highest chance of success was to access genetic resources that were out of scope of the NP (Section 3.3.2). Three years of trying to negotiate access to material in scope of the NP failed, whilst a two week time-frame delivered out of scope genetic resources which were free of restrictions. However, this route of access did not provide benefits to original providers through NP-based mechanisms. Further, it must be considered that accessing material through this out of scope route was only possible by the specifications of EU-user legislation, which are currently under review.

If scientists were expected to only access genetic resource which are in scope of the NP, then at present, this would create a clear barrier to research. This is in contrast to early optimism that the NP could help remedy the uncertainty of biology-based research post the Convention on Biological Diversity, by providing clear rules to follow (169, 208). The impacts of such a barrier are two fold. Difficulties in access to genetic resources may prevent the development of novel products that could benefit humanity whilst simultaneously reducing the amount of benefits flowing back to provider countries and communities. It has been suggested that the NP negotiations may indeed have raised unrealistic expectations of what benefits could arise from individual genetic resources (209), further compounding the problem. There is also a threat that this barrier could direct research away from the most interesting species and towards the most accessible species. This project is an example of such a redirection as, due to barriers created by the NP, investigation of azadirachtin biosynthesis is being performed in *M. azedarach*, a species which is reported to produce precursors (123), but not azadirachtin itself (128).

The counter argument to these points is that the NP is still in its infancy (Section 3.2.3) and in the future these teething issues will be resolved. However, rapid uptake and dissemination of the NP seems unlikely when considering the current speed that legislation is being uploaded to the ABSCH website (Table 3.2.1).

To conclude, cases of past unfair bioprospecting and public opinion have shown that there is a clear need for legal rules and a framework to allow fair access to genetic resources. However, the NP, as it stands, fails to achieve its goals either because of legislation enabling the bypassing of the benefit sharing cycle (Figure 3.3.4) or due to creation of a regulatory barrier which could slow or divert the progress of biology-based research (Figure 3.3.2).

4

Profiling of limonoid biosynthesis in *M. azedarach*

Synopsis

In contrast to the well-known and studied neem tree (*A. indica*) there is a lack of knowledge regarding its close relative chinaberry (*M. azedarach*). Over 100 limonoids have been isolated from *M. azedarach*, however attempts to profile the tissues they occur within is currently limited to one limonoid, toosedanin, which was detected in all tissues assessed. This chapter describes the profiling of *M. azedarach* to identify sites of limonoid biosynthesis. The accumulation of the *seco*-C-ring limonoid salannin, along with the early protolimonoids melianol and dihydroniloticin, have been compared between individuals and tissues of *M. azedarach*. The expression of known melianol biosynthetic genes (*MaOSC1*, *MaCYP71CD2* and *MaCYP71BQ5*) have also been evaluated across tissues. Together these experiments revealed a variation in limonoid accumulation between individual *M. azedarach* plants and higher levels of accumulation in petiole and root tissues than leaves. This informed the design of a RNA-Seq experiment to aid the identification of candidate genes involved in post melianol biosynthesis.

4.1   Introduction

Chinaberry (*M. azedarach*) was selected for this work due to its availability for research (Chapter 3), the presence of a pre-existing transcriptome (Chapter 2) and its production of the second highest number of limonoids within the Meliaceae family (5). A recent phylogenetic analysis of 254 species from across the genera of the Meliaceae family (31) (summarised in Figure 4.1.1) shows *M. azedarach* to be phylogenetically close to *A. indica* (neem) occurring within the Melioideae subfamily. Although emu apple (*Owenia acidula*) is phylogenetically closer to *A. indica*, metabolite analysis and sequencing data are not available for members of this Australia-specific Owenia genera (210). Therefore, *M. azedarach* represents the best species for characterisation of the biosynthetic pathway of the anti-insect *seco*-C-ring limonoids, over 40 of which have been isolated from *M. azedarach* (5, 6).
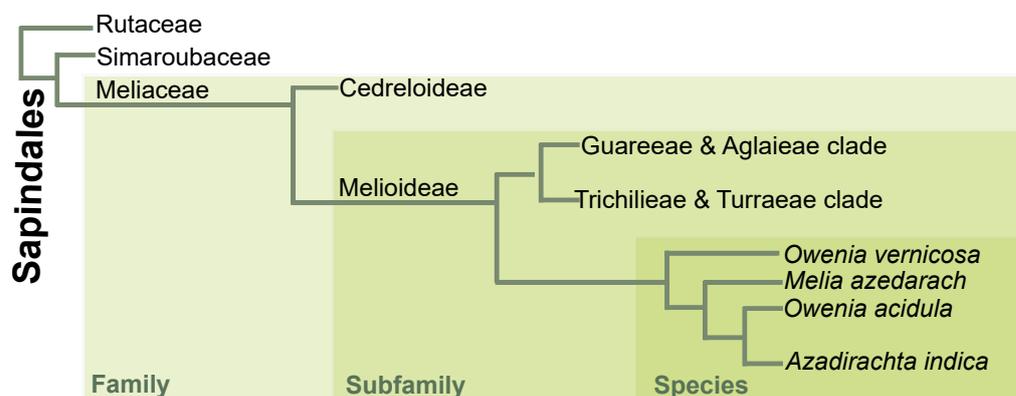
Figure 4.1.1: **Phylogeny of subfamilies within the Meliaceae.**
Positioning of genera containing *M. azedarach* and *A. indica* within the Meliaceae family, Sapindales. The phylogeny is based on predictions using internal transcribed spacer (ITS) sequences of the *rbcL, matK, rps16* and *ycf1* genes (31).

While *A. indica*, which produces ~150 limonoids, is by far the most prolific limonoid producing plant species, *M. azedarach* is also reported to produce a relatively high number of limonoids (>100) in comparison to other members of the Meliaceae. *M. azedarach* has a similar profile of limonoids to that of *A. indica*, largely consisting of *seco*-C-ring and ring-intact limonoids (Figure 4.1.2). The similarity in the types of limonoids produced by *A. indica* and *M. azedarach* is significant when the vast diversity of limonoids within the Meliaceae family as a whole is considered. Meliaceae species in different subfamilies can produce markedly different limonoids to the *seco*-C-ring limonoids that are the focus of this project. For instance, the *seco*-B-ring limonoids of the Cedreloideae subfamily and the *seco*-A,B-ring limonoids of the Guareeae and Aglaieae clades are distinctly different from the *seco*-C-ring limonoids of the Azadirachta and Melia genera (Figure 4.1.3).
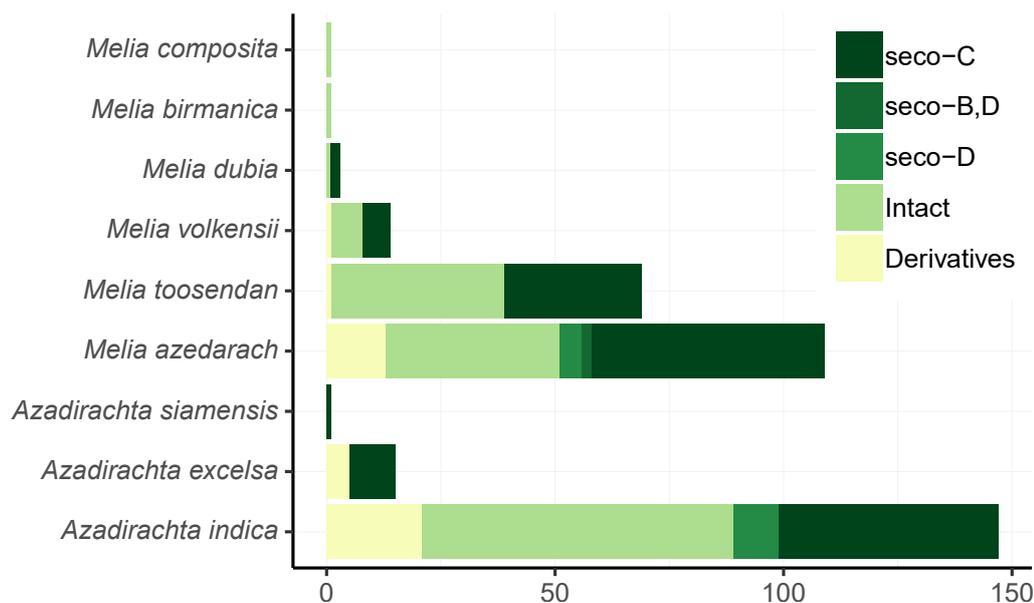
Figure 4.1.2: **Profiles of limonoid-type from *Melia* and *Azadirachta* species.**
Reported occurrence of limonoids within *Melia* and *Azadirachta* species based on recent reviews (5, 6). Limonoid counts are shaded by type (*seco*-C-ring, *seco*-B,D-ring, *seco*-D-ring, ring-intact limonoids or derivatives).

There are contrasting reports as to whether azadirachtin itself is produced by *M. azedarach* (128, 211). However, the meliacarpins, isolated from *M. azedarach*, have a highly similar structure to azadirachtin and share its anti-insect activity. For example, 1-cinnamoyl-3-acetyl-11-hydroxymeliacarpin has a reported LC50 of 0.48 ppm±0.21 on neonate larvae of egyptian cotton leafworms (*Spodoptera littoralis*), which is comparable to the reported LC50 of 0.32 ppm±0.13 for azadirachtin (212). Further, the difference in structure of azadirachtin and 1-cinnamoyl-3-acetyl-11-hydroxymeliacarpin (Figure 4.1.4) is only in the decoration of the A-ring (the C1 tiglate group is replaced by a cinnamate and the C29 position lacks a methoxycarbonyl group) (212).

Whilst the production of most meliacarpins appears to be specific to *M. azedarach* (5), the less complex *seco*-C-ring limonoid, salannin, is produced by *A. indica*, *M. azedarach* and other species in this genera (5). A feeding study has proven that salannin can act as precursor to azadirachtin, as ^{14}C-labelled salannin fed to *A. indica* seed kernels was converted to azadirachtin (0.8 % incorporation rate). A feasible biosynthetic conversion of salannin to azadirachtin would be possible in ~9 biosynthetic steps. Further, salannin itself has reported anti-insect effects, with feeding inhibition reported in species from the Lepidoptera, Diptera and Coleoptera Orders (as reviewed (14)).

Arguably for identification of the site of biosynthesis, the reporting of pathway intermediates (protolimonoids) is more crucial than limonoids themselves. The
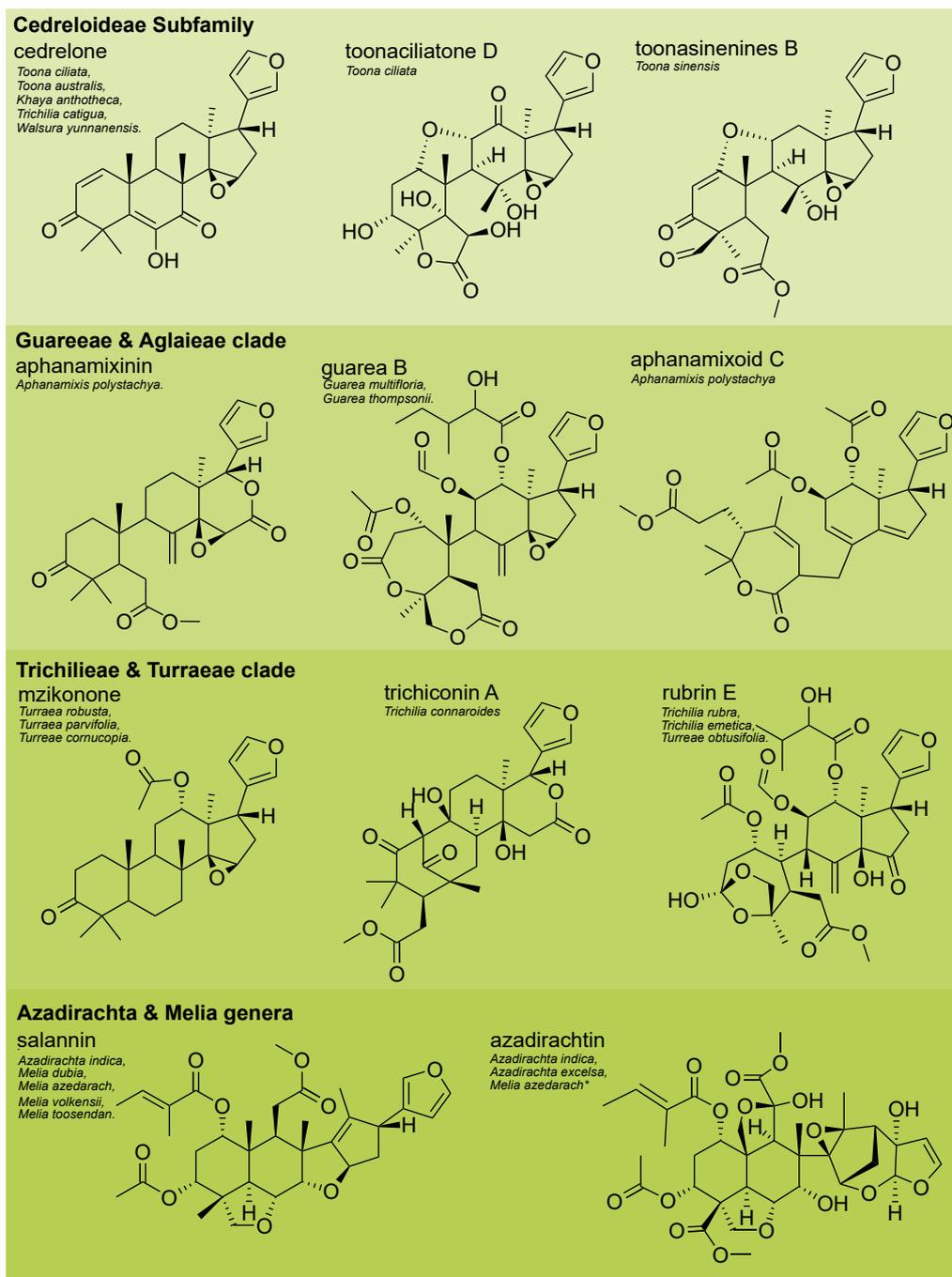
Figure 4.1.3: **Example structures of subfamily-specific Meliaceae limonoids.**
Examples of limonoid structures from the Cedreloidaeae subfamily, Guareeae/Agliaeae clade, Trichilieae/Turraeae clade and Melia/Azadirachta genera (phylogeny summarised in Figure 4.1.1). Information regarding occurrence and species of isolation is based on recent reviews (5, 6). Asterix (*) indicates reported isolation of azadirachtin from *M. azedarach* (211), which has been recently disputed (128).
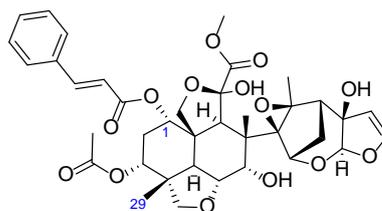
Figure 4.1.4: **Structure of 1-cinnamoyl-3-acetyl-11-hydroxymeliacarpin.**

isolation and occurrence of protolimonoids has been subject to less investigation than limonoids. Species highly investigated for limonoid content, such as *M. azedarach*, may have comparatively few reports of protolimonoids. Nonetheless, both melianol and its C3 ketone melianone have been reported from *M. azedarach* (Appendix B.10), along with seven additional protolimonoid structures (kulinone, 3-α-tigloylmelianol, 21-β-acetoxymelianone, methyl kulonate, 16β-hydroxytirucalla-7,24(25)-dien-3-oxo-21,23-olide, 21α-methylmelianodiol and 21β-ethylmelianodiol) from the fruit and bark of this species (213–216).

Despite >60 studies identifying 109 limonoids and numerous other secondary metabolites in *M. azedarach*, there is a lack of understanding regarding where these limonoids occur. This is in contrast to *A. indica* where traditional knowledge and use dictates that limonoids occur in greatest abundance in the developing fruit and seed (55). The only limonoid profiled in *M. azedarach* is the ring-intact limonoid toosendanin, which was present in all tissues investigated (leaf and fruit of varying ages) (217).

The current lack of tissue specific knowledge of limonoids in *M. azedarach* limits the ability to characterise plant natural product pathways, as the site of biosynthesis (spatial or temporal) is crucial to inform RNA-Seq experiments and allow identification of candidate genes by differential expression analysis. Investigations of this type have elucidated the podophyllotoxin pathway in *Podophyllum hexandrum* (mayapple) (218) and mogroside V pathway in *Siraitia grosvenorii* (monk fruit) (219).

4.1.1   Aims

The accumulation of *seco*-C-ring limonoids and the availability of both sequencing data and material for research (Chapter 3) recommendeds *M. azedarach* as the species with which to continue investigations of limonoid biosynthesis. The availability of standards (Figure 4.1.5) created an opportunity to profile the accumulation of two protolimonoids and the three limonoids within this species. Further, the characterisation of three melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5* (Chapter 2)) enabled expression levels to be profiled. Therefore the aims of this chapter are:

1. Analysis of the limonoid content of different *M. azedarach* individuals.

2. Analysis of the limonoid content of different tissues of *M. azedarach.*

3. Design of a RNA-Seq experiment comparing low to high limonoid tissues.
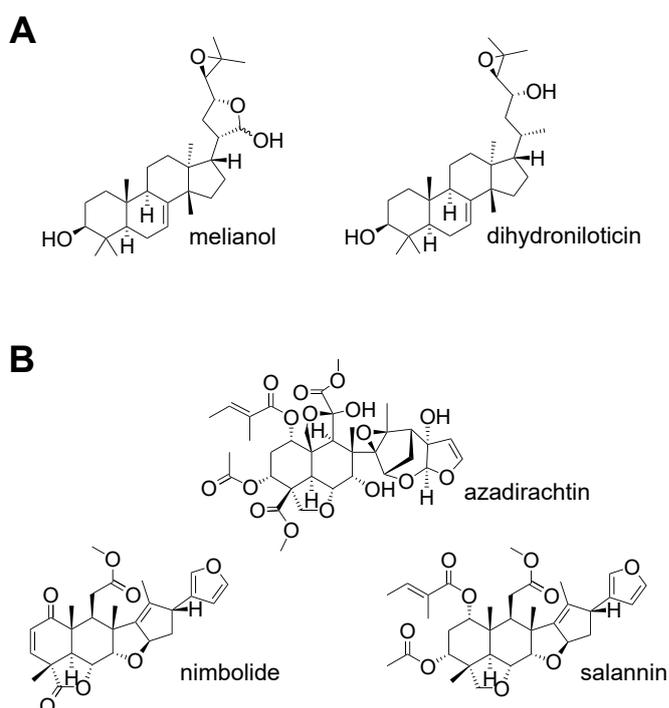
Figure 4.1.5: Structures of standards available for metabolite profiling.
(A) The protolimonoids melianol and dihydroniloticin purified from heterologous expression in *N. benthamiana* (Chapter 2). (B) *Seco*-C-ring limonoids azadirachtin and nimbolide (purchased from Sigma-Aldrich), and salannin (purchased from Greyhound Chromatography).

4.2   Results and Discussion

4.2.1   Analysis of the limonoid content of different *M. azedarach* individuals

The leaves (Figure 4.2.1) of 16 young individual *Melia azedarach* plants (< 30cm tall, 4-5 leaves/plant) were used for metabolite analysis. Leaves were harvested, freeze-dried and limonoids were extracted using a method developed for extraction from *N. benthamiana* (Chapter 9). The resultant limonoid extracts were analysed by UHPLC-IT-TOF using available standards (Figure 4.1.5) to confirm retention times and mass spectra (example trace provided in Appendix B.13).

Figure 4.2.1: **Example young (~6 months) *M. azedarach* leaf.**

Salannin, melianol and dihydroniloticin were detectable in the leaf material of each of the 16 individual *M. azedarach* plants (Figure 4.2.2). Identification of salannin was expected based on its reported isolation from *M. azedarach* and other Melia species (*M. dubia, M. volkensii* and *M. toosendan*) (5). The identification of melianol from *M. azedarach* had also been previously reported (220), however the identification of dihydroniloticin within this species has not previously been reported. Azadirachtin was not detected in *M. azedarach* under the conditions used in this analysis. Reports of azadirachtin production in *M. azedarach* vary (128, 211) and therefore its production may be accession, tissue or age specific. It seems most likely that azadirachtin production in *M. azedarach* is tissue-specific as the only reported isolation utilised fruits (211). Nimbolide was also not detectable, consistent with its reported isolation from *A. indica* and *A. excelsa* alone (5).

Relative quantification of salannin, melianol and dihydroniloticin was performed using an internal standard (podophyllotoxin) (Figure 4.2.2). Estimated levels of all three metabolites varied between individuals with values of salannin ranging from 0.05-0.5 mg/g DW, melianol from 0.005-0.075 mg/g DW and dihydroniloticin from
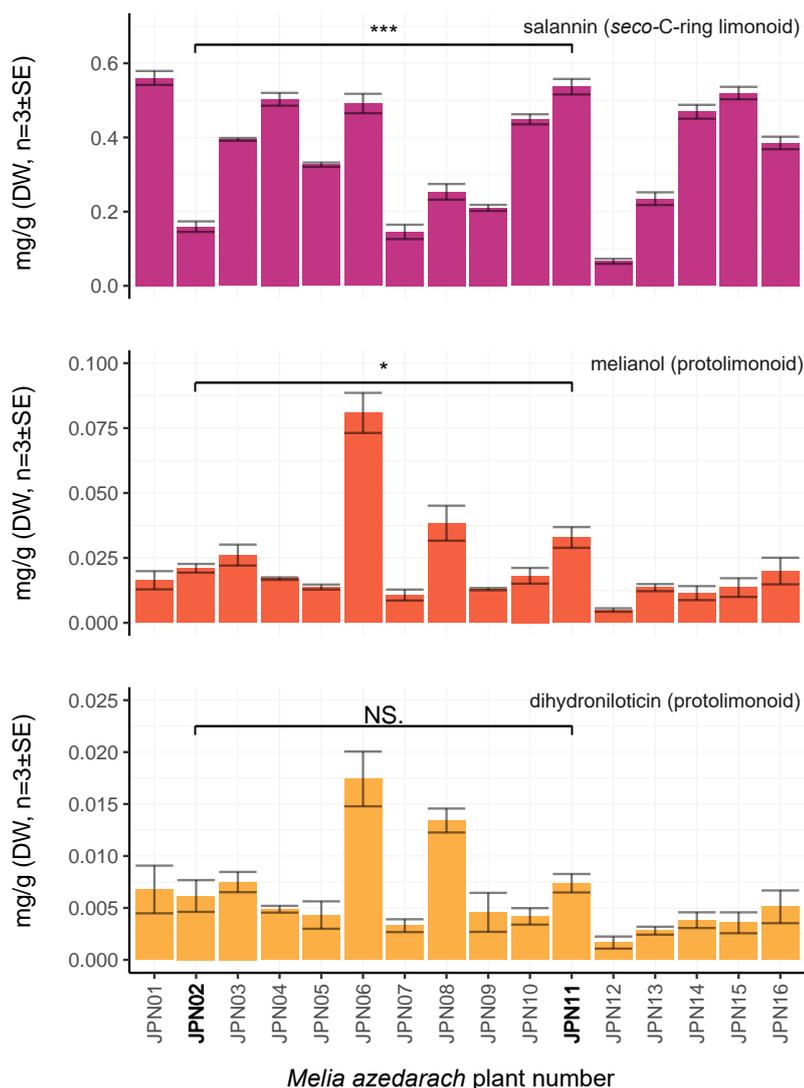
Figure 4.2.2: **Relative accumulation of salannin (purple), melianol (orange) and dihydroniloticin (yellow) in *M. azedarach* individuals.**
Estimated limonoid content (mg/g, DW, n=3±SE) of leaves from 16 *M. azedarach* individuals (~6 months old). T-test significance values are indicated for the individuals chosen to use in further experimentation (JPN02 and JPN11): not significant (NS) and P-value ≤ 0.05 ($*$), 0.01 ($**$), or ≤ 0.001 ($***$).

0.001-0.0015 mg/g DW. Accumulation of protolimonoids (melianol and dihydroniloticin) was in general low, with the exception of individuals JPN06 and JPN08, which accumulated slightly higher levels. It was clear from this analysis that variation in limonoid and protolimonoid levels is high between different *M. azedarach* individuals, even those with a shared provenance. Based on salannin content and on the health and size of plants, JPN02 and JPN11 were selected for further work as representative low and high salannin producers, respectively.

4.2.2 Analysis of the limonoid content of different tissues of *M. azedarach* individual JPN11

Within the selected high salannin individual, JPN11, tissue-specific levels of limonoids and protolimonoids were also determined via the same methods. At the time of experimentation the tissue types available to profile were leaves (leaflet only), petioles (including rachis) and roots. The highest relative levels of salannin were detected in the root tissue followed by the petiole tissue, which both accumulated significantly higher levels than leaf tissues (Figure 4.2.3). Further, the highest levels of both the protolimonoids (melianol and dihydroniloticin) were observed in the petioles which were significantly higher protolimonoid accumulators than the leaves (Figure 4.2.3). Azadirachtin and nimbolide were again not detectable in these samples corroborating previous results and literature reports.
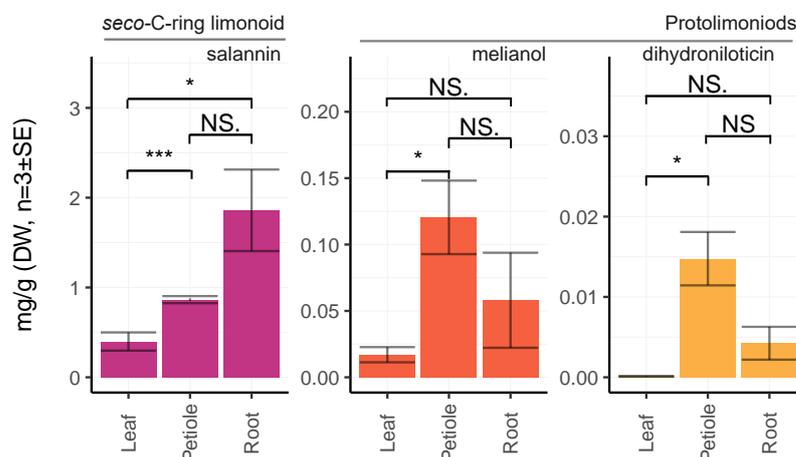
Figure 4.2.3: **Relative accumulation of salannin (purple), melianol (orange) and dihydroniloticin (yellow) in *M. azedarach* JPN11 tissues.**
Estimated limonoid content (mg/g DW, n=3±SE) of extracts from leaf, petiole and root tissue of *M. azedarach* JPN11 ($\sim$ 9 months old). T-test significance values are indicated: not significant (NS), P-value ≤ 0.05 ($*$), 0.01 ($**$), or ≤ 0.001 ($***$).

The detection of salannin in all tissues assessed is consistent with the the only other limonoid profiled in *M. azedarach*, toosedanin (217), which also was detectable in all tissues. The ideal accumulation of salannin from the perspective of uncovering biosynthetic pathways would be for one tissue to be accumulating high levels and another displaying a clear absence of production. This would allow differential gene expression analysis to identify the presence of limonoid biosynthetic genes by using the non-producing tissue as a negative comparator. In this case, the significant difference in levels of salannin and protolimonoid accumulation between tissues (Figure 4.2.3) suggests that leaf-root and the leaf-petiole comparisons should provide sufficient differences for identifying limonoid biosynthetic of genes.

4.2.3   Design of a RNA-Seq experiment comparing low to high limonoid tissues from *M. azedarach*

Based on the knowledge gained from basic profiling of *M. azedarach* an experimental design was constructed for a RNA-Seq experiment to elucidate limonoid biosynthetic gene candidates by differential expression analysis (Table 4.2.1). The design aimed to generate the most possible negative to positive comparisons between low and high salannin and protolimonoid containing tissues and individuals.

Table 4.2.1: **Overview of RNA-Seq experimental design.**

	Individual	Tissue	Description	*n*
1	JPN11	Upper Leaf	Middle leaflets from uppermost fully open leaf	4
2	JPN11	Lower Leaf	Middle leaflets from fourth fully open leaf	4
3	JPN11	Petiole	Petiole and rachis of uppermost leaf	4
4	JPN11	Root	Thin young furthermost roots	4
5	JPN02	Upper Leaf	Middle leaflets from uppermost fully open leaf	4
6	JPN02	Lower Leaf	Middle leaflets from fourth fully open leaf	4
7	JPN02	Petiole	Petiole and rachis of uppermost leaf	4

The individual, tissue, number of biological replicates (n) and a description of sampling are detailed. Image of example leaf sampling is provided in Appendix B.14.

Samples were collected and RNA extracted based on this experimental design. A proportion of the extracted RNA was sent to the Earlham Institute (EI) for sequencing (as discussed in Chapter 5) and a proportion retained for expression analysis by Quantitative Real-Time Reverse Transcription PCR (qRT-PCR). In parallel, the limonoid content was estimated for the RNA-Seq samples (Figure 4.2.4).

Significant differences in salannin and protolimonoid accumulation were observed between the RNA-Seq samples (Figure 4.2.4), suggesting that the corresponding sequencing data should be useful for identifying differentially expressed limonoid biosynthetic genes. T-test comparisons between individuals revealed that the levels of salannin in JPN11 and JPN02 leaf tissues were significantly different (P-value = 0.0000075), as already confirmed (Figure 4.2.2). This was also reflected in the differing accumulation of salannin between the petioles of these individuals (P-value = 0.0075). A significantly higher accumulation of protolimonoids was observed in the petioles compared to the leaves of both individuals (Figure 4.2.4), as previously seen for JPN11 (Figure 4.2.3), with the exception of highly variable dihydroniloticin levels in the petioles of JPN02. Within JPN11 there was significantly higher accumulation of salannin in root compared to leaf tissues (Figure 4.2.4), as previously described (Figure 4.2.3).
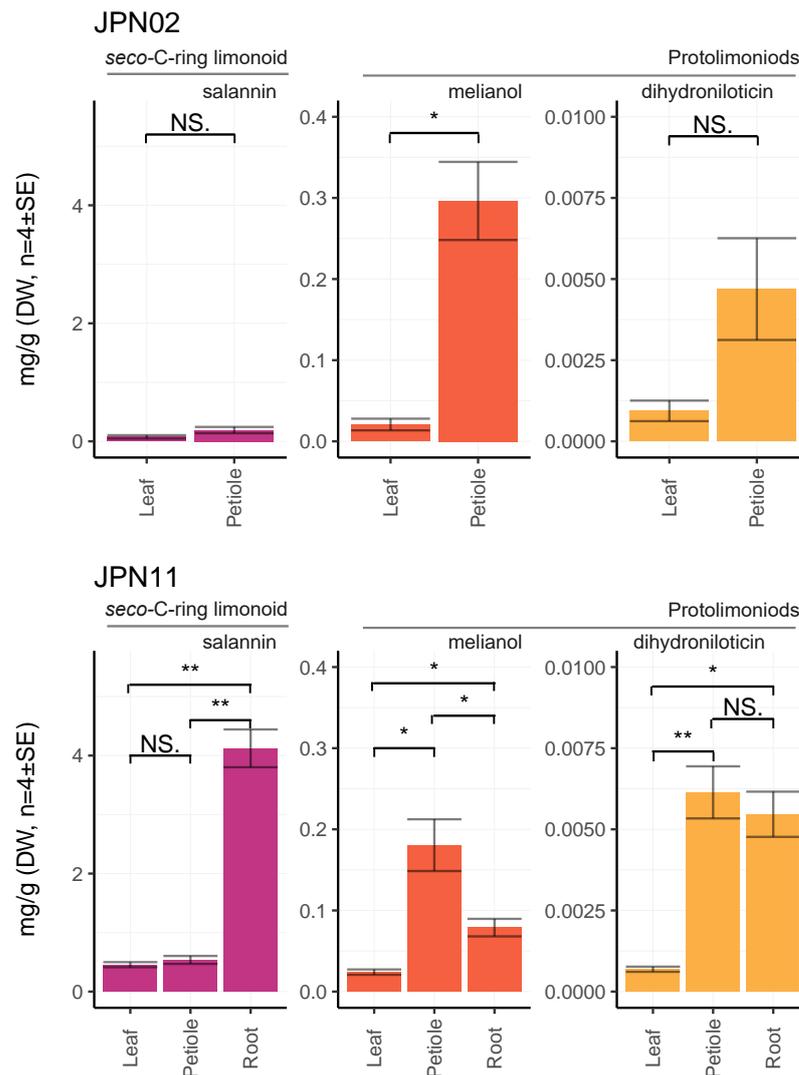
Figure 4.2.4: **Relative accumulation of salannin (purple), melianol (orange) and dihydroniloticin (yellow) across *M. azedarach* RNA-Seq samples.**
Estimated limonoid content (mg/g DW, n=4± SE) in *M. azedarach* tissue used for RNA-Seq analysis (leaf and petiole of JPN02 and leaf, petiole and root of JPN11). T-test significance values are indicated: not significant (NS), P-value ≤ 0.05 (∗), 0.01 (∗∗), or ≤ 0.001 (∗ ∗ ∗).

A subset of RNA was used for cDNA synthesis and qRT-PCR to determine the expression levels of known melianol biosynthetic genes (Chapter 2). The resulting expression pattern of melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*) (Figure 4.2.5) is reflected in the observations of limonoid and protolimonoid accumulation (Figure 4.2.4).

In both JPN11 and JPN02, expression of all three melianol biosynthetic genes was significantly higher in the petioles compared to the leaves (Figure 4.2.5). This
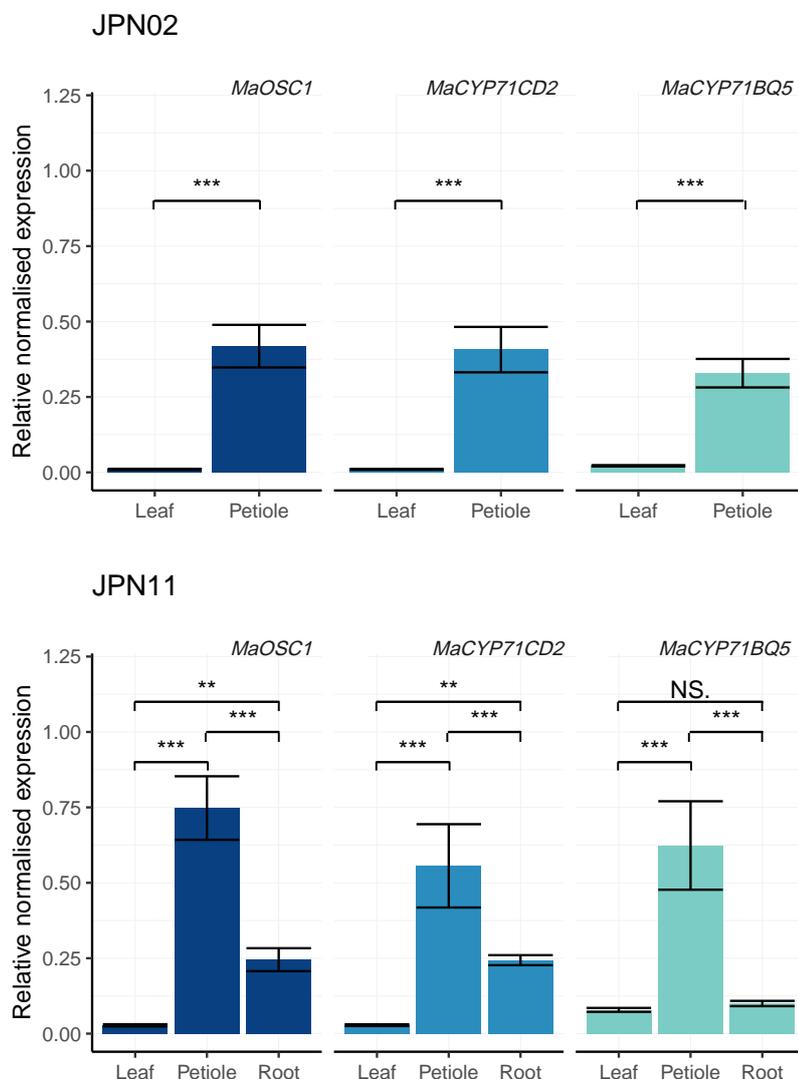
Figure 4.2.5: **Relative expression levels of melianol biosynthetic genes.**
Normalised expression of melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and
MaCYP71BQ5), relative to *Maβactin*, in RNA from *M. azedarach* tissues used for RNA-
Seq. Tissues include the leaf and petiole tissues of JPN02 and leaf, petiole and root
of JPN11. Three biological replicates and three technical replicates were included
(n= 9±SE) and relative expression levels were calculated by ΔΔCq method. T-test
significance values are indicated: not significant (NS), P-value ≤ 0.05 (∗), 0.01 (∗∗),
or ≤ 0.001 (∗ ∗ ∗).

indicates that the high level of protolimonoid accumulation in petioles is reflected, and potentially caused by, high melianol biosynthetic gene expression. This further supports the usefulness of petiole-leaf comparisons, particularly for protolimonoid and early limonoid biosynthetic pathway elucidation. Additionally, with the exception of *MaCYP71BQ5*, there was a significant difference in the level of expression of melianol biosynthetic genes between leaf and root tissues (Figure 4.2.5), which corroborates the occurrence of salannin being high in root tissues and supports to use of root-leaf comparisons.

Finally, between leaf tissues of JPN11 and JPN02 there were significant differences in expression of all melianol biosynthetic genes (P-value < 0.001) (Figure 4.2.5), which is inline with the significant difference in salannin observed between leaves of JPN11 and JPN02 and suggests that this could also be used as a comparison if required.

4.3   Conclusion

Here profiling of available individuals and tissues of *M. azedarach* plants has been performed. The previously reported production of the *seco*-C-ring limonoid salannin and protolimonoid melianol by *M. azedarach* has been confirmed in the individuals investigated. Along with the additional detection of low levels of the protolimonoid dihydroniloticin.

The variation in salannin and protolimonoid content was found to be high between these individuals, despite their shared provenance and age. Further, comparison of available *M. azedarach* tissues revealed significant differences that were consistently identifiable (Chapter 2). Together these significant differences suggested that petiole-leaf, root-leaf and JPN11:leaf-JPN02:leaf may all represent useful comparisons for the identification of candidate genes by differentially expression.

These significant differences provided a basis for designing a RNA-Seq experiment to maximise low-high comparisons of limonoids and protolimonoids within available *M. azedarach* tissues. Analysis of the limonoid content and relative expression of melianol biosynthetic genes within RNA-Seq samples, suggests this dataset has successfully captured differences in limonoid accumulation and expression of relevant biosynthetic genes. Therefore, this RNA-Seq dataset will represent a crucial tool for the identification of limonoid biosynthetic genes by differential expression analysis.

4.4   Chapter-specific materials and methods

4.4.1   Limonoid extraction and analysis from *M. azedarach*

Limonoid extraction from *M. azedarach* was performed using the methanol extraction methods described for *N. benthamiana* (Chapter 9). For more fibrous tissue, such as roots and petioles, the homogenisation step (Tungsten Carbide Beads (3 mm; Qiagen) in TissueLyser (1000 rpm, 1 min)) was repeated until a fine powder was yielded.

Analysis of extracts was performed by UHPLC-IT-TOF using the 'limonoid' gradient and quantified as described (Chapter 9).

4.4.2   RNA extraction

For RNA-Seq, four replicates of each sample were harvested and RNA was extracted following the MacKenzie hybrid method (Chapter 9). Quality control was performed by NanoDrop and gel electrophoresis to ensure samples met the required standard (> 3ng, $260/280 \geq 2$, $260/230 \geq 1.8$) for sequencing by EI.

4.4.3   qRT-PCR

To design intron-spanning primers in the absence of a *M. azedarach* genome an assumption was made that the intron patterning would be similar to homologous genes in the close relative *A. indica*. Subsequent alignment of genes to the draft genome of *A. indica* (PRJNA176672:AMY00000000.1, (127)) was used to design primers. A number of primers were designed to test the efficacy of different control genes (*MaEF, Ma18srRNA, MaGADPH* and *Maβ-actin*). However, when assayed, the *MaEF* primers were unsuccessful (no detection) and both *Ma18srRNA* and *MaGADPH* primers produced high Cq values. Therefore the *Maβ-actin* primers, which produced lower Cq values, were deemed the most effective primers and used for future qRT-PCRs.

Lightcycler 480 SYBR Green I Mastermix (Roche) was used for qRT-PCR which was performed on CFX96 real-time system and C1000 touch thermal cycler (BioRad). R (221) was used to calculate relative expression of genes compared to Maβ-actin using the $\Delta\Delta$Cq method (222) and results were plotted using ggplot2 (223).

5

Generation of a high-quality *M. azedarach* genome sequence

Synopsis

Genome sequencing within the Meliaceae family has thus far been confined to neem (*A. indica*). Of the two separate attempts to sequence *A. indica*, one was limited by a low level of assembly and the other has only very recently been made publicly available. To address this, here the generation of the first pseudochromosome level assembly of a Meliaceae species (chinaberry; *M. azedarach*) is reported. The data from previous RNA-Seq experiments within the Meliaceae family is of limited use for expression based identification of candidate genes because these experiments either lack biological replicates or focus on only one tissue type. Here, the generation of RNA-Seq datasets for different *M. azedarach* tissues has addressed this. The RNA-Seq data has also been utilised to support a structural and functional genome annotation (EIv1) of *M. azedarach*. Together the genome and transcriptome resources generated for *M. azedarach* will aid the identification of further candidate limonoid biosynthetic genes.

5.1   Introduction

A comprehensive and high-quality sequence resource for a Meliaceae species has thus far not been publicly available. Three protolimonoid biosynthetic genes have been characterised from the available sequencing resources within the Meliaceae family (Chapter 2). However, this was reliant on *de novo* transcriptome assemblies with a lack of RNA-Seq replicates, which limited the comprehensiveness of the analysis and reduced the capacity to identify further limonoid biosynthetic genes.

Currently, the only limonoid producing species with publicly available and comprehensive sequence resources are members of the the Rutaceae (Citrus) family. The Citrus Genome Database (139) houses the genomes of 13 citrus species, of which seven are highly assembled and annotated. Utilising citrus resources to investigate limonoid biosynthesis may be of value for identifying genes responsible for converting protolimonoids to shared ring-intact limonoid structures, as these steps that are likely to be conserved between Rutaceae and Meliaceae species. However, citrus species will not be able to inform on subsequent downstream pathway steps to the bioactive *seco*-C-ring limonoids (e.g. azadirachtin and nimbolide) that are Meliaceae-specific. This highlights the need for a high-quality sequencing resource within the Meliaceae family.

The only genome sequence from the Meliaceae family to date, is for *A. indica* (neem). The substantial and sustained scientific interest in *A. indica* (16), along with its cultural significance in India (224), led to it being the thirty-forth plant species to be sequenced (126, 225). The initial *A. indica* genome, completed by the Ganit Labs (Bengaluru, India), was generated using four different sequencing technologies, but relied heavily on Illumina short-reads for the final assembly by SOAPdenovo (226).

In 2016 Ganit Labs revisited this assembly, incorporating longer Pacific Bioscience (PacBio) sequencing reads and using the Platanus assembler, which is more suitable for heterogeneous genomes (152). This yielded a genome of 225 Mbp (~60% of the predicted genome size of *A. indica* (227)) assembled to a relatively high degree (N50 of 2 Mbp). However, this genome assembly is not available on a public database. The Ganit Labs website hosts a Neem Genome Portal (228), which, although not publicised in either publication (126, 152), is reported to contain the assembled *A. indica* genome. All attempts to access this website were unsuccessful. However, in July 2019, three years after the publication of the improved *A. indica* assembly (152), an assembled genome was deposited (in raw FASTA format) as a correction to this publication (152). Although the timing of this resulted in the assembly not being used for mining of candidate genes within this project, the genome assembly has been used for comparative purposes in this chapter.

In parallel to the sequencing efforts of *A. indica* by the Ganit Labs, the National Centre for Biological Sciences (NCBS) (Bangalore, India) conducted a separate sequencing study (127). The genome assembly produced by NCBS is a similar size (261 Mbp, 70% of the predicted genome size (227)) to the Ganit Labs assembly and was made available on a public database (NCBI; PRJNA176672; AMWY00000000.1). However, as this assembly was based on short read technology (Illumina and Roche 454) it is highly fragmented and has a N50 of only 3 Kbp. This prevented the identification of full-length candidate genes from this assembly. However, it has been used for comparative purposes in this chapter.

In addition to the above limitations of the *A. indica* genomes, neither of the assemblies have been released in an annotated form, which reduced their usability for candidate gene identification. Further, there have been no attempts to generate positional information which could convert these draft assemblies to pseudochromosome-level assemblies. Although for the purposes of identifying candidate genes contig level assemblies are sufficient, the lack of chromosome-level assemblies limits downstream analyses such as evaluating synteny and identification of candidate biosynthetic gene clusters (BGCs). BGCs are found in a number of triterpene biosynthetic pathways (11) and, if BGC were involved in limonoid biosynthesis, tools such as plantiSMASH (229) could be used to identify their presence and associated candidate genes.

The draft nature and the availability issues of the two *A. indica* genomes has necessitated the use of transcriptomic data to identify the first three genes in limonoid biosynthesis (Chapter 2). Transcriptomic studies within the Meliaceae are more widespread, with raw RNA-Seq reads available on NCBI for *A. indica*, *M. azedarach* and *T. sinensis* (as discussed in Chapter 2). Again, the usefulness of these resources is limited. It has been established that for accurate differential expression analysis, biological replicates are crucial (230). However, of the Meliaceae RNA-Seq datasets available only one includes replicates and unfortunately these were for leaf tissues alone (128), which prevented comparative analysis of gene expression profiles in different tissues. The incorporation of RNA-Seq data into gene annotation pipelines is also known to be greatly beneficial to establishing the correct annotation of genes, by improving the accuracy of assignment and levels of identification (231).

5.1.1   Aims

This chapter reports the generation of a pseudochromosome-level assembly of the genome of *M. azedarach*. The aims of this chapter are as follows:

1. Generation of a draft genome assembly of *M. azedarach* (JPN11).

2. Generation of RNA-Seq data for *M. azedarach* (JPN11 and JPN02).

3. Annotation of the draft *M. azedarach* genome.

4. Generation a pseudochromosome-level *M. azedarach* assembly.

5. PlantiSMASH analysis of *M. azedarach* genome.

5.2   Results and Discussion

5.2.1   Generation of a draft genome assembly of *M. azedarach* (JPN11)

High molecular weight genomic DNA (HMW gDNA) with an average size of 58 Kbp was extracted from the leaves of *M. azedarach* accession JPN11 using the cetyl trimethylammonium bromide (CTAB) method (232). A 20-30 Kbp Pacific Bioscience (PacBio) shotgun library was constructed by the Earlham Institute (EI) using this HMW gDNA, which was subsequently sequenced on 10 single molecule real-time (SMRT) cells of the Sequel system (PacBio). This sequencing yielded over 2 million filtered subreads with an average length of 13 Kbp. The hierarchical genome assembly process 4 (HGAP-4, PacBio) tool was utilised by the Swarbreck group (EI) to perform *de novo* genome assembly. The resultant draft genome assembly had a total length of 230 Mbp and a total of 550 contigs (Table 5.2.1). This length represents only 55% of the estimated genome size (421 Mbp) reported in the literature (227). However, this estimate was based on a accession originating from Lucknow (India) (227) and therefore this may be due to a large variation in genome size within the *M. azedarach* species, as indeed genome size estimates vary widely within the Meliaceae family (227).

Table 5.2.1: **Statistics for PacBio genome assembly of *M. azedarach***

Polished contigs	550
Maximum contig length	7,574,565
N50 contig length	3,132,033
Sum of contig lengths	230,844,410

Comparison of this *M. azedarach* genome assembly to the two previous assemblies of the close relative *A. indica* (Ganit Labs (152) and NCBS (127)) was performed by QUAST (233), a quality assessment tool for genome assemblies (Table 5.2.2). Although *M. azedarach* and *A. indica* are estimated to have different genome sizes (421 Mbp and 384 Mbp, respectively (227)) comparisons of the level of assembly could still be drawn (Table 5.2.2).

The genome assembly produced by NCBS (127) appears to have the largest total size of all three genomes (Table 5.2.2), however this is likely to be an artefact caused by the high-level of fragmentation within this assembly (N50 = 3.6 Kbp). Fragmentation of draft genomes has previously been associated with errors such as falsely inflated number of genes (234). The fragmentation of the NCBS genome is clear with over 20,000 contigs being < 500 bp and therefore not included in the main QUAST analysis. Such fragmentation severely limits the identification of large candidate genes such as OSCs which have an average length of 2.5 Kbp and can contain > 10 introns.

The recently available Ganit Labs assembly of *A. indica* (152) appears to be a similar size and assembly level to the *M. azedarach* draft genome generated here, the latter having a slightly superior N50 and L50. However, the main advantage of the *M. azedarach* draft genome over both *A. indica* assemblies is the lack of uncalled bases (Ns). Both the Ganit Labs (152) and NCBS (127) assemblies have a high number of Ns with the Ganit Labs assembly containing an average of ~50 Ns per 1 Kbp, which could feasibly limit the identification of full-length, accurate sequences of genes.

Table 5.2.2: **Assembly statistics of *M. azedarach* draft assembly compared to *A. indica.***

Assembly	*M. azedarach* (this study)	*A. indica* (Ganit Labs)	*A. indica* (NCBS)
Number of contigs	548	2,689	105,122
Largest contig	7,574,565	**7,894,471**	47,608
Total length	230,843,876	221,441,386	**254,234,150**
GC (%)	32.21	31.90	32.11
N50	**3,132,033**	2,629,187	3,634
N75	**1,834,773**	1,149,129	1,769
L50	**24**	27	18,204
L75	**48**	60	43,631
Ns per 100 Kbp	**0.00**	5,414.21	174.90

All statistics were generated by QUAST V.4.6.3 (233) and are based on contigs of size ≥ 500 bp. Assemblies included are *M. azedarach* (this study), *A. indica* (Ganit Labs) (152) and *A. indica* (NCBS) (127). Optimal value for each statistic is highlighted in bold.

5.2.2   Generation of RNA-Seq data for *M. azedarach* (JPN11 and JPN02)

RNA was extracted from seven tissues of *M. azedarach* plants JPN02 and JPN11, which had been selected based on their differing limonoid content (Chapter 4). Extracted RNA (four replicates for each tissue (n=4)) was sent to the EI for quality control and sequencing. Following validation of quality, a high-throughput stranded RNA library was constructed for each replicate and all libraries were multiplexed. Sequencing of the resultant library (containing 28 multiplexed replicates) was performed over two lanes of HiSeq 4000 (Illumina) to generate over 635 million paired end reads, averaging 91 million per tissue (Appendix B.15).

The RNA-Seq data was utilised by the Swarbreck group (EI) to support structural annotation of genes in the draft *M. azedarach* genome (Section 5.2.3).   Once constructed, the EIv1 annotation was used as a reference sequence to align RNA-Seq reads for each sample and count the occurrence of reads per gene. Raw read counts were scaled by library size to account for differences in number of reads sequenced for each replicate (Appendix B.15) and avoid over-inflation of read counts in samples with larger libraries (235, 236).   During the quality control process, hierarchical clustering performed on read counts of all 28 replicates showed a clear segregation of replicates by sample (Figure 5.2.1). The samples from roots were clearly different from the above-ground samples, and those from the petiole tissues (from JPN11 and JPN02) were distinct from the leaves.   Read counts were subsequently used in differential expression analysis to examine expression patterns of candidate genes (Chapter 6).
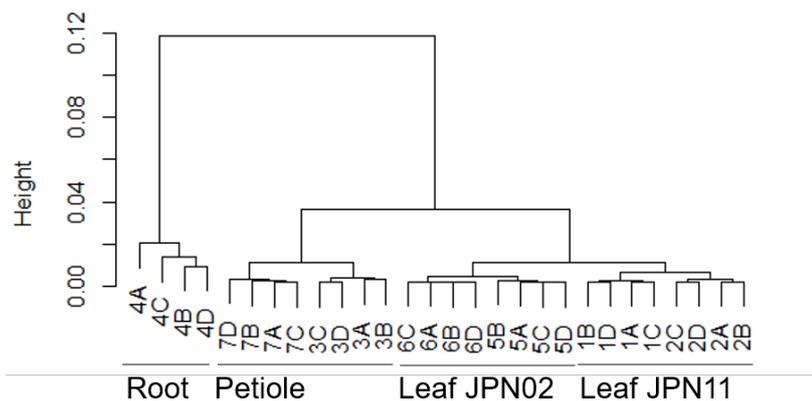
Figure 5.2.1: **Hierarchical clustering of RNA-Seq samples.**
Dendrogram of hierarchical clustering of rlog-transformed read counts (generated by DEseq2 V1.22.1 (237)) of all replicates of RNA-Seq samples calculated using Pearson correlation (238) and complete linkage. Replicates (labelled A-D) are of following samples: JPN11 upper leaf (1), lower leaf (2), petiole (3), root (4) and JPN02 upper leaf (5), lower leaf (6), petiole (7).

5.2.3 Annotation of the *M. azedarach* genome

A high quality structural genome annotation (EIv1) was generated for *M. azedarach* accession JPN11 by the Swarbreck group (EI), using their specialist plant genome annotation pipeline. The pipeline incorporates repeat identification, RNA-Seq mapping and alignment to protein predictions from related species, as well as utilising EI-developed tools such as: Portcullis (239), to filter out false splice junctions; and Mikado (240), to incorporate RNA-Seq alignment data generated by multiple methods. This pipeline has previously been used to successfully annotate the genomes of diverse plant species, e.g. ash (*Fraxinus excelsior*) (241) and bread wheat (*Triticum aestivum*) (242). Over 26,000 genes were identified within the *M. azedarach* genome using this pipeline (Table 5.2.3). This is in line with the 25,379 genes reported for *C. sinensis* (145) and 20,169 reported by the Ganit lab for their *A. indica* genome annotation (152). Although the number of predicted genes is reported to be almost twice this value in the NCBS *A. indica* annotation (127), as previously stated this is likely to be due to the fragmented genome assembly (234).

Using the *M. azedarach* JPN11 (EIv1) annotation, genome completeness was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool (243). BUSCO assessments identify lineage-specific sets of genes within a genome. These sets of genes include highly conserved genes expected to be present as a single copy within all species in the lineage. The lack of release of annotations from the two *A. indica* genomes prevented direct comparison by BUSCO to the *M. azedarach* JPN11 (EIv1) annotation. However, of the species included in this comparison, the *M. azedarach* assembly had the highest number of complete BUSCO genes, with the exception of the gold standard *A. thaliana* (Thale Cress) and also *Capsella grandifolia* (Figure 5.2.2). The *M. azedarach* genome annotation had a lower number of missing and fragmented BUSCO genes compared to the annotations of citrus species assessed.

This BUSCO assessment (Figure 5.2.2) indicates that the *M. azedarach* JPN11 (EIv1) annotation is an accurate resource for future identification of candidate genes. Additionally, the BUSCO assessment (Figure 5.2.2) indicates that this 230 Mbp assembly represents a high level of completeness. If the genome sequenced was indeed the estimated size of 421 Mbp, then the BUSCO assessment would indicate a high level of missing genes. Therefore the discrepancy between the *M. azedarach* estimated and sequenced genome size is either due to a high-level of within species variation in the genome size of different *M. azedarach* accessions or an inaccuracy in previous 1C-based estimations of genome size (227), which were calculated based on amount of nuclear DNA at mid-prophase in an individual accession.

Table 5.2.3: **Summary statistics of the *M. azedarach* JPN11 genome annotation (EIv1).**

Genes	
Total number of genes	26,738
Protein coding (high)	22,785
Transposable element (high)	1,250
Predicted (low)	230
Protein coding (low)	1,651
Transposable element (low)	822
Transcripts	
Transcripts per gene	1.16
Total number of transcripts	31,048
CDS	
Transcript mean size CDS (bp)	1,309.11
Min CDS	78
Max CDS	15,903
CDS mean size (bp)	245.97
Exon mean size (bp)	312.11
Exons per transcript	5.71
Total exons	177,227
Monoexonic transcripts	5,473
cDNA	
Transcript mean size cDNA (bp)	1,781.55
Min cDNA	114
Max cDNA	16,537
Intron mean size (bp)	392.02
5UTR mean size (bp)	186.24
3UTR mean size (bp)	286.21

All statistics were generated by EI based on the EIv1 annotation. Genes are classified as either: protein coding, predicted (limited homology support <30%) or transposable element (>40% overlap with interspersed repeats). Genes are assigned a confidence classification of high or low based on their ability to meet specified criteria (>80% coverage to reference proteins or >60% protein coverage with >40% of the structure supported by transcriptome data). Statistics for coding sequences (CDS) and complementary DNA (cDNA) as also included.
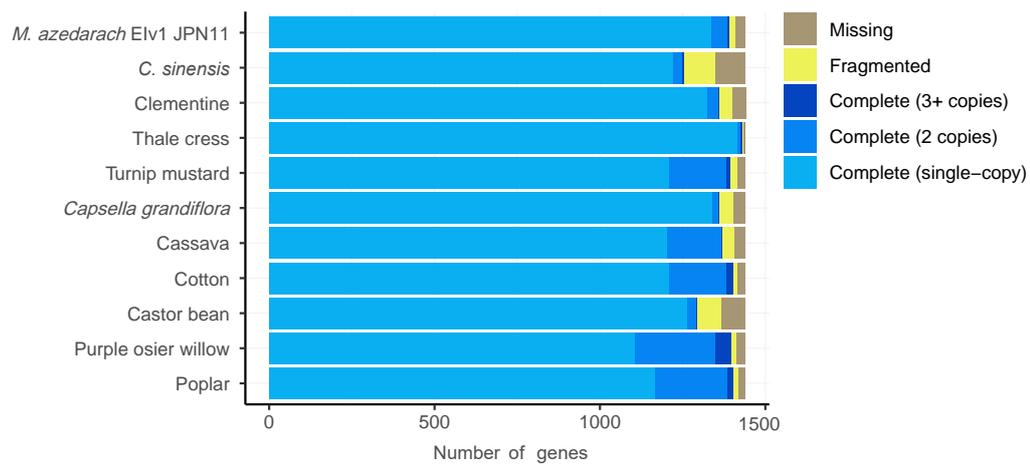
Figure 5.2.2: **BUSCO assessment of *M. azedarach* (EIv1) annotation and comparison with other sequenced species.**
BUSCO (Benchmarking Universal Single-Copy Orthologs) (243) assessment of protein annotation of *M. azedarach* EIv1. The BUSCO assessment was performed by the Swarbreck group (EI). The closest related species *C. sinensis* (sweet orange) with an annotated sequence has been included in this analysis.

5.2.4   Generation a pseudochromosome-level *M. azedarach* assembly

Hi-C (244) sequencing was performed by Phase Genomics to further assemble the genome to a pseudochromosome-level. This technique involves cross-linking of chromatin within a genome, followed by digestion and re-ligation of the DNA based on proximity. Subsequent sequencing enables the quantification of chromatin interactions between regions of DNA within the genome and so provides inferred positional information (as chromatin interactions are more frequent between locations on the same chromosome (244)).

Leaf material from *M. azedarach* (JPN11) was sent to a collaborator (Dr. Alex Harkess, Donald Danforth Plant Science Center) who performed sample preparation for subsequent Hi-C sequencing by Phase Genomics. The proximal tool was used to produce a pseudochromosome-level assembly based on chromatin interactions from the Hi-C analysis and the draft *M. azedarach* genome (EI). This process identified a potential misassembly within the draft genome (contig 000011F) which was subsequently split. This allowed the generation of 14 pseudochromosomes consisting of 218 contigs from the draft assembly, with the addition of 100 Ns between each contig scaffolded. The Hi-C interactions of this pseudochromosome-level assembly are visible in the post-scaffolding heatmap generated by Phase Genomics (Figure 5.2.3).

The reported chromosome number of 1n = 14 based on this pseudochromosome assembly (Figure 5.2.3) is in line with previous reports for *M. azedarach* of 2n = 28
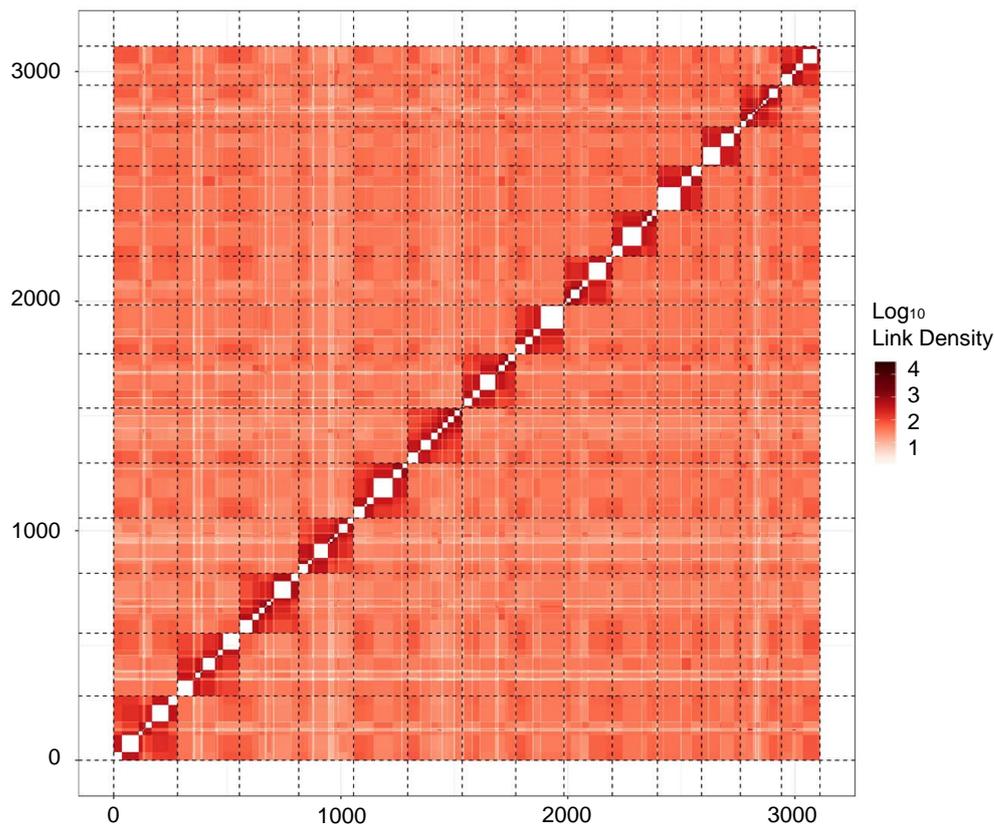
Figure 5.2.3: **Hi-C post-scaffolding heatmap of *M. azedarach* genome.**
Analysis and generation of heatmap was performed by Phase Genomics. The genome
has been divided into 3,000 bins (length = 75,470bp) for this analysis. The density
of Hi-C links is plotted (red). Links between the same contig are not shown (white).
White boxes therefore indicate draft assembly contigs.

(227). However, due to the discrepancy between the predicted and experimentally
determined genome size of *M. azedarach*, karyotyping was performed by Dr.
Azahara Martin (JIC) using young root tissue from *M. azedarach* JPN11. The
karyotyping confirmed the previously reported 2n = 28 (Figure 5.2.4).

The development of a pseudochromosome-level assembly allows comparison to the
chromosome level assembly of *C. sinensis* (145). QUAST results show that the N50
and L50 values are comparable between these two assemblies. Even with the
introduction of Ns into the *M. azedarach* genome during pseudochromosome
assembly, the occurrence of Ns is still almost 1000-fold less than in *C. sinensis*
genome assembly used as a comparison (Table 5.2.4).
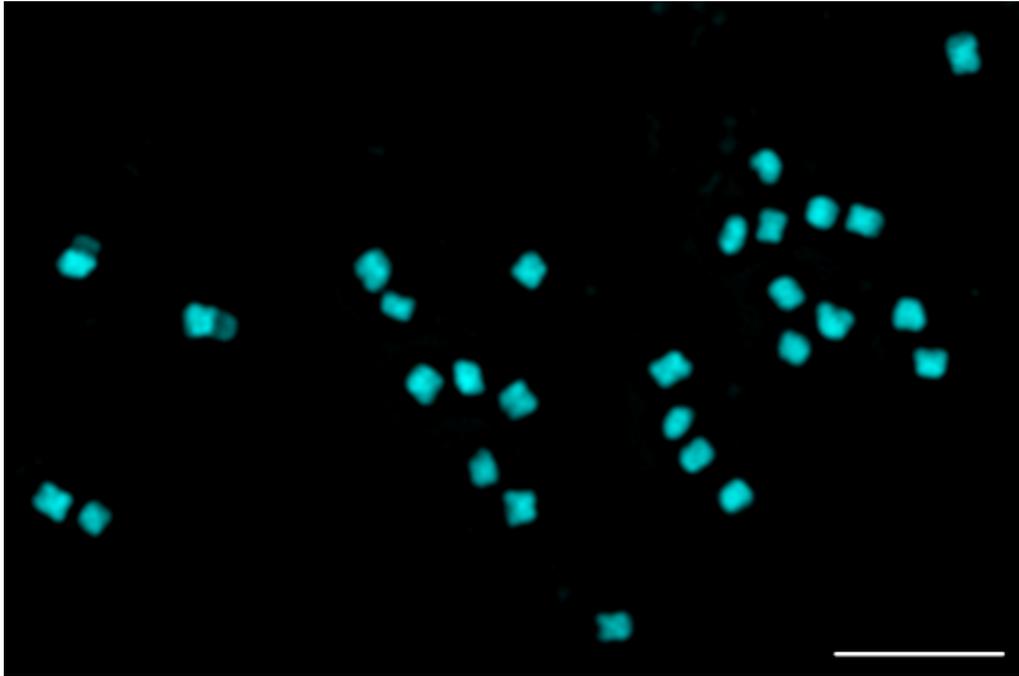
Figure 5.2.4: **Karyotyping of *M. azedarach*.**
Representative image of karyotyping of *M. azedarach* JPN11 kindly performed by Dr.
Azahara Martin. Scale bar = 5μm.

Table 5.2.4: **Assembly statistics of *M. azedarach* Hi-C assembly compared to draft
M. azedarach assembly and *C. sinensis* assembly.**

Assembly	*M. azedarach* (Draft)	*M. azedarach* (P.Chr)	*C. sinensis*
Number of contigs	548	346	4617
Largest contig	7,574,565	20,704,184	**36,147,563**
Total length	230,843,876	230865674	**327,586,883**
GC (%)	32.21	32.21	34.06
N50	3,132,033	16,923,081	**22,711,823**
N75	1,834,773	**14,637,465**	3,133,023
L50	24	7	**6**
L75	48	**10**	11
Ns per 100 kbp	**0.00**	9.44	8,119.38

All statistics were generated by QUAST V.4.6.3 (233) and are based on contigs of size ≥
500 bp. Assemblies are labelled as follows: *M. azedarach* draft genome generated by
PacBio (Draft), *M. azedarach* pseudochromosome-level assembly generated by Hi-C
(P.Chr) and *C. sinensis* asssembly (145). Optimal value for each statistic is highlighted
in bold.

5.2.5   PlantiSMASH analysis of *M. azedarach* genome

The three melianol biosynthetic genes identified so far, *MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*, occur on pseudochromosomes Chr 3, Chr 13 and Chr 0, respectively. Therefore the pseudochromosome-level assembly of *M. azedarach* confirms these initial limonoid biosynthetic genes do not occur within a Biosynthetic gene cluster (BGC). However, it is still possible that other as yet uncharacterised limonoid biosynthetic genes maybe clustered.

PlantiSMASH is a bioinformatic tool used for the rapid identification of putative BGCs within plant genomes (229).   PlantiSMASH analysis of the *M. azedarach* genome (Figure 5.2.5) predicted 47 putative BGCs, 15 of which were annotated by plantiSMASH as 'terpene', one of which (Cluster 032) contained *MaOSC1* (Appendix B.16). The plantiSMASH annotation of 'terpene' is assigned based on the occurrence of initial cyclisation enzymes alone. Therefore clusters not annotated as 'terpene' are still be of interest for later biosynthetic pathway steps. The 47 putative clusters identified by plantiSMASH appear to be evenly spread across the genome and occur on all pseudochromosomes with the exception of pseudochromosome Chr 12, which lacks any putative BGC (Figure 5.2.5).
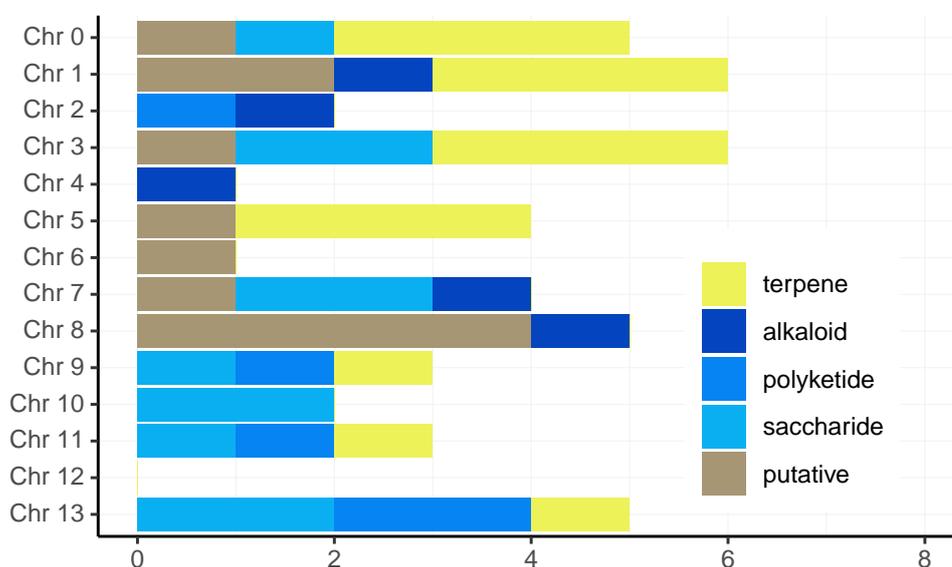
Figure 5.2.5: **Summary of plantiSMASH analysis of *M. azedarach* (EIv1).**
Barplot of the 47 putative gene clusters identified by plantiSMASH (229) (performed by Charlotte Owen (JIC)), organised by which pseudochromosome the clusters occur on.   The plantiSMASH cluster type is indicated by different colours with mixed clusters assigned to a group. Barplot produced in R (221) by ggplot (223).

5.3   Conclusion

The genome sequence of *M. azedarach* reported here represents the first pseudochromosome-level assembly for any species within the Meliaceae family. Further, whilst the annotations for the previous *A. indica* genomes (127, 152) remain unreleased, the *M. azedarach* annotation EIv1 will represent the first structural and functional annotations for a genome within this family. Currently the only annotation within the Meliaceae are from the OneKP 2019 release of transcriptomes (142, 143) which are limited in coverage (e.g. the *M. azedarach* OneKP transcriptome annotations contains less than 13,500 annotated genes).

This reference genome and annotation will greatly benefit the identification of candidate genes involved in Meliaceae-specific limonoid biosynthesis, such as the commercially relevant *seco*-C-ring limonoids. Indeed the generation of this genome sequence has already enabled the identification of 47 putative BGCs by plantiSMASH (229), one of which includes the characterised melianol biosynthetic gene *MaOSC1*.

The usefulness of the *M. azedarach* genome has been enhanced further by the generation of RNA-Seq datasets. These datasets have supported the annotation of the genome and will enable identification of candidate limonoid biosynthetic genes by differential expression.

Comparisons of the *M. azedarach* assembly and annotation with the reference citrus species, *C. sinensis* (145), indicate that the *M. azedarach* assembly is a suitable reference genome within the Meliaceae, and should provide a valuable contribution to wider analyses within the Sapindales order.

5.4   Chapter-specific materials and methods

5.4.1   Extraction of nucleic acids

HMW gDNA (required for PacBio sequencing) was extracted from *M. azedarach* JPN11 leaves using the a modified CTAB protocol which includes the addition of proteinase K and RNase A (Qiagen) (232).

RNA was extracted from tissues of *M. azedarach* individuals JPN11 and JPN02 as described in Chapter 9. All tissues were harvested on the same day, flash frozen in liquid N_2 and stored at -80°C prior to extraction. RNA extractions were performed in technical replicates (four separate extractions, each including one replicate of each tissue type).

Leaf material form JPN11 was sent to Dr. Alex Harkess (Donald Danforth Plant Science Center) to perform chromatin cross-linking and extraction using the Proximo Hi-C Plant Kit (Phase Genomics).

5.4.2 Sequencing and assembly

Preparation and sequencing of a 20-30 Kb PacBio shotgun HMW gDNA library and high-throughput Illumina stranded RNA library (150bp, paired end) was performed by EI and sequenced on PacBio Sequel system and Illumina HiSeq4000 respectively. Assembly of the PacBio reads using HGAP-4 smrtlink V5.0.1.9585 was performed by the Swarbreck group (EI), who additionally performed structural annotations for this assembly using a specialised pipeline. Functional annotation of EIv1 was performed the Swarbreck group (EI) using the Assignment of Human Readable Descriptions (AHRD) V.3.3.3 (245) tool. AHRD was provided with results of BLAST V2.6.0 (158) searches (e-value = 1e-5) against reference proteins from TAIR (246), UniProt, Swiss-Prot and TREMBL datasets (247), along with interproscan (248) results.

Library preparation for Hi-C was kindly performed by Dr. Alex Harkess (Donald Danforth Plant Science Center) using Proximo Hi-C Kit (Plant) and sequenced (Illumina technology) by Phase Genomics. Quantitative analysis of chromatin interactions and subsequent pseudochromosome-level assembly was performed by Phase Genomics using the Proximo analysis tool.

5.4.3 Karyotyping

Karyotyping of *M. azedarach* was kindly performed by Dr. Azahara Martin. Briefly, root tips were obtained from >1 year old *M. azedarach* plants and the preparation of mitotic metaphase spreads was carried out as described previously (249). Chromosomes were counterstained with DAPI (1 μg/ml). Images were acquired using a Leica DM5500B microscope equipped with a Hamamatsu ORCA-FLASH4.0 camera and controlled by Leica LAS X software v2.0.

5.4.4 plantiSMASH analysis of *M. azedarach*

Analysis of *M. azedarach* genome by plantiSMASH (229) was kindly performed by Charlotte Owen (JIC). Due to the very recent completion of the pseudochromosome assembly of *M. azedarach*, plantiSMASH analysis was performed based on the EIv1 annotation and draft (PacBio) genome assembly. Occurrence of clusters on

pseudochromosomes was based on this analysis and the reported positioning of draft contigs into the pseudochromosomes.

5.4.5 Processing RNA-seq data

To generate expression data from the raw RNA-Seq reads relative to the EIv1 annotation a basic methodology outlined in 'Intro2RNAseq' (produced by Weill Cornell Medical College (250)) was followed. All tools and quality control steps were performed with parameters specified in this protocol (250).

Quality control of all samples was assessed by FastQC V.0.10.1 (251). STAR V.2.5 (252) was used to align all reads (pooling all reads per replicate (directional and lane)) to the EIv1 annotation. Samtools V.1.7 (253) was utilised to index the subsequent alignment. The featureCounts tool of subread V1.6.0 (254) was used to generate raw read counts by counting the number of reads overlapping with EIv1 genes in each alignment. Raw read counts were analysed in R (221) using DEseq2 v1.22.1 (237). Genes with zero counts were removed from the analysis, normalisation was performed based on library size and subsequent counts were $\log_2$ transformed with a pseudo count of one. The resultant library-normalised $\log_2$ read counts were used for downstream analyses.

6

Identification of new limonoid biosynthetic genes from the *M. azedarach* genome

Synopsis

The newly generated *M. azedarach* genome (EIv1) and accompanying RNA-Seq datasets have been utilised to identify limonoid biosynthetic genes for steps downstream of melianol. Candidate genes were selected based on their annotation and shared expression with melianol biosynthetic genes. A subset of genes were assessed for activity by co-expression with melianol biosynthetic gene in *N. benthamiana*. This approach has led to the identification of four diverse melianol-modifying enzymes thus far. Preliminary analysis suggests that together a CYP (MaCYP88A108) and a sterol isomerase (MaIsom1-I) may be capable of controlling structural rearrangements of the internal scaffold of melianol to form a true limonoid scaffold, here termed 'melianol B'. Further two tailoring enzymes, a short chain dehydrogenase/reductase (MaSDR1) and an acyltransferase (MaBAHD1), are predicted to function on melianol-type scaffolds by dehydrogenation and acetylation, respectively. These predicted functions require confirmation by structural elucidation, however their identification indicates that the *M. azedarach* genome is a crucial resource for the identification of further limonoid biosynthetic genes.

6.1   Introduction

The discovery of three melianol biosynthetic genes (Chapter 2) has suggested that melianol is an intermediate in the biosynthesis of limonoids and therefore has enabled more specific predictions to be made of the next likely limonoid biosynthetic steps. These predictions in turn inform speculation of which enzyme classes may be involved in such transformations. However, as discussed below a number of different enzyme classes are potentially implicated in the next predicted steps.

The next stage of limonoid biosynthesis, beyond melianol, is the production of simple ring-intact limonoids such as 7-deacetylazadirone (Figure 6.1.1). Two major biosynthetic transformations are required to convert the protolimonoid melianol to such a ring intact limonoid: the formation of a furan ring, accompanied by loss of four carbons; and internal scaffold rearrangements involving the loss of the C7 alkene. An additional transformation needed is the conversion of the C3 hydroxy to a carbonyl group. The mechanisms and order of these transformations remain speculative (Figure 6.1.1).

The occurrence of furan rings within triterpenes appears to be rare outside of the limonoids, and subsequently the mechanism of furan ring formation remains entirely speculative. A Paal-Knoor-like route (131, 132) may convert the hemiacetal ring of melianol to the furan ring of true limonoids. This could feasibly proceed by a dehydration to remove the C21 hydroxy group, followed by opening of the epoxide, loss of the four carbon epoxide-containing fragment, and aromatisation by a form of desaturation reaction (Figure 6.1.1).

The characterisation of CYPs capable of melianol biosynthesis (Chapter 2), along with the well-known diverse functions of this family of enzymes in triterpene biosynthesis (118), suggests that additional CYPs could be involved in such transformations. For example, CYP710A enzymes are known to desaturate the side chain of sterols at the C22 position in plants (255).

Although CYPs are implicated, there are alternative enzyme classes that could also carry out these transformations. Although not yet characterised in triterpene biosynthesis, 2-oxoglutarate Fe(II)-dependant oxygenases (OGDs) are responsible for a number of oxidations in plant secondary metabolism (256, 257), such as their established roles in coumarin biosynthesis (258). Known activities of OGDs also include desaturation reactions, e.g. the final reaction in carbapenem biosynthesis performed by CarC (259, 260).

Furan ring formation could involve rearrangement of the side chain epoxide to a
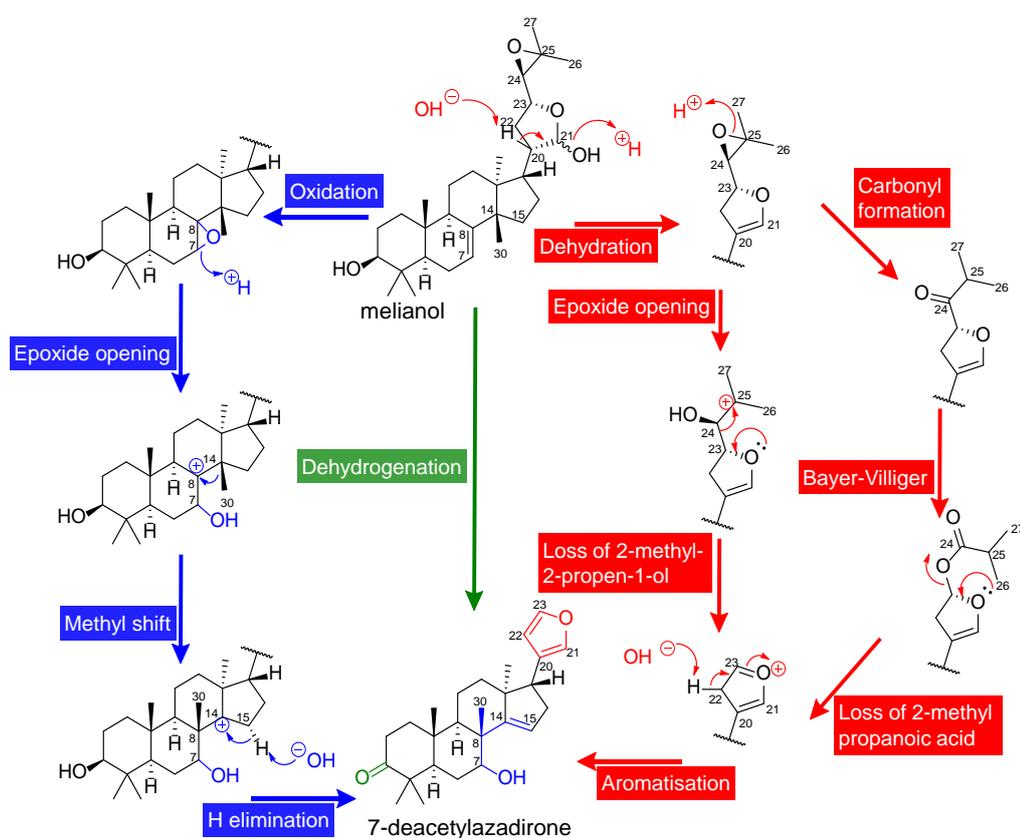
Figure 6.1.1: **Prediction of post-melianol biosynthetic steps.**
Predictions of enzymatic steps required for conversion of the protolimonoid (melianol) to a simple ring intact limonoid (7-deacetylazadirone). The major steps required are; furan ring formation (red); scaffold rearrangement (blue); and C3 carbonyl formation (green). Furan ring formation may involve modification of the C23 epoxide containing fragment to create a better leaving group.

ketone followed by a Baeyer-Villiger-type reaction to introduce an ester, which would increase the leaving potential of this group (Figure 6.1.1). Baeyer-Villiger reactions have also been predicted to be involved in later limonoid biosynthetic steps to afford *seco*-ring limonoids (5). Characterised enzymes capable of Baeyer-Villiger reactions in higher plants are limited to two CYPs, AtCYP85A2 from *A. thaliana* and SlCYP85A3 from *Solanum lycopersicum* (tomato), that both facilitate the formation of the *seco*-B-ring in brassinosteroid biosynthesis (261, 262). In microbes, Baeyer-Villiger Monooxygenases (BVMOs) carry out highly selective Baeyer-Villiger oxidations in secondary metabolism (263). BVMOs have not as yet been identified from higher plants. However, broad-substrate BVMOs acting on a range of ketone substrates have been characterised from moss and red algae (264). Alternatively, certain OGDs are capable of oxidative ring expansion, e.g. the deacetoxycephalosporin C synthase (DAOCS) which catalyses oxidative ring expansion of penicillin N (265). Therefore, CYPs (particularly of the CYP85A family), OGDs, BVMOs and other flavin monooxygenases all represent enzyme classes that may be involved in furan ring

formation as well as later *seco*-ring formation in limonoid biosynthesis.

The majority of ring-intact limonoids do not contain the C7 double bond and C14 methyl group of tirucalla-7,24-dien-3β-ol and instead contain a C14 double bond, a C8 methyl group, and C7 oxidation. Biosynthesis of this modified scaffold could feasibly be initiated through epoxidation of the C7 double bond and then proceed by the subsequent opening of the epoxide forming a carbocation, which could allow a methyl shift from C14 to C8 and elimination of hydrogen at C15 resulting in the formation of C14 double bond (Figure 6.1.1). Again, although CYPs could mediate the epoxidation of the C7 double bond, members of the OGD superfamily, of plant (266) and bacterial (267, 268) origin, are also capable of epoxidation.

Once epoxidation has occurred an additional enzyme may be required to open the epoxide and trigger a rearrangement (Figure 6.1.1). Alternatively this may happen spontaneously. An epoxide hydrolase could be responsible for the opening of the C7 epoxide, as demonstrated for the mogroside triterpene biosynthetic pathway in *Siraitia grosvenorii* (219). Alternatively, isomerisation of the double bond could occur in the absence of epoxidation. For instance, enzyme classes such as short-chain dehydrogenase/reductases (SDRs) have diverse functions in plants (269). One of the most extreme examples of this is the SDR enzyme from *S. lycopersicum* (SlGAME25) which isomerises the C5,6 alkene to a C4,5 alkene, while simultaneously dehydrogenating the C3 hydroxy group, during steroidal glycoalkaloid biosynthesis (270). Therefore, SDRs could be involved in the scaffold isomerisations required for limonoid biosynthesis. Further, SDRs are the most likely enzyme class to mediate the conversion of the C3 alcohol of protolimonoids to a carbonyl. In addition to the activity of GAME25, such a C3 dehydrogenation has been demonstrated by the SDR AtTHAR1, which catalyses C3 carbonyl formation in *A. thaliana* triterpenes (121).

Subsequently, when selecting candidate genes for potential involvement in post-melianol biosynthesis, a broad range of enzyme classes will need to be considered as candidates. The newly generated genomic resources from *M. azedarach* (Chapter 5) are ideal for this purpose. The RNA-Seq datasets, generated from tissues that differentially accumulate limonoids, will enable the capture of genes with similar expression patterns to known melianol biosynthetic genes via differential expression analysis and hierarchical clustering. Functional annotation (EIv1) will then enable co-expressed genes to be selected as candidates based on the relevance of their biosynthetic annotations to predicted next steps. This will allow the selection of candidates from a broad range of enzyme classes that are supported by expression data.

6.1.1 Aims

The aim of this chapter is to identify and characterise candidate genes that may be involved in the immediate steps of limonoid biosynthesis beyond melianol. Therefore the aims of this chapter are as follows:

1. Selection of candidate genes based on co-expression and annotation.

2. Characterisation of candidate genes by functional expression in *N. benthamiana.*

6.2 Results and discussion

6.2.1 Selection of candidate genes based on co-expression and annotation

A subset of 18,151 differentially expressed genes (P-value < 0.05) was extracted from the 31,048 genes annotated in the *M. azedarach* genome (EIv1). Hierarchical clustering was performed to identify which of these differentially expressed genes shared expression patterns with characterised melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*), which resulted in a list of of 283 candidate genes (Appendix B.17). This list was manually searched to identify genes annotated with biosynthetic functions of interest. As the next steps in limonoid biosynthesis could proceed by a number of routes, a broad range of enzyme classes were considered to be potential candidates. This includes, but is not limited to: hydrolase, oxygenase, dehydratase and isomerase enzymes. Subsequently, 26 candidates were selected for further analysis (Table 6.2.1), all of which showed strong co-expression (Figure 6.2.1). Expression of melianol biosynthetic genes, and selected candidates, is highest in the petiole tissues and lowest in the leaves (Figure 6.2.1), in line with previous profiling of this species by metabolite analysis and qRT-PCR (Chapter 4).

In an attempt to identify additional candidates, the occurrence of *MaOSC1* within a putative plantiSMASH BGC (Cluster 32, pseudochromosome 3) was investigated. It was thought that the genes within this cluster, which included two SDRs, may represent strong limonoid biosynthetic candidates. However, with the exception of *MaOSC1*, the expression of genes within Cluster 32 did not correspond with the high expression in petioles and roots observed for the other melianol biosynthetic genes (Appendix B.16). Therefore, at present, only the 28 genes identified as differentially expressed (Table 6.2.1, Figure 6.2.1) were investigated further.

Table 6.2.1: **Annotation of differentially expressed candidate genes from *M. azedarach* (EIv1).**

Name	EIv1 ID	Human readable annotations
MaOGD1	MELAZ155640_EIv1_0015190.1	2-Oxoglutarate and Fe(II)-dependent oxygenase
MaOGD2	MELAZ155640_EIv1_0015210.1	2-Oxoglutarate and Fe(II)-dependent oxygenase
MaOGD2.1	MELAZ155640_EIv1_0015210.1	2-Oxoglutarate and Fe(II)-dependent oxygenase
MaOGD2.2	MELAZ155640_EIv1_0015210.1	2-Oxoglutarate and Fe(II)-dependent oxygenase
MaAKR1	MELAZ155640_EIv1_0015320.1	NADPH-dependent codeinone reductase-like
MaAKR2	MELAZ155640_EIv1_0015350.1	Aldo/keto reductase
MaEst1	MELAZ155640_EIv1_0015800.1	GDSL esterase/lipase
MaCYP716-3	MELAZ155640_EIv1_052990.1	Cytochrome P450
MaCYPU-1	MELAZ155640_EIv1_0061950.1	Cytochrome P450 family ent-kaurenoic acid
MaCY88A108*	MELAZ155640_EIv1_0061960.1	Cytochrome P450 family ent-kaurenoic acid
MaAD1	MELAZ155640_EIv1_0078490.1	Alcohol dehydrogenase-like
MaEH1	MELAZ155640_EIv1_0095290.1	Bifunctional epoxide hydrolase 2-like
MaDT1	MELAZ155640_EIv1_0120380.1	2-Hydroxyisoflavanone dehydratase-like
MaCYP735-1	MELAZ155640_EIv1_0128570.1	Cytochrome P450, putative
MaEst2	MELAZ155640_EIv1_0135650.1	Carboxylesterase 1-like
MaOR2	MELAZ155640_EIv1_0141190.1	Oxidoreductases, acting on NADH or NADPH
MaBAHD1	MELAZ155640_EIv1_0142070.1	Vinorine synthase-like
MaTF2	MELAZ155640_EIv1_0148250.1	Methyltransferase-like protein
MaBAHD3	MELAZ155640_EIv1_0164450.1	Vinorine synthase-like
MaCYP714-1	MELAZ155640_EIv1_0170970.1	Cytochrome P450
MaIsom1	MELAZ155640_EIv1_0192980.1	Sterol-8,7-isomerase
MaSDR1	MELAZ155640_EIv1_0198190.1	Short-chain dehydrogenase/reductase
MaSDR2	MELAZ155640_EIv1_0198240.1	Secoisolariciresinol dehydrogenase-like
MaCYPU-2	MELAZ155640_EIv1_0217180.1	Beta-amyrin 28-oxidase-like
MaAKR3	MELAZ155640_EIv1_0223780.1	NADPH-dependent codeinone reductase-like
MaBAHD2	MELAZ155640_EIv1_0235630.1	Vinorine synthase-like
MaBAHD4	MELAZ155640_EIv1_0238770.1	Vinorine synthase-like
MaOR1	MELAZ155640_EIv1_0242600.1	NAD(P)-linked oxidoreductase superfamily

Name were manually assigned for simplicity. EIv1 identifiers and simplified human readable annotations are given. A complete list of the differentially expressed genes identified by hierarchical clustering (including InterPro domains) can be found in Appendix B.17 along with protein sequences in Appendix C.1. A previously cloned gene, MaCYP88A108* was also included. A partial version of this gene was previously identified from a *de novo* assembled *M. azedarach* transcriptome (Ma1) (Chapter 2). The predicted full-length gene was represented in EIv1 and therefore included here. The annotation of *MaOGD2* was predicted as one gene. However closer analysis revealed that this gene could be annotated as two separate full-length genes (*MaOGD2.1, MaOGD2.2*), therefore these genes were also selected as candidates.
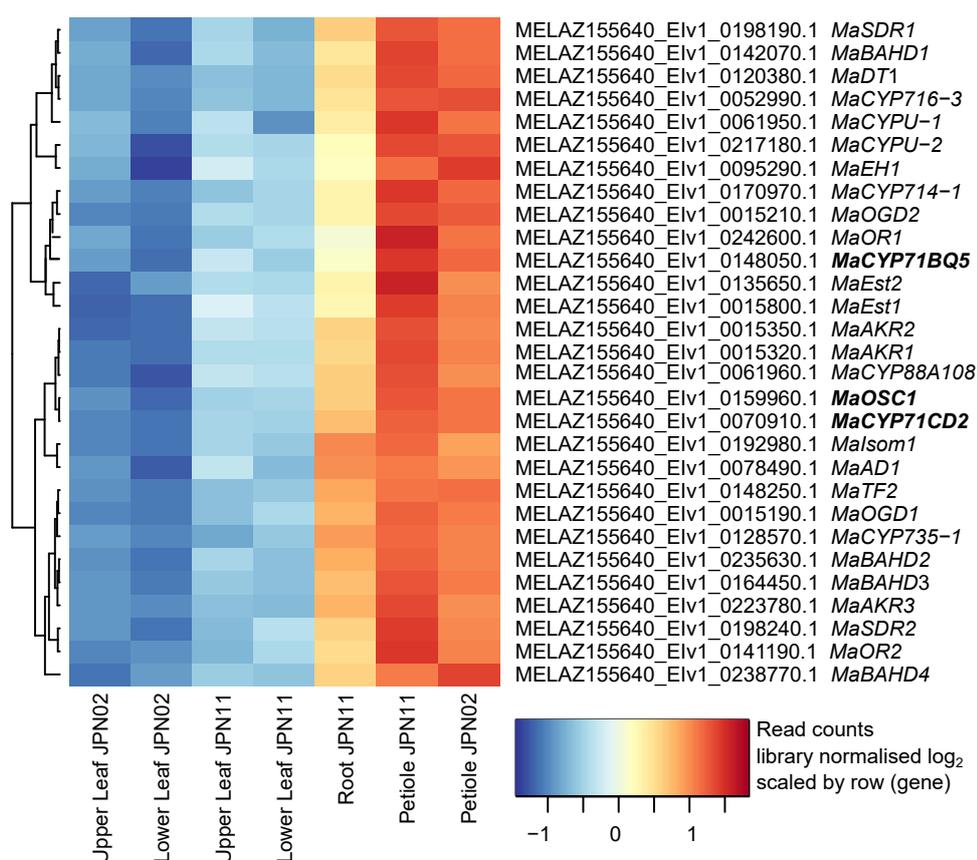
Figure 6.2.1: **Expression patterns of differentially expressed candidate genes from *M. azedarach* (EIv1).**
Genes were selected based on annotation (Table 6.2.1), from a larger subset of differentially expressed genes identified as co-expressed based on hierarchical clustering (Appendix B.17). Genes are labelled with EIv1 identifier and given name (based on human readable annotation). The melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*) are included for comparison (bold). Read counts used for hierarchical clustering were normalised by library size and $\log_2$ transformed. The heatmap was constructed by Heatmap3 V1.1.1 (271) with scaling performed by row (gene) to emphasise pattern of expression.

6.2.2 Characterisation of candidate genes by functional expression in *N. benthamiana*

The candidate genes selected based on differential expression analysis were cloned into pEAQ-HT-DEST1 vectors to allow their functional characterisation by transient expression in *N. benthamiana*. Expression in combination with melianol biosynthetic genes revealed that four candidate genes (Appendix A.8, Appendix A.9) had activity on melianol-type scaffolds. The observed activity of MaCYP88A108, MaIsom1-I, MaSDR1 and MaBAHD1 are discussed in this section.

MaCYP88A108 activity

MaCYP88A108 was originally identified as a homolog of *AiCYP88A108*, which was co-expressed with *AiOSC1* in *A. indica* (Chapter 2). It was reselected as a candidate here because the newly generated *M. azedarach* genome revealed an extended 5' sequence (141 bp) within the coding sequence of this gene. Therefore, *MaCYP88A108* was transiently expressed in *N. benthamiana* in combination with melianol biosynthetic genes *AiOSC1, MaCYP71CD2* and *MaCYP71BQ5* (Figure 6.2.2).

Expression of melianol biosynthetic genes, in the absence of *MaCYP88A108*, results in the production of melianol and additionally three peaks **(1-3)** with a mass of 511.3384, equivalent to the sodium adduct of melianol with an additional oxygen (Figure 6.2.2.A). These are believed to be the result of modification of the melianol scaffold by endogenous *N. benthamiana* enzymes.

However, when *MaCYP88A108* is expressed in combination with these genes, there is a ~80% reduction in melianol accumulation and almost complete reduction of these the endogenous peaks **(1-3)** (Figure 6.2.2.A). This is accompanied by the accumulation of three clear new peaks **(4,5,7)** and one broad peak **(6)**, all of which have an identical mass equivalent to the sodium adduct of melianol with an additional oxygen (Figure 6.2.2.A). Both copies of *MaCYP88A108* (from the Ma1 transcriptome assembly and the new genome (EIv1)) give similar results.

The production of peaks **(4-7)**, all with a mass suggesting the addition of one oxygen without the loss of hydrogen, could be explained by either the addition of a hydroxy group to a carbon or the conversion of an alkene to an epoxide. MaCYP88A108 may introduce a hydroxy group to multiple positions on the melianol scaffold. However, this seems unlikely given that no peaks with multiple oxidations were identified. Previously characterised triterpene biosynthetic CYPs that are capable of multiple oxidations, such as MaCYP71CD2 from *M. azedarach* (Chapter 2) and AsCYP51H10 from *A. strigosa* (272), oxidise both positions simultaneously as opposed to oxidising positions individually.

A more likely explantation may be that melianol is being oxidised in only one position and a lack of stability in the resultant structure may lead to rearrangements that produce multiple isomers of the same mass. This theory could be in line with the predicted next steps in the pathway (Figure 6.1.1), whereby MaCYP88A108 is capable of oxidising the C7 alkene of melianol to form an epoxide (7,8-epoxymelianol) the mass of which would be 511.3384 (Figure 6.2.2.B). Subsequent epoxide opening would form a carbocation intermediate, which could be rearranged into a number of different structures with a mass of 511.3384.

Figure 6.2.2: **Expression of and *MaCYP88A108* in *N. benthamiana.***
(A) UHPLC-IT-TOF generated EICs of extracts of *N. benthamiana* leaves expressing melianol biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*) with and without *MaCYP88A108*. The UHPLC-IT-TOF 'limonoid' gradient was used. The EICs shown are for the following adducts [melianol+Na]$^+$=495.3433 (red) and [melianol+O+Na]$^+$=511.3384 (blue). Melianol peaks are labelled along with newly identified peaks **(1-7)**. Mass spectra of newly identified peaks are available (Appendix A.7). (B) Predictions of structures, exact masses and mechanisms of formation of oxidised melianol peaks **(4-7)**. Structure of 8-methyl-melian-14,15-ene-3,7-diol has been termed melianol B for the rest of this chapter for simplicity.

One such rearrangement (depicted in Figure 6.2.2.B) could involve a methyl shift from C14 to C8, elimination of hydrogen from C15, and formation of a C14-C15 alkene. This proposed rearrangement would produce a very similar structure to melianol except with a mature limonoid internal scaffold rather than a protolimonoid internal arrangement. For the purposes of this chapter this structure will be termed melianol B. It is feasible that one of the MaCYP88A108 peaks **(4-7)** represents 7,8-epoxymelianol and the remaining are rearranged structures, such as

melianol B. Alternatively, if the epoxide was unstable, all four of these peaks could represent carbocation-derived rearranged structures.

Structural characterisation is needed to confirm the position and type of oxidation carried out by MaCYP88A108. However, the efficiency of consumption of melianol (Figure 6.2.2.A), in combination with the close co-expression of *MaCYP88A108* with melianol biosynthetic genes (Figure 6.2.1), strongly suggests that this gene is involved in limonoid biosynthesis. Therefore, the other candidate genes identified by differential expression were screened for activity against this putative melianol B scaffold, by co-expression of candidates with melianol biosynthetic genes and *MaCYP88A108* (Appendix A.8, Appendix A.9).

MaIsom1-I activity

MaIsom1 is annotated in EIv1 as sterol-8,7-isomerase (IPR007905, IPR033118) and was selected based on the structural similarity of protolimonoids to sterols and the potential requirement of isomerase function during limonoid scaffold rearrangement. The sequence amplified using MaIsom1 primers was not identical to the predicted sequence, despite multiple attempts to amplify and clone this sequence, and therefore is termed MaIsom1-I (Appendix C.1).

Transient expression of melianol biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*) and *MaCYP88A108* in *N. benthamiana*, results in production of four peaks **(4-7)** with a mass equivalent to the sodium adduct of oxidised melianol (Figure 6.2.3.A), as discussed above. However, when *MaIsom1-I* is co-expressed with these genes only one of these peaks is detectable **(6)** and there is an absence of all other peaks with this mass **(4,5,7)** (Figure 6.2.3.A).

Structural characterisation of the remaining oxidised melianol peak **(6)** is required to confirm the function of MaIsom1-I. However, the most likely explanation is that MaIsom1-I is capable of controlling the rearrangement of the carbocation intermediate towards the formation of only one isomer (Figure 6.2.3.B). MaCYP88A108 and MaIsom1-I together convert melianol to this sole oxidised melianol peak **(6)** in a relatively efficient manner. This may suggest that, based on predictions of limonoid biosynthesis, the most likely structure for the product of MaIsom1-I is melianol B (Figure 6.2.3.B). This would suggest that all four MaCYP88A108 mediated peaks **(4-7)** (Figure 6.2.3.A) represent structures that have been generated via the carbocation intermediate, rather than 7,8-epoxymelianol, or alternatively that if one of these peaks does represent 7,8-epoxymelianol, MaIsom1-I is capable of its full conversion.
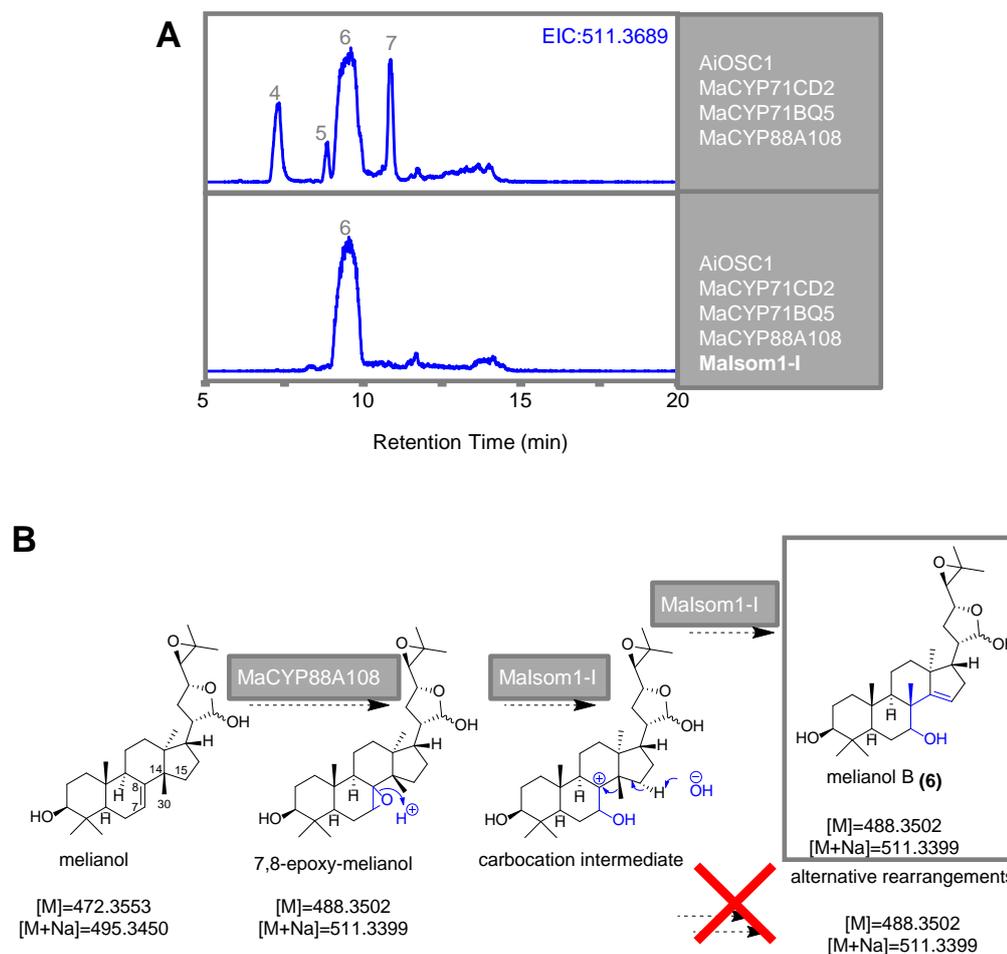
Figure 6.2.3: **Expression of *MaIsom1-I* in *N. benthamiana*.**
(A) UHPLC-ITOF generated EIC of extracts from *N. benthamiana* leaves expressing melianol biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*) and *MaCYP88A108* with and without *MaIsom1-I*. The UHPLC-IT-TOF 'limonoid' gradient was used. The EIC trace shown is for the [melianol+O+Na]$^+$=511.3369 adduct (blue) and corresponding peaks are labelled **(4-7)**. (B) Predictions of structures, exact masses and mechanism of formation of peak **(6)**, thought to be melianol B.

An alternative suggestion is that MaIsom1-I may function in the absence of oxidation and MaCYP88A108 could subsequently oxidise the C7 position of the new scaffold. This is feasible since the substrates of other plant sterol-8,7-isomerases (characterised from *A. thaliana* and *Zea mays*) do not require prior oxygenation in order to perform their isomerisations (273, 274). Similarly, the sterol-24-isomerase (Ws24ISO), which initiates the biosynthesis of withanolides in *Withania somnifera* (Ashwaganda), does not require an oxygenated substrate (275). However, when *MaIsom1-I* is expressed in *N. benthamiana* in the absence of *MaCYP88A108*, melianol is not consumed and no new products are detected (Appendix A.10), suggesting this is not the case.

Therefore, it appears that MaIsom1-I may function by a novel and as yet not understood mechanism, which may involve controlling the rearrangement of the carbocation to yield one product, most likely melianol B. Full structural characterisation is needed in order to confirm this. For simplicity for the rest of this chapter this peak **(6)** will tentatively be referred to as melianol B. It is possible that different chromatographic methods may resolve this broad peak **(6)** to multiple structures. If so, then MaIsom1-I may control rearrangement to a degree, but not efficiently enough to produce a single product. Based on the simpler background afforded by the expression of the oxidised melianol biosynthetic genes (*AiOSC1*, *MaCYP71CD2* and *MaCYP71BQ5*) in combination with both *MaCYP88A108* and *MaIsom1-I*, the other genes discussed in this chapter (*MaSDR1* and *MaBAHD1*) were re-tested for activity on this proposed melianol B scaffold.

MaSDR1 activity

MaSDR1 is annotated as an SDR (IPR002347) and was selected based on its potential ability to convert the C3 hydroxy group of melianol to a ketone. In absence of expression of *MaSDR1*, when melianol biosynthetic genes (*AiOSC1*, *MaCYP71CD2* and *MaCYP71BQ5*) are expressed with *MaCYP88A108* and *MaIsom1-I* there is accumulation of a small background peak **(8)** with a mass of 509.3254 (Figure 6.2.4.A), that would be equivalent to the sodium adduct of doubly dehydrogenated melianol B. Peak **(8)** is thought to be a *N. benthamiana* derived modification. When *MaSDR1* is co-expressed with this combination of genes melianol B is almost completely consumed and a new peak **(9)** with a mass of 509.3254 is detectable (Figure 6.2.4.A). This peak **(9)** shares the broad nature of its precursor peak **(6)** and further chromatographic method development is required to determine whether this peak represents one or multiple compounds.

The mass of 509.3254 **(9)** would be equivalent to the sodium adduct of doubly dehydrogenated melianol B and therefore the new peak feasibly represents melianone B, after dehydrogenation of the C3 hydroxy group of melianol B (Figure 6.2.4.C). However, based on mass spectra alone it is not possible to confirm which of the three hydroxy groups (C3,C7,C21) of the melianol B scaffold has been dehydrogenated, or whether MaSDR1 is alternatively catalysing the formation of an alkene rather than a carbonyl. Based on predicted pathway steps (Figure 6.1.1) and the relatively high conversion efficiency (Figure 6.2.4.A), the C3 position seems the most likely target of MaSDR1. Further, the promiscuous AtTHAR1 from *A. thaliana*, which performs this function on a range of triterpene scaffolds, shares 44% protein identity to MaSDR1. MaSDR1 can also doubly dehydrogenate a melianol scaffold (Appendix A.11), suggesting it is promiscuous within this pathway.
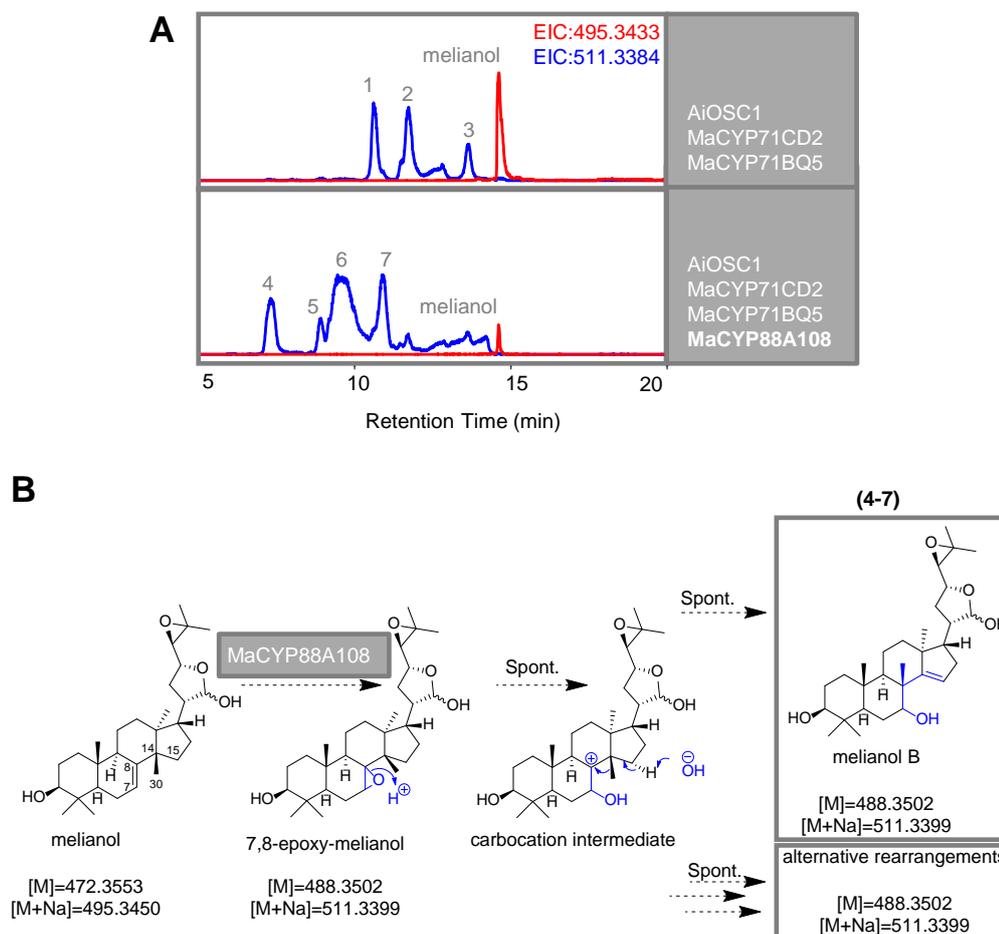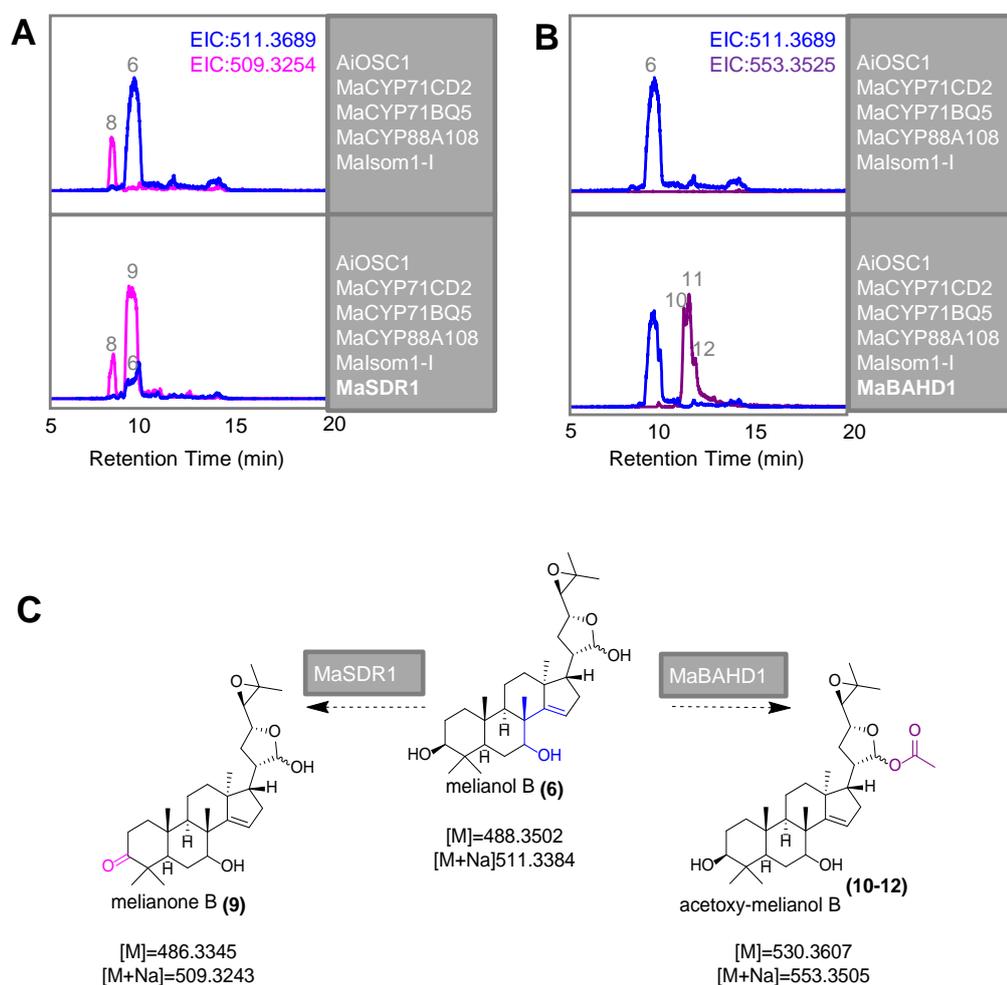
Figure 6.2.4: **Expression of *MaSDR1* and *MaBAHD1* in *N. benthamiana*.**
UHPLC-IT-TOF EICs of extracts from *N. benthamiana* leaves exptressing melianol
B biosynthetic genes (*AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108* and
MaIsom1-I) with and without *MaSDR1* (A) and with and without *MaBAHD1* (B). The
UHPLC-IT-TOF 'limonoid' gradient was used. The EICs shown are for the following
adducts: $[melianol+O+Na]^+$=511.3369 (blue), $[melianol+O-2H+Na]^+$=509.3254 (pink)
and $[melianol+O+CH_2CO+Na]^+$=553.3525 (purple). Melianol peaks are labelled as
follows: melianol B **(6)**; dehydrogenated oxidised melianol B **(8-9)**; oxidised melianol
B acetate **(10-12)**. The mass spectra for newly identified peaks can be found in
(Appendix A.13). (C) Predictions of structures, exact masses and mechanisms of
formation of the new peaks **(8-12)**.

MaBAHD1 activity

MaBAHD1 was selected based on its annotation as a 'vinorine synthase'. Vinorine synthases are a specific enzyme-type within the Benzylalcohol acetyl-, anthocyanin-*O*-hydroxy-cinnamoyl-, anthranilate-*N*-hydroxy-cinnamoyl/benzoyl-, deacetylvindoline acetyltransferase (BAHD) superfamily, which transfer acetyl groups in monoterpenoid indole alkaloid biosynthetic pathways (276). Such acetyl decorations are common in protolimonoid and limonoid structures. Many complex *seco*-C-limonoids, such as azadirachtin and salannin, contain acetate groups along with simple ring intact limonoids such as azadirone, which contains an acetate group at C7, and salimuzzalin, which contains two acetate groups on the furan ring (Figure 6.2.5). Protolimonoids such as turreanthin and 21-β-acetoxymelianone also contain acetate groups at the C3 position and hemiacetal ring respectively (Figure 6.2.5).
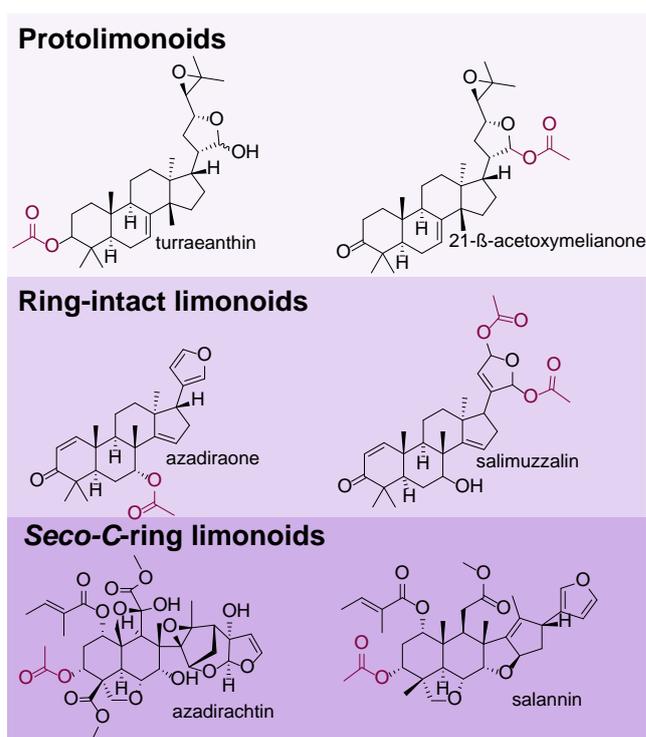
Figure 6.2.5: **Examples of protolimonoid and limonoid structures with acetate groups.**
Structures of protolimonoid, ring intact and *seco*-C-ring limonoid structures with acetate groups (purple). Structures have been isolated from a variety of species: azadiraone, salimuzzalin, salannin, azadirachtin are all reported from *A. indica* (5), turreanthin from *Turreanthus africanus* (156) and 21-β-acetoxymelianone along with salannin from *M. azedarach* (277).

In the absence of *MaBAHD1*, when melianol biosynthetic genes (*AiOSC1,*

MaCYP71CD2 and *MaCYP71BQ5*) are expressed with *MaCYP88A108* and *MaIsom1-I*, there is no observations of peaks with a mass suggesting acetylation has occurred (Figure 6.2.4.B). However, when *MaBAHD1* is expressed in combination with these genes, the melianol B peak is partially consumed and a new broad peak with a mass of 553.3525 is detected, which appears may consist of three distinct peaks **(10-12)** (Figure 6.2.4.B).

This mass of these new peaks **(10-12)** is consistent with a sodium adduct of melianol B acetate, suggesting MaBAHD1 functions by acetylation of a hydroxy group on the melianol B scaffold (Figure 6.2.4.C). The occurrence of three peaks **(10-12)** could indicate two possible scenarios. MaBAHD1 may be capable of acetylating each of the three different hydroxy groups of the melianol B scaffold. Alternatively, MaBAHD1 may act at only one position, but on an unresolved melianol B peak **(6)** that may represent multiple structures. Further purification and structural characterisation are required to address this.

The low level of consumption of melianol B by MaBAHD1 (Figure 6.2.4.B) may suggest that melianol B is not its preferred substrate and that in *M. azedarach* this enzyme may act later in the limonoid biosynthetic pathway. Acyl decoration are often added late in triterpene biosynthesis, such as the acylation of avenacin which occurs at a late stage of the pathway (278). Co-expression of *MaBAHD1* with melianol biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*) alone indicated that MaBAHD1 is also able to acetylating the melianol scaffold (Appendix A.11).

Expression of both *MaSDR1* and *MaBAHD1* in combination with the melianol B biosynthetic genes (*AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108* and *MaIsom1-I*) resulted in the production of new peaks with masses equivalent to sodium adducts of double dehydrogenated and acetylated melianol B (Appendix A.12).

6.3 Conclusion

In conclusion, initial attempts to investigate and utilise the newly generated *M. azedarach* sequencing data have been highly successful, yielding four functional genes. The enzymes encoded by two of these newly identified genes, MaCYP88A108 and MaIsom1-I, are predicted to perform critical steps in the conversion of protolimonoids to limonoids. MaCYP88A108 is predicted to epoxidise the C7,8 alkene of melianol, while MaIsom1-I, which is annotated as a sterol-7,8-isomerase, is predicted to control the rearrangement of this structure to melianol B (via a novel mechanism for this class of enzyme).

Additionally a SDR (MaSDR1) and a BAHD (MaBAHD1) have been identified as dehydrogenating and an acetylating melianol-type scaffolds. MaBAHD1 appears to be less efficient at converting its substrate than the previously characterised protolimonoid biosynthetic enzymes (Chapter 2) and therefore may be a tailoring rather than a core biosynthetic enzyme. Large-scale expression experiments in *N. benthamiana* in the future will enable purification and structural confirmation of the protolimonoids that these new enzymes produce.

In addition to the identification of candidate genes, the RNA-Seq datasets generated for *M. azedarach* have demonstrated again the closely linked expression patterns of the melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*). The four genes identified in this chapter (*MaCYP88A108, MaIsom1-I, MaBAHD1* and *MaSDR1*) also share this expression pattern which further supports their role in limonoid biosynthesis. If the protolimonoids that these genes produce can be profiled in *M. azedarach* tissue (as was successful for dihydroniloticin and melianol (Chapter 4)) this will further support their proposed role.

6.4   Chapter-specific materials and methods

6.4.1   Differential gene expression analysis

Following methodology and best practice given by 'Intro2RNAseq' (Weill Cornell Medical College) (250) three methods of differential expression analysis (DEseq2 (237), EdgeR (165), Limma (279)) were performed on raw read counts in R (221). The number of differentially expressed genes (P-value < 0.05) identified by each method were compared (Figure 6.4.1) and as the greatest number of differentially expressed genes were identified by EdgeR V3.22.5 (165) this method was used for further analysis.

Briefly, raw read counts were imported into an EdgeR object and genes with low coverage (with less than one count per million in more than four samples) were discarded. Normalisation (by library size) was performed using the 'trimmed mean of M-values' method. To identify differentially expressed genes, a genewise negative binomial generalized linear model (glmQLFit) was used with pairwise comparisons between all sample types. DEseq2 V1.22.1 (237) was used to produce read counts for hierarchical clustering, by removing read counts of zero, normalising by library size and performing $\log_2$ transformation with a pseudo count of one.
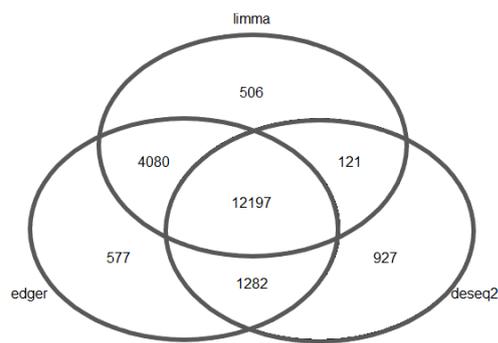
Figure 6.4.1: **Number of genes identified as differentially expressed via different methods.**
Number of genes identified as differentially expressed in *M. azedarach* EIv1 using DEseq2 (237), EdgeR (165) and Limma (279) (P-value < 0.05)

.

6.4.2 Hierarchical clustering

The $\log_2$-library-normalised read counts were used for hierarchical clustering and plotting as described in Chapter 9. For the identification of co-expressed genes discussed in this chapter, a number of methods were used to capture the widest possible pool of co-expressed genes. The criteria used were as follows: genes clustering with melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*) when using read counts from all 28 RNA-Seq tissues or using a mean value for each of the seven tissues and genes clustering with the latest melianol biosynthetic gene at time of analysis (*MaCYP71BQ5*) based on read counts from all 28 RNA-Seq tissues or using a mean value for each of the seven tissues.

7

Screening limonoid biosynthetic genes for their contribution to anti-insect activity

Synopsis

A heavy focus on known antifeedant limonoids, such as azadirachtin, has led to a lack of knowledge regarding the anti-insect effect of more minor limonoid structures, including protolimonoids. Thus there is relatively little understanding of how the structure-activity relationships of limonoids develop throughout their biosynthesis. An assay has been developed to investigate this by transiently expressing early limonoid pathway genes in *N. benthamiana* and evaluating the effects of their expression on tobacco hornworm (*Manduca sexta*) feeding. These experiments indicate that production of melianol significantly reduces *M. sexta* feeding. This is consistent with earlier literature reports that certain protolimonoids can cause feeding inhibition, despite the fact that they represent very early stages in limonoid biosynthesis. Moreover, the insect feeding assay developed here will enable step-wise elucidation of the complicated structure-activity relationships of the limonoids, by enabling suites of structurally related analogs to be tested. Therefore, this assay may conceivably reveal as yet overlooked intermediates with comparable bioactivity to well known anti-insect limonoids such as azadirachtin.

7.1   Introduction

Despite nearly a century of research, the structure-activity of limonoids is far from fully understood. Although the *seco*-C-ring limonoid azadirachtin is renowned as the most famous antifeedant limonoid, an investigation of ~60 limonoids and derivatives identified five limonoid structures that gave a greater feeding inhibition of *Spodoptera litura* (tobacco cutworm) than azadirachtin (67). These included diverse limonoid structures such as photooxidation products of ring-intact and *seco*-C-ring limonoids and azadirachtol, which differs from azadirachtin only in modifications to the A-ring (Figure 7.1.1). These results suggest that antifeedant effect does not simply increase incrementally as the pathway progresses towards azadirachtin. Azadirachtin has most probably received such considerable attention due to its abundance in the seed kernel of *A. indica*, rather than potency alone (16).
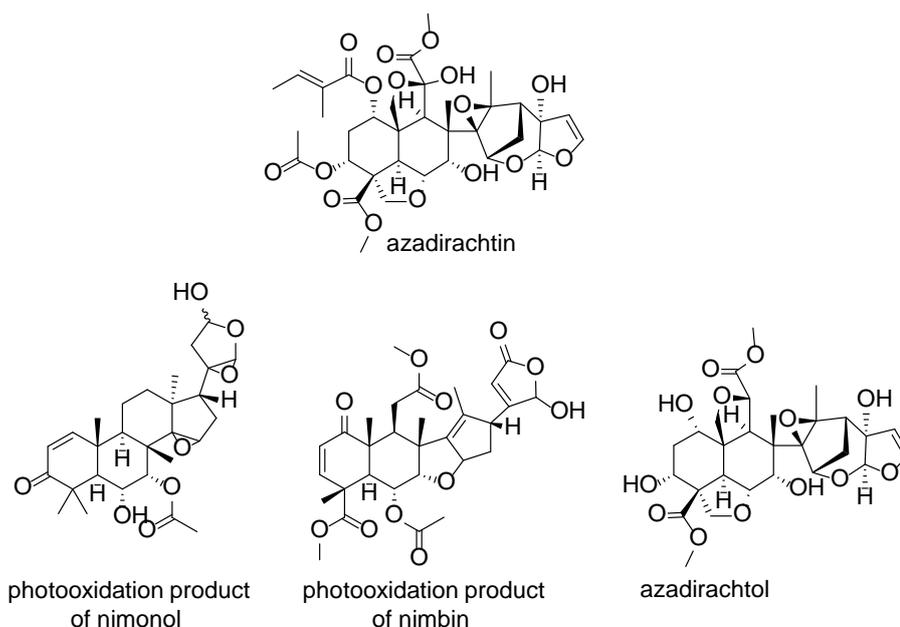
Figure 7.1.1: **Structures with greater anti-insect activity than azadirachtin.**
Structure of azadirachtin along with structure of the photooxidation product of nimonol (280) and of nimbin (281) and azadirachtol (282), all of which have reported greater antifeedant activities towards *S. litura* than azadirachtin (67).

Further, as an assay of just ~60 limonoids identified a number of potent bioactivities, there is a great potential for discovery of other potent anti-insect activities amongst the ~2,000 limonoid structures isolated to date (4–9). Many of these structures will not yet have been accessible in quantities sufficient for such assays.

Structure-activity studies of the limonoids have also revealed that, after the *seco*-C-ring limonoids, the second most active group are ring-intact limonoids.

Ring-intact limonoids have markedly different 3D-structures when compared to their *seco*-derivatives (14, 67). The ring-intact and complex *seco*-C-ring limonoids are effectively at different ends of the spectrum of limonoid structures, raising questions about how anti-insect effects may build and vary throughout limonoid biosynthetic pathways.

Protolimonoids have been isolated less frequently (Chapter 4) and so the biological activity data for these compounds, particularly relevant to anti-insect activity, is scant. For example, although the two protolimonoids, dihydroniloticin and melianol (produced by transient expression in Chapter 2) are a common occurrence in limonoid producing species (Appendix B.10), a literature search to identify anti-insect data for these compounds yielded no results. However, a 1992 review reported anti-insect activity for two protolimonoids with similar structures to melianol (14). Melianone (a C3 ketone derivative of melianol) and meliantriol (a melianol degradation product) have reported antifeedant activity towards Mexican bean beetles (*Epilachna varivestis*) (14) and desert locusts (*Schistocerca gregaria*) (134), respectively. This suggests that other protolimonoids (the very earliest of limonoid structures) may also have anti-insect activities that have not as yet been identified.

The characterisation of three protolimonoid biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*) and identification of a further four genes encoding enzymes with activity towards protolimonoids (*MaCYP88A108, MaIsom1-I, MaSDR1* and *MaBAHD1*) has enabled the production, via transient expression in *N. benthamiana*, of nine different protolimonoid structures. These protolimonoids could in principle be purified in sufficient quantities for use in biological activity assays, as has been recently demonstrated for β-amyrin type triterpenes using this system (112). However, this would require large-scale transient plant expression followed by purification, which can be time-consuming. The Bak group at Copenhagen University recently reported a protocol for directly evaluating the anti-insect effects of engineered hederagenin-type triterpenoid saponins in *N. benthamiana*, following transient expression (283). This involved presenting *N. benthamiana* leaves transiently expressing different combinations of genes to insects in a no-choice feeding assay. Prior to the development of this assay, assessment of insect activity through heterologous expression in *N. benthamiana* had been limited to evaluating the effects of single proteins (284, 285), rather than the effects of metabolites produced by a series of transiently expressed enzymes. Therefore, this presented an opportunity to develop a method for assessing the anti-insect effects of engineered limonoid pathways, in parallel with the discovery of the biosynthetic genes required to produce them.

The Bak group utilised larvae of tobacco hornworm (*M. sexta*), an insect species that

has been shown to be sensitive to limonoids. Direct injection of azadirachtin causes a developmental response in *M. sexta* (286). Furthermore, azadirachtin, salannin, nimbin and 6-desacetylnimbin all inhibit the *M. sexta* enzyme ecdysone-20-monooxygenase (the enzyme responsible for the creation of the crucial moulting hormone 20-hydroxy-ecdysone (287)) *in vitro* (84). Since *M. sexta* is sensitive to a broad range of limonoids, the feasibility of using transient expression in *N. benthamiana* to investigate the structure-activity relationship of limonoids in direct feeding experiments with *M. sexta* was investigated.

7.1.1 Aims

Here, a strategy to directly assess the anti-insect effects of limonoids following transient expression in *N. benthamiana* is developed, to determine the anti-insect impact of the limonoid biosynthetic genes characterised thus far (Chapter 2, Chapter 6). To achieve this, a *M. sexta* colony was established in the insectary at JIC (by Darrell Bean), which allowed assays to be performed. The aims of this chapter are as follows:

1. Development and validation of a *N. benthamiana*-based assay for evaluating the antifeedancy of limonoids.

2. Evaluation of the anti-insect activity of functionally characterised limonoid biosynthetic genes.

7.2 Results and Discussion

7.2.1 Development and validation of a *N. benthamiana*-based assay for evaluating the antifeedancy of limonoids

Using *M. sexta* eggs kindly provided by the Baldwin group (Max-Planck Institue, Jena), the JIC Insectary established a healthy *M. sexta* colony, which produced eggs for a two-week period every six weeks. The antifeedant assay was developed at JIC using one day-old larvae from this colony based on the protocol kindly shared by Qing Liu (Bak Group, Copenhagen University).

Several trials were conducted to establish suitability and optimal conditions for this system. The assay preparation involved treating (by infiltration or painting) *N. benthamiana* leaves at least four days prior to the assay being performed. Further, before use in the assay, *M. sexta* larvae had to be acclimatised on uninfiltrated *N.*

benthamiana for 24 hours. To perform the assay, leaf discs were taken from treated leaves and placed in 24-well plates. An acclimatised day-old *M. sexta* larvae was then transferred onto each leaf disc. Images of the 24-well plates were taken before and after the assay, which was terminated after for 38-42 hours.

These preliminary trials demonstrated that one day-old *M. sexta* larvae were sensitive to the azadirachtin containing insecticide NeemAZAL T/S (Figure 7.2.1) at recommended concentrations (~0.005% azadirachtin), which corroborates previous literature (84, 286, 287). Further, although NeemAZAL T/S formulation is intended to be topically applied to plant surfaces, trials indicated that both topical application and leaf infiltration resulted in the same anti-insect effects. Therefore, NeemAZAL T/S was infiltrated in subsequent assays as this represented a closer experimental condition to agroinfiltration of limonoid biosynthetic genes.

negative   NeemAzal T/S
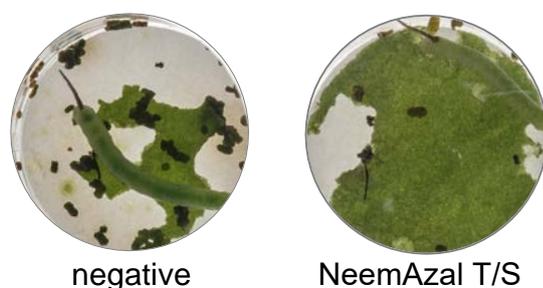
Figure 7.2.1: **Representative negative control and NeemAZAL T/S-treated *N. benthamiana* leaf discs after *M. sexta* feeding.**

Preliminary assays also revealed that the use of the triterpene-yield boosting gene, *AstHMGR*, resulted in leaf necrosis 5-6 days post-infiltration, dependent on the combination of other genes co-expressed. For triterpene extraction and analysis this necrosis is not problematic (112, 166). However, necrotic leaves are not a suitable feeding source for the larvae and associated leaf discolouration may interfere with downstream image analysis. Therefore for this assay, *N. benthamiana* leaves were harvested and used four days-post infiltration. Metabolite analysis indicated that limonoid intermediates were detectable in leaf extracts at this earlier timepoint.

To enable quantification of levels of feeding, an image analysis script was developed by Dr. Matthew Hartley (JIC informatics) to enable calculation of the percentage leaf disc area before and after feeding. Once a quantitative protocol had been established, further experiments were undertaken to confirm appropriate controls for investigating limonoids anti-insect activity using this system (Figure 7.2.2). Negative controls assessed included: no infiltration (negative A), infiltration with MMA solution only (negative B) and infiltration with an *A. tumefaciens* strain transformed with an empty pEAQ-HT-DEST1 vector (negative C). Positive controls assessed included: infiltration of NeemAZAL T/S (~0.005% azadirachtin)

immediately before the assay (positive A), infiltration of NeemAZAL T/S (~0.005% azadirachtin) in parallel to infiltration with *A. tumefaciens* strains four days prior to the assay (positive B) and infiltration of NeemAZAL T/S (~0.05% azadirachtin) immediately before the assay (positive C).

NeemAZAL T/S caused a significant reduction (~40%) in the feeding of *M. sexta* larvae in comparison to all negative controls (T-test, P-value < 0.05 (Figure 7.2.2)). The negative control selected for use in future assays was infiltration with an *A. tumefaciens* strain transformed with an empty pEAQ-HT-DEST1 vector (negative C), as this was considered to be the closest experimentally to infiltration of strains harbouring limonoid biosynthetic genes. The positive control selected for for use in future assays was infiltration of NeemAZAL T/S (~0.005% azadirachtin) four days prior to the assay (positive B), as this utilised the recommended concentration of this insecticide and involved infiltration in parallel to that of the *A. tumefaciens* strains.
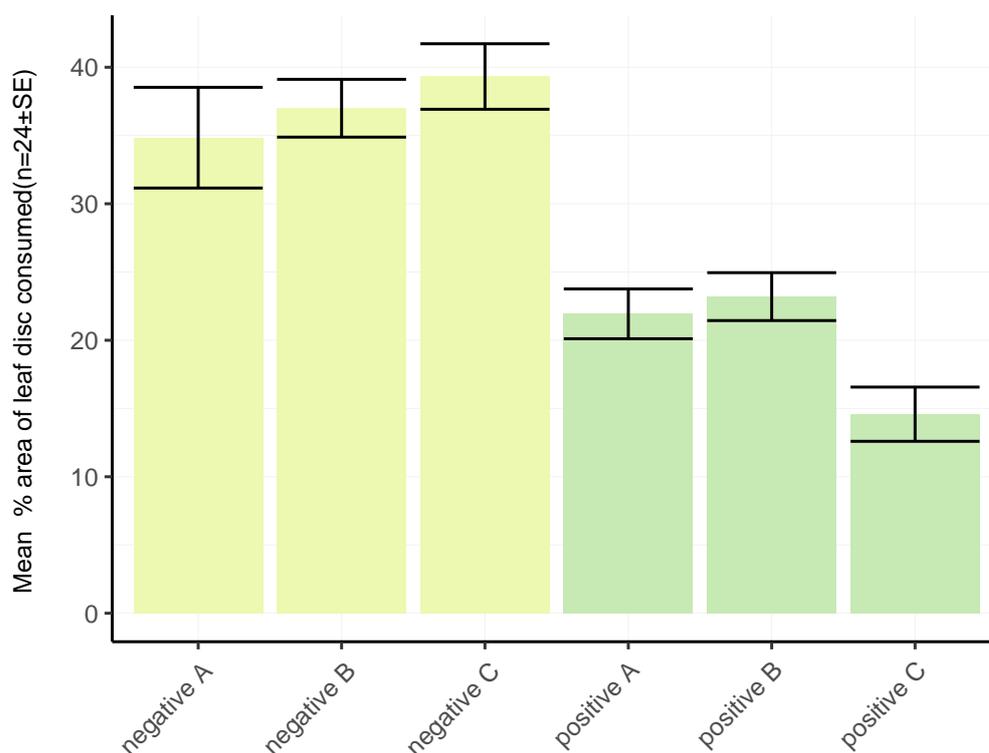
Figure 7.2.2: **Assessment of controls in *M. sexta* feeding assay**.
Barplot showing percentage of infiltrated *N. benthamiana* leaf disc consumed (n=24±SE) by *M. sexta* in feeding assay. Negative controls (yellow) and positive controls (green) are detailed in the main text.

7.2.2 Evaluation of the anti-insect activity of functionally characterised limonoid biosynthetic genes

The establishment of an *M. sexta* assay with appropriate controls enabled the anti-insect effect of limonoid biosynthetic genes identified thus far to be assessed (Figure 7.2.3).
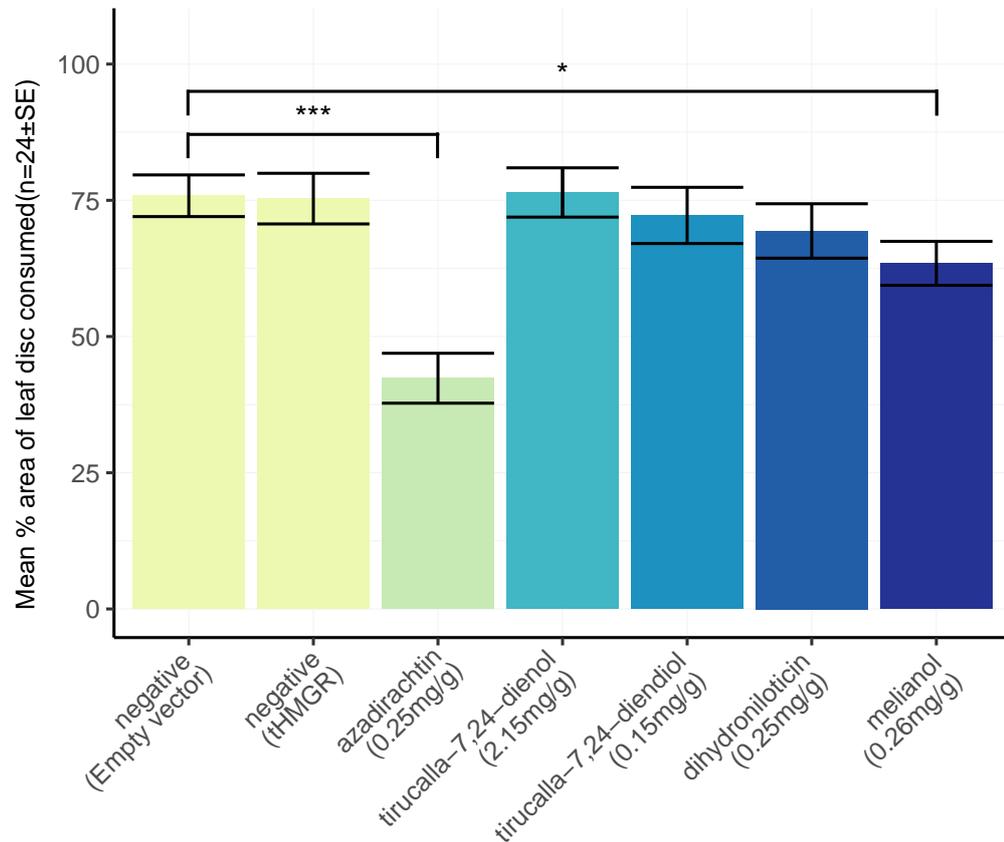
Figure 7.2.3: **Assessment of melianol biosynthetic genes in *M. sexta* feeding assay.** Barplot showing percentage of infiltrated *N. benthamiana* leaf disc consumed (n=24±SE) by *M. sexta* in feeding assay. Negative controls (yellow) are empty vector and *AstHMGR* infiltration and positive control (green) is infiltration with NeemAZAL T/S. Different combinations of co-expressed genes have been labelled by the protolimonoid of interest they produce and an estimated mean value of it's accumulation in leaf tissue used (n=3, mg/g DW). Full metabolite analysis is available in Appendix B.18. T-test significance values are indicated as follows: not significant (NS), P-value ≤ 0.05 (∗), 0.01 (∗∗), or ≤ 0.001 (∗ ∗ ∗).

An additional negative control was included in this assay, which was infiltration of leaves with *A. tumefaciens* strains harbouring the triterpene-yield boosting gene, *AstHMGR* alone. The combinations of genes tested included the three protolimonoid biosynthetic genes characterised in Chapter 2. Different combinations of genes were co-expressed by co-infiltration of *A. tumefaciens* strains harbouring the genes (Figure 7.2.3). This allowed the assessment of the anti-insect effect of the following

combination of genes: *AstHMGR* and *AiOSC1* (which together encode enzymes that produce tirucalla-7,24-dien-3β-ol); *AstHMGR, AiOSC1* and *MaCYP71BQ5* (which together encode enzymes that produce tirucalla-7,24-dien-3β,21-diol); *AstHMGR, AiOSC1* and *MaCYP71CD2* (which together encode enzymes that produce dihydroniloticin); and *AstHMGR, AiOSC1, MaCYP71CD2* and *MaCYP71BQ5* (which together encode enzymes that produce melianol).

Although there were no significant differences in feeding between the negative controls the majority of combinations assessed (Figure 7.2.3). Leaf discs from leaves expressing the genes required to produce melianol (*AstHMGR, MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*) showed a significant (P-value ≤ 0.05) reduction in feeding (~15%) in comparison to the negative control (Figure 7.2.3).

Although the anti-insect activity of melianol has not been previously reported, the activity observed in this work is in line with previous reports of the anti-insect effects of protolimonoids with similar structures (melianone and meliantriol (14, 134)). Further, this activity is consistent with the earlier observation that oxygenation of the furan ring, in the form of hydroxyl or epoxide groups, is associated with increased antifeedant effects (67). This suggests that the decrease in feeding activity observed for melianol may be due to its C21 hydroxyl containing hemiacetal ring, which is absent in the other protolimonoid structures assessed here. Indeed this may also explain the reported activity for melianone and meliantriol.

7.3   Conclusion

Here an *in planta* assay has been developed to assess the anti-insect effects of limonoids and limonoid pathway intermediates. This provides a new dimension to the system of transient production of triterpenes in *N. benthamiana* (112, 166, 288), since infiltrated leaves can be used directly in biological activity assays. This enables anti-insect activity to be assessed using a single infiltrated *N. benthamiana* plant, rather than the tens to hundreds of agro-infiltrated plants needed for large-scale compound purification.

This assay has enabled the anti-insect activity of new limonoid structures to be assessed in parallel with the identification of new limonoid biosynthetic genes, and revealed a previously unreported anti-insect effect of a protolimonoid. Transient expression of the genes required for melianol biosynthesis in *N. benthamiana* was sufficient to cause significant reduction in feeding of *M. sexta.* This is consistent with the antifeedant effects observed by others investigating similar protolimonoid scaffolds. Activity may be associated with the presence of a hydroxy group on the

hemiacetal ring, since hydroxylation of furan rings has been linked with increased anti-insect activity of limonoid structures (67).

The identification of antifeedant effects as early as three steps into the limonoid biosynthetic pathway confirms that the anti-insect activity of limonoids is not confined to *seco*-C-ring limonoids alone. Future experiments will enable the structure activity relationships of simple and more complex limonoid structures to be investigated using this metabolic engineering approach. Such investigations may reveal further examples of previously unreported limonoid bioactivity that may have so far been overlooked due to their limited accumulation in natural sources.

7.4 Chapter-specific materials and methods

7.4.1 *N. benthamiana* leaf disc based *M. sexta* feeding assay

The *M. sexta* assay was developed based on the protocol kindly provided by Qing Liu (Bak Group, Copenhagen University) (283).

M. sexta colony establishment and maintenance

JIC insectary team received eggs from Baldwin group (Max Planck Institute) and established a healthy *M. sexta* colony kept under controlled environmental conditions (26°C 10 h light, 20°C 14h dark, 48% humidity).

The JIC *M. sexta* colony produces eggs for a two week period every six weeks.

Assay preparation (infiltration and acclimatisation)

When the *M. sexta* colony was producing a sufficient number of eggs for experimentation (>200), *A. tumefaciens* liquid cultures were inoculated with strains harbouring genes of interest (from pre-prepared plates). Infiltration of *N. benthamiana* leaves with *A. tumefaciens* and NeemAZAL T/S solution (diluted 1:200 in ddH$_2$O to achieve recommended concentration) was performed the following day. *N. benthamiana* plants infiltrated had been grown in insecticide-free soil.

M. sexta larvae require acclimatisation before use in the assay to ensure feeding on infiltrated leaves. Therefore, three days post-lay *M. sexta* eggs were transferred onto uninfiltrated *N. benthamiana* leaves in petri-dishes containing moistened blue roll.

These were sealed with micropore and incubated under colony maintenance conditions.

Assay implementation

Three discs of filter paper (Whatman qualitative Grade 1 15 mm) were placed in each well of a 24-well plate (CLS3738 Corning Costar). 120 μl of sterile water was added to each well. Discs of infiltrated *N. benthamiana* leaves (four days post-infiltration) were prepared using a 10 mm diameter leaf cutter and placed into wells following a computer generated randomised experimental design (across all plates being used in the assay). Tweezers wrapped in parafilm were used to transfer leaves into well to avoid damage to the leaf discs. Remaining leaf material was freeze-dried and quantitative metabolite analysis was performed as described (Chapter 9).

A paintbrush was used to transfer an acclimatised five day post-lay (one day post-hatch) *M. sexta* larvae into each well, ensuring no contact between paintbrush and leaf. Plates were sealed with micropore tape and stacked, with a blank plate placed at the top avoid formation of condensation. Plates were incubated for 32-48 hours (dependant on the level of feeding) in a controlled environment identical to that used for *M. sexta* colony maintenance. Images of plates were taken by JIC photography before and after the assay.

Assay image analysis

Post-assay images were aligned to pre-assay images using the landmark correspondence plug-in in Fiji (289) to generate comparable post-assay images. Python was used to run the 'find green area differences' script, developed by Dr. Matthew Hartley (JIC informatics), to calculate the percentage of green area remaining after the assay. Example inputs and outputs of this script are provided in Figure 7.4.1. To calculate the percentage of feeding this value was subtracted from 100. Subsequent analysis and generation of plots was performed in R (221) using ggplot (223).
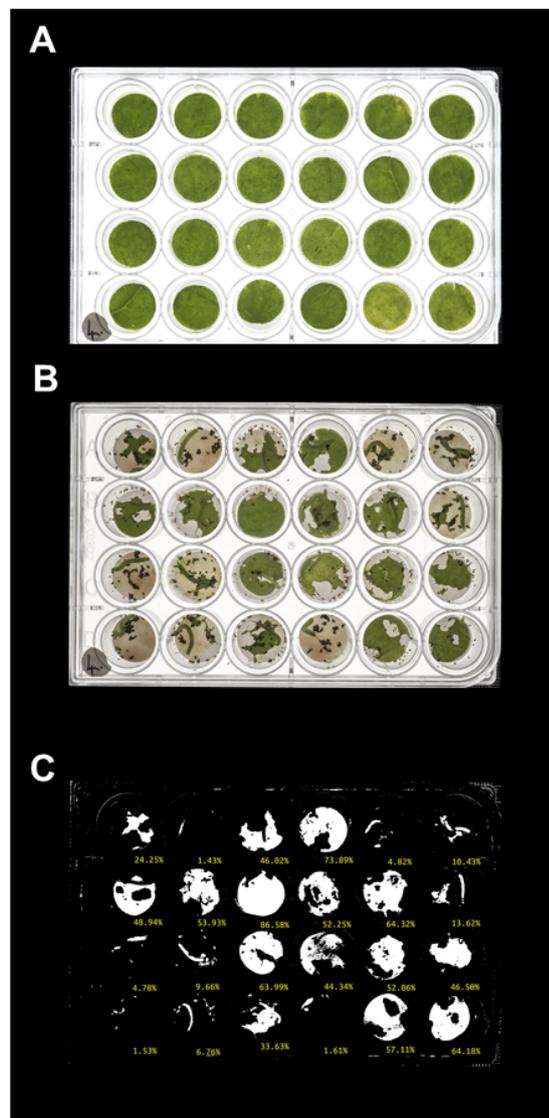
Figure 7.4.1: **Example image processing for *M. sexta* assays.**
(A) Pre-assay image; (B) post-assay image; (C) output of 'find green area differences'
script (percentage of green area remaining).

8

General Discussion

Limonoids (tetranortriterpenes) are a diverse and bioactive class of plant natural products. A near century of scientific interest in these compounds has led to the isolation of thousands of limonoid structures with diverse biological activities, along with the development of commercial crop protection formulations. The commercial inviability of chemical synthesis has meant that extraction from plant material is currently the sole source of limonoids (13, 16). The associated costs of this process have limited commercial use of limonoids (16). Furthermore the lack of availability of pure limonoids and structural variants there of have prevented in-depth investigations of their bioactivity (288). Metabolic engineering of limonoids could offer an alternative and viable source of limonoids, but first requires knowledge of limonoid biosynthesis which has remained entirely uncharacterised. The aim of this project was to address this by initiating elucidation of limonoid biosynthesis, to enable future metabolic engineering of limonoids for crop protection and other applications.

At the commencement of this project the OSC and associated triterpene scaffold implicated in initiating limonoid biosynthesis were unknown. Retrosynthetic analysis of protolimonoids, the presumed precursors to limonoids, implicated tirucalla-7,24-dien-3β-ol as a likely triterpene scaffold. However, early feeding studies with *A. indica* had suggested euphol to be the precursor to limonoids (150), which created ambiguity within the literature. To identify the OSC that carries out this first step, available genome resources for *C. sinensis* were mined along with *de novo* assembled transcriptomes for the Meliaceae family (Chapter 2). Closely homologous OSCs were identified from three limonoid-producing species, *C. sinensis, M. azedarach* and *A. indica*. Functional characterisation of these three OSCs (CsOSC1, MaOSC1 and AiOSC1) confirmed that all three produced tirucalla-7,24-dien-3β-ol. These results suggest that this first step in limonoid biosynthesis is conserved across the Meliaceae and Rutaceae family, and resolves the ambiguity of the scaffold initiating limonoid biosynthesis (Chapter 2).

The Meliaceae transcriptomes were further investigated, using a combination of differential expression analysis and phylogenetic analysis, leading to the identification of two functional CYPs (MaCYP71CD2 and MaCYP71BQ5) which together perform three oxidations on the side-chain of tirucalla-7,24-dien-3β-ol to produce the known protolimonoid, melianol (Chapter 2). Melianol has been previously reported from a number of Meliaceae species and its C3 carbonyl derivative, melianone, has been reported from a number of limonoid-producing species belonging to the Meliaceae, Rutaceae and Simaroubaceae families. These three enzymatic steps for melianol production were therefore postulated to represent a shared start in limonoid biosynthesis within the Sapindales. The identification of close homologs of all three genes in *C. sinensis* further supported this (Chapter 2). Collaborators at Stanford University were subsequently able to show that these *C. sinensis* CYPs were indeed functionally equivalent to those characterised from *M. azedarach* (122).

The initial characterisations of enzymes from Meliaceae species relied on publicly available RNA-Seq data generated by others. However, a full understanding of limonoid biosynthesis requires access to plant material, to enable in-depth analysis of correlations between metabolite content and transcriptome data in different tissues. Attempts were made to secure material in scope of the Nagoya Protocol through an existing collaboration with a Vietnamese University, but were unsuccessful (Chapter 3). However, limonoid-producing *M. azedarach* plants, originating from Japan, were secured through a source that was out of scope of the Nagoya Protocol and therefore available for use in this project (Chapter 3). This enabled metabolite and expression profiling of *M. azedarach* to be performed (Chapter 4). All tissues assessed in *M. azedarach* accumulated the limonoid salannin. However differing levels of salannin between tissues were identified, with high levels in the petioles and roots and low in the leaves, an observation which was reflected in the expression of melianol biosynthetic genes (Chapter 4).

This *M. azedarach* material was also used to generate a pseudochromosome-level genome assembly with accompanying RNA-Seq experiments designed to capture differences in limonoid content (Chapter 5). The genome was annotated using the same Earlham Institute pipeline and tools used for the hexaploid wheat genome (242, 290). This high-quality annotation has greatly progressed both the identification and cloning of candidate genes within this project compared to using transcriptome data alone.

The newly generated *M. azedarach* RNA-Seq data enabled exploitation of the strong co-expression of limonoid biosynthetic genes. This was not previously possible because, of the available RNA-Seq data within the Meliaceae family, all bar one (128) lacked repeats within their dataset, which are crucial for performing differential

expression analysis with statistical significance (230). Further, all previous RNA-Seq investigations of *M. azedarach* utilised leaf tissue only (128, 142, 143) which has now been demonstrated to be the lowest limonoid accumulating tissue in this species. The RNA-Seq data generated here (Chapter 5), when analysed in the context of the *M. azedarach* (EIv1) genome annotation, proved the close co-expression of the melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*). This linked expression and limonoid occurrence further supports the role of these genes in limonoid biosynthesis.

The melianol biosynthetic genes were then used as bait genes to identify (via differential expression and hierarchical clustering) a subset of genes with a shared expression pattern. Cloning and *N. benthamiana*-based expression of a selection of these genes, yielded four genes which, when co-expressed with melianol biosynthetic genes, were capable of modifying melianol-type scaffolds (Chapter 6). A CYP, *MaCYP88A108*, was identified which is predicted to epoxidise the C7-C8 alkene of melianol and lead to its spontaneously rearrangement into four different isomers. A second gene, *MaIsom1-I*, when expressed in combination with melianol biosynthetic genes and *MaCYP88A108*, appears to be capable of controlling this rearrangement and leading to the formation of a sole isomer. This implicates that the MaIsom1-I is functioning by an unknown mechanism different to its annotated function (as a sterol-8,7-isomerase). Based on the expected route of limonoid biosynthesis it is predicted that this structure produced by MaCYP88A108 and MaIsom1-I is similar to melianol, but with a mature limonoid scaffold (C14-C15 alkene, C7 hydroxyl, C8 methyl) rather than a protolimonoid scaffold (C7,8 alkene, C14 methyl). For simplicity, this structure has been termed melianol B in this work.

Two further melianol-acting genes (*MaBAHD1* and *MaSDR1*) have also been characterised from this subset via expression in *N. benthamiana* (Chapter 6). Based on MS data, MaBAHD1 and MaSDR1 are thought to modify melianol-type scaffolds by acetylation and double dehydrogenation, respectively. The position which MaBAHD1 acts is difficult to predict as acetate decorations occur at multiple positions on both protolimonoid and limonoid scaffolds. Further, as the efficiency of conversion by MaBAHD1 is relatively low, this enzyme may be a tailoring enzyme which acts later in the biosynthetic pathway. MaSDR1 is predicted to be acting at the C3 position, to convert the hydroxy group to a carbonyl. This is a required biosynthetic step to convert protolimonoids to limonoids and an SDR from *A. thaliana* has recently been characterising as performing this transformation at the C3 position of diverse triterpene scaffolds (121).

A priority for future work will be to use large-scale *N. benthamiana* infiltration (112, 166), as done for melianol and dihydroniloticin, to purify the yet to be confirmed products and determine their structure by NMR. The characterisation of

these structure will reveal whether predictions of C7-C8 epoxidation by MaCYP88A108 and subsequent conversion to a mature limonoid scaffold by MaIsom-1-I are correct and reveal the position of MaBAHD1 and MaSDR1 mediated modifications.

Triterpene biosynthetic gene clusters have been reported from a number of diverse plant species (120, 291). Although analysis of the newly generated *M. azedarach* genome using the plantiSMASH search algorithm (229) identified 47 putative biosynthetic gene clusters, these do not contain any of the seven functional genes identified thus far, which are spatially separated across six pseudochromosomes. *MaOSC1* does fall within a putative biosynthetic cluster. However, examination of the expression patterns of genes within this putative cluster revealed a lack of co-expression with melianol biosynthetic genes. There is still a possibility that later limonoid biosynthetic steps may be clustered. Therefore future work could investigate whether any of 47 gene clusters identified in *M. azedarach* may be relevant to later stages of limonoid biosynthesis. Both expression analysis and comparative genomics (using the newly released *A. indica* draft genome (126)) would be useful tools to investigate this.

Thus, this project has identified seven functional limonoid biosynthetic enzymes (Figure 8.0.1), three of which are capable of the production of the protolimonoid melianol (Chapter 2) and four further enzymes which modify this scaffold (Chapter 6). This represents significant progress in understanding limonoid biosynthesis, which at the commencement of this project remained completely uncharacterised. This elucidation, as well as providing a foundation for continuing investigations of limonoid biosynthesis, means that for the first time metabolic engineering of limonoids is becoming a possibility. Further, there are conceivably only a handful of enzymatic steps required to convert melianol B into a simple ring-intact limonoid such as azadirone (Figure 8.0.1). Therefore there is a potential that in the future heterologous systems could be used to produce high volumes of simple ring-intact limonoids, which could become new substrates for semi-synthesis of more valuable limonoids.

Further, using this new knowledge of limonoid biosynthesis (Figure 8.0.1) metabolic engineering can immediately begin to be used to investigate the structure activity relationships of limonoids. Expression of the newly identified limonoid biosynthetic genes in *N. benthamiana* can be used to screen for anti-insect activity using the *M. sexta* assay (developed in Chapter 7). Melianol-type structures have a reported anti-insect activity (14, 134), which was observed in this assay when expressing melianol biosynthetic genes in *N. benthamiana* (Chapter 7). This confirms that anti-insect activity occurs early within the limonoid biosynthetic pathway. Future assays, using the newly identified genes, will be able to determine whether the
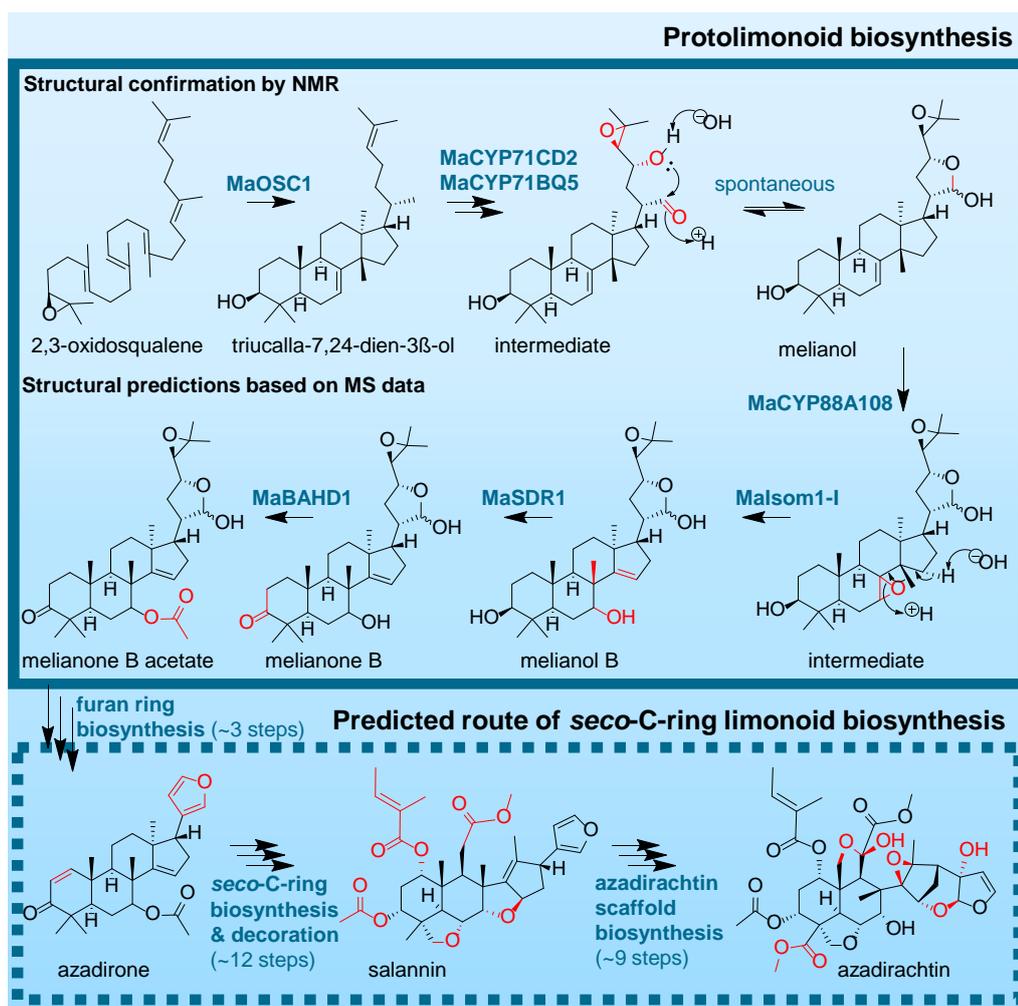
Figure 8.0.1: **Progress towards characterising *seco*-C-ring limonoid biosynthesis.**
Functions of MaOSC1, MaCYP71CD2 and MaCYP71BQ5 have been confirmed by
NMR. Functions of MaCYP88A108, MaIsom1-I, MaBAHD1 and MaSDR1 are based on
MS data and awaiting confirmation.

internal scaffold rearrangements, predicted to be performed by MaCYP88A108 and
MaIsom1-I, and modifications by MaBAHD1 and MaSDR1 change the anti-insect
activity of melianol. Therefore, the future use of this assay creates an opportunity to
gain a thorough understanding of the structure activity relationships of limonoids,
where previous studies have been limited by a lack of availability of structural
analogs of limonoids (14, 67).

8.1   Concluding remarks

This project has made inroads into elucidating the biosynthetic pathway of the limonoids, a class of plant natural product with important agronomical and potential pharmaceutical uses. The committing three enzymatic steps have been characterised, in both Meliaceae and Rutaceae species (122), and can turn 2,3-oxidosqualene into a multiply oxidised protolimonoid, melianol (Chapter 2). This knowledge will provide a foundation for all future investigations of limonoid biosynthesis. Beyond this, this project sourced (Chapter 3), profiled (Chapter 4) and sequenced (Chapter 5) *M. azedarach*, the second most prolific producer of the commercially relevant *seco*-C-ring limonoids. Using the newly generated *M. azedarach* resource, two limonoid biosynthetic genes have been identified which are speculated to produce a mature limonoid and a further two genes identified which can decorate this scaffold (Chapter 6). Such biosynthetic knowledge already has functional applications in metabolic biology. For instance, the development of a *N. benthamiana*-based *M. sexta*-assay has enabled the anti-insect activity of protolimonoids such as melianol to be assessed (Chapter 7). Therefore, in addition to providing a foundation of knowledge for limonoid biosynthesis, this project has developed resources which should enable the elucidation of later limonoid biosynthetic steps and simultaneously provide an understanding of which structures and steps are associated with the highest anti-insect effect.

9

General materials and methods

Some of the methods described in this chapter have been previously included in authors publication (122) and are reproduced here.

9.1 General bioinformatic methods

9.1.1 *De novo* transcriptome assembly

Raw RNA-Seq reads from two studies of *A. indica* (55, 126, 140) and one study of *M. azedarach* (128) were downloaded from NCBI-SRA (Table 2.1.1). Within each dataset, tissues were pooled and a reference transcriptome was assembled using Trinity *de novo* assembler V.R0140717 (144) following a standard protocol (292) (Table 2.2.1).

For protein annotation Augustus V3.2.2 (160) was used in intron-less mode with an *A. thaliana* training model and untranslated region (UTR) identification turned off.

9.1.2 Hierarchical clustering and heatmap generation

Hierarchical clustering was performed in R using $\log_2$ library-normalised reads generated by DEseq2 V1.22.1 (237). Correlation matrices were calculated using the Person (238) method and where correlation between samples was performed the Spearman method was used. Conversion of correlation matrices to distance matrices was performed based on complete linkages.

Heatmap3 V1.1.1 (271) was used to visualise this data. Where specified scaling by 'row' has been performed using the inbuilt Heatmap3 function, to aid visualisation of expression patterns rather than levels of expression. The dendrogram of clustered genes was cut iteratively to identify suitable sized clusters of co-expressed genes.

9.2 General molecular-biology methods

9.2.1 Maintenance of *M. azedarach* plants

16 young (<1 year) *Melia azedarach* plants was purchased from Crûg Farm Plants (UK) in summer 2017 and maintained in a John Innes Centre greenhouse (24 ºC, 16 h light, grown in John Innes Cereal mix). Plants are maintained in greenhouse conditions by performing periodic coppicing (to around 20cm in height) and spraying with Dynamec to control glasshouse red spider mite (*Tetranychus urticae*).

The provenance of these plants is Chikugogawa Prefectural Natural Park (Japan). Seeds were collected by Crûg Farm Plants in autumn 2015 from an area of the park with no sampling restrictions. Confirmation that the material was out of scope of the Nagoya protocol and Access and Benefit Sharing legislation was given by the National Focal Point of Japan.

9.2.2 RNA extraction and cDNA synthesis

To harvest samples for RNA extraction sterile scissors were used to isolate material which was placed in an Eppendorf tube and immediately flash-frozen in liquid N_2. Tissues were ground to a fine powder using a pestle and mortar, which had been pre-cooled to -80°C, with continuous addition of liquid N_2. Samples were stored at -80°C.

RNA was extracted from *M. azedarach* tissues using the MacKenzie-modified RNAeasy Plant Mini Kit (Qiagen) protocol developed for extraction from woody plants (293). Care was taken to ensure during washing steps the wash buffer reached all sides of the column to remove impurities. If required, purity ratios were improved by re-diluting the RNA in loading buffer and repeating the on column washing process. DNAase treatment was performed 'on-column' using DNAse (Promega) following the manufacturer's instructions.

First-strand cDNA synthesis was performed using the GoScript Reverse transcription system (Promega) following the manufacturer's instructions.

9.2.3 PCR conditions and primers

All PCRs were performed on a Mastercycler Pro PCR machine (Eppendorf). All primers were ordered from Sigma-Aldrich. Primers used in this project are displayed in Table 9.2.1.

PCR reactions for cloning were performed using Phusion polymerase (Promega) following the manufacturer's instructions. The annealing temperature, extension time, number cycles and quantity of template used were adjusted based on the type of template being used and size of target for amplification (based on manufacturer's recommendations). Gel electrophoresis was used to confirm the sizes of PCR fragments. For cloning, a 50 μl reaction was performed to ensure enough amplified PCR product.

Colony PCRs were performed using GoTaq Green Mastermix (Promega) following manufacturer's instructions. Cells from selected colonies were transferred using a sterile pipette tip into 20 μl of ddH$_2$O, 10 μl of which was retained for subsequent growth and 10 μl was denatured by heating to 95°C for 10 minutes. 1 μl of template was used as template material for 10 μl PCR reaction.

Table 9.2.1: **Table of primers**

Name	Sequence
Sanger sequencing	
GAL1-F	AATATACCTCTATACTTTAACGTC
pYES2-R	GCGTGAATGTAAGCGTGAC
AttL1-F	TCGCGTTAACGCTAGCAT
AttL2-R	GTAACATCAGAGATTTTGAGACAC
AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTA
AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTA
AiOSC1_middle-F	GGAAAGAATTGGCTTTCG
CsOSC1_middle-F	TGGGGAAAGAATTGGCTTTC
CsOSC2_middle-F	CTTCCTGGGGAAAGACTTGG
CsOSC3_middle-F	CATCTTGGGGAAAGACTTGG
Yeast recombination of OSCs (Ai1, Ma1 and Cs)	
AiOSC1_pYES2-F	ACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTAATGTGGAAGCTGAAGATTG
AiOSC1_join-R	TTCAGGATCCTCCACCCAACAAGCAAGCATACACAGC
AiOSC1_join-F	GCTGTGTATGCTTGCTTGTTGGGTGGAGGATCCTGAA
Ai/MaOSC1_pYES2R	GAATGTAAGCGTGACATAACTAATTACATGATGCGGCCCTTTAATTAGGCAATGGAAC
MaOSC1_pYES2-F	ACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTAATGTGGAAGCTGAAGGTTGCAGAG
CsOSC1_PYES2-F	ACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTAATGTGGAGGCTGAAGGTTGC
CsOSC1_join-R	CGTTTGGATCTTCAACCCAACAAGCAAGCATACACAACG
CsOSC1_join-F	CGTTGTGTATGCTTGCTTGTTGGGTTGAAGATCCAAACG
CsOSC1_PYES2-R	GAATGTAAGCGTGACATAACTAATTACATGATGCGGCCCTTTAATTAGGCAGTGGAACTC
CsOSC2_PYES2-F	ACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTAATGTGGAAGCTAAAGGTAGGAG
CsOSC2_join-R	CTTCCAAAGCTCTGCATCTTCATTCCATCCTCAGCAACCC
CsOSC2_join-F	GGGTTGCTGAGGATGGAATGAAGATGCAGAGCTTTGGAAG
CsOSC2_PYES2-R	GAATGTAAGCGTGACATAACTAATTACATGATGCGGCCCTTCAAGAAAGAGTAACCTGATG
CsOSC3_PYES2-F	ACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTAATGTGGAGGCTTAAGATTGG
CsOSC3_join-R	CCATTAGGATCTTCCGCCCAACAGGAGAGCATGTTCAGCG
CsOSC3_join-F	CGCTGAACATGCTCTCCTGTTGGGCGGAAGATCCTAATGG
CsOSC3_PYES2-R	GAATGTAAGCGTGACATAACTAATTACATGATGCGGCCCTTCAGAAAATCTTGGACGATTG
AtLUP5_PYES2-F	ACTACTAGCAGCTGTAATACGACTCACTATAGGGAATATTAATGTGGAGGTTAAAGGTAG
AtLUP5_PYES2-R	GAATGTAAGCGTGACATAACTAATTACATGATGCGGCCCTCTATAGATCTGCGTGATGT

Table 9.2.1: (continued)

Name	Sequence
Gateway cloning of OSCs (Ai1, Ma1 and Cs)	
AiOSC1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTGGAAGCTGAAGATTG
Ai/MaOSC1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAATTAGGCAATGGAAC
MaOSC1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTGGAAGCTGAAGGTTGCAGAGGG
CsOSC1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAATGTGGAGGCTGAAGGTTGC
CsOSC1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAATTAGGCAGTGGAACTC
CsOSC2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTGGAAGCTAAAGGTAGGAG
CsOSC2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAAGAAAGAGTAACCTGATG
CsOSC3_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTGGAGGCTTAAGATTGG
CsOSC3_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTCAGAAAATCTTGGACGATTG
Gateway cloning of CYPs from *M. azedarach* (Ma1)	
CYP72A720_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGTTATCTCTGAAATCGG
CYP72A720_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTATAATTTCTTTAAAATCAAG
CYP71BQ5_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGTTCAGACTGCCTGTTC
CYP71BQ5_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTTCTGAAAAGGAATACGAG
CYP88A108_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAGCTAAATTTCCTGTGG
CYP88A108_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAGAAGTTCTTGACCTTGATG
CYP71BQ6_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATTCTTCAATATCCC
CYP71BQ6_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCACTTCTGTAAAGGAAAACG
CYP71D557_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGGTTCAATTTGTTTCC
CYP71D557_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCATGGATGATCAGCCGG
CYP71CD2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAATCTCCAACTCGATTAC
CYP71CD2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAATTTTCCACCTCAATGTTG
CYP71BE124_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGTACCAACTTCCATC
CYP71BE124_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTAGACTAGTTTGGAGTTATTG
qRT-PCR primers from *M. azedarach* (Ma1)	
EF_qP-F	TTCACCCTTGGTGTCAAGCAGATC
EF_qP-R	ACTTCCTTAACTATTTCATCGTACG
18srRNA_qP-F	GCACGCGCGCTACACTGATGT
18srRNA_qP-R	GTGTGTACAAAGGGCAGGGACG
GADPH_qP-F	TCGGAATCAACGGTTTTGGAA
GADPH_qP-R	CCATTGACCGTGAACACTGT
ßactin_qP-F	CCAAGCAGCATGAAGATTAAGG
ßactin_qP-R	ATCTGCTGGAAGGTGCTGAG
OSC1_qP-F	GTTGAGCATACTTATGTGGAATGC
OSC1_qP-R	CAATCTCCTTCTTGCGATGATGTGGG
CYP71CD2_qP-F	TGTGATTTGGAGTGTATTTGC
CYP71CD2_qP-R	GGATTTCTCATCAATTCTGAC
CYP71BQ5_qP-F	CTGTCACTCTGGACATTTTCTCAG
CYP71BQ5_qP-R	GCTTGTCGGACCTCTTCTTGTGCC
Gateway cloning of candidate genes from *M. azedarach* (EIv1)	
OGD1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGGATCATCTGACTGC
OGD1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATGCTTTCTTTCCCACAG
OGD2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCAGAACGGATTGATGG
OGD2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAATATTTTGTGACGTCTATTAC
OGD2.1_AttB1	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAATATTTTTTGACTTCTACTATTAC
OGD2.1_AttB2	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCAGAACAGATTGAAGG

Table 9.2.1: (continued)

Name	Sequence
OGD2.2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAATCTCCGAATCACTTCC
OGD2.2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAAGTATGGTGGGATTAGG
BAHD1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCCTGAAATAATTTCC
BAHD1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCACATGGGACTTGGG
BAHD2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCTCAAGATTGTTTCTTC
BAHD2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAATTATGCTTGTATCAACAGAGG
BAHD3_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGCGACTCATCATCTG
BAHD3_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGCGAACTTTGGTCTTG
Isom1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGAACAAAAGTGCATCG
Isom1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAATTTAGTTTATCATTATCCATCC
AD1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGTTCTTCTTCTTATGTGG
AD1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTAATTGGTCTGGTCTACG
CYPU-1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTGCACTTCTTTAACTTTGGGG
CYPU-1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAATCCTCTTGACATGGAG
CYP88A108_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCTAAATTTCCTGTGG
CYP88A108_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGAAGTTCTTGACCTTGATG
CYPU-2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCTATGGTGGTGTTGGC
CYPU-2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGGGATTGATGGGCTTCAAG
CYP714-1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGTTTCATGTCATTTGCCG
CYP714-1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGAAGTGCTTCTGGAAG
CYP735-1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGACAGTGAAGAGAATC
CYP735-1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTACGAAGGTTTCCATGAAG
EH1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAGCTCTTCCTACCC
EH1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATTTGTTGTAGGGATATAGGCG
OR1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAAGTGCAAATCATTTCC
OR1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATGAAACTGGGTTTGG
CYP716-3_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCATCAAGTGGCTCTG
CYP716-3_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTACGACTCTCGTTTGATGAAGG
BAHD4_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGTGCAGTGCCTGAG
BAHD4_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATAACTCTGCATCAAGCTG
DT1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGAAGCTTTGCATCTTGG
DT1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTATAACTCTGCATCAAGCTG
AKR1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAACAGTTATTCATCCGCG
AKR1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAATTGATAAGATTATAAGCTTTC
AKR2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATAAAAGAGCAAGGATG
AKR2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAATGGAGAGTCATATTGAATAC
SDR1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGGATCATTTGAATGC
SDR1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTAAGCGAGACCACGATC
Est1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCAGATCATTTGAATGC
Est1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAGCTTCGATCTTGGG
Est2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCCAGTAACACCACC
Est2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTATTCATCCCAAAGCTC
TF2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTTTTCTGAGACATTG
TF2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAATCAGATTTTTCTGAG
AKR3_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCAACCTCATCTATTCTGTC
AKR3_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCAAGAATCAGTGTACTTAAAC
OR2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCGGATCATGCAGC

Table 9.2.1: (continued)

Name	Sequence
OR2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATTAAGCATCTGTGGCGTC
Hyd1_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATATGGATTCAAGCCAC
Hyd1_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTAGTCCCGATTGATGAAATGAACC
Hyd2_AttB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGATTCAACCAAGATTGAG
Hyd2_AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTACTATTCCTGCTTTATGAAATCG

Table 9.2.1: The sequences (nucleotide) of all primers used in this project are listed. Header of each section specifies purpose of primers, species and resource used to design the primers. Ai1 and Ma1 refer to the transcriptomes for *A. indica* and *M. azedarach*, respectively, generated from previously published RNA-Seq data (Table 2.2.1). Cs refers the *C. sinensis* protein annotation (GCF_000317415, (145)) and EIv1 refers to the *M. azedarach* genome assembly and annotation generated in Chapter 5. For gateway cloning primers AttB1 represent the forward primer and AttB2 represent the reverse. Where species prefix is not given species is specified in header.

9.2.4   Gateway cloning

The coding sequences of candidate genes (including OSCs, CYPs (Chapter 2) and a variety of other enzyme classes (Chapter 6)) were amplified by PCR with a forward primer containing 5' AttB1 site and a reverse containing 5' AttB2. PCR products were purified using QIAquick Gel Extraction Kit or QIAquick PCR Purification Kit (Qiagen).

To transfer the coding sequences into expression vectors, gateway technology (Invitrogen) was used following the manufacturer's instructions. Briefly, purified PCR fragments were transferred into donor vector pDNR207 by performing a BP recombination reaction followed by transformation into *Escherichia coli* (DH5α (ThermoFisher Scientific)). A colony PCR was performed, using AttL primers, to select positively transformed colonies, which were subsequently grown overnight to enable plasmid extraction using QIAprep Spin Miniprep Kit (Qiagen). Plasmids were sequenced using AttL1 and AttL2 primers (Sanger, Eurofins) to check for successful recombination and correct coding sequence.

An LR recombination reaction was performed to transfer the coding sequence of candidate genes from pDNR207 to the desired expression vector following the manufacturer's instructions. For expression in yeast, PYES2-DEST52 (Thermo-Fisher Scientific), pAG423GAL and pAG425GAL (Addgene) were used as expression vectors. The pEAQ-HT-DEST1 vector (294) (kindly provided by Lomonossoff laboratory) was used as an expression vector for *A. tumefaciens*-mediated transient expression in *N. benthamiana*.

9.2.5 Heterologous expression in yeast

Yeast strains and maintenance

S. cerevisiae strains GIL77 (295) and Y21900 (*MATa/α ura3Δ0 leu2Δ0 his3Δ1 met15Δ0/MET15 LYS2/lys2Δ0 ERG7/ERG7::kanMX4*) (EuroScarf) were used for expression of candidate OSC and CYP genes. Both strains have either partial (Y21900) or full (GIL77) loss of function of ERG7. All media used for GIL77 strain were supplemented with 20 μg/ml ergosterol (Fluka), 13 μg/ml hemin (Sigma-Aldrich) and 5 mg/ml Tween 80 (Sigma-Aldrich) with appropriate selection media (Table 9.2.2). Strains were grown at 30°C with liquid culture shaking at 200 rpm.

Table 9.2.2: **Media used for selection of Yeast**

Plasmid	Selection	Medium
No plasmid	NA	YPD
pYES2	URA3	SD-URA
pAG423GAL	HIS3	SD-HIS
pAG425GAL	LEU2	SD-LEU
pYES2,	URA3	
pAG423GAL,	HIS3	SD-URA -HIS -LEU
pAG425GAL	LEU2	

Supplement drop-out (SD) media used to select for each *S. cerevisiae* plasmid. Selection amino acids include uracil (URA), histidine (HIS) and leucine (LEU). Table has been reproduced from the author's previous publication (122).

Yeast recombination

Heterologous recombination was performed in the *S. cerevisiae* strain GIL77 (*MATa/α gal2 hem3-6 erg7 ura3-176*) (295). Sequences were amplified by PCR using primers with homologous sequences to the ends of linearised pYES2 vector cut at restriction sites XbaI and HindIII. The 5' of the forward primer overlapped with GAL1 promoter and that of the reverse primer with the CYC1 terminator sequence. Coding sequences of OSC candidates from *A. indica* and *C. sinensis* were synthesized by Integrated DNA Technologies (IDT) in two fragments. Fragment 1 was amplified with a pYES2-specific forward primer as described and a reverse primer with the 3' end complementary to the 5' end of fragment 2. Fragment 2 was amplified using a forward primer with the 5' end complementary to the 3' of fragment 1 and a pYES2-specific reverse primer as described above.

MaOSC1 was cloned from *M. azedarach* cDNA into pYES2, and the cloned gene for *AtLUP5*, a previously characterised OSC (149), was sourced through TAIR (Stock: U16880).

Yeast transformation

Competent cells of both GILL77 and Y21900 strains were generated by inoculating 50 ml of YPD media with 500 μl of an overnight culture and incubating for approximately 4 hours until OD_{600} was between 0.6 and 0.8. Cells were pelleted by centrifugation (3000 g, 5 min), washed in 0.5 volume of sterile water and re-suspended in 0.01 volume of ddH_2O. Cells were transferred to a sterile centrifuge tube, pelleted (3000 g, 5 min) and re-suspended in 0.01 volume of competent cell solution (5% v/v glycerol, 10% v/v DMSO). 50 μl aliquots of cells were frozen at -80°C with care being taken to ensure cells froze slowly by padding using styrofoam rack and blue roll.

Transformation of GIL77 and Y21900 was performed using a previously described 'high-efficiency' protocol (296). Briefly, competent cells were pelleted and the following reagents were consecutively added: 240 μl polyethylene glycol (MW 3350, 50% w/v), 35 μl 1M lithium acetate and 8 μl denatured salmon sperm (ssDNA, 2 mg/ml, 5-7 Kbp) (Sigma-Aldrich). DNA for transformation was then added to the mixture. For intact plasmids 0.3-1.5 μg was added. For recombination a minimum of 0.6 μg of each PCR product was added and linearised PYES2 vector was added at a 3:1 ratio. Cells were vortexed, incubated at 30°C for 30 minutes and heat-shocked at 42°C for 20 minutes before transferring onto an agar plate containing the relevant media (Table 9.2.2).

To confirm successful recombination and correct coding sequence of PCR fragments, plasmids were extracted using Zymoprep Yeast Plasmid Miniprep (Zymo Research), transformed into *E. coli* for propagation and extracted for sequencing.

Heterologous expression in Yeast

For expression of candidate genes, strains were first pre-cultured to saturation (~48 h) in SD + glucose (2% wt/vol) + [supplements]. Cells were then pelleted by centrifugation, washed in ddH_2O, resuspended in SD + galactose (2 % wt/vol) + [supplements] and cultured for a further 48 h before being pelleted for extraction. Small-scale experiments were performed in 10ml cultures.

9.2.6   *A. tumefaciens*-mediated transient expression in *N. benthamiana*

A. tumefaciens maintenance

A. tumefaciens strain LBA4404 was grown in LB media with 50 μg/ml rifampicin and 100 μg/ml streptomycin. Strains which had been transformed with pEAQ-HT-DEST1 vectors were additionally selected for using 50 μg/ml kanamycin.

Strains were grown at 28°C and liquid culture with shaking at 200 rpm.

A. tumefaciens transformation

Competent *A. tumefaciens* (LBA4404) cells were made by resuspension of cells in exponential phase of growth in ice-cold 20 mM aqueous $CaCl_2$, followed by flash-freezing in liquid N_2. For transformation, aliquots (50 μl) of competent cells were thawed and mixed with ~300 ng of pEAQ-HT-DEST1 vector. A cold shock was then performed by submerging in liquid N_2 for 5 minutes. After thawing at room temperature, cells were incubated in 1 ml of SOC media (28°C, 2 h) before plating.

Transformation of chemically competent *A. tumefaciens* (LBA4404) cells was performed as previously described (166).

A. tumefaciens-mediated transient expression in *N. benthamiana*

Small-scale agro-infiltration of *N. benthamiana* was performed as previously described (112, 294). Briefly, liquid cultures of transformed LBA4404 strains were incubated for ~24 hours. Cells were pelleted by centrifugation and resuspended in MMA buffer (10 mM MES (2-[N-morpholino]ethanesulfonic acid) pH 5.6, 10 mM $MgCl_2$, 100 μM acetosyringone). The OD_{600} values (1:10) were measured for each resuspended strain. This was then used to calculate and aliquot the volume of each resuspended strain which was required for the strain to have a final OD_{600} of 0.2 in 10 ml of infiltrating solution (MMA).

Leaves of ~5-7 week old *N. benthamiana* plants were used for infiltration. A guide hole was made with a pipette tip in each leaf before infiltrating using a needless-syringe. Leaves were harvested for analysis 3-6 days post infiltration.

Large-scale agro-infiltration was performed using vacuum-infiltration as previously described (166).

9.3 General limonoid extraction and analysis methods

9.3.1 Extraction from Yeast

After expression experiments, Yeast cells were saponified by resuspension in 250 μl of saponification reagent (20% (wt/vol) KOH in 50% (vol/vol) EtOH) and incubation for 2 h at 65°C. Triterpenes were then extracted in an equal volume of hexane and the hexane extracts were pooled and dried down. Analysis was performed by GC-MS.

9.3.2 Extraction from *N. benthamiana* (and *M. azedarach*)

Material was freeze-dried prior to extraction and ~10 mg of sample was weighed. For quantitative experiments weight of leaf disc was recorded. Homogenisation was performed using Tungsten Carbide Beads (3 mm; Qiagen) with a TissueLyser (1000 rpm, 1 min). Samples were agitated at 18°C for 20 min in 550 μl 100% methanol (5 μg/ml podophyllotoxin and 5 μg/ml coprostanol (Sigma-Aldrich)). 400 μl of supernatant was added to 140 μl ddH$_2$O. De-fatting by addition of hexane (400 μl) and subsequent removal of the upper phase (300 μl) was performed in triplicate. Both the hexane and methanol extracts were evaporated to dryness under N$_2$ gas or using Genevac EZ-2 evaporator (SP Scientific) and stored until required for analysis by GC-MS or UHPLC IT-TOF, respectively.

9.3.3 GC-MS sample preparation

Samples were derivitised before GC-MS analysis. Briefly, a desired volume of sample (dependant on concentration) was transferred to a glass autosampler vial and dried under N$_2$ gas or using Genevac EZ-2 evaporator (SP Scientific). 50 μl of 1-(Trimethylsilyl)imidazole - Pyridine mixture (Sigma-Aldrich) was added and samples were heated to 65°C for approximately 30 minutes. 1 μl of each sample was injected.

9.3.4 GC-MS analysis '20 minute' gradient

GC-MS analysis was undertaken using a 7890B GC (Agilent) and an electron-impact (EI) 5977A MSD (Agilent) fitted with a Zebron ZB5-HT Inferno column (Phenomenex) using methods previously described (112). Briefly, 1 μl of sample was injected (inlet 250°C) in pulse spiltless mode (pulse pressure 30 psi) with program of oven temperature at 2 min 170°C, ramp of 20°C/min to 300°C, 11.5 min at 300°C.

Detection was performed in scan mode (60-800 mass units) set to 7.2 after a solvent delay of 8 minutes. Data was analysed using MassHunter workstation (Agilent).

9.3.5 UHPLC-IT-TOF sample preparation

Dried extracts were re-suspended in 100 μl of methanol and filtered by centrifugation in Spin-X Centrifuge Filter Centrifuge Tubes (pore size 0.22 μm, Corning Costar). 50 μl of eluate was diluted in 50 μl of methanol and transferred to a glass insert inside a glass autosampler vial. 5 μl of sample was injected.

9.3.6 UHPLC-IT-TOF 'limonoid' gradient

LC-MS analysis was undertaken based on modifications to previously described methods (55). This utilised positive mode electrospray LC-MS using a Nexera/Prominence UHPLC equipped with an ion-trap ToF mass spectrometer (Shimadzu). Separation was on a 100×2.1 mm 100 Å2.6 μm Kinetex EVO C18 column (Phenomenex) using 0.1% formic acid in water (A) versus methanol (B) run at 500 μl/min, 40°C and the following gradient of solvent B: 32-60% from 0-3 min, 60-65% from 7-13 min, 65-90% from 13-13.5 min, 90% from 13.5-16.5 min, 90-40% from 16.5-17 min and 40% 17-20 min.

Full MS spectra were collected (m/z 200-2000) with a maximum ion accumulation time of 20msec, and automatic sensitivity control set to a target of 70% optimal base peak intensity. The instrument also collected data-dependent MS2 (m/z 50-2000) of the most abundant precursor ions, with an isolation width of m/z 3.0, 50% collision energy and 50% collision gas, and a fixed ion accumulation time of 10msec. Spray chamber conditions were 300°C heat block, 250°C curved desorbation line, 1.5 L/min nebuliser gas, and drying gas 'on'. The instrument was calibrated using sodium trifluoroacetate cluster ions according to the manufacturer's instructions, immediately before use based on modifications of previously described methods. LC-MSsolutions V3 (Shimadzu) was utilised to analyse chromatograms.

9.3.7 UHPLC-IT-TOF 'protolimonoid' gradient

To identify less polar protolimonoids with later elution times the gradient of the 'limonoids' LC-MS program was modified. The modified gradient used 0.1% formic acid in water (A) versus methanol (B) run at 500 μl/min, 40°C and the following gradient of solvent B; 70-95% from 0-10 min, 95% from 10-11 min, 95-70% from

11-11.1 min and 70% from 11.1-14.5 min. All other aspects of the program and analysis were unaltered.

9.3.8   Internal standard based quantification

LC-MSsolutions V3 (Shimadzu) was utilised for UHPLC-IT-TOF chromatograms and MassHunter workstation (Agilent) for GC-MS to perform peak identification and extract peak area of internal standard and target analyte. Azadirachtin, nimbolide (Sigma-Aldrich), salannin (Greyhound Chromatography), melianol and dihydroniloticin (Chapter 2) analytical standards were used to confirm retention times and mass adducts of target analytes. The internal standards (podophyllotoxin (UHPLC-IT-TOF) or coprostanol (GC-MS)) were used to calculate a relative concentration of target in sample, which was subsequently scaled by recorded dry weight to estimate amount of analyte in starting material. Downstream analysis was performed in R (221) and barplots constructed in ggplot (223).

9.3.9   General considerations for NMR

All NMR experiments and analyses were kindly performed by Dr. Michael Stephenson. NMR spectra were recorded in Fourier transform mode at a nominal frequency of 400 MHz for 1H NMR, and 100 MHz for 13C NMR (unless specified otherwise), using the specified deuterated solvent. Chemical shifts were recorded in ppm and referenced to the residual solvent peak or to an internal TMS standard. Multiplicities are described as, s = singlet, d = doublet, dd = doublet of doublets, dt = doublet of triplets, t = triplet, q = quartet, quint = quintet, tquin = triplet of quintets, m = multiplet, br = broad, appt = apparent; coupling constants are reported in hertz as observed and not corrected for second order effects.

Bibliography

[1] A. E. Osbourn and V. Lanzotti. *Plant-derived natural products.* Springer, 2009.

[2] C. R. Pye, M. J. Bertin, R. S. Lokey, W. H. Gerwick, and R. G. Linington. Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences*, 114(22):5601–5606, 2017.

[3] C. Owen, N. J. Patron, A. Huang, and A. Osbourn. Harnessing plant metabolic diversity. *Current Opinion in Chemical Biology*, 40(Supplement C):24–30, 2017.

[4] S. Hasegawa and M. Miyake. Biochemistry and biological functions of citrus limonoids. *Food Reviews International*, 12(4):413–435, 1996.

[5] Q. G. Tan and X. D. Luo. Meliaceous limonoids: chemistry and biological activities. *Chem Rev*, 111, 2011.

[6] Y. Zhang and H. Xu. Recent progress in the chemistry and biology of limonoids. *RSC Advances*, 7(56):35191–35220, 2017.

[7] R. A. Hill and J. D. Connolly. Triterpenoids. *Natural product reports*, 34(1):90–122, 2017.

[8] R. A. Hill and J. D. Connolly. Triterpenoids. *Natural product reports*, 32(2):273–327, 2015.

[9] R. A. Hill and J. D. Connolly. Triterpenoids. *Natural product reports*, 30(7):1028–1065, 2013.

[10] F. Chassagne, G. Cabanac, G. Hubert, B. David, and G. Marti. The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products. *Phytochemistry Reviews*, 18(3):601–622, 2019.

[11] R. Thimmappa, K. Geisler, T. Louveau, P. O'Maille, and A. Osbourn. *Triterpene Biosynthesis in Plants*, volume 65 of *Annual Review of Plant Biology*, pages 225–257. Palo Alto, 2014.

[12] A. Roy and S. Saraf. Limonoids: overview of significant bioactive triterpenes distributed in plants kingdom. *Biological and Pharmaceutical Bulletin*, 29(2):191–201, 2006.

[13] G. D. Manners. Citrus limonoids: analysis, bioactivity, and biomedical prospects. *Journal of Agricultural and Food Chemistry*, 55(21):8285–8294, 2007.

[14] D. E. Champagne, O. Koul, M. B. Isman, G. G. Scudder, and G. N. Towers. Biological activity of limonoids from the Rutales. *Phytochemistry*, 31(2):377–394, 1992.

[15] A. Mordue and A. Blackwell. Azadirachtin: an update. *Journal of insect physiology*, 39 (11):903–924, 1993.

[16] E. D. Morgan. Azadirachtin, a scientific gold mine. *Bioorganic & medicinal chemistry*, 17(12):4096–4105, 2009.

[17] S. C. Gupta, S. Prasad, A. K. Tyagi, A. B. Kunnumakkara, and B. B. Aggarwal. Neem

(*Azadirachta indica*): An indian traditional panacea with modern molecular basis. *Phytomedicine*, 34:14–20, 2017.

[18] J. F. Ayafor, B. L. Sondengam, J. D. Connolly, D. S. Rycroft, and J. I. Okogun. Tetranortriterpenoids and related compounds. Part 26. Tecleanin, a possible precursor of limonin, and other new tetranortriterpenoids from *Teclea grandifolia* Engl.(Rutaceae). *Journal of the Chemical Society, Perkin Transactions 1*, pages 1750–1753, 1981.

[19] B. Heasley. Synthesis of limonoid natural products. *European Journal of Organic Chemistry*, 2011(1):19–46, 2011.

[20] Dictionary of natural products (version 28.1). Published on the internet-date accessed:21.11.2019. URL http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml.

[21] V. Domingo, J. F. Arteaga, J. F. Q. del Moral, and A. F. Barrero. Unusually cyclized triterpenes: occurrence, biosynthesis and chemical synthesis. *Natural product reports*, 26(1):115–134, 2009.

[22] W. J. Baas. Naturally occurring *seco*-ring-A-triterpenoids and their possible biological significance. *Phytochemistry*, 24(9):1875–1889, 1985.

[23] J.-Q. Zhao, Y.-M. Wang, H.-T. Zhu, D. Wang, S.-H. Li, R.-R. Cheng, C.-R. Yang, Y.-F. Wang, M. Xu, and Y.-J. Zhang. Highly oxygenated limonoids and lignans from *Phyllanthus flexuosus*. *Natural products and bioprospecting*, 4(4):233–242, 2014.

[24] J.-Q. Zhao, Y.-M. Wang, H.-P. He, S.-H. Li, X.-N. Li, C.-R. Yang, D. Wang, H.-T. Zhu, M. Xu, and Y.-J. Zhang. Two new highly oxygenated and rearranged limonoids from *Phyllanthus cochinchinensis*. *Organic letters*, 15(10):2414–2417, 2013.

[25] B. Jabeen and N. Riaz. Isolation and characterization of limonoids from *Kigelia africana*. *Zeitschrift fuer Naturforschung B*, 68(9):1041–1048, 2013.

[26] D. S. Seigler. Limonoids, quassinoids, and related compounds. In *Plant Secondary Metabolism*, pages 473–485. Springer, 1998.

[27] G. Fiaschetti, M. A Grotzer, T. Shalaby, D. Castelletti, and A. Arcaro. Quassinoids: From traditional drugs to new cancer therapeutics. *Current medicinal chemistry*, 18(3):316–328, 2011.

[28] K. K. Purushothaman, K. Duraiswamy, J. D. Connolly, and D. S. Rycroft. Triterpenoids from *Walsura piscidia*. *Phytochemistry*, 24(10):2349–2354, 1985.

[29] Y. Ozaki, C. H. Fong, Z. Herman, H. Maeda, M. Miyake, Y. Ifuku, and S. Hasegawa. Limonoid glucosides in citrus seeds. *Agricultural and biological chemistry*, 55(1):137–141, 1991.

[30] V. De Feo, L. De Martino, E. Quaranta, and C. Pizza. Isolation of phytotoxic compounds from tree-of-heaven (*Ailanthus altissima* Swingle). *Journal of Agricultural and Food Chemistry*, 51(5):1177–1180, 2003.

[31] E. J. Koenen, J. J. Clarkson, T. D. Pennington, and L. W. Chatrou. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist*, 207(2):327–339, 2015.

[32] A. Muellner, D. Vassiliades, and S. Renner. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Plant Systematics and Evolution*, 266(3-4):233–252, 2007.

[33] File:orangebloss wb.jpg- wikamedia commons. URL https://commons.wikimedia.org/wiki/File:OrangeBloss_wb.jpg.

[34] File:ailanthus altissima 004.jpg- wikamedia commons, . URL https://commons.wikimedia.org/wiki/File:Ailanthus_altissima_004.JPG.

[35] File:khaya senegalensis leaf 05.jpg- wikamedia commons. URL https://commons.wikimedia.org/wiki/File:Khaya_senegalensis_leaf_05.JPG.

[36] File:neem (azadirachta indica) in hyderabad w img 6976.jpg- wikamedia commons, . URL https://commons.wikimedia.org/wiki/File:Neem_(Azadirachta_indica)_in_Hyderabad_W_IMG_6976.jpg.

[37] S. Arnott, A. Davie, J. M. Robertson, G. Sim, and D. Watson. 818. The structure of limonin: X-ray analysis of epilimonol iodoacetate. *Journal of the Chemical Society (Resumed)*, pages 4183–4200, 1961.

[38] S. Hasegawa, M. Miyake, Y. Ozaki, and M. Berhow. Limonin: A non-volatile bitter principle in citrus juice. 1996.

[39] S. Hasegawa, C. G. Suhayda, W.-J. Hsu, and G. H. Robertson. Purification of limonoid glucosyltransferase from navel orange albedo tissues. *Phytochemistry*, 46(1):33–37, 1997.

[40] M. Kita, Y. Hirata, T. Moriguchi, T. Endo-Inagaki, R. Matsumoto, S. Hasegawa, C. G. Suhayda, and M. Omura. Molecular cloning and characterization of a novel gene encoding limonoid UDP-glucosyltransferase in citrus. *Febs Letters*, 469(2-3):173–178, 2000.

[41] S. Hasegawa, Z. Herman, E. Orme, and P. Ou. Biosynthesis of limonoids in Citrus: sites and translocation. *Phytochemistry*, 25(12):2783–2785, 1986.

[42] P. Ou, S. Hasegawa, Z. Herman, and C. H. Fong. Limonoid biosynthesis in the stem of Citrus limon. *Phytochemistry*, 27(1):115–118, 1988.

[43] S. Hasegawa and Z. Herman. Biosynthesis of obacunone from nomilin in Citrus limon. *Phytochemistry*, 24(9):1973–1974, 1985.

[44] V. Severino, S. de Freitas, P. Braga, M. Forim, M. da Silva, J. Fernandes, P. Vieira, and T. Venâncio. New limonoids from *Hortia oreadica* and unexpected coumarin from *H. superba* using chromatography over cleaning sephadex with sodium hypochlorite. *Molecules*, 19(8):12031–12047, 2014.

[45] The plant list (2013). URL http://www.theplantlist.org/.

[46] S. Siddiqui. A note on the isolation of three new bitter principles from the nim oil. *Current Science*, 11(7):278–279, 1942.

[47] M. Harris, R. Henderson, R. McCrindle, K. Overton, and D. Turner. Tetranortriterpenoids—VIII: The constitution and stereochemistry of nimbin. *Tetrahedron*, 24(3):1517–1523, 1968.

[48] J. H. Butterworth and E. Morgan. Isolation of a substance that suppresses feeding in locusts. *Chemical Communications (London)*, (1):23–24, 1968.

[49] J. Butterworth and E. Morgan. Investigation of the locust feeding inhibition of the seeds of the neem tree, *Azadirachta indica. Journal of Insect Physiology*, 17(6):969–977, 1971.

[50] H. B. Broughton, S. V. Ley, A. M. Slawin, D. J. Williams, and E. D. Morgan. X-ray crystallographic structure determination of detigloyldihydroazadirachtin and reassignment of the structure of the limonoid insect antifeedant azadirachtin. *Journal of the Chemical Society, Chemical Communications*, (1):46–47, 1986.

[51] J. N. Bilton, H. B. Broughton, P. S. Jones, S. V. Ley, H. S. Rzepa, R. N. Sheppard, A. M. Slawin, D. J. Williams, Z. Lidert, and E. D. Morgan. An X-ray crystallographic, mass spectroscopic, and NMR study of the limonoid insect antifeedant azadirachtin and

related derivatives. *Tetrahedron*, 43(12):2805–2815, 1987.

[52] D. Taylor. Azadirachtin: A study in the methology of structure determination. *Tetrahedron*, 43(12):2779–2787, 1987.

[53] C. J. Turner, M. S. Tempesta, R. B. Taylor, M. G. Zagorski, J. S. Termini, D. R. Schroeder, and K. Nakanishi. An NMR spectroscopic study of azadirachtin and its trimethyl ether. *Tetrahedron*, 43(12):2789–2803, 1987.

[54] W. Kraus, M. Bokel, A. Bruhn, R. Cramer, I. Klaiber, A. Klenk, G. Nagl, H. Pöhnl, H. Sadlo, and B. Vogler. Structure determination by NMR of azadirachtin and related compounds from *Azadirachta indica* A. Juss (Meliaceae). *Tetrahedron*, 43(12):2817–2830, 1987.

[55] A. Pandreka, D. S. Dandekar, S. Haldar, V. Uttara, S. G. Vijayshree, F. A. Mulani, T. Aarthy, and H. V. Thulasiram. Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*). *Bmc Plant Biology*, 15:14, 2015.

[56] A. S. J. David and P. Dayanandan. Identification of terpenoid storage cells of Neem (*Azadirachta indica* A. Juss.). *Allelopathy Journal*, 4:303–312, 1997.

[57] X.-R. Yang, N. Tanaka, D. Tsuji, F.-L. Lu, X.-J. Yan, K. Ito, D.-P. Li, and Y. Kashiwada. Limonoids from the aerial parts of *Munronia pinnata*. *Tetrahedron*, page 130779, 2019.

[58] X. Fan, P. Lin, X. Lu, and J. Zi. A new spiro-type limonoid from *Azadirachta indica* A. Juss. *Tetrahedron letters*, 60(16):1158–1160, 2019.

[59] G. R. Pettit, D. H. Barton, C. L. Herald, J. Polonsky, J. M. Schmidt, and J. D. Connolly. Evaluation of limonoids against the murine P388 lymphocytic leukemia cell line. *Journal of natural products*, 46(3):379–390, 1983.

[60] M. Isman, J. Arnason, G. Towers, et al. Chemistry and biological activity of ingredients of other species of Meliaceae. The neem tree: *Azadirachta indica* A. Juss and other meliaceous plants: sources of unique natural products for integrated pest management, medicine, industry and other purposes. 1995.

[61] T. Govindachari, N. Narasimhan, G. Suresh, P. Partho, G. Gopalakrishnan, and G. K. Kumari. Structure-related insect antifeedant and growth regulating activities of some limonoids. *Journal of chemical ecology*, 21(10):1585–1600, 1995.

[62] H. Rembold and I. Puhlmann. Phytochemistry and biological activity of metabolites from tropical Meliaceae. In *Phytochemical Potential of Tropical Plants*, pages 153–165. Springer, 1993.

[63] W. Kraus. Biologically active compounds from Meliaceae. *Studies in organic chemistry*, 1984.

[64] A. Jennifer Mordue, M. S. Simmonds, S. V. Ley, W. M. Blaney, W. Mordue, M. Nasiruddin, and A. J. Nisbet. Actions of azadirachtin, a plant allelochemical, against insects. *Pesticide Science*, 54(3):277–284, 1998.

[65] NeemAZAL T/S. Published on the internet- date accessed: 27.11.2019. URL http://www.trifolio-m.de/portfolio/neemazal-ts-2/.

[66] C. M. Boursier, D. Bosco, A. Coulibaly, and M. Negre. Are traditional neem extract preparations as efficient as a commercial formulation of azadirachtin A? *Crop protection*, 30(3):318–322, 2011.

[67] G. Suresh, G. Gopalakrishnan, S. D. Wesley, N. Pradeep Singh, R. Malathi, and S. Rajan. Insect antifeedant activity of tetranortriterpenoids from the Rutales. A perusal of structural relations. *Journal of agricultural and food chemistry*, 50(16):4484–4490, 2002.

[68] R. Preeti, M. Mani, R. Swati, et al. Neem-the gift of nature. *Food Science Research*

Journal, 3(1):111–114, 2012.

[69] K. Girish and B. S. Shankara. Neem-a green treasure. *Electronic journal of Biology*, 4(3): 102–111, 2008.

[70] Y. Wang. A comprehensive report on clinical ascaris anthelmintic therapeutic effect of toosendanin pills. *J. Tradit. Chin. Med.*, 262:46–49, 1959.

[71] M. Kim, S. Kim, B. Park, K. Lee, G. Min, J. Seoh, C. Park, E. S. Hwang, C. Cha, and Y.-H. Kook. Antiviral effects of 28-deacetylsendanin on herpes simplex virus-1 replication. *Antiviral research*, 43(2):103–112, 1999.

[72] L. Wang, J. Z. Do Dang Khoa Phan, P.-S. Ong, W. L. Thuya, R. Soo, A. L.-A. Wong, W. P. Yong, S. C. Lee, P. C.-L. Ho, G. Sethi, et al. Anticancer properties of nimbolide and pharmacokinetic considerations to accelerate its development. *Oncotarget*, 7(28): 44790, 2016.

[73] R. V. Priyadarsini, P. Manikandan, G. H. Kumar, and S. Nagini. The neem limonoids azadirachtin and nimbolide inhibit hamster cheek pouch carcinogenesis by modulating xenobiotic-metabolizing enzymes, DNA damage, antioxidants, invasion and angiogenesis. *Free radical research*, 43(5):492–504, 2009.

[74] G. H. Kumar, R. V. Priyadarsini, G. Vinothini, P. V. Letchoumy, and S. Nagini. The neem limonoids azadirachtin and nimbolide inhibit cell proliferation and induce apoptosis in an animal model of oral oncogenesis. *Investigational new drugs*, 28(4):392–401, 2010.

[75] R. V. Priyadarsini, R. S. Murugan, P. Sripriya, D. Karunagaran, and S. Nagini. The neem limonoids azadirachtin and nimbolide induce cell cycle arrest and mitochondria-mediated apoptosis in human cervical cancer (HeLa) cells. *Free radical research*, 44(6): 624–634, 2010.

[76] P. Manikandan, P. V. Letchoumy, M. Gopalakrishnan, and S. Nagini. Evaluation of *Azadirachta indica* leaf fractions for in vitro antioxidant potential and *in vivo* modulation of biomarkers of chemoprevention in the hamster buccal pouch carcinogenesis model. *Food and chemical toxicology*, 46(7):2332–2343, 2008.

[77] A. Koul, N. Mukherjee, and S. C. Gangar. Inhibitory effects of *Azadirachta indica* on DMBA-induced skin carcinogenesis in Balb/c mice. *Molecular and cellular biochemistry*, 283(1-2):47–55, 2006.

[78] R. Dubey, K. Patil, S. C. Dantu, D. M. Sardesai, P. Bhatia, N. Malik, J. D. Acharya, S. Sarkar, S. Ghosh, R. Chakrabarti, et al. Azadirachtin inhibits amyloid formation, disaggregates pre-formed fibrils and protects pancreatic β-cells from human islet amyloid polypeptide/amylin-induced cytotoxicity. *Biochemical Journal*, 476(5):889–907, 2019.

[79] E. G. Miller, R. Fanous, F. Rivera-Hidalgo, W. H. Binnie, S. Hasegawa, and L. K. Lam. The effects of citrus limonoids on hamster buccal pouch carcinogenesis. *Carcinogenesis*, 10 (8):1535–1537, 1989.

[80] E. Kurowska, N. Borradaile, J. Spence, and K. Carroll. Hypocholesterolemic effects of dietary citrus juices in rabbits. *Nutrition research*, 20(1):121–129, 2000.

[81] E. M. Kurowska, C. Banh, S. Hasegawa, and G. D. Manners. Regulation of apo B production in HepG2 cells by citrus limonoids. ACS Publications, 2000.

[82] G. D. Manners, R. A. Jacob, P. Andrew III, T. K. Schoch, and S. Hasegawa. Bioavailability of citrus limonoids in humans. *Journal of agricultural and food chemistry*, 51(14):4156–4161, 2003.

[83] C. N. Ruscoe. Growth disruption effects of an insect antifeedant. *Nature New Biology*,

236(66):159, 1972.

[84] M. Mitchell, S. Smith, S. Johnson, and E. Morgan. Effects of the neem tree compounds azadirachtin, salannin, nimbin, and 6-desacetylnimbin on ecdysone 20-monooxygenase activity. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, 35(1-2):199–209, 1997.

[85] H. Rembold and R. Annadurai. Azadirachtin inhibits proliferation of Sf 9 cells in monolayer culture. *Zeitschrift für Naturforschung C*, 48(5-6):495–499, 1993.

[86] J.-F. Huang, K.-J. Shui, H.-Y. Li, M.-Y. Hu, and G.-H. Zhong. Antiproliferative effect of azadirachtin A on *Spodoptera litura* Sl-1 cell line through cell cycle arrest and apoptosis induced by up-regulation of p53. *Pesticide biochemistry and physiology*, 99(1):16–24, 2011.

[87] J. Huang, C. Lv, M. Hu, and G. Zhong. The mitochondria-mediate apoptosis of Lepidopteran cells induced by azadirachtin. *PLoS One*, 8(3):e58499, 2013.

[88] A. Salehzadeh, A. Akhkha, W. Cushley, R. Adams, J. Kusel, and R. Strang. The antimitotic effect of the neem terpenoid azadirachtin on cultured insect cells. *Insect biochemistry and molecular biology*, 33(7):681–689, 2003.

[89] A. Anuradha, R. S. Annadurai, and L. Shashidhara. Actin cytoskeleton as a putative target of the neem limonoid azadirachtin A. *Insect biochemistry and molecular biology*, 37(6):627–634, 2007.

[90] S. L. Robertson, W. Ni, T. S. Dhadialla, A. J. Nisbet, C. McCusker, S. V. Ley, W. Mordue, and A. J. Mordue. Identification of a putative azadirachtin-binding complex from Drosophila Kc167 cells. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, 64(4):200–208, 2007.

[91] X.-Y. Huang, O.-W. Li, and H.-H. Xu. Induction of programmed death and cytoskeletal damage on *Trichoplusia ni* BTI-Tn-5B1-4 cells by azadirachtin. *Pesticide biochemistry and physiology*, 98(2):289–295, 2010.

[92] B. Shu, J. Zhang, G. Cui, R. Sun, X. Yi, and G. Zhong. Azadirachtin affects the growth of *Spodoptera litura* Fabricius by inducing apoptosis in larval midgut. *Frontiers in physiology*, 9:137, 2018.

[93] K. Xie, L. Tian, X. Guo, K. Li, J. Li, X. Deng, Q. Li, Q. Xia, Y. Zhong, Z. Huang, et al. BmATG5 and BmATG6 mediate apoptosis following autophagy induced by 20-hydroxyecdysone or starvation. *Autophagy*, 12(2):381–396, 2016.

[94] J. N. Spradlin, X. Hu, C. C. Ward, S. M. Brittain, M. D. Jones, L. Ou, M. To, A. Proudfoot, E. Ornelas, M. Woldegiorgis, et al. Harnessing the anti-cancer natural product nimbolide for targeted protein degradation. *Nature chemical biology*, page 1, 2019.

[95] T. Abbas and A. Dutta. p21 in cancer: intricate networks and multiple activities. *Nature Reviews Cancer*, 9(6):400, 2009.

[96] T. Tokoroyama, Y. Fukuyama, T. Kubota, and K. Yokotani. Synthetic studies on terpene compounds. Part 13. Total synthesis of fraxinellone. *Journal of the Chemical Society, Perkin Transactions 1*, pages 1557–1562, 1981.

[97] J. Jauch. Total synthesis of azadirachtin—finally completed after 22 years. *Angewandte Chemie International Edition*, 47(1):34–37, 2008.

[98] G. E. Veitch, E. Beckmann, B. J. Burke, A. Boyer, S. L. Maslen, and S. V. Ley. Synthesis of azadirachtin: a long but successful journey. *Angewandte Chemie International Edition*, 46(40):7629–7632, 2007.

[99] S. V. Ley, A. Abad-Somovilla, J. C. Anderson, C. Ayats, R. Bänteli, E. Beckmann, A. Boyer,

M. G. Brasca, A. Brice, H. B. Broughton, et al. The synthesis of azadirachtin: a potent insect antifeedant. *Chemistry–A European Journal*, 14(34):10683–10704, 2008.

[100] S. Yamashita, A. Naruko, Y. Nakazawa, L. Zhao, Y. Hayashi, and M. Hirama. Total synthesis of limonin. *Angewandte Chemie International Edition*, 54(29):8538–8541, 2015.

[101] E. Corey and R. W. Hahl. Synthesis of a limonoid, azadiradione. *Tetrahedron letters*, 30 (23):3023–3026, 1989.

[102] J. M. Faber and C. M. Williams. A concise total synthesis of (±)-cipadonoid B from synthetic azedaralide. *Chemical Communications*, 47(8):2258–2260, 2011.

[103] J. M. Faber, W. A. Eger, and C. M. Williams. Enantioselective total synthesis of the mexicanolides: khayasin, proceranolide, and mexicanolide. *The Journal of organic chemistry*, 77(20):8913–8921, 2012.

[104] C. Lv, X. Yan, Q. Tu, Y. Di, C. Yuan, X. Fang, Y. Ben-David, L. Xia, J. Gong, Y. Shen, et al. Isolation and asymmetric total synthesis of perforanoid A. *Angewandte Chemie International Edition*, 55(26):7539–7543, 2016.

[105] J. Manosroi, Y. Chisti, and A. Manosroi. Biotransformation of cortexolone to hydrocortisone by molds using a rapid color-development assay. *Applied Biochemistry and Microbiology*, 42(5):479–483, 2006.

[106] S. Haldar, F. A. Mulani, T. Aarthy, and H. V. Thulasiram. Whole-cell mediated 11β-hydroxylation on the basic limonoid skeleton by *Cunninghamella echinulata*. *The Journal of organic chemistry*, 80(12):6490–6495, 2015.

[107] S. Haldar, S. P. Kolet, D. S. Dandekar, B. S. Kale, R. G. Gonnade, and H. V. Thulasiram. Biocatalyst mediated functionalization of salannin, an insecticidal limonoid. *RSC Advances*, 4(53):27661–27664, 2014.

[108] T. P. Howard, S. Middelhaufe, K. Moore, C. Edner, D. M. Kolak, G. N. Taylor, D. A. Parker, R. Lee, N. Smirnoff, S. J. Aves, et al. Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 110(19):7636–7641, 2013.

[109] S. Galanie, K. Thodey, I. J. Trenchard, M. F. Interrante, and C. D. Smolke. Complete biosynthesis of opioids in yeast. *Science*, 349(6252):1095–1100, 2015.

[110] N. D. Gold, E. Fossati, C. C. Hansen, M. DiFalco, V. Douchin, and V. J. Martin. A combinatorial approach to study cytochrome P450 enzymes for *de novo* production of steviol glucosides in baker's yeast. *ACS synthetic biology*, 7(12):2918–2929, 2018.

[111] P. Beyer, S. Al-Babili, X. Ye, P. Lucca, P. Schaub, R. Welsch, and I. Potrykus. Golden rice: Introducing the β-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. *The Journal of nutrition*, 132(3):506S–510S, 2002.

[112] J. Reed, M. J. Stephenson, K. Miettinen, B. Brouwer, A. Leveau, P. Brett, R. J. Goss, A. Goossens, M. A. O'Connell, and A. Osbourn. A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metabolic engineering*, 42:185–193, 2017.

[113] S. Racolta, P. B. Juhl, D. Sirim, and J. Pleiss. The triterpene cyclase protein family: a systematic analysis. *Proteins: Structure, Function, and Bioinformatics*, 80(8):2009–2019, 2012.

[114] I. Abe. Enzymatic synthesis of cyclic triterpenes. *Natural product reports*, 24(6):1311–1331, 2007.

[115] A. W. Munro, H. M. Girvan, A. E. Mason, A. J. Dunford, and K. J. McLean. What makes a P450 tick? *Trends in biochemical sciences*, 38(3):140–150, 2013.

[116] K. Jensen and B. L. Møller. Plant NADPH-cytochrome P450 oxidoreductases. *Phytochemistry*, 71(2-3):132–141, 2010.

[117] S. Bak, F. Beisson, G. Bishop, B. Hamberger, R. Höfer, S. Paquette, and D. Werck-Reichhart. Cytochromes P450. *The Arabidopsis Book/American Society of Plant Biologists*, 9, 2011.

[118] S. Ghosh. Triterpene structural diversification by plant cytochrome P450 enzymes. *Frontiers in plant science*, 8:1886, 2017.

[119] A. Leveau, J. Reed, X. Qiao, M. J. Stephenson, S. T. Mugford, R. E. Melton, J. C. Rant, R. Vickerstaff, T. Langdon, and A. Osbourn. Towards take-all control: a C-21β oxidase required for acylation of triterpene defence compounds in oat. *New Phytologist*, 221(3): 1544–1555, 2019.

[120] Z. Liu, H. G. Suarez Duran, Y. Harnvanichvech, M. J. Stephenson, M. E. Schranz, D. Nelson, M. H. Medema, and A. Osbourn. Drivers of metabolic diversification: how dynamic genomic neighborhoods generate new biosynthetic pathways in the brassicaceae. *New Phytologist*, n/a(n/a).

[121] A. C. Huang, T. Jiang, Y.-X. Liu, Y.-C. Bai, J. Reed, B. Qu, A. Goossens, H.-W. Nützmann, Y. Bai, and A. Osbourn. A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science*, 364(6440):eaau6389, 2019.

[122] H. Hodgson, R. De La Peña, M. J. Stephenson, R. Thimmappa, J. L. Vincent, E. S. Sattely, and A. Osbourn. Identification of key enzymes responsible for protolimonoid biosynthesis in plants: Opening the door to azadirachtin production. *Proceedings of the National Academy of Sciences*, 116(34):17096–17104, 2019.

[123] A. Akhila, M. Srivastava, and K. Rani. Production of radioactive azadirachtin in the seed kernels of *Azadirachta indica* (the Indian neem tree). *Natural Product Letters*, 11(2): 107–110, 1998.

[124] T. Aarthy, F. A. Mulani, A. Pandreka, A. Kumar, S. S. Nandikol, S. Haldar, and H. V. Thulasiram. Tracing the biosynthetic origin of limonoids and their functional groups through stable isotope labelling and inhibition in neem tree (*Azadirachta indica*) cell suspension. *BMC plant biology*, 18(1):230, 2018.

[125] S. Bhambhani, D. Lakhwani, P. Gupta, A. Pandey, Y. V. Dhar, S. K. Bag, M. H. Asif, and P. K. Trivedi. Transcriptome and metabolite analyses in *Azadirachta indica*: identification of genes involved in biosynthesis of bioactive triterpenoids. *Scientific reports*, 7(1):5043, 2017.

[126] N. M. Krishnan, S. Pattnaik, P. Jain, P. Gaur, R. Choudhary, S. Vaidyanathan, S. Deepak, A. K. Hariharan, P. B. Krishna, J. Nair, L. Varghese, N. K. Valivarthi, K. Dhas, K. Ramaswamy, and B. Panda. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genomics*, 13, 2012.

[127] N. A. Kuravadi, V. Yenagi, K. Rangiah, H. B. Mahesh, A. Rajamani, M. D. Shirke, H. Russiachand, R. M. Loganathan, C. S. Lingu, S. Siddappa, A. Ramamurthy, B. N. Sathyanarayana, and M. Gowda. Comprehensive analyses of genomes, transcriptomes and metabolites of neem tree. *Peerj*, 3, 2015.

[128] Y. Wang, X. Chen, J. Wang, H. Xun, J. Sun, and F. Tang. Comparative analysis of the terpenoid biosynthesis pathway in *Azadirachta indica* and *Melia azedarach* by RNA-Seq. *SpringerPlus*, 5(1):1–9, 2016.

[129] S. Bhambhani, D. Lakhwani, T. Shukla, A. Pandey, Y. V. Dhar, M. H. Asif, and P. K. Trivedi. Genes encoding members of 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) gene family from *Azadirachta indica* and correlation with azadirachtin biosynthesis. *Acta physiologiae plantarum*, 39(2):65, 2017.

[130] R. Xu, G. C. Fazio, and S. P. Matsuda. On the origins of triterpenoid skeletal diversity. *Phytochemistry*, 65(3):261–291, 2004.

[131] C. Paal. Ueber die Derivate des Acetophenonacetessigesters und des Acetonylacetessigesters. *Berichte der deutschen chemischen Gesellschaft*, 17(2): 2756–2767, 1884.

[132] L. Knorr. Synthese von Furfuranderivaten aus dem Diacetbernsteinsäureester. *Berichte der deutschen chemischen Gesellschaft*, 17(2):2863–2870, 1884.

[133] S. Siddiqui, T. Mahmood, B. S. Siddiqui, and S. Faizi. Isolation of a triterpenoid from *Azadirachta indica*. *Phytochemistry*, 25(9):2183–2185, 1986.

[134] D. Lavie, M. K. Jain, and S. Shpan-Gabrielith. A locust phagorepellent from two Melia species. *Chemical Communications (London)*, (18):910–911, 1967.

[135] F. S. Wang, M. Wang, X. N. Liu, Y. Y. Xu, S. P. Zhu, W. X. Shen, and X. C. Zhao. Identification of putative genes involved in limonoids biosynthesis in Citrus by comparative transcriptomic analysis. *Frontiers in Plant Science*, 8, 2017.

[136] B. S. Siddiqui, S. T. Ali, and S. K. Ali. Chemical wealth of *Azadirachta indica* (neem). *Neem: A Treatise, IK International Publishing House Pvt Ltd., New Delhi*, page 171, 2009.

[137] L. K. Narnoliya, R. Rajakani, N. S. Sangwan, V. Gupta, and R. S. Sangwan. Comparative transcripts profiling of fruit mesocarp and endocarp relevant to secondary metabolism by suppression subtractive hybridization in *Azadirachta indica* (neem). *Molecular biology reports*, 41(5):3147–3162, 2014.

[138] R. Rajakani, L. Narnoliya, N. S. Sangwan, R. S. Sangwan, and V. Gupta. Subtractive transcriptomes of fruit and leaf reveal differential representation of transcripts in *Azadirachta indica*. *Tree genetics & genomes*, 10(5):1331–1351, 2014.

[139] Citrus genome database. Published on the internet- date accessed: 29.09.2019. URL https://www.citrusgenomedb.org/.

[140] N. M. Krishnan, S. Pattnaik, S. Deepak, A. K. Hariharan, P. Gaur, R. Chaudhary, P. Jain, S. Vaidyanathan, P. B. Krishna, and B. Panda. De novo sequencing and assembly of *Azadirachta indica* fruit transcriptome. *Current Science*, pages 1553–1561, 2011.

[141] S. Wang, H. Zhang, X. Li, and J. Zhang. Gene expression profiling analysis reveals a crucial gene regulating metabolism in adventitious roots of neem (*Azadirachta indica*). *RSC Advances*, 6(115):114889–114898, 2016.

[142] J. H. Leebens-Mack, M. S. Barker, E. J. Carpenter, M. K. Deyholos, M. A. Gitzendanner, S. W. Graham, I. Grosse, Z. Li, M. Melkonian, S. Mirarab, et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, pages Epub–ahead, 2019.

[143] E. J. Carpenter, N. Matasci, S. Ayyampalayam, S. Wu, J. Sun, J. Yu, F. R. Jimenez Vieira, C. Bowler, R. G. Dorrell, M. A. Gitzendanner, et al. Access to RNA-Sequencing data from 1,173 plant species: The 1000 plant transcriptomes initiative (1KP). *GigaScience*, 8(10): giz126, 2019.

[144] M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. Full-

length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, 29, 2011.

[145] Q. Xu, L.-L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.-B. Jiao, B.-H. Hao, M. P. Lyon, et al. The draft genome of sweet orange (*Citrus sinensis*). *Nature genetics*, 45 (1):59, 2013.

[146] M. N. Price, P. S. Dehal, and A. P. Arkin. FastTree 2-approximately maximum-likelihood trees for large alignments. *PloS one*, 5(3):e9490, 2010.

[147] I. Letunic and P. Bork. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*, 44(W1):W242–W245, 2016.

[148] P. Morlacchi, W. K. Wilson, Q. Xiong, A. Bhaduri, D. Sttivend, M. D. Kolesnikova, and S. P. T. Matsuda. Product profile of PEN3: The last unexamined oxidosqualene cyclase in *Arabidopsis thaliana*. *Organic Letters*, 11(12):2627–2630, 2009.

[149] Y. Ebizuka, Y. Katsube, T. Tsutsumi, T. Kushiro, and M. Shibuya. Functional genomics approach to the study of triterpene biosynthesis. *Pure and applied chemistry*, 75(2-3): 369–374, 2003.

[150] D. E. U. Ekong, S. A. Ibiyemi, and E. O. Olagbemi. The Meliacins (limonoids). biosynthesis of nimbolide in the leaves of *Azadirachta indica*. *J. Chem. Soc. D*, pages 1117–1118, 1971.

[151] Arabidopsis cytochrome P450s homepage. Published on the internet- date accessed: 13.10.2019. URL http://www.p450.kvl.dk.

[152] N. M. Krishnan, P. Jain, S. Gupta, A. K. Hariharan, and B. Panda. An improved genome assembly of *Azadirachta indica* A. Juss. *G3: Genes|Genomes|Genetics*, 6(7):1835–1840, 2016.

[153] D. R. Nelson. The cytochrome P450 homepage. *Human genomics*, 4(1):59, 2009.

[154] W.-Y. Zhao, J.-J. Chen, C.-X. Zou, Y.-Y. Zhang, G.-D. Yao, X.-B. Wang, X.-X. Huang, B. Lin, and S.-J. Song. New tirucallane triterpenoids from *Picrasma quassioides* with their potential antiproliferative activities on hepatoma cells. *Bioorganic chemistry*, 84:309–318, 2019.

[155] T. Nakanishi, A. Inada, and D. Lavie. A new tirucallane-type triterpenoid derivative, lipomelianol from fruits of *Melia toosendan* Sieb. et Zucc. *Chemical and pharmaceutical bulletin*, 34(1):100–104, 1986.

[156] C. Bevan, D. Ekong, T. Halsall, and P. Toft. West African timbers. Part XX. The structure of turraeanthin, an oxygenated tetracyclic triterpene monoacetate. *Journal of the Chemical Society C: Organic*, pages 820–828, 1967.

[157] J. Polonsky, Z. Baskevitch-Varon, and B. C. Das. Triterpenes tetracycliques du *Simarouba amara*. *Phytochemistry*, 15(2):337–339, 1976.

[158] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. BLAST+: architecture and applications. *BMC bioinformatics*, 10(1):421, 2009.

[159] S. Racolta, P. B. Juhl, D. Sirim, and J. Pleiss. The triterpene cyclase protein family: a systematic analysis. *Proteins: Structure, Function, and Bioinformatics*, 80(8):2009–2019, 2012.

[160] M. Stanke and B. Morgenstern. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33:W465–W467, 2005.

[161] D. R. Nelson. Cytochrome P450 nomenclature, 2004. In *Cytochrome P450 protocols*,

pages 1–10. Springer, 2006.

[162] R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–1797, 2004.

[163] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3):R25, 2009.

[164] B. Li and C. N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12(1):323, 2011.

[165] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1): 139–140, 2010.

[166] M. J. Stephenson, J. Reed, B. Brouwer, and A. Osbourn. Transient expression in *Nicotiana benthamiana* leaves for triterpene production at a preparative scale. *Journal of visualized experiments: JoVE*, (138), 2018.

[167] Convention on biological diversity. Published on the internet- date accessed: 21.09.19. URL http://www.cbd.int.

[168] C. Aubertin and G. Filoche. The Nagoya Protocol on the use of genetic resources: one embodiment of an endless discussion. *Sustentabilidade em Debate*, 2(1):51–63, 2011.

[169] M. J. Oliva. Sharing the benefits of biodiversity: a new international protocol and its implications for research and development. *Planta medica*, 77(11):1221–1227, 2011.

[170] M. A. Bagley and A. K. Rai. The Nagoya Protocol and synthetic biology research: A look at the potential impacts. *Virginia Public Law and Legal Theory Research Paper*, (2014-05), 2014.

[171] M. A. Bagley. Digital dna: The nagoya protocol, intellectual property treaties, and synthetic biology. *Intellectual Property Treaties, and Synthetic Biology (February 2016). Virginia Public Law and Legal Theory Research Paper*, (11), 2016.

[172] Like-minded megadiverse countries. Published on the internet- date accessed: 01.10.19. URL https://www.environment.gov.za/likeminded_megadiversecountries_lmmc.

[173] Access and benefit sharing cleating house. Published on the internet- date accessed:10.10.2019. URL https://absch.cbd.int/.

[174] R. K. Joseph. International regime on access and benefit sharing: where are we now? *Asian Biotechnology and Development Review*, 12(3):77–94, 2010.

[175] P. Oldham, S. Hall, and O. Forero. Biological diversity in the patent system. *PLOS ONE*, 8(11):1–16, 11 2013.

[176] K. C. Justi, J. V. Visentainer, N. Evelázio de Souza, and M. Matsushita. Nutritional composition and vitamin c stability in stored camu-camu (*Myrciaria dubia*) pulp. *Archivos latinoamericanos de nutrición*, 50(4):405–408, 2000.

[177] J. W. Penn JR. The cultivation of camu camu (*Myrciaria dubia*): a tree planting programme in the Peruvian Amazon. *Forests, Trees and Livelihoods*, 16(1):85–101, 2006.

[178] W. Feng. Appropriation without benefit-sharing: Origin-of-resource disclosure requirements and enforcement under TRIPS and the Nagoya Protocol. *Chi. J. Int'l L.*, 18:245, 2017.

[179] H. Cardello, M. Da Silva, and M. Damasio. Measurement of the relative sweetness of stevia extract, aspartame and cyclamate/saccharin blend as compared to sucrose at different concentrations. *Plant Foods for Human Nutrition*, 54(2):119–129, 1999.

[180] S. Singh and G. Rao. Stevia: The herbal sugar of 21 st century. *Sugar tech*, 7(1):17–24, 2005.

[181] M. Carakostas, L. Curry, A. Boileau, and D. Brusick. Overview: the history, technical function and safety of rebaudioside A, a naturally occurring steviol glycoside, for use in food and beverages. *Food and Chemical Toxicology*, 46(7):S1–S10, 2008.

[182] The bitter sweet taste of stevia. Published on the internet-date accessed: 10.10.2019. URL https://www.publiceye.ch/fileadmin/doc/Biopiraterie/ 2015_PublicEye_Stevia_E_Report.pdf.

[183] Eversweet (cargill). Published on the internet- date accessed: 04.10.2019. URL https: //www.cargill.com/food-bev/na/eversweet-sweetener.

[184] L. Nordling. Rooibos tea profits will be shared with indigenous communities in landmark agreement. *Nature*, 575(7781):19, 2019.

[185] E. Norten. *Neem: India's miraculous healing plant.* Inner Traditions Bear & Co, 1999.

[186] H. S. Puri. *Neem: the divine tree Azadirachta indica.* CRC Press, 1999.

[187] J. Conrick. *Neem: the ultimate herb.* Lotus Press, 2001.

[188] C. Sheridan. EPO neem patent revocation revives biopiracy debate, 2005.

[189] 29101683994- flickr. URL https://www.flickr.com/photos/maskus/ 29101683994/.

[190] File:stevia plant.jpg- wikamedia commons. URL https://commons.wikimedia.org/ wiki/File:Stevia_plant.jpg.

[191] File:rooibos (aspalathus linearis).jpg- wikamedia commons. URL https://commons. wikimedia.org/wiki/File:Rooibos_(Aspalathus_linearis).jpg.

[192] P. Gepts. Who owns biodiversity, and how should the owners be compensated? *Plant physiology*, 134(4):1295–1307, 2004.

[193] D. J. Newman and G. M. Cragg. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *Journal of natural products*, 75(3):311–335, 2012.

[194] C. L. Cantrell, F. E. Dayan, and S. O. Duke. Natural products as sources for new pesticides. *Journal of natural products*, 75(6):1231–1242, 2012.

[195] K. Kariyawasam and M. Tsai. Access to genetic resources and benefit sharing: Implications of Nagoya Protocol on providers and users. *The Journal of World Intellectual Property*, 21(5-6):289–305, 2018.

[196] E. Commission. Commission notice — Guidance document on the scope of application and core obligations of regulation (EU) No 511/2014 of the European Parliament and of the council on the compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilisation in the Union. *Official Journal of the European Union*, 59:C 313/01, 2016.

[197] M. E. Watanabe. The Nagoya Protocol: The conundrum of defining digital sequence information. *BioScience*, 69(6):480–480, 2019.

[198] B. S. Manheim. Regulation of synthetic biology under the Nagoya Protocol. *Nature biotechnology*, 34(11):1104, 2016.

[199] Crûg Farm Plants. Published on the internet- date accessed 08.10.2019. URL http: //www.crug-farm.co.uk/.

[200] R. Mounce, P. Smith, and S. Brockington. Ex situ conservation of plant diversity in the world's botanic gardens. *Nature plants*, 3(10):795, 2017.

[201] Cambridge University Botanical Gardens. Published on the internet- date accessed 08.10.2019. URL https://www.botanic.cam.ac.uk/.

[202] Kew Royal Botanical Gardens. Published on the internet- date accessed 08.10.2019. URL https://www.kew.org/.

[203] A. Osbourn. SAW: breaking down barriers between art and science. *PLoS biology*, 6(8): e211, 2008.

[204] Science Arts Writing trust. Published on the internet- date accessed 04.10.2019. URL https://www.sawtrust.org.

[205] J. Rant and J. Tait. Building foundations for an open perspective on synthetic biology research and innovation. *New Phytologist*, 200(1):13–15, 2013.

[206] The Brilliant Club. Published on the internet- date accessed 04.10.2019. URL https://thebrilliantclub.org/.

[207] S. Laird and R. Wynberg. Study on access and benefit-sharing arrangements. In *Ad-Hoc Open-Ended Working Group on Access and Benefit Sharing, 6th Meeting, Geneva, Switzerland*, pages 21–25, 2008.

[208] J. Soares. The Nagoya Protocol and natural product-based research, 2011.

[209] J. Overmann and A. H. Scholz. Microbiological research under the Nagoya Protocol: facts and fiction. *Trends in microbiology*, 25(2):85–88, 2017.

[210] D. A. Mulholland and D. A. Taylor. Limonoids from Australian members of the Meliaceae. *Phytochemistry*, 31(12):4163–4166, 1992.

[211] E. Morgan and M. Thornton. Azadirachtin in the fruit of *Melia azedarach*. *Phytochemistry*, 12(2):391–392, 1973.

[212] F. Bohnenstengel, V. Wray, L. Witte, R. Srivastava, and P. Proksch. Insecticidal meliacarpins (C-*seco* limonoids) from *Melia azedarach*. *Phytochemistry*, 50(6):977–982, 1999.

[213] N. G. Ntalli, F. Cottiglia, C. A. Bueno, L. E. Alche, M. Leonti, S. Vargiu, E. Bifulco, U. Menkissoglu-Spiroudi, and P. Caboni. Cytotoxic tirucallane triterpenoids from *Melia azedarach* fruits. *Molecules*, 15(9):5866–5877, 2010.

[214] M. Díaz, L. Castillo, C. Elisa Diaz, A. González-Coloma, and C. Rossini. Limonoids from *Melia azedarach* with deterrent activity against insects. *The Natural Products Journal*, 2(1):36–44, 2012.

[215] Q.-G. Tan, X.-N. Li, H. Chen, T. Feng, X.-H. Cai, and X.-D. Luo. Sterols and terpenoids from *Melia azedarach*. *Journal of natural products*, 73(4):693–697, 2010.

[216] S.-B. Wu, Q.-Y. Bao, W.-X. Wang, Y. Zhao, G. Xia, Z. Zhao, H. Zeng, and J.-F. Hu. Cytotoxic triterpenoids and steroids from the bark of *Melia azedarach*. *Planta medica*, 77(09): 922–928, 2011.

[217] Y. M'Rabet, N. Rokbeni, S. Cluzet, A. Bonlila, T. Richard, S. Krisa, L. Marzouki, H. Casabianca, and K. Hosni. Profiling of phenolic compounds and antioxidant activity of *Melia azedarach* L. leaves and fruits at two stages of maturity. *Industrial Crops and Products*, 107:232–243, 2017.

[218] W. Lau and E. S. Sattely. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science*, 349(6253):1224–1228, 2015.

[219] M. Itkin, R. Davidovich-Rikanati, S. Cohen, V. Portnoy, A. Doron-Faigenboim, E. Oren, S. Freilich, G. Tzuri, N. Baranes, S. Shen, M. Petreikov, R. Sertchook, S. Ben-Dor, H. Gottlieb, A. Hernandez, D. R. Nelson, H. S. Paris, Y. Tadmor, Y. Burger, E. Lewinsohn, N. Katzir, and A. Schaffer. The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. *Proc Natl Acad Sci U S A*, 113(47): E7619–E7628, 2016.

[220] J. Han, W. Lin, R. Xu, W. Wang, and S. Zhao. Studies on the chemical constituents of *Melia azedarach* L. *Yao xue xue bao= Acta pharmaceutica Sinica*, 26(6):426–429, 1991.

[221] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org/.

[222] K. J. Livak and T. D. Schmittgen. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\delta\delta$CT method. *methods*, 25(4):402–408, 2001.

[223] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL https://ggplot2.tidyverse.org.

[224] A. Tomar, K. Singh, S. Phogat, and R. Dhillon. Neem: An introduction. In *Neem: A Treatise*, page 546. IK International Publishing House Pvt. Ltd, 2008.

[225] T. P. Michael and S. Jackson. The first 50 plant genomes. *The Plant Genome*, 6(2), 2013.

[226] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, 1(1):18, 2012.

[227] D. Ohri, A. Bhargava, and A. Chatterjee. Nuclear DNA amounts in 112 species of tropical hardwoods - new estimates. *Plant Biology*, 6(5):555–561, 2004.

[228] Neem genome portal. Published on the internet- date accessed: 29.09.2019. URL https://www.ganitlabs.in/neem-genome-portal.

[229] S. A. Kautsar, H. G. Suarez Duran, K. Blin, A. Osbourn, and M. H. Medema. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Research*, 45(W1):W55–W63, 2017.

[230] J. A. Robles, S. E. Qureshi, S. J. Stephen, S. R. Wilson, C. J. Burden, and J. M. Taylor. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC genomics*, 13(1):484, 2012.

[231] B. J. Haas, N. Volfovsky, C. D. Town, M. Troukhan, N. Alexandrov, K. A. Feldmann, R. B. Flavell, O. White, and S. L. Salzberg. Full-length messenger RNA sequences greatly improve genome annotation. *Genome biology*, 3(6):research0029–1, 2002.

[232] M. Giolai, P. Paajanen, W. Verweij, L. Percival-Alwyn, D. Baker, K. Witek, F. Jupe, G. Bryan, I. Hein, J. D. Jones, et al. Targeted capture and sequencing of gene-sized DNA molecules. *Biotechniques*, 61(6):315–322, 2016.

[233] A. Gurevich, V. Saveliev, N. Vyahhi, and G. Tesler. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075, 2013.

[234] J. F. Denton, J. Lugo-Martinez, A. E. Tucker, D. R. Schrider, W. C. Warren, and M. W. Hahn. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS computational biology*, 10(12):e1003998, 2014.

[235] M. D. Robinson and A. Oshlack. A scaling normalization method for differential expression analysis of RNA-Seq data. *Genome biology*, 11(3):R25, 2010.

[236] C. Evans, J. Hardin, and D. M. Stoebel. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in bioinformatics*, 19(5):776–792, 2017.

[237] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome biology*, 15(12):550, 2014.

[238] K. Pearson. Notes on regression and inheritance in the case of two parents proceedings of the royal society of london, 58, 240-242, 1895.

[239] D. Mapleson, L. Venturini, G. Kaithakottil, and D. Swarbreck. Efficient and accurate detection of splice junctions from RNA-Seq with Portcullis. *GigaScience*, 7(12):giy131,

2018.

[240] L. Venturini, S. Caim, G. G. Kaithakottil, D. L. Mapleson, and D. Swarbreck. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*, 7(8):giy093, 2018.

[241] E. S. Sollars, A. L. Harper, L. J. Kelly, C. M. Sambles, R. H. Ramirez-Gonzalez, D. Swarbreck, G. Kaithakottil, E. D. Cooper, C. Uauy, L. Havlickova, et al. Genome sequence and genetic diversity of European ash trees. *Nature*, 541(7636):212, 2017.

[242] B. J. Clavijo, L. Venturini, C. Schudoma, G. G. Accinelli, G. Kaithakottil, J. Wright, P. Borrill, G. Kettleborough, D. Heavens, H. Chapman, et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome research*, 27(5):885–896, 2017.

[243] F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212, 2015.

[244] J.-M. Belton, R. P. McCord, J. H. Gibcus, N. Naumova, Y. Zhan, and J. Dekker. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*, 58(3): 268–276, 2012.

[245] A. Hallab. *Protein Function Prediction Using Phylogenomics, Domain Architecture Analysis, Data Integration, and Lexical Scoring*. PhD thesis, Universitäts-und Landesbibliothek Bonn, 2015.

[246] T. Z. Berardini, L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, and E. Huala. The Arabidopsis information resource: making and mining the 'gold standard' annotated reference plant genome. *genesis*, 53(8):474–485, 2015.

[247] U. Consortium. UniProt: a hub for protein information. *Nucleic acids research*, 43(D1): D204–D212, 2014.

[248] P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, et al. Interproscan 5: genome-scale protein function classification. *Bioinformatics*, 30(9):1236–1240, 2014.

[249] M.-D. Rey, G. Moore, and A. C. Martín. Identification and comparison of individual chromosomes of three accessions of *Hordeum chilense, Hordeum vulgare*, and *Triticum aestivum* by FISH. *Genome*, 61(6):387–396, 2018.

[250] RNA-Seq analysis workshop course material (Weill Cornell Medical College). Published on the internet- date accessed: 06.11.19. URL http://chagall.med.cornell.edu/RNASEQcourse/.

[251] S. Andrews, F. Krueger, A. Segonds-Pichon, L. Biggins, C. Krueger, and S. Wingett. FastQC. Babraham Institute, 2012.

[252] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics*, 29(1): 15–21, 2013.

[253] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25 (16):2078–2079, 2009.

[254] Y. Liao, G. K. Smyth, and W. Shi. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*, 41(10):e108–e108, 2013.

[255] T. Morikawa, M. Mizutani, N. Aoki, B. Watanabe, H. Saga, S. Saito, A. Oikawa, H. Suzuki,

N. Sakurai, D. Shibata, et al. Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato. *The Plant Cell*, 18(4):1008–1022, 2006.

[256] S. Martinez and R. P. Hausinger. Catalytic mechanisms of Fe (ii)-and 2-oxoglutarate-dependent oxygenases. *Journal of Biological Chemistry*, 290(34):20702–20711, 2015.

[257] M. S. Islam, T. M. Leissing, R. Chowdhury, R. J. Hopkinson, and C. J. Schofield. 2-oxoglutarate-dependent oxygenases. *Annual review of biochemistry*, 87:585–620, 2018.

[258] B.-I. Shimizu. 2-oxoglutarate-dependent dioxygenases in the biosynthesis of simple coumarins. *Frontiers in plant science*, 5:549, 2014.

[259] R. Li, A. Stapon, J. T. Blanchfield, and C. A. Townsend. Three unusual reactions mediate carbapenem and carbapenam biosynthesis. *Journal of the American Chemical Society*, 122(38):9296–9297, 2000.

[260] W.-c. Chang, Y. Guo, C. Wang, S. E. Butch, A. C. Rosenzweig, A. K. Boal, C. Krebs, and J. M. Bollinger. Mechanism of the C5 stereoinversion reaction in the biosynthesis of carbapenem antibiotics. *Science*, 343(6175):1140–1144, 2014.

[261] T.-W. Kim, J.-Y. Hwang, Y.-S. Kim, S.-H. Joo, S. C. Chang, J. S. Lee, S. Takatsuto, and S.-K. Kim. Arabidopsis CYP85A2, a cytochrome P450, mediates the Baeyer-Villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis. *The Plant Cell*, 17(8): 2397–2412, 2005.

[262] T. Nomura, T. Kushiro, T. Yokota, Y. Kamiya, G. J. Bishop, and S. Yamaguchi. The last reaction producing brassinolide is catalyzed by cytochrome P-450s, CYP85A3 in tomato and CYP85A2 in Arabidopsis. *Journal of Biological Chemistry*, 280(18):17873–17879, 2005.

[263] M. L. Mascotti, W. J. Lapadula, and M. J. Ayub. The origin and evolution of Baeyer—Villiger monooxygenases (BVMOs): An ancestral family of flavin monooxygenases. *PloS one*, 10(7):e0132689, 2015.

[264] E. Beneventi, M. Niero, R. Motterle, M. Fraaije, and E. Bergantino. Discovery of Baeyer-Villiger monooxygenases from photosynthetic eukaryotes. *Journal of Molecular Catalysis B: Enzymatic*, 98:145–154, 2013.

[265] H. Tarhonskaya, A. Szöllössi, I. K. Leung, J. T. Bush, L. Henry, R. Chowdhury, A. Iqbal, T. D. Claridge, C. J. Schofield, and E. Flashman. Studies on deacetoxycephalosporin C synthase support a consensus mechanism for 2-oxoglutarate dependent oxygenases. *Biochemistry*, 53(15):2483–2493, 2014.

[266] J. Li, M. J. van Belkum, and J. C. Vederas. Functional characterization of recombinant hyoscyamine 6β-hydroxylase from *Atropa belladonna*. *Bioorganic & medicinal chemistry*, 20(14):4356–4363, 2012.

[267] S. K. Nair and W. A. van der Donk. Structure and mechanism of enzymes involved in biosynthesis and breakdown of the phosphonates fosfomycin, dehydrophos, and phosphinothricin. *Archives of biochemistry and biophysics*, 505(1):13–21, 2011.

[268] W.-c. Chang, M. Dey, P. Liu, S. O. Mansoorabadi, S.-J. Moon, Z. K. Zhao, C. L. Drennan, and H.-w. Liu. Mechanistic studies of an unprecedented enzyme-catalysed 1,2-phosphono-migration reaction. *Nature*, 496(7443):114, 2013.

[269] H. Moummou, Y. Kallberg, L. B. Tonfack, B. Persson, and B. Van Der Rest. The plant short-chain dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. *BMC plant biology*, 12(1):219, 2012.

[270] P. D. Sonawane, U. Heinig, S. Panda, N. S. Gilboa, M. Yona, S. P. Kumar, N. Alkan, T. Unger, S. Bocobza, M. Pliner, et al. Short-chain dehydrogenase/reductase governs

steroidal specialized metabolites structural diversity and toxicity in the genus solanum. *Proceedings of the National Academy of Sciences*, 115(23):E5419–E5428, 2018.

[271] S. Zhao, Y. Guo, Q. Sheng, and Y. Shyr. Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC bioinformatics*, 15(S10):P16, 2014.

[272] K. Geisler, R. K. Hughes, F. Sainsbury, G. P. Lomonossoff, M. Rejzek, S. Fairhurst, C.-E. Olsen, M. S. Motawia, R. E. Melton, A. M. Hemmings, et al. Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proceedings of the National Academy of Sciences*, 110(35):E3360–E3367, 2013.

[273] R. J. Grebenok, T. E. Ohnmeiss, A. Yamamoto, E. D. Huntley, D. W. Galbraith, and D. Della Penna. Isolation and characterization of an *Arabidopsis thaliana* C-8,7 sterol isomerase: functional and structural similarities to mammalian C-8, 7 sterol isomerase/emopamil-binding protein. *Plant molecular biology*, 38(5):807–815, 1998.

[274] A. Rahier, S. Pierre, G. Riveill, and F. Karst. Identification of essential amino acid residues in a sterol 8, 7-isomerase from zea mays reveals functional homology and diversity with the isomerases of animal and fungal origin. *Biochemical Journal*, 414 (2):247–259, 2008.

[275] E. Knoch, S. Sugawara, T. Mori, C. Poulsen, A. Fukushima, J. Harholt, Y. Fujimoto, N. Umemoto, and K. Saito. Third DWF1 paralog in Solanaceae, sterol δ24-isomerase, branches withanolide biosynthesis from the general phytosterol pathway. *Proceedings of the National Academy of Sciences*, 115(34):E8096–E8103, 2018.

[276] A. Bayer, X. Ma, and J. Stöckigt. Acetyltransfer in natural product biosynthesis-functional cloning and molecular analysis of vinorine synthase. *Bioorganic & medicinal chemistry*, 12(10):2787–2795, 2004.

[277] P. Caboni, N. Ntalli, C. Bueno, and L. Alche. Isolation and chemical characterization of components with biological activity extracted from *Azadirachta indica* and *Melia azedarach*. In *Emerging Trends in Dietary Components for Preventing and Combating Disease*, pages 51–77. ACS Publications, 2012.

[278] S. T. Mugford, X. Qi, S. Bakht, L. Hill, E. Wegel, R. K. Hughes, K. Papadopoulou, R. Melton, M. Philo, F. Sainsbury, et al. A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *The Plant Cell*, 21(8):2473–2484, 2009.

[279] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.

[280] G. Gopalakrishnan, N. Pradeep Singh, and V. Kasinath. Photooxygenation of nimonol, a tetranortriterpenoid from *Azadirachta indica* A. Juss. *Molecules*, 7(2):112–118, 2002.

[281] A. P. Jarvis, S. Johnson, E. D. Morgan, M. S. Simmonds, and W. M. Blaney. Photooxidation of nimbin and salannin, tetranortriterpenoids from the neem tree (*Azadirachta indica*). *Journal of chemical ecology*, 23(12):2841–2860, 1997.

[282] R. Malathi, S. Rajan, G. Gopalakrishnan, and G. Suresh. Azadirachtol, a tetranortriterpenoid from neem kernels. *Acta Crystallographica Section C: Crystal Structure Communications*, 58(12):o708–o710, 2002.

[283] Q. Liu, B. Khakimov, P. D. Cárdenas, F. Cozzi, C. E. Olsen, K. R. Jensen, T. P. Hauser, and S. Bak. The cytochrome P450 CYP72A552 is key to production of hederagenin-based saponins that mediate plant defense against herbivores. *New Phytologist*, 222(3):1599–

1609, 2019.

[284] A. C. Smigocki, S. Ivic-Haymes, H. Li, and J. Savić. Pest protection conferred by a *Beta vulgaris* serine proteinase inhibitor gene. *PloS one*, 8(2):e57303, 2013.

[285] P. A. Rodriguez, S. A. Hogenhout, and J. I. Bos. Leaf-disc assay based on transient over-expression in *Nicotiana benthamiana* to allow functional screening of candidate effectors from aphids. In *Plant-Pathogen Interactions*, pages 137–143. Springer, 2014.

[286] U. Schlüter, H. Bidmon, and S. Grewe. Azadirachtin affects growth and endocrine events in larvae of the tobacco hornworm, *Manduca sexta*. *Journal of insect physiology*, 31(10):773–777, 1985.

[287] G. F. Weirich, J. Svoboda, and M. Thompson. Ecdysone 20-monooxygenases. In *Biosynthesis, metabolism and mode of action of invertebrate hormones*, pages 227–233. Springer, 1984.

[288] J. Reed and A. Osbourn. Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. *Plant cell reports*, 37(10):1431–1441, 2018.

[289] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676, 2012.

[290] R. Appels, K. Eversole, C. Feuillet, B. Keller, J. Rogers, N. Stein, C. J. Pozniak, F. Choulet, A. Distelfeld, J. Poland, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403):eaar7191, 2018.

[291] H.-W. Nützmann, A. Huang, and A. Osbourn. Plant metabolic clusters-from genetics to genomics. *New phytologist*, 211(3):771–789, 2016.

[292] B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A. Regev. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8): 1494–1512, 2013.

[293] D. J. MacKenzie, M. A. McLean, S. Mukerji, and M. Green. Improved RNA extraction from woody plants for the detection of viral pathogens by reverse transcription-polymerase chain reaction. *Plant Disease*, 81(2):222–226, 1997.

[294] F. Sainsbury, E. C. Thuenemann, and G. P. Lomonossoff. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant biotechnology journal*, 7(7):682–693, 2009.

[295] T. Kushiro, M. Shibuya, K. Masuda, and Y. Ebizuka. Mutational studies on triterpene synthases: engineering lupeol synthase into β-amyrin synthase. *Journal of the American Chemical Society*, 122(29):6816–6824, 2000.

[296] R. D. Gietz and R. H. Schiestl. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols*, 2(1):31, 2007.

[297] F. Sievers and D. G. Higgins. Clustal omega. *Current protocols in bioinformatics*, 48(1): 3–13, 2014.

[298] A. I. Gray, P. Bhandari, and P. G. Waterman. New protolimonoids from the fruits of *Phellodendron chinense*. *Phytochemistry*, 27(6):1805–1808, 1988.

[299] D. A. Mulholland, M. Kotsos, H. A. Mahomed, and D. A. H. Taylor. Triterpenoids from *Owenia cepiodora*. *Phytochemistry*, 49(8):2457–2460, 1998.

[300] C. Wattanapiromsakul and P. G. Waterman. Flavanone, triterpene and chromene

derivatives from the stems of *Paramignya griffithii*. *Phytochemistry*, 55(3):269–273, 2000.

[301] E. R. Fo, J. a. B. Fernandes, P. C. Vieira, and M. F. D. G. F. Da Silva. Isolation of secoisolariciresinol diesters from stems of *Simaba cuneata*. *Phytochemistry*, 31(6): 2115–2116, 1992.

[302] X.-D. Luo, S.-H. Wu, Y.-B. Ma, and D.-G. Wu. Tirucallane triterpenoids from *Dysoxylum hainanense*. *Phytochemistry*, 54(8):801–805, 2000.

[303] H. Liu, J. Heilmann, T. Rali, and O. Sticher. New tirucallane-type triterpenes from *Dysoxylum variabile*. *Journal of Natural Products*, 64(2):159–163, 2001.

[304] V. Kumar, N. M. M. Niyaz, D. B. M. Wickramaratne, and S. Balasubramaniam. Tirucallane derivatives from *Paramignya monophylla* fruits. *Phytochemistry*, 30(4): 1231–1233, 1991.

[305] G. Jayakumar, M. D. Ajitha Bai, and Y. Fujimoto. Beddomeilactone: A new triterpene from *Dysoxylum Beddomei* AU - hisham, a. *Natural Product Research*, 18(4):329–334, 2004.

[306] J. Gu, S.-Y. Qian, G.-G. Cheng, Y. Li, Y.-P. Liu, and X.-D. Luo. Chemical components of *Dysoxylum densiflorum*. *Natural products and bioprospecting*, 3(2):66–69, 2013.

[307] S. N. J. Grosvenor, K. Mascoll, S. McLean, W. F. Reynolds, and W. F. Tinto. Tirucallane, apotirucallane, and octanorapotirucallane triterpenes of *Simarouba amara*. *Journal of Natural Products*, 69(9):1315–1318, 2006.

[308] K. Mohamad, M.-T. Martin, M. Litaudon, C. Gaspard, T. Sévenet, and M. Pas. Tirucallane triterpenes from *Dysoxylum macranthum*. *Phytochemistry*, 52(8):1461–1468, 1999.

[309] A. T. Orisadipe, A. A. Adesomoju, M. D'Ambrosio, A. Guerriero, and J. I. Okogun. Tirucallane triterpenes from the leaf extract of *Entandrophragma angolense*. *Phytochemistry*, 66(19):2324–2328, 2005.

[310] J. Chen, J. Chen, Y. Sun, Y. Yan, L. Kong, Y. Li, and M. Qiu. Cytotoxic triterpenoids from *Azadirachta indica*. *Planta medica*, 77(16):1844–1847, 2011.

[311] C. Y. Ragasa, O. B. Torres, L. O. Bernardo, E. H. Mandia, M.-J. Don, and C.-C. Shen. Glabretal-type triterpenoids from *Dysoxylum mollissimum*. *Phytochemistry Letters*, 6 (4):514–518, 2013.

[312] A. Inada, M. Konishi, H. Murata, and T. Nakanishi. Structures of a new limonoid and a new triterpenoid derivative from pericarps of *Trichilia connaroides*. *Journal of natural products*, 57(10):1446–1449, 1994.

[313] J. I. Vieira, O. De Aquino Azevedo, J. J. De Souza, R. Braz-Filho, M. Dos Santos Gonçalves, and F. M. De Araújo. Hirtinone, a novel cycloartane-type triterpene and other compounds from Trichilia hirta L. (Meliaceae). *Molecules*, 18(3), 2013.

[314] V. F. Rodrigues, H. M. Carmo, R. F. Braz, L. Mathias, and I. Vieira. Two new terpenoids from *Trichilia quadrijuga* (Meliaceae). *Natural product communications*, 5(2):179–184, 2010.

[315] W. W. Harding, H. Jacobs, P. A. Lewis, S. McLean, and W. F. Reynolds. Cycloartanes, protolimonoids, a pregnane and a new ergostane from *Trichilia reticulata*. *Natural product letters*, 15(4):253–260, 2001.

[316] P. Ketwaru, J. Klass, W. F. Tinto, S. McLean, and W. F. Reynolds. Pregnane steroids from *Trichilia schomburgkii*. *Journal of natural products*, 56(3):430–431, 1993.

[317] W. F. Tinto, P. K. Jagessar, P. Ketwaru, W. F. Reynolds, and S. McLean. Constituents of

Trichilia schomburgkii. Journal of Natural Products, 54(4):972–977, 1991.

[318] G.-C. Wang, J.-H. Yu, Y. Shen, Y. Leng, H. Zhang, and J.-M. Yue. Limonoids and triterpenoids as 11β-HSD1 inhibitors from *Walsura robusta*. *Journal of Natural Products*, 79(4):899–906, 2016.

[319] J.-Q. Liu, C.-F. Wang, Y. Li, J.-C. Chen, L. Zhou, and M.-H. Qiu. Limonoids from the leaves of *Toona ciliata* var. yunnanensis. *Phytochemistry*, 76:141–149, 2012.

[320] K. Kishi, K. Yoshikawa, and S. Arihara. Limonoids and protolimonoids from the fruits of *Phellodendron amurense. Phytochemistry*, 31(4):1335–1338, 1992.

[321] H. Itokawa, E. Kishi, H. Morita, and K. Takeya. Cytotoxic quassinoids and tirucallane-type triterpenes from the woods of *Eurycoma longifolia. Chemical & Pharmaceutical Bulletin*, 40(4):1053–1055, 1992.

[322] R. d. C. G. Saraiva, A. C. Pinto, S. M. Nunomura, and A. M. Pohlit. Triterpenes and a canthinone alkaloid from the stems of *Simaba polyphylla* (Cavalcante) WW Thomas (Simaroubaceae). *Química Nova*, 29(2):264–268, 2006.

[323] C. O. Esimone, C. O. Esimone, G. Eck, T. N. Duong, K. Überla, P. Proksch, and T. Grunwald. Potential anti-respiratory syncytial virus lead compounds from Aglaia species. *Die Pharmazie - An International Journal of Pharmaceutical Sciences*, 63(10):768–773, 2008.

[324] A. Benosman, P. Richomme, T. Sevenet, G. Perromat, A. H. A. Hadi, and J. Bruneton. Tirucallane triterpenes from the stem bark of *Aglaia leucophylla. Phytochemistry*, 40(5):1485–1487, 1995.

[325] B. N. Irungu, N. Adipo, J. A. Orwa, F. Kimani, M. Heydenreich, J. O. Midiwo, P. Martin Björemark, M. Håakansson, A. Yenesew, and M. Erdélyi. Antiplasmodial and cytotoxic activities of the constituents of *Turraea robusta* and *Turraea nilotica. Journal of Ethnopharmacology*, 174:419–425, 2015.

[326] J.-R. Wang, H.-L. Liu, T. Kurtán, A. Mándi, S. Antus, J. Li, H.-Y. Zhang, and Y.-W. Guo. Protolimonoids and norlimonoids from the stem bark of *Toona ciliata* var. pubescens. *Organic & Biomolecular Chemistry*, 9(22):7685–7696, 2011.

[327] M. Ahsan, J. A. Armstrong, A. I. Gray, and P. G. Waterman. Boronialatenolide: a novel pentanortriterpene from the aerial parts of *Boronia alata* (Rutaceae). *Australian Journal of Chemistry*, 47(9):1783–1787, 1994.

[328] M. Ahsan, J. A. Armstrong, A. I. Gray, and P. G. Waterman. Terpenoids, alkaloids and coumarins from *Boronia inornata* and *Boronia gracilipes. Phytochemistry*, 38(5):1275–1278, 1995.

[329] A. D. Reegan, M. R. Gandhi, M. G. Paulraj, K. Balakrishna, and S. Ignacimuthu. Effect of niloticin, a protolimonoid isolated from *Limonia acidissima* L. (Rutaceae) on the immature stages of dengue vector *Aedes aegypti* L. (Diptera: Culicidae). *Acta Tropica*, 139:67–76, 2014.

[330] T. P. Lien, C. Kamperdick, J. Schmidt, G. Adam, and T. Van Sung. Apotirucallane triterpenoids from *Luvunga sarmentosa* (Rutaceae). *Phytochemistry*, 60(7):747–754, 2002.

[331] J. Kiplimo, S. Islam, and N. Koorbanally. Ring A, D-*seco* limonoids and flavonoid from the Kenyan *Vepris uguenensis* Engl. and their antioxidant activity. *Planta Medica*, 78(11):PI111, 2012.

[332] Z.-L. Hong, J. Xiong, S.-B. Wu, J.-J. Zhu, J.-L. Hong, Y. Zhao, G. Xia, and J.-F. Hu. Tetracyclic triterpenoids and terpenylated coumarins from the bark of *Ailanthus*

altissima ('Tree of Heaven'). *Phytochemistry*, 86:159–167, 2013.

[333] P. A. Grieco, J. Haddad, M. M. Piñeiro Núñez, and J. C. Huffman. Quassinoids from the twigs and thorns of *Castela polyandra*. *Phytochemistry*, 50(4):637–645, 1999.

[334] J. Wang, Y. Zhang, J. Luo, and L. Kong. Complete 1H and 13C NMR data assignment of protolimonoids from the stem barks of *Aphanamixis grandifolia*. *Magnetic Resonance in Chemistry*, 49(7):450–457, 2011.

[335] J. R. Wang, Q. Shen, L. Fang, S. Y. Peng, Y. M. Yang, J. Li, H. L. Liu, and Y. W. Guo. Structural and stereochemical studies of five new pregnane steroids from the stem bark of *Toona ciliata* var. pubescens. *Steroids*, 76(6):571–576, 2011.

[336] H. L. Huang, C. M. Wang, Z. H. Wang, M. J. Yao, G. T. Han, J. C. Yuan, K. Gao, and C. S. Yuan. Tirucallane-type triterpenoids from *Dysoxylum lenticellatum*. *Journal of Natural Products*, 74(10):2235–2242, 2011.

[337] W. Hayasida, L. Oliveira, A. Ferreira, and M. Lima. Ergostane steroids, tirucallane and apotirucallane triterpenes from *Guarea convergens*. *Chemistry of natural compounds*, 53(2):312–317, 2017.

[338] A. Jimenez, C. Villarreal, R. A. Toscano, M. Cook, J. T. Arnason, R. Bye, and R. Mata. Limonoids from *Swietenia humilis* and *Guarea grandiflora* (Meliaceae) taken in part from the PhD and MS theses of C. Villarreal and M. A. Jiménez, respectively. *Phytochemistry*, 49(7):1981–1988, 1998.

[339] C. H. Miguita, C. D. S. Barbosa, L. Hamerski, U. C. Sarmento, J. N. d. Nascimento, W. S. Garcez, and F. R. Garcez. 3β-o-tigloylmelianol from *Guarea kunthiana*: A new potential agent to control *Rhipicephalus (Boophilus) microplus*, a cattle tick of veterinary significance. *Molecules*, 20(1):111, 2015.

[340] P. H. Coombes, D. A. Mulholland, and M. Randrianarivelojosia. Mexicanolide limonoids from the Madagascan Meliaceae *Quivisia papinae*. *Phytochemistry*, 66(10):1100–1107, 2005.

[341] R. Kaur and S. Arora. Chemical constituents and biological activities of *Chukrasia tabularis* A. Juss.-a review. *Journal of Medicinal Plants Research*, 3(4):196–216, 2009.

[342] J. Polonsky, Z. Varon, R. M. Rabanal, and H. Jacquemin. 21, 20-anhydromelianone and melianone from *Simarouba amara* (Simaroubaceae); Carbon13 NMR spectral analysis of δ7tirucallol-type triterpenes. *Israel Journal of Chemistry*, 16(1):16–19, 1977.

[343] R. Su, M. Kim, H. Kawaguchi, T. Yamamoto, K. Goto, T. Taga, Y. Miwa, M. Kozuka, and S. Takahashi. Triterpenoids from the fruits of *Phellodendron chinense*: The stereostructure of niloticin. *Chemical and Pharmaceutical Bulletin*, 38(6):1616–1619, 1990.

[344] M. W. Biavatti, P. C. Vieira, M. F. G. F. da Silva, J. B. Fernandes, S. Albuquerque, C. M. I. Magalhães, and F. C. Pagnocca. Chemistry and bioactivity of *Raulinoa echinata* Cowan, an endemic Brazilian Rutaceae species. *Phytomedicine*, 8(2):121–124, 2001.

[345] S.-P. Yang, G. Ni, Y.-C. Gu, and J.-M. Yue. Triterpenoids from *Aglaia odorata* var. microphyllina AU - liu, jia. *Journal of Asian Natural Products Research*, 14(10):929–939, 2012.

A

Heterologous expression experiments

A.1   Expression and characterisation of CsOSC2
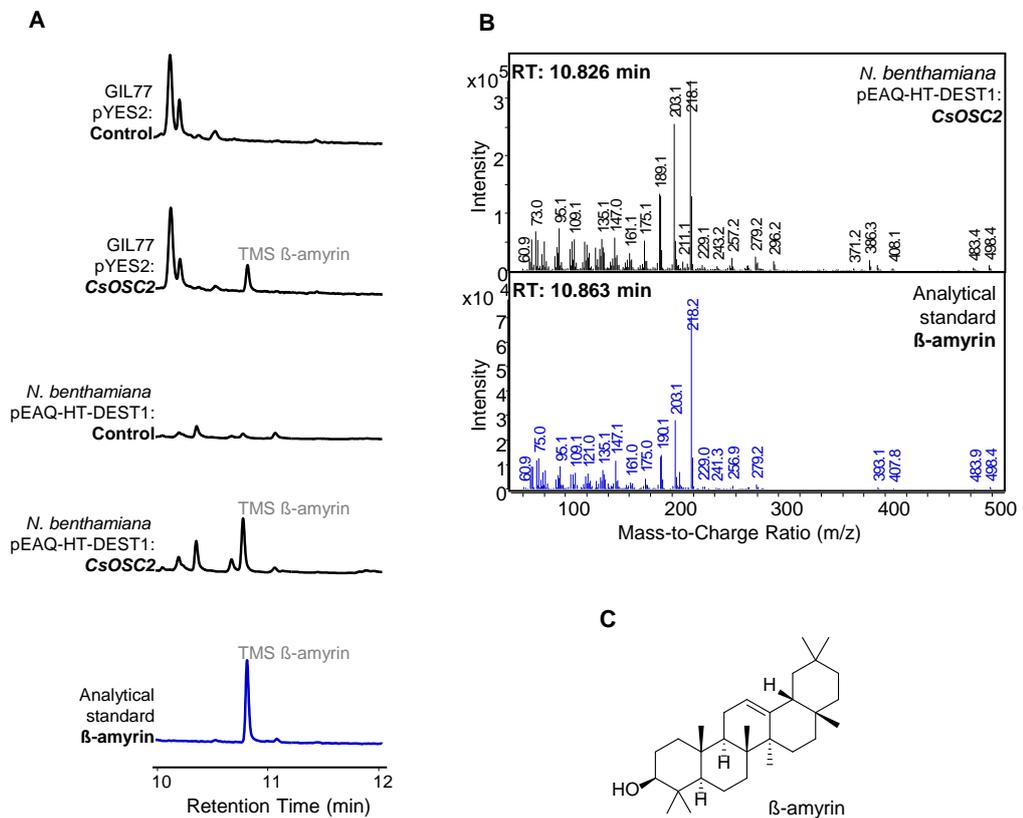
Figure A.1.1: **Characterisation of CsOSC2 as a β-amyrin synthase.**
(A) GC-MS generated TICs of derivatised hexane extracts from yeast strains and
ethyl acetate extracts from agroinfiltrated *N. benthamiana* leaves expressing CsOSC2
in pYES2 and pEAQ-HT-DEST1 vectors, respectively. Traces for the empty vector
controls and for the derivatised analytical β-amyrin standard (Sigma-Aldrich) are
included. (B) GC-MS mass spectra of TMS β-amyrin from agroinfiltrated *N.
benthamiana* leaves (black) and the β-amyrin analytical standard (blue). (C)
Structure of β-amyrin. Figure has been reproduced from the author's previous
publication (122).

A.2   Expression and characterisation of CsOSC3
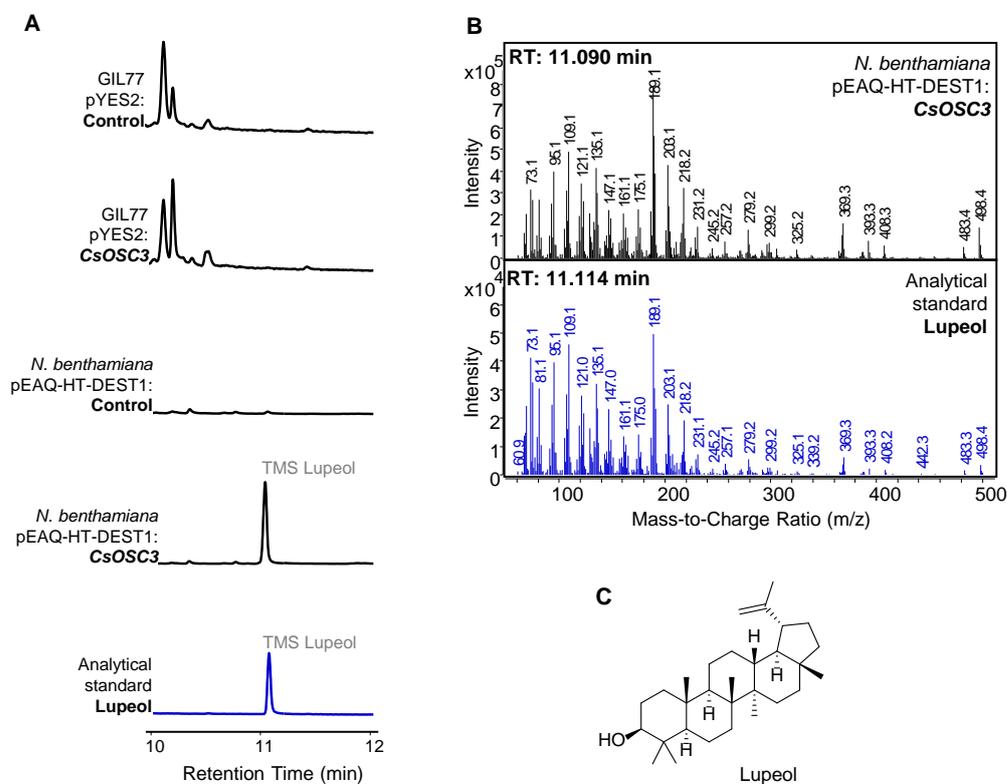
Figure A.2.1: **Characterisation of CsOSC3 as a lupeol synthase.**
(A) GC-MS generated TICs of derivatised hexane extracts from yeast strains and ethyl acetate extracts from agroinfiltrated *N. benthamiana* leaves expressing CsOSC3 in pYES2 and pEAQ-HT-DEST1 vectors, respectively. Traces for the empty vector controls and the derivatised analytical lupeol standard (Sigma-Aldrich) are included. (B) GC-MS mass spectra of TMS lupeol from agroinfiltrated *N. benthamiana* leaves (black) and analytical standard (blue). (C) Structure of lupeol. Figure has been reproduced from the author's previous publication (122).

A.3 Expression of candidate CYPs from *M.azedarach* (Ma1)
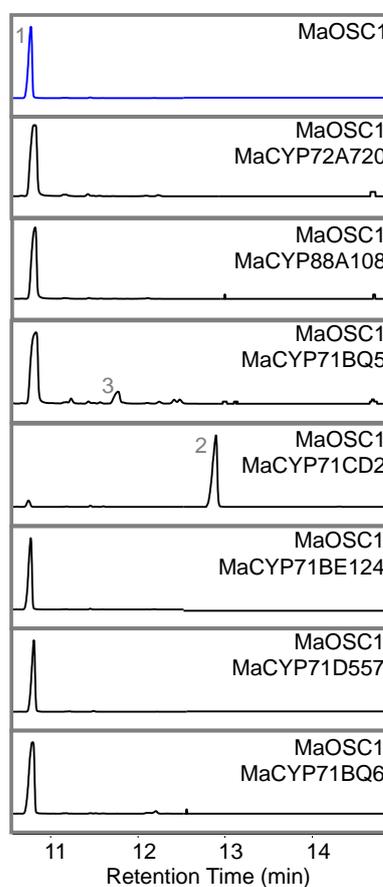
Figure A.3.1: **Transient expression of candidate CYPs from Ma1 in *N. benthamiana.*** GC-MS generated EICs of derivatised hexane extracts from agroinfiltrated *N. benthamiana* leaves expressing candidate *M. azedarach* (Ma1) CYPs in pEAQ-HT-DEST1 vectors. CYP candidates *MaCYP71BE123* and *MaCYP71BE125* were not amplifiable (Appendix B.4) and therefore are not included in this analysis. EICs represent combined traces for the following ions: [tirucalla-7,24-dien-3β-ol + TMS - H] = 498.6, [tirucalla-7,24-dien-3β-ol + O + TMS - H] = 586.4 and [tirucalla-7,24-dien-3β-ol + 2O + TMS - H] = 602.6. Traces for the control *MaOSC1* expression alone are indicated (blue).

A.4 GC-MS mass spectra of novel peaks produced by MaCYP71CD2 and MaCYP71BQ5
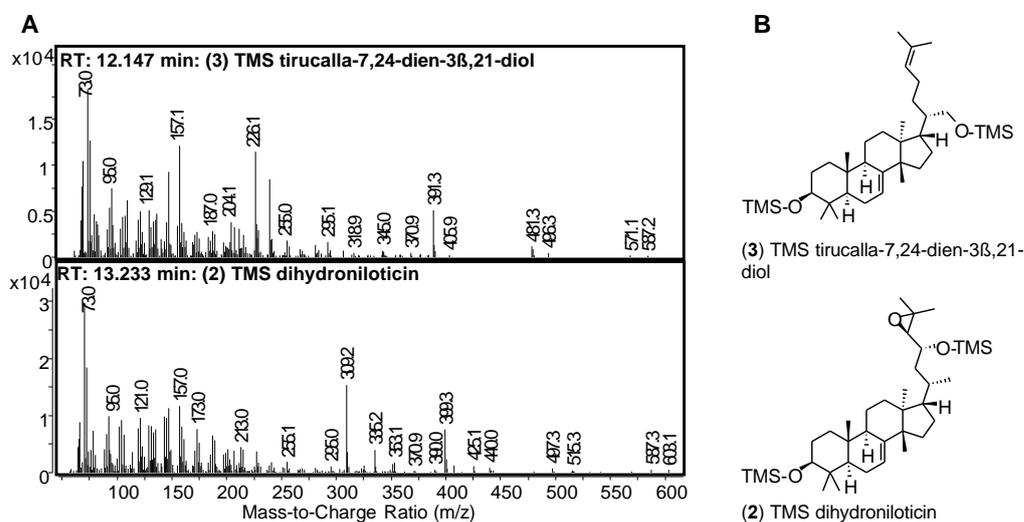
Figure A.4.1: **Mass spectra of tirucalla-7,24-dien-3β,21-diol (3) and dihydroniloticin (2).**
(A) Mass spectra corresponding to GC-MS total ion chromatograms, depicted in Figure 2.2.5, of derivatised triterpene extracts from agroinfiltrated *N. benthamiana* leaves expressing candidate genes in pEAQ-HT-DEST1. (B) Structures of TMS-dihydroniloticin **(2)** and TMS-tirucalla-7,24-dien-3β,21-diol **(3)**. Figure has been reproduced from the author's previous publication (122).

A.5 Mass spectra of novel peaks produced by MaCYP71CD2 and MaCYP71BQ5
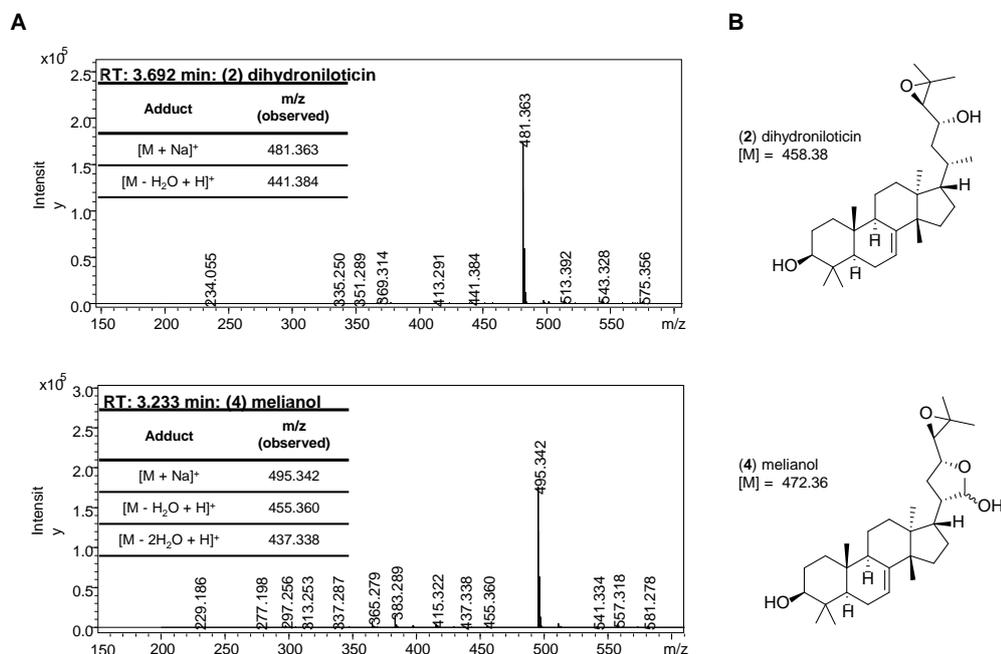
Figure A.5.1: **Mass spectra of dihydroniloticin (2) and melianol (4).**
(A) Mass spectra corresponding to UHPLC-IT-TOF EICs, depicted in Figure 2.2.5, of triterpene extracts from agroinfiltrated *N. benthamiana* leaves expressing candidate genes in pEAQ-HT-DEST1. Observed adducts are listed. (B) Structures and exact mass of dihydroniloticin **(2)** and melianol **(4)**. UHPLC-IT-TOF performed using 'protolimonoid' method. Figure has been reproduced from the author's previous publication (122).

A.6 GC-MS mass spectra of peak (*) co-eluting with tirucalla-7,24-dien-3ß-ol
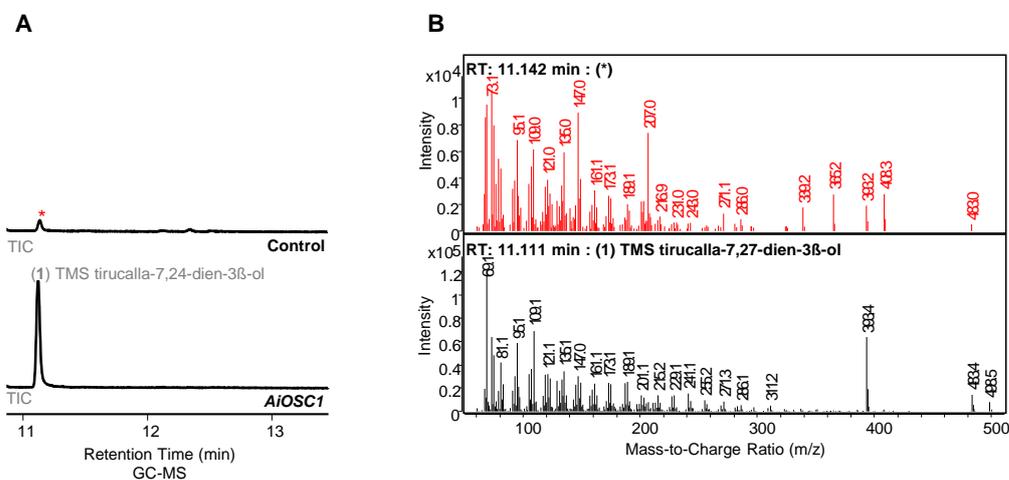
Figure A.6.1: **Mass spectra of tirucalla-7,24-dien-3β-ol and co-eluting peak.**
(A) GC-MS TICs of extracts from *N. benthamiana* expressing *AiOSC1* and a negative control (expression of *AsTHMGR* alone). Tirucalla-7,24-dien-3ß-ol **(1)** (black) and co-eluting peak (red asterisk) are indicated. (B) Mass spectra generated by GC-MS of tirucalla-7,24-dien-3β-ol **(1)** and co-eluting peak (red). Referring to Figure 2.2.5. Figure has been reproduced from the author's previous publication (122).

A.7 LC-MS mass spectra of novel peaks produced by MaCYP88A108
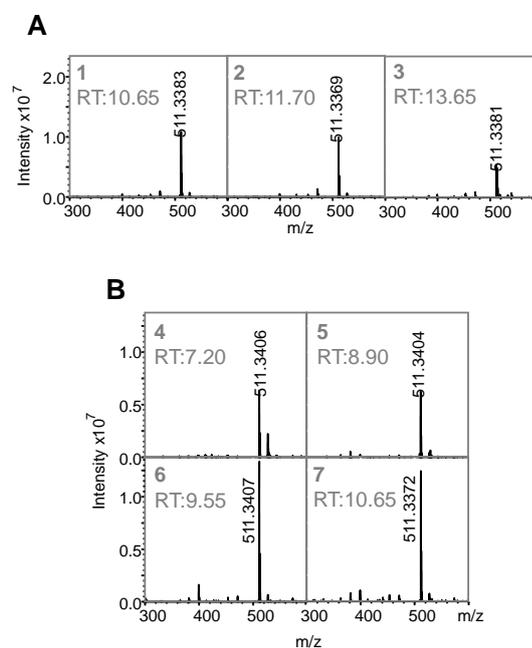
Figure A.7.1: **Mass spectra of oxidised melianol peaks (1-7).**
Mass spectra corresponding to UHPLC-IT-TOF analysis, depicted in Figure 6.2.2, of triterpene extracts from agroinfiltrated *N. benthamiana* leaves expressing candidate genes in pEAQ-HT-DEST1. (A) Mass spectra of peaks **(1-3)** present when melianol biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*) alone are expressed in *N. benthamiana* and thought to be the result of endogenous *N. benthamiana* modification. (B) Mass spectra of peaks **(4-7)** present when melianol biosynthetic genes are expressed with *MaCYP88A108*. UHPLC-IT-TOF performed using 'limonoid' method.

A.8   Expression of candidate genes from *M. azedarach* (EIv1) (batch 1 of 2)
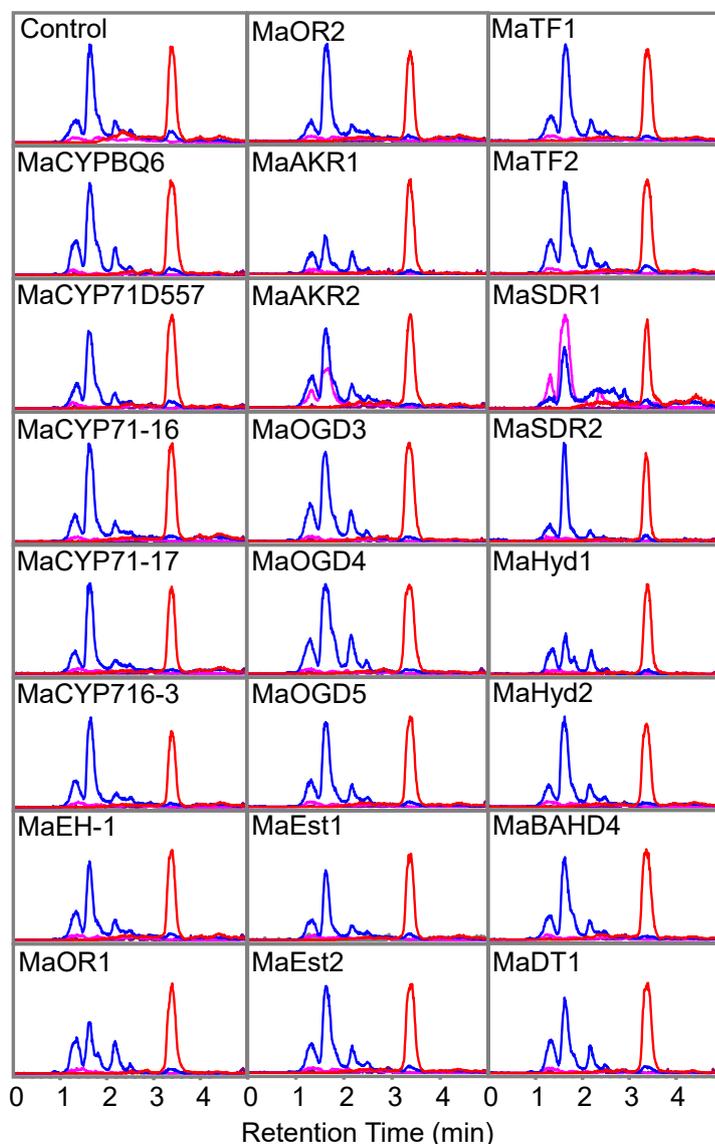
Figure A.8.1: **Transient expression of candidate genes from EIv1 in *N. benthamiana*.**
UHPLC-IT-TOF generated EICs of methanol extracts from agroinfiltrated *N. benthamiana* leaves expressing candidate genes (Table 6.2.1) from *M. azedarach* (EIv1). *MaAKR3* and *MaOGD2* were unamplifiable and therefore are not included in this analysis. Candidate genes were expressed in combination with *AiOSC1*, *MaCYP71CD2*, *MaCYP71BQ5* and *MaCYP88A108*. Expression of these genes alone is labelled (control). EICs displayed are for the following adducts: [melianol+Na]$^+$ 495.3461 (red), [melianol+O+Na]$^+$=511.3412 (blue), [melianol+O-2H+Na]$^+$=509.3240 (pink) and [melianol+O+CH$_2$CO+Na]$^+$=553.3245 (purple). UHPLC-IT-TOF analysis was performed using 'protolimonoid' method.

A.9 Expression of candidate genes from *M. azedarach* (EIv1) (batch 2 of 2)
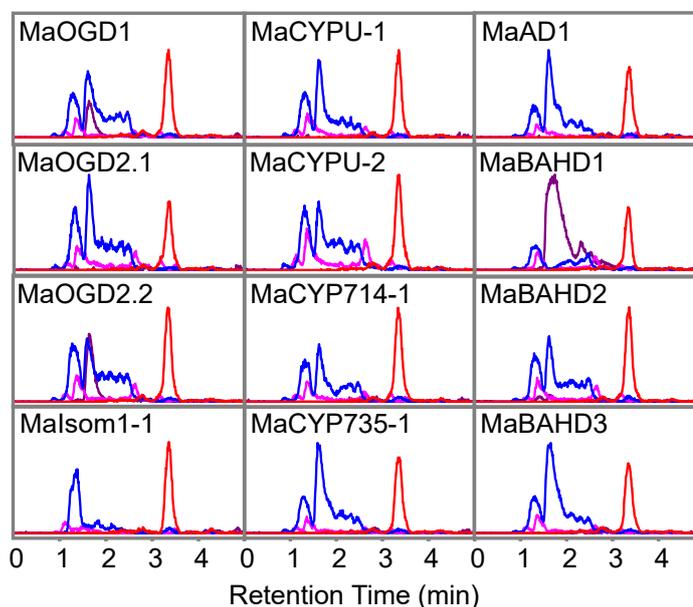
Figure A.9.1: **Transient expression of candidate genes from EIv1 in *N. benthamiana*.** UHPLC-IT-TOF generated EICs of methanol extracts from agroinfiltrated *N. benthamiana* leaves expressing candidate genes (Table 6.2.1) from *M. azedarach* (EIv1). *MaAKR3* and *MaOGD2* were unamplifiable and therefore are not included in this analysis. Candidate genes were expressed in combination with *AiOSC1, MaCYP71CD2, MaCYP71BQ5* and *MaCYP88A108*. EICs displayed are for the following adducts: $[\text{melianol+Na}]^+$=495.3489 (red), $[\text{melianol+O+Na}]^+$=511.3411 (blue), $[\text{melianol+O-2H+Na}]^+$=509.3240 (pink) and $[\text{melianol+O+CH}_2\text{CO+Na}]^+$=553.3531 (purple). UHPLC-IT-TOF analysis was performed using 'protolimonoid' method.

A.10 Activity of MaIsom1-I on melianol
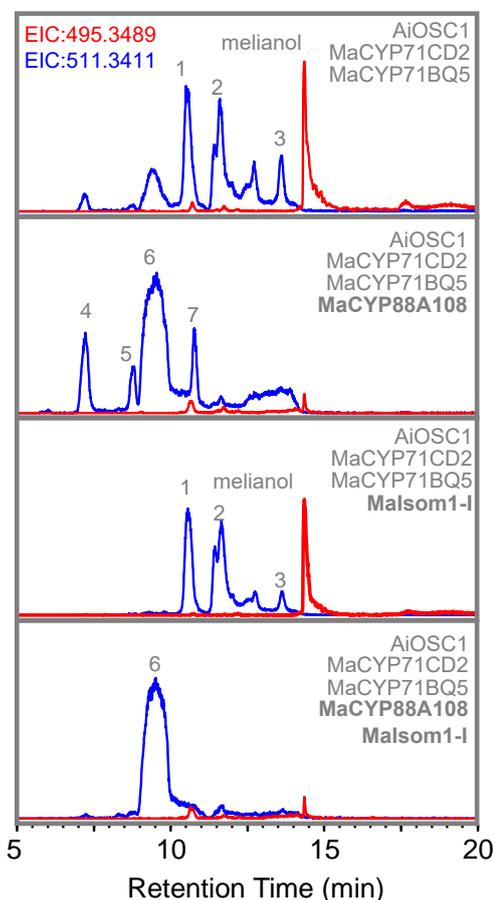
Figure A.10.1: **Combinatorial transient expression of *MaIsom1-I* in *N. benthamiana.***
UHPLC-IT-TOF generated EICs of methanol extracts from agroinfiltrated *N. benthamiana* leaves expressing *MaIsom1-I* in combination with melianol biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*), with and without *MaCYP88A108*. EICs displayed are for the adducts [melianol+Na]$^+$=495.3489 (red) and [melianol+O+Na]$^+$=511.3411 (blue). UHPLC-IT-TOF analysis performed using 'limonoid' method.

A.11 Activity of MaSDR1 and MaBAHD1 on melianol
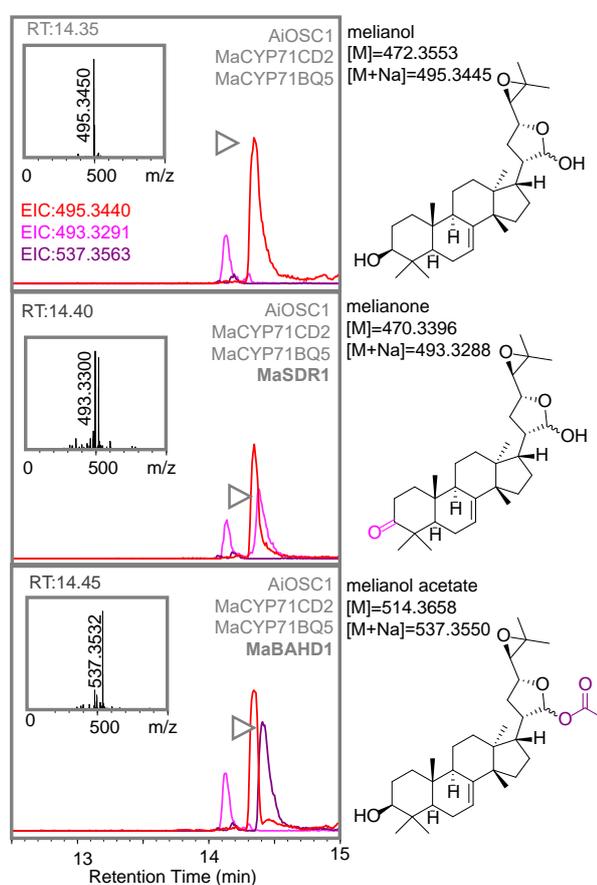
Figure A.11.1: **Combinatorial transient expression of *MaSDR1* and *MaBAHD1* in *N. benthamiana.***
UHPLC-IT-TOF generated EICs of methanol extracts from agroinfiltrated *N. benthamiana* leaves expressing *MaSDR1* and *MaBAHD1* in combination with melianol biosynthetic genes (*AiOSC1, MaCYP71CD2* and *MaCYP71BQ5*). EICs displayed are for the following adducts: [melianol+Na]$^+$=495.3440 (red), [melianol-2H+Na]$^+$=493.3291 (pink) and [melianol+CH$_2$CO+Na]$^+$=537.3563 (purple). Mass spectra of new peaks (highlighted with grey arrow) are given. UHPLC-IT-TOF analysis performed using 'limonoid' method. Predicted structures for labelled peaks are also provided (with exact mass and sodium adduct), however NMR confirmation is required.

A.12 Co-expression of all seven limonoid biosynthetic genes
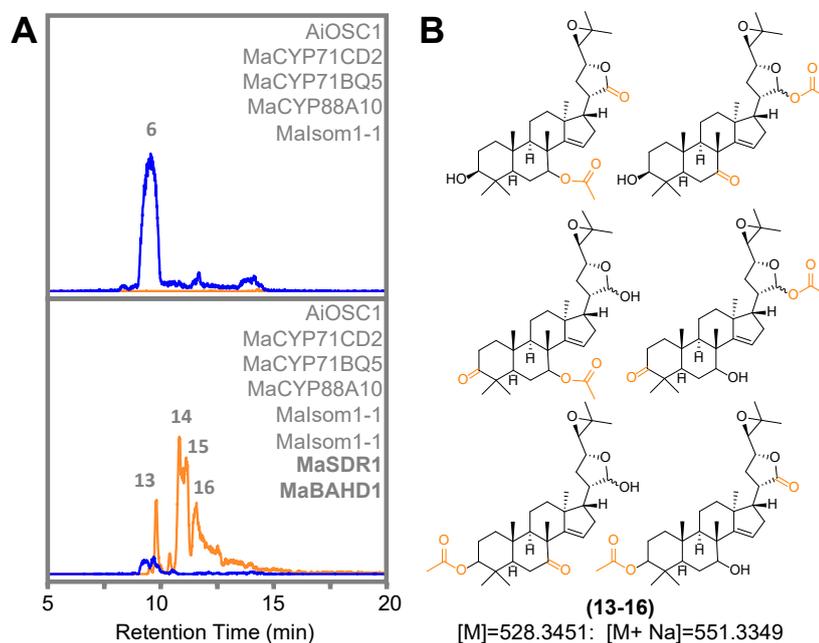
Figure A.12.1: **Transient expression of melianol B biosynthetic genes in combination with *MaBAHD1* and *MaSDR1* in *N. benthamiana*.**
(A) UHPLC-IT-TOF generated EICs of methanol extracts from agroinfiltrated *N. benthamiana* leaves expressing melianol B biosynthetic genes (*AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108* and *MaIsom1-I*) with and without both tailoring genes (*MaSDR1* and *MaBAHD1*). EICs displayed are for the adducts $[\text{melianol}+\text{O}+\text{Na}]^+ = 511.3369$ (blue) and $[\text{melianol}+\text{O}-2\text{H}+\text{CH}_2\text{CO}]^+$ 551.3358 (orange). Newly identified peaks are labelled **(12-16)** and mass spectra are available (Appendix A.13). (B) Postulated structures which new peaks **(12-16)** could represent based on mass spectra. Structural confirmation is required.

A.13 LC-MS mass spectra of novel peaks produced by MaIsom1-I, MaBAHD1 and MaSDR1
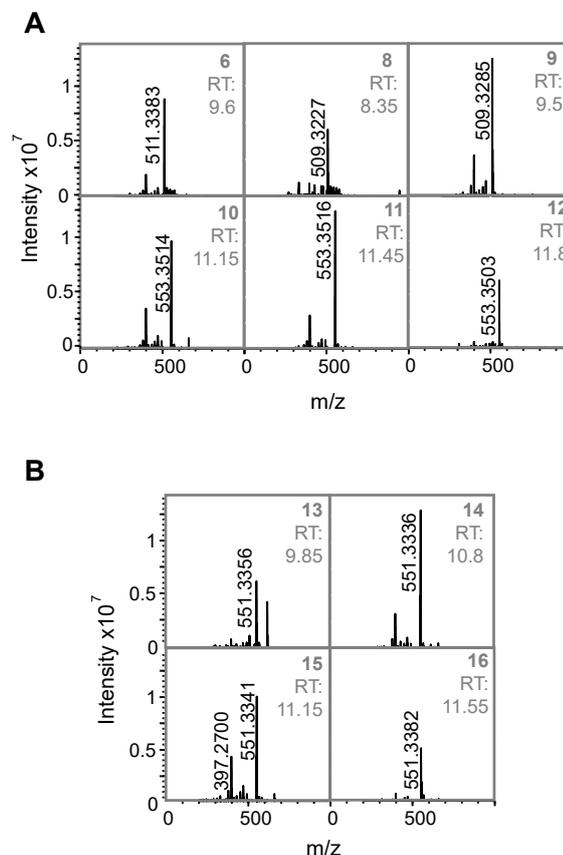
Figure A.13.1: **Mass spectra of newly identified peaks (6-16).**
Mass spectra, corresponding to UHPLC-IT-TOF analysis of *N. benthamiana* leaves expressing candidate genes in pEAQ-HT-DEST1. (A) Mass spectra for peaks: **(6)** produced by MaIsom1-I (Figure 6.2.3), **(8)** produced endogenously in *N. benthamiana*, **(9)** produced by MaSDR1 and **(10-12)** produced by MaBAHD1 (Figure 6.2.4). (B) Mass spectra for peaks **(13-16)** produced by co-expression of *AiOSC1, MaCYP71CD2, MaCYP71BQ5, MaCYP88A108, MaIsom1-I, MaSDR1* and *MaBAHD1* (Appendix A.12). UHPLC-IT-TOF analysis was performed using 'limonoid' method.

B

Miscellaneous

B.1 Details of RNA-Seq data used for transcriptome assembly

Table B.1.1: **SRA identifiers of RNA-Seq data used to assemble transcriptomes (Table 2.2.1).**

Species	ID	BioSample	SRA	Tissue	Reference
A. indica	Ai1	SAMN01084665	SRX1074135	Fruit	(126, 140)
		SAMN01084665	SRX157673	Root	
		SAMN01084665	SRX157674	Leaf	
		SAMN01084665	SRX157675	Stem	
		SAMN01084665	SRX157676	Flower	
A. indica	Ai2	SAMN03941262	SRX1122777	Leaf	(127)
				Flower	
				Fruit	
M. azedarach	Ma1	SAMN04510263	SRX1597253	Leaf	(128)
		SAMN04510265	SRX1597254	Leaf	
		SAMN04510267	SRX1597256	Leaf	

Table adapted from author's previous publication (122).

B.2   Homologs of melianol biosynthetic genes
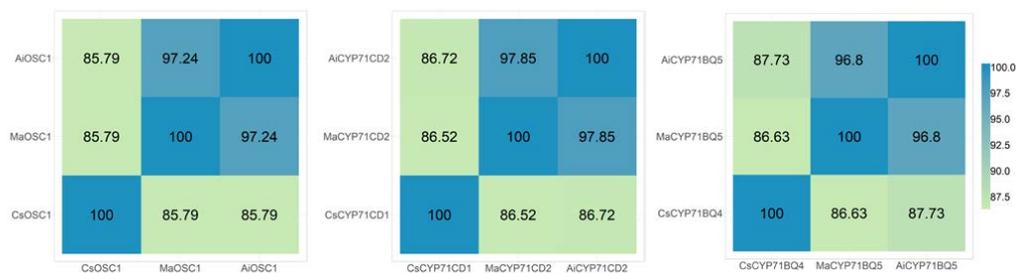
Figure B.2.1: **Protein sequence identity of melianol biosynthetic genes.**
Percentage protein identities calculated by Clustal Omega (297) for MaOSC1, MaCYP71CD2 and MaCYP71BQ5 and their orthologs from *A. indica* and closest homologs from *C. sinensis*. Figure has been reproduced from the author's previous publication (122).

B.3   Functionally characterised OSCs

Table B.3.1: **Summary of functionally characterised OSCs.**

Gene	Species	ID	Product
AiOSC1 (MK803262)	*A. indica*	Ai2:c26798_g1_i1	Tirucalla-7,24-dien-3β-ol
MaOSC1 (MK803261)	*M. azedarach*	Ma1:c5546_g1_i1	Tirucalla-7,24-dien-3β-ol
CsOSC1	*C. sinensis*	Cs:XP_006468116.1	Tirucalla-7,24-dien-3β-ol
CsOSC2	*C. sinensis*	Cs:XP_024957905.1	β-Amyrin
CsOSC3	*C. sinensis*	Cs:XP_015382484.1	Lupeol
AtLUP5	*A. thaliana*	TAIR:AT1G66960.1	Tirucalla-7,24-dien-3β-ol

Assigned gene name is provided along with GenBank accession numbers for *M. azedarach* and *A. indica* sequences. ID represents datasets and relevant identifier from transcriptomes (Ai1, Ai2 and Ma1 (Table 2.2.1)) and *C. sinensis* protein annotation (GCF_000317415.1 (Cs)). Evidence of product identification is provided in Figure 2.2.2, Appendix A.1 and Appendix A.2. Table has been reproduced from author's previous publication (122).

B.4 Functionally characterised CYPs from *M. azedarach* (Ma1)

Table B.4.1: **Summary of candidate CYPs from *M. azedarach* (Ma1) and their homologs.**

M. azedarach (Ma1)	*A. indica* (Ai1, Ai2)	*C. sinensis* (Cs)	Evidence	Cloning
MaCYP72A720 (MK803263) Ma1:c42086_g1_i1	**AiCYP72A721**** (MK803273) Ai1:c10232_g3_i1** Ai2:c60329_g1_i1** Ai2:c45683_g1_i1**	NA	DGE	2 SNPs
MaCYP88A108 (MK803265) Ma1:c5845_g1_i1 Ma1:c4476_g1_i1	**AiCYP88A108**** (MK803277) Ai1:c72218_g1_i1** Ai2:c44421_g1_i1**	**CsCYP88A51** (1 indel) XP006485427.1	DGE	0 SNPs
MaCYP71BQ5 (MK803264) Ma1:c121026_g1_i1 Ma1:c9658_g1_i1	**AiCYP71BQ5*** (MK803272) Ai1:c1285_g1_i1* Ai2:c23690_g1_i1* Ai2:c42255_g1_i1*	**CsCYP71BQ4** XP006469495.1	DGE CYP71	0 SNPs
MaCYP71CD2 (MK803271) Ma1:c87349_g1_i1 Ma1:c123757_g1_i1	**AiCYP71CD2** (MK803275) Ai1:c49952_g1_i1 Ai1:c62263_g1_i1 Ai2:c16061_g1_i1**	**CsCYP71CD1** XP006467299.1	CYP71	0 SNPs
MaCYP71BE124 (MK803267) Ma1:c17683_g1_i1	**AiCYP71BE127** (MK803276) Ai1:c62481_g1_i1	**CsCYP71BE38** XP006475119.1	CYP71	1 SNPs
MaCYP71D557 (MK803270) Ma1:c54548_g1_i1	NA	**CsCYP71D416** XP015382368.2	CYP71	0 SNPs
MaCYP71BQ6 (MK803269) Ma1:c54333_g1_i1 Ma1:c127686_g1_i1	NA	NA	CYP71	0 SNPs
MaCYP71BE123 (MK803266) Ma1:c1469_g1_i1	NA	NA	CYP71	NA
MaCYP71BE125 (MK803268) Ma1:c40628_g1_i1	**AiCYP71BE126*** (MK803274) Ai1:c14914_g1_i1* Ai2:c39917_g1_i1* Ai2:c59390_g1_i1*	NA	CYP71	NA

Table B.4.1: (continued)

Candidate CYPs from *M. azedarach* transcriptome (Ma1) are listed with their orthologs or closet homologs from *A. indica* transcriptomes (Ai1, Ai2) and *C. sinensis* protein annotation (GCF_000317415.1 (Cs) (145)). For each candidate, the CYP nomenclature is listed (bold) along with identifiers in datasets and GenBank accession numbers for *M. azedarach* and *A. indica* sequences. Instances where were not full-length candidates (*), or were under 300 amino acids and therefore considered a fragment (**) are indicated. Candidate selection was based on differential expression analysis performed on an Ai1 (DGE) (126, 140) or occurrence in the unique CYP71 subclade (CYP71). Candidates from *M. azedarach* (Ma1) were cloned and number of single nucleotide polymorphisms (SNPs) are given. Table has been reproduced from the authors previous publication (122).

B.5   NMR of AiOSC1 product

Table B.5.1: ^{13}C and ^{1}H δ assignments for tirucalla-7,24-dien-3β-ol.

Carbon numbering scheme	Selected COSY and HMBC
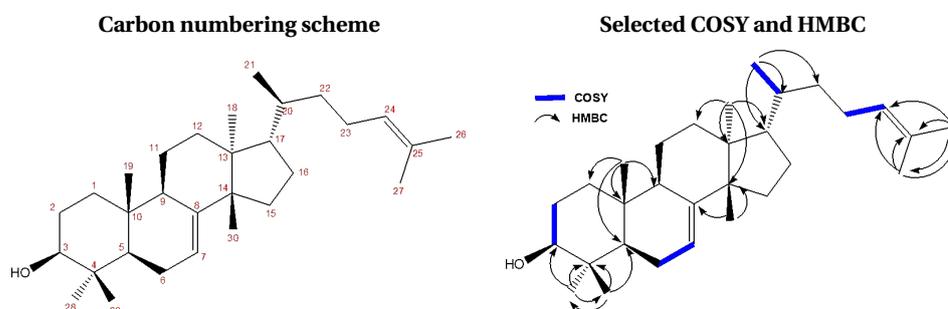
Carbon	^{13}C δ	^{1}H δ	Carbon	^{13}C δ	^{1}H δ
8	145.90	/	15	34.03	1.45 (2H, m)
25	130.94	/	12	33.79	1.79 (1H, m) 1.61 (1H, m)
24	125.22	5.01	16	28.22	1.93 (1H, m) 1.27 (1H, m)
		(1H, appt tquin, J=7.1, 1.4)			
7	117.80	5.25 (1H, m)	2	27.70	1.64 (2H, m)
3	79.27	3.25	28	27.61	0.97 (3H, s)
		(1H, dd, J=11.2, 4.3)			
17	52.95	1.47 (1H, m)	30	27.27	0.97 (3H, s)
14	51.15	/	27	25.73	1.68 (3H, s)
5	50.63	1.32 (1H, m)	23	25.02	2.04 (1H, m) 1.86 (1H, m)
9	48.95	2.20 (1H, m)	6	23.94	2.14 (1H, m) 1.96 (1H, m)
13	43.52	/	18	21.91	0.81 (3H, s)
4	38.96	/	21	18.33	0.88 (3H, d, J=6.4)
1	37.21	1.68 (1H, m) 1.14 (1H, m)	11	18.13	1.51 (2H, m)
22	36.19	1.46 (1H, m) 1.03 (1H, m)	26	17.64	1.60 (3H, s)
20	35.97	1.36 (1H, m)	29	14.73	0.86 (3H, s)
10	34.95	/	19	13.12	0.75 (3H, s)

NMR spectra were recorded using CDCl$_3$ and referenced to TMS. Coupling constants are reported as observed and not corrected for second order effects. Assignments were made via a combination of ^{1}H, ^{13}C, DEPT-edited HSQC, HMBC and 2D NOESY experiments. Where signals overlap ^{1}H δ is reported as the centre of the respective HSQC crosspeak. Assignments were consistent with previous literature assignments for tirucalla-7,24-dien-3β-ol (148). Table has been reproduced from the author's previous publication.

B.6   Phylogenetic tree of candidate CYPs from *M. azedarach* (Ma1)
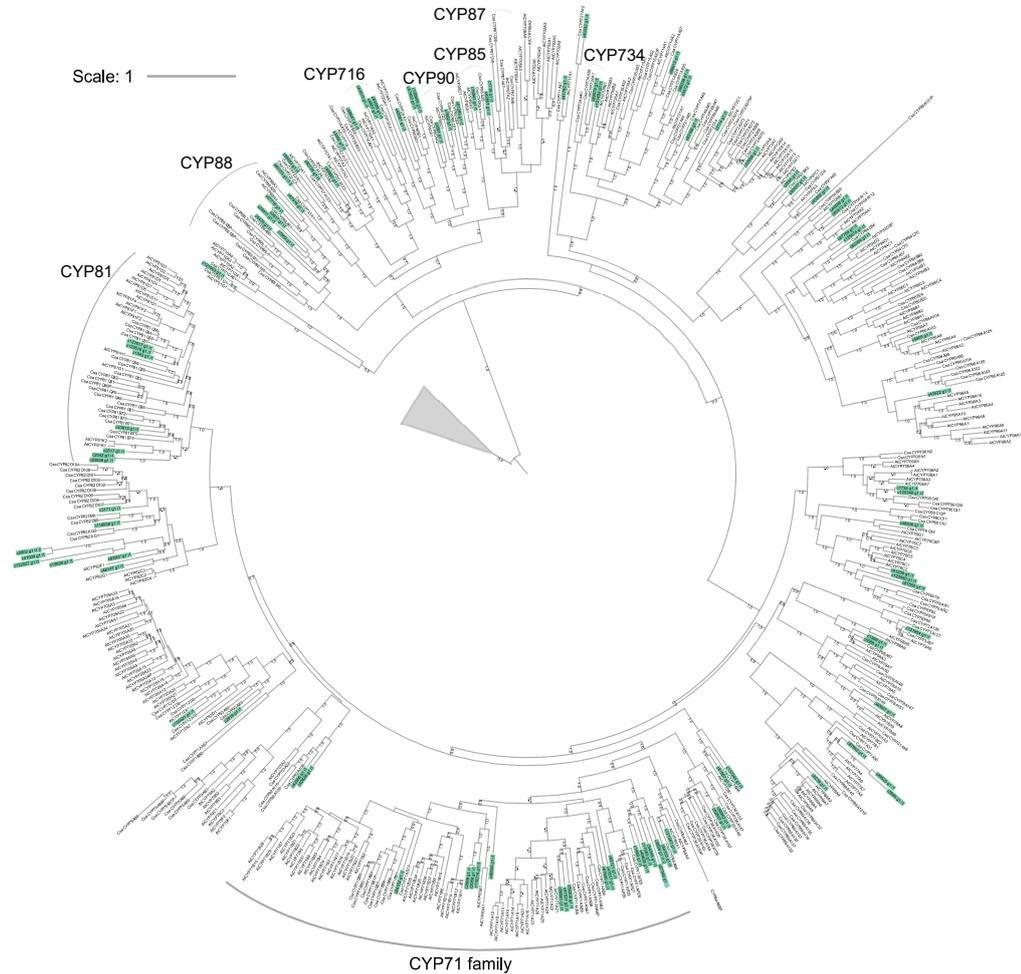
Figure B.6.1: **Full-version of phylogenetic tree (Figure 2.2.4).**
Phylogenetic tree of candidate CYPs from *M. azedarach* (green) with previously identified CYPs from *A. thaliana* (151) and *Cucumis sativus* (153) (black) included. The CYP74 family is used as an outgroup (grey triangle) and the CYP81, CYP71, CYP734, CYP87, CYP85, CYP90, CYP716 and CYP88 clades are labelled based on previously described phylogeny (117). The phylogenetic tree was constructed using FastTree V2.1.7 (146) and formatted with iTOL (147). Local support values from FastTree Shimodaira-Hasegawa (SH) test (between 0.6 and 1.0) are indicated at the nodes. The scale bar indicates estimated number of amino acid substitutions per site. Figure has been adapted from the author's previous publication (122)

B.7 NMR of AiOSC1 and MaCYP71CD2 product

Table B.7.1: ^{13}C and ^{1}H δ assignments for dihydroniloticin.

Carbon numbering scheme   **Selected COSY and HMBC**
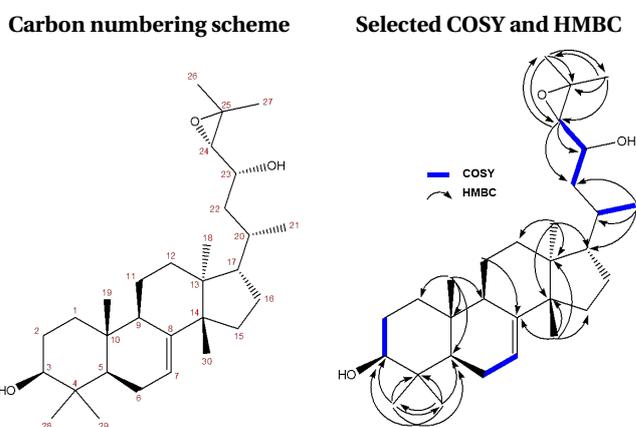
Carbon	^{13}C δ	^{1}H δ	Carbon	^{13}C δ	^{1}H δ
8	145.59	/	15	34	1.49 (2H, m)
7	118.07	5.26 (1H, m)	12	33.75	1.81 (1H, m) 1.61 (1H, m)
3	79.25	3.24 (1H, dd, J=11.1, 4.2)	20	33.6	1.40 (1H, m)
23	69.32	3.57 (1H, td, J=8.2, 5.2)	16	28.78	2.04 (1H, m) 1.22 (1H, m)
24	68.45	2.66 (1H, d, J=8.2)	2	27.7	1.65 (1H, m)
25	60.25	/	28	27.62	0.97 (3H, s)
17	53.28	1.56 (1H, m)	30	27.24	0.99 (3H, s)
14	51.21	/	26	24.88	1.33 (3H, s)
5	50.65	1.32 (1H, m)	6	23.95	2.15 (1H, m) 1.97 (1H, m)
9	48.95	2.19 (1H, m)	18	21.72	0.82 (3H, s)
13	43.61	/	21	19.94	0.96 (3H, d, J=6.2)
22	40.76	1.66 (1H, m) 1.41 (1H, m)	27	19.82	1.32 (3H, s)
4	38.98	/	11	18.1	1.53 (2H, m)
1	37.22	1.69 (1H, m) 1.14 (1H, m)	29	14.73	0.86 (3H, s)
10	34.96	/	19	13.12	0.75 (3H, s)

NMR spectra were recorded using CDCl$_3$ and referenced to TMS. Coupling constants are reported as observed and not corrected for second order effects. Assignments were made via a combination of ^{1}H, ^{13}C, DEPT-edited HSQC, HMBC and 2D NOESY experiments. Where signals overlap ^{1}H δ is reported as the centre of the respective HSQC crosspeak. Assignments were consistent with previous literature assignments for dihydroniloticin (tirucalla-7-ene-24,25-epoxy-3β,23-diol) (298). Table has been reproduced from the author's previous publication (122).

B.8 NMR of AiOSC1 and MaCYP71BQ5 product

Table B.8.1: [13] C and [1] H δ assignments for tirucalla-7,24-dien-3β,21-diol.

Carbon numbering scheme	Selected COSY and HMBC
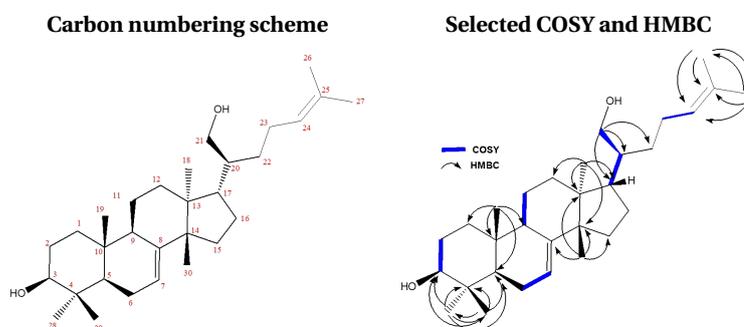
C	[13] $C\,\delta$	[1] $H\,\delta$	C	[13] $C\,\delta$	[1] $H\,\delta$
8	145.61 / 15 33.91 1.49 (2H, m)				
25	131.48	/	12	32.83	1.87 (1H, m)
					1.61 (1H, m)
24	124.86 (1H, tq, J=7.1, 1.3)	5.12	22	29.47	1.44 (2H, m)
21	62.5	3.73 (1H, dd, J=11.0, 2.9)	29	27.61	0.97 (3H, s)
		3.6 (1H, J=11.0, 4.9)			
14	51.23	/	30	27.35	0.99 (3H, s)
5	50.62	1.31 (1H, m)	26	25.72	1.69 (3H, s)
9	48.84	2.20 (1H, m)	23	24.87	2.06 (1H, m)
					1.95 (1H, m)
17	47.04	1.82 (1H, m)	6	23.95	2.15 (1H, m)
					1.97 (1H, m)
13	43.35	/	18	22.14	0.83 (3H, s)
20	42.55	1.46 (1H, m)	11	18.03	1.53 (2H, m)
4	38.96	/	27	17.73	1.62 (3H, s)
1	37.19	1.68 (1H, m) 1.14	28	14.72	0.86 (3H, s)
		(1H, td, J=12.5, 4.3)			
10	34.96	/	19	13.12	0.75 (3H, s)

NMR spectra were recorded using CDCl$_3$ and referenced to TMS. Coupling constants are reported as observed and not corrected for second order effects. Assignments were made via a combination of ^{1}H, ^{13}C, DEPT-edited HSQC, HMBC and 2D NOESY experiments. Where signals overlap [1]H δ is reported as the centre of the respective HSQC crosspeak. Table has been reproduced from the author's previous publication (122).

B.9 NMR of AiOSC1, MaCYP71CD2 and MaCYP71BQ5 product

Table B.9.1: ^{13}C δ **comparison to the literature for melianol C21 epimeric mixture.**
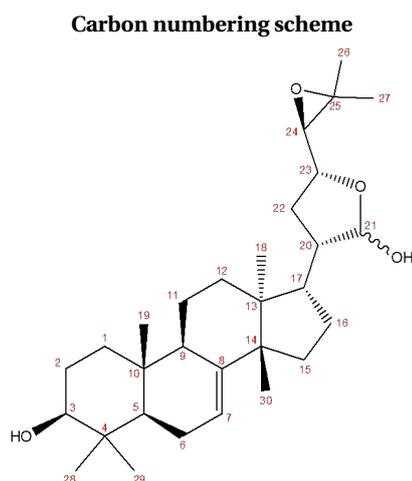
Carbon numbering scheme

	Literature ^{13}C δ	This work 13C δ		Literature ^{13}C δ	This work 13C δ
C	(100 MHz)	(150 MHz)	C	(100 MHz)	(150 MHz)
9	145.71 145.53	145.59 145.42	10	35.11 /	35.05 /
7	118.28 118.17	118.28 118.18	15	34.32 /	34.25 /
21	101.83 97.8	101.86 97.86	20	33.9 31.9	33.83 31.87
3	79.24 /	79.25 79.21	22	31.61 31.49	31.52 31.47
23	78.5 77.02	78.53 77.12	2	27.7 /	27.7 /
24	67.85 65.49	67.76 65.34	28	27.55 /	27.48 27.61
25	57.97 57.24	58.03 57.25	16	27.4 27.17	27.37 27.12
5	50.85 50.82	50.78 50.75	26	25.05 24.99	25.05 24.95
14	50.79 50.52	50.68 50.44	19	24.06 /	23.99 23.98
9	49.67 48.88	49.71 48.83	6	23.27 /	23.23 /
17	47.17 45.24	47.15 45.25	30	22.64 /	22.6 22.71
13	43.87 43.69	43.82 43.63	27	19.51 19.3	19.46 19.22
4	39.03 /	38.99 /	11	17.64 /	17.56 17.55
1	37.28 /	37.17 37.15	29	14.8 /	14.72 /
12	35.23 /	35.24 /	18	13.15 /	13.08 /

NMR spectra were recorded in Fourier transform mode at a nominal frequency of 150 MHz using CDCl$_3$ and referenced to TMS. Assignments were consistent with previous literature assignments for melianol (tirucalla-7-ene-23,21,24,25-diepoxy-3δ,21-diol)(155). Table has been reproduced from the author's previous publication (122).

B.10 Reported isolation of melianol-type protolimonoids

Table B.10.1: **Previously reported isolation of melianol-type protolimonoids.**

Protolimonoids	Family	Species	Tissue	Ref.
tirucalla-7,24-dien-3β,21-diol	Simaroubaceae	*Picrasma quassioides*	S	(154)
3-oxotirucalla-7,24-dien-21-al	Meliaceae	*Owenia cepiodora*	B	(299)
	Rutaceae	*Paramignya griffithii*	S	(300)
	Simaroubaceae	*Simaba cuneata*	S	(301)
		Simarouba amara	B	(157)
tirucalla-7,24-dien-23,3β-diol	Meliaceae	*Dysoxylum hainanense*	B	(302)
		Dysoxylum variabile	B	(303)
	Rutaceae	*Paramignya monophylla*	F	(304)
	Simaroubaceae	*Picrasma quassioides*	S	(154)
3-oxotirucalla-7,24-dien-23-ol	Meliaceae	*Dysoxylum beddomei*	L	(305)
		Dysoxylum densiflorum	S	(306)
	Rutaceae	*Paramignya monophylla*	F	(304)
	Simaroubaceae	*Picrasma quassioides*	S	(154)
		Simarouba amara	B	(307)
tirucalla-7,24-dien-3,23-dione	Meliaceae	*Dysoxylum macranthum*	B	(308)
3,23-dioxotirucalla-7,24-dien-21-al	Meliaceae	*Entandrophragma angolense*	L	(309)
dihydroniloticin	Meliaceae	*Azadirachta indica*	LS	(310)
		Dysoxylum hainanense	B	(302)
		Dysoxylum mollissimum	L	(311)
		Dysoxylum variabile	B	(303)
		Trichilia connaroides	F	(312)
		Trichilia hirta	F	(313)
		Trichilia quadrijuga	LS	(314)
		Trichilia reticulata	LS	(315)
		Trichilia schomburgkii	L	(316, 317)
		Walsura robusta	L	(318)
		Toona ciliata	L	(319)
	Rutaceae	*Phellodendron amurense*	F	(320)
		Phellodendron chinense	F	(298)
	Simaroubaceae	*Eurycoma longifolia*	S	(321)
		Picrasma quassioides	S	(154)
		Simaba polyphylla	S	(322)
23-oxodihydroniloticin	Meliaceae	*Azadirachta indica*	LS	(310)
niloticin	Meliaceae	*Aglaia cucullata*	S	(323)
		Aglaia leucophylla	B	(324)
		Dysoxylum beddomei	L	(305)
		Dysoxylum variabile	B	(303)
		Trichilia hirta	F	(313)

Table B.10.1: (continued)

Protolimonoids	Family	Species	Tissue	Ref.
		Trichilia quadrijuga	LS	(314)
		Trichilia schomburgkii	L	(316, 317)
		Turraea nilotica	R	(325)
		Walsura robusta	L	(318)
		Toona ciliata	B	(326)
	Rutaceae	*Boronia alata*	LBS	(327)
		Boronia inornata	NA	(328)
		Limonia acidissima	L	(329)
		Luvunga sarmentosa	L	(330)
		Phellodendron amurense	F	(320)
		Phellodendron chinense	F	(298)
		Vepris uguenensis	NA	(331)
	Simaroubaceae	*Ailanthus altissima*	B	(332)
		Castela polyandra	S	(333)
		Eurycoma longifolia	NA	(321)
		Picrasma quassioides	S	(154)
		Simarouba amara	B	(307)
		Simaba polyphylla	S	(322)
24,25-epoxytirucalla-7-ene-3,23-dione	Meliaceae	*Aphanamixis polystachya*	B	(334)
		Dysoxylum laxiracemosum	B	(335)
		Dysoxylum lenticellatum	LS	(336)
melianone	Meliaceae	*Dysoxylum beddomei*	L	(305)
		Guarea convergens	L	(337)
		Guarea grandiflora	Se	(338)
		Guarea kunthiana	F	(339)
		Melia azedarach	F	(213, 220)
		Melia toosendan	F	(155)
		Trichilia connaroides	F	(312)
		Trichilia hirta	F	(313)
		Trichilia reticulata	LS	(315)
		Quivisia papinae	Se	(340)
		Chukrasia tabularis	NA	(341)
		Swietenia mahagoni	NA	(342)
	Rutaceae	*Luvunga sarmentosa*	L	(330)
		Phellodendron chinense	F	(343)
		Raulinoa echinata	S	(344)
	Simaroubaceae	*Eurycoma longifolia*	NA	(321)
		Picrasma quassioides	S	(154)
		Simarouba amara	R	(157)
melianol	Meliaceae	*Aglaia odorata*	S	(345)
		Guarea kunthiana	F	(339)
		Melia azedarach	NA	(220)
		Melia toosendan	F	(155)
		Trichilia connaroides	F	(312)
		Trichilia hirta	F	(313)
		Quivisia papinae	Se	(340)

Table B.10.1: (continued)

Protolimonoids	Family	Species	Tissue	Ref.
		Toona ciliata	B	(326)
melianone lactone	Meliaceae	*Trichilia hirta*	F	(313)
turraeanthin	Meliaceae	*Turraeanthus africanus*	S	(156)

The family, species, tissue of isolation and relevant publication are given. The tissue of isolation is indicated as follows Stem (S), Bark (B), Fruit (F), Leaf (L), Seed (Se) and Root (R). Table has been reproduced from the author's previous publication (122).
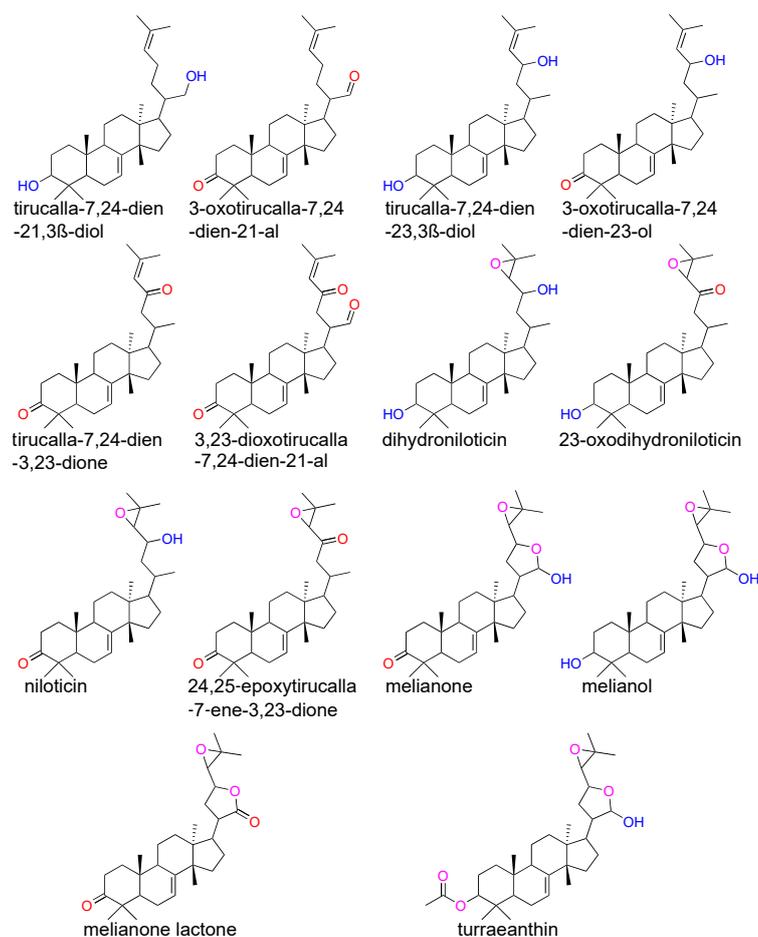
Figure B.10.1: **Structures of previously isolated protolimonoids**
Protolimonoid structures with melianone-, niloticin- or tirucalla-7,24-dien-3β-ol-type scaffolds. Oxidations of each protolimonoid are indicated as follows: ketone or aldehyde (red), hydroxylation (blue) and other oxidations (pink). To encompass reported isolation of structures where stereochemistry has not been specified, the stereochemistry at C3, C5, C9, C20, C21, C23 and C24 was not specified in Reaxys searches. Figure has been reproduced from the author's previous publication (122).

B.11   Template email for provenance enquiries

Hello,

I am a scientist at the John Innes Centre (Norwich, UK) studying chemical pathways in the *XXXXX* family of plants. I found online that you sell the species *XXXXX* which I am interested in purchasing to use in my research. However, to ensure the research is conducted fairly and in line with current access and benefit sharing legislation- I need to know the details of the 'provenance' of any plants I use (i.e. the trail of information of when and where it was sourced). Further, to be out of scope of the Nagoya Protocol it must have been sourced previously to the Nagoya Protocol coming in to force on the 12th October 2014).

I was wondering whether you can provide this type of background information about the source of this species.
Apologies that this is a bit of a strange and difficult request, any information you could provide would be greatly appreciated!

Furthermore, if this information is not available for this species- do you sell any other members of the *XXXXX* family which you might have this type of information for.

Many thanks in advance for your time,
Best wishes

B.12   Correspondence with Japanese National Focal Point

From: Hannah Griffiths (JIC)
Sent: Monday, July 10, 2017 10:19 PM
To: OKUDA SEISHU
Subject: Use of Melia azedarach plants from UK supplier in research
Dear Mr. Seishu Okuda,

I am contacting you as you are listed as the ABS national focal point for Japan on the access and benefit sharing clearing house (ABSCH) and I was hoping you could answer a related question for me. I am a PhD student at John Innes Centre (UK) and I am studying the species Melia azedarach. I have been looking for suppliers of this species in the UK and found a small plant nursery that are selling individual plants for horticultural use. Before purchasing any plants from them I asked for information about the individual's provenance (i.e. where they sourced the seeds from and when) to ensure that I would not be working on material in-scope of the Nagoya protocol. Their response was the following;

'The seed of Melia azedarach was collected by ourselves in Japan in the autumn of 2015. The Japanese do not require us to apply for permits to collect, as long as it is not in a National Park or other protected areas. This seed was collected near Fukuoka on Kyushu Island, just north of the town of Kurume in the low mountains there at 450 m edge of forest.'

As there is no legislative information currently available on the ABSCH and Japan is not yet party to the Nagoya protocol, I am unsure whether this material is considered by yourselves in-scope of the Nagoya Protocol as it was sourced after 12th October 2014. Further I do not know whether using these individuals for research-purposes breaches any other legislation or requires a form of permit. I was hoping that you could answer the above questions for me and let me know as soon as possible whether I would be able to use this material in my research. Please let me know if there is any other information I need to provide.

Thank you in advance for your time,
Very Best Wishes,
Hannah Griffiths

From: OKUDA SEISHU
Sent: Tuesday, July 11, 2017 4:14:59 AM
To: Hannah Griffiths (JIC)
Subject: RE: Use of Melia azedarach plants from UK supplier in research
Dear Ms Hannah Griffiths,

Thank you for your email. Japan ratified the Nagoya Protocol on 22 May this year and will be a party from 20 August. As for your question, with regard to the Nagoya Protocol, you don't need a permit for the utilization of the plant which you collected because Japan doesn't require prior informed consent (PIC) as provided in Article 6.1 of the Protocol.

Please note that there is no mountain which has more than 450 meters high in the north of Kurume city as long as I see a map of Kyushu island, therefore I'm not sure where the seed was collected and I have no way to determine whether the place where the seed was collected is in protected areas or not. The redline above shows the area of Kurume city. Perhaps, the seed was collected in Tosu city located north of Kurume city and south of Fukuoka city, I guess. Anyway, I hope your study will be fruitful and contribute to the conservation and sustainable use of biodiversity.

Best regards,
Seishu Okuda

From: Hannah Griffiths (JIC)
Sent: Tuesday, July 11, 2017 5:16 PM
To: OKUDA SEISHU
Subject: Re: Use of Melia azedarach plants from UK supplier in research
Dear Mr. Seishu Okuda,

Thank you very much for responding so quickly and efficiently to my email- it is very much appreciated. Regarding the exact sampling location- I see how the description where I was send could cause some ambiguity. I believed the "450m" was referring to distance from edge of the forest rather than altitude. However as I cannot be sure I have contacted the supplier and asked them to confirm for me. Also from looking at google maps there were some regions in the lowland mountain circled below that had altitude of up to 600m- I suspect the seeds were collected from this region but I will email you in the near future to confirm. Are there any national parks/protected areas in this region?

Thank you for confirming that no permit/PIC is required, are you able to confirm- for my records that this means I am able to use the material in my research with no further documentation providing we determine the seed was not collected in a protected area?

Many thanks for your time and best wishes,
Hannah Griffiths

From: OKUDA SEISHU
Sent: Tuesday, July 11, 2017 10:16:47 AM
To: Hannah Griffiths (JIC)
Subject: RE: Use of Melia azedarach plants from UK supplier in research
Dear Ms Hannah Griffiths,

There are two prefectural natural parks, "Sefuri-Hokusan Prefectural Natural Park" and "Sefuri-Raizan Prefectural Natural Park". The dotted areas(blue, green, purple and red (orange?) colored. The difference of these colors shows the difference of the levels of regulation) in the map below are their areas. The red circle shows almost the same area that you descripted in your mail. I'm not the authority of these prefectural natural parks, but as long as I checked, there is no regulation of these Prefectural Natural Parks that prohibits to collect seed of Melia azedarach, fortunately. (cf. It is prohibited to cut trees without permission in green, purple and red(orange?) dotted areas.)

Best regards,
Seishu Okuda

From: Hannah Griffiths (JIC)
Sent: Tuesday, July 11, 2017 7:15 PM
To: OKUDA SEISHU
Subject: Re: Use of Melia azedarach plants from UK supplier in research
Dear Mr. Seishu Okuda,

Thank you again for such a detailed email and apologies for causing any confusion regarding this. I know have confirmed with the suppliers the exact location- I believe we had both misunderstood the description as the sampling site was further east than we thought (I have attached the map edited by the supplier below). Are you able to confirm the state and any regulations of any national parks in this region? If there is none- are you able to confirm that I can use these plants in my research.

Many Thanks again for your help- it is much appreciated.
Hannah Griffiths

From: OKUDA SEISHU
Sent: Wednesday, July 12, 2017 10:49 AM
To: 'Hannah Griffiths (JIC)'
Subject: RE: Use of Melia azedarach plants from UK supplier in research
Dear Ms Hannah Griffiths,

The place you pointed seems ordinary zone of Chikugogawa Prefectural Natural Park in Fukuoka Prefecture. The green circle shows almost the same area that you described. Blue dotted area shows ordinary zone. The red point below shows the point of altitude 450 meters. I believe this map will help you. You can see detailed map from the following link.
http://www.pref.fukuoka.lg.jp/uploaded/life/106719_17825731_misc.pdf

Google translate will help you understand the Chikugogawa Prefectural Natural Park. The link below is a webpage of Fukuoka Prefectural Government regarding the Chikugogawa Prefectural Natural Park
http://translate.google.com/translate?hl=en&sl=ja&tl=en&u=http%3A%2F%2Fwww.pref.fukuoka.lg.jp%2Fcontents%2Fchikugogawa.html&sandbox=1

If you are interested in Natural Park Act of Japan, click the following link.
http://www.env.go.jp/en/laws/nature/law_np.pdf

Chapter III (Article 73-81) is about Prefectural Natural Parks. Article 33 shall apply mutatis mutandis to Ordinary Zone of Prefectural Natural Park. Ordinary Zone has no regulation of collecting seed.

Best regards,
Seishu Okuda

B.13 Example LC-MS EIC from *M. azedarach* profiling
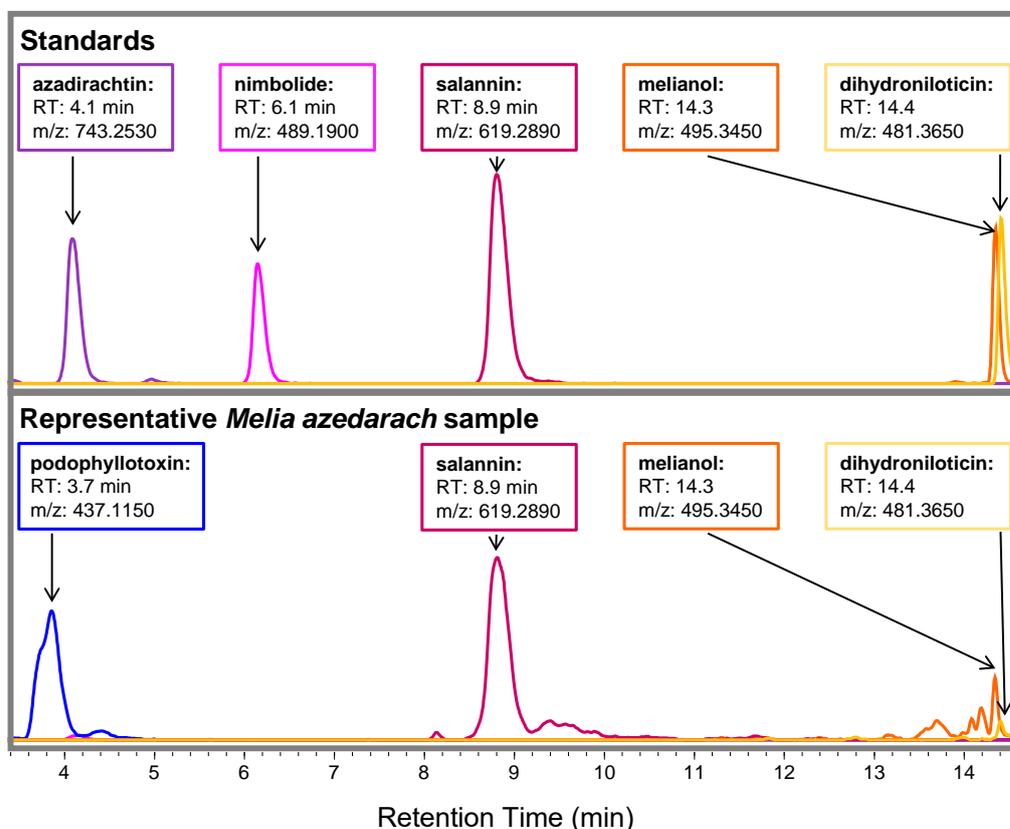
Figure B.13.1: **Example chromatograms from profiling of *M. azedarach.***
UHPLC-IT-TOF generated EIC of analytical standards and a representative *M. azedarach* extract. EICs displayed are for the following adducts: [azadirachtin+Na]$^+$=743.2530 (purple), [nimbolide+Na]$^+$=489.1900 (pink), [salannin+Na]$^+$=619.2890 (maroon), [melianol+Na]$^+$=495.3450 (orange), [dihydroniloticin+Na]$^+$=481.3650 (yellow) and the internal standard [podophyllotoxin+Na]$^+$=437.1150 (blue). Within the representative *M. azedarach* trace, melianol and dihydroniloticin traces have been scaled by a factor of five to allow clearer visibility of these peaks. UHPLC-IT-TOF analysis performed using 'limonoid' method.

B.14   Example sampling of *M. azedarach* leaf for RNA-Seq
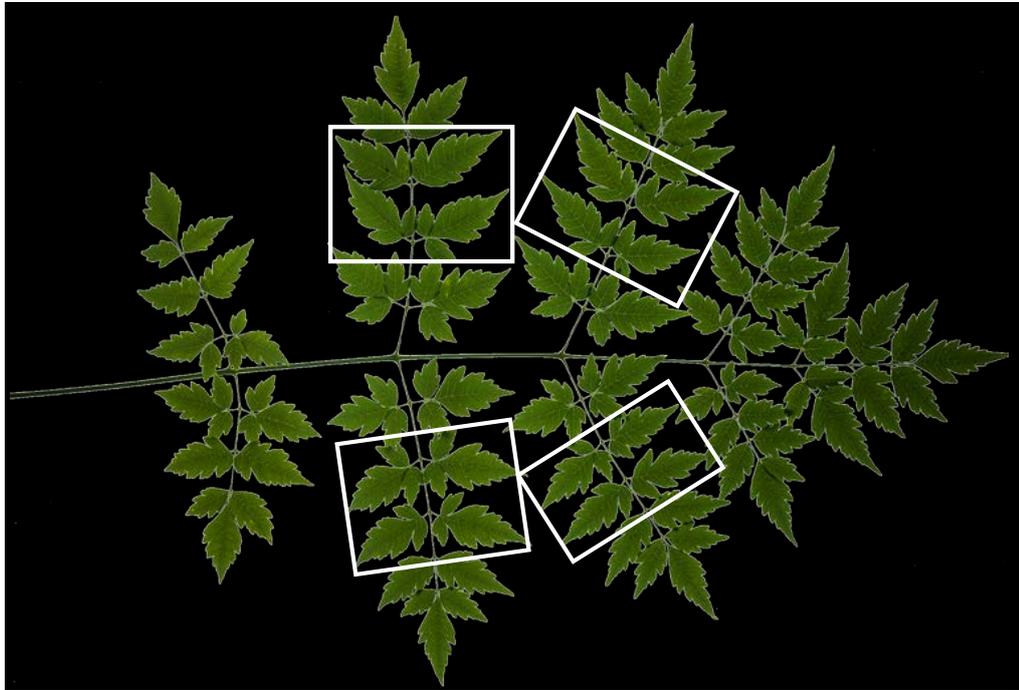
Figure B.14.1: ***M. azedarach* leaf (~ 1 year old) used for RNA-Seq.**
White boxes indicate leaflets sample, with each box treated as a a biological replicate.
Four leaflets were collected to ensure enough material for both RNA extraction and
metabolite analysis. The morphology of the individuals at the time of sampling
prevented the collection of different leaves as biological replicates. Each individual
consisted of one to two long thin stems with a small number of fully formed leaves in
the upper region of the stem. The production of leaves occurring only at the growing
tip of the stem resulted in all of the leaves being of different ages and therefore
unsuitable biological replicates. Image taken by Andrew Davis (JIC photography).

B.15   Summary of raw *M. azedarach* RNA-Seq data

Table B.15.1: **Number of paired end reads generated by EI per lane, replicate and sample.**

Sample	Rep.	Lane 1	Lane 2	Total (per rep.)	Total (per sample)
JPN11	1A	7,312,258	7,957,007	15,269,265	78,676,364
Upper Leaf	1B	12,440,818	13,367,863	25,808,681	
	1C	9,677,158	10,402,466	20,079,624	
	1D	8,501,858	9,016,936	17,518,794	
JPN11	2A	14,706,042	15,713,081	30,419,123	95,833,402
Lower Leaf	2B	9,952,003	10,506,690	20,458,693	
	2C	9,995,844	10,724,057	20,719,901	
	2D	11,759,629	12,476,056	24,235,685	
JPN11	3A	11,225,462	12,293,851	23,519,313	82,662,893
Petiole	3B	8,518,447	9,151,386	17,669,833	
	3C	8,723,766	9,267,735	17,991,501	
	3D	11,360,248	12,121,998	23,482,246	
JPN11	4A	12,795,130	13,497,456	26,292,586	107,736,216
Root	4B	9,430,278	10,235,484	19,665,762	
	4D	14,075,197	14,780,596	28,855,793	
	4F	15,951,734	16,970,341	32,922,075	
JPN02	5A	8,425,596	8,942,230	17,367,826	86,473,401
Upper Leaf	5B	7,483,256	7,905,622	15,388,878	
	5C	15,588,782	16,252,245	31,841,027	
	5D	10,597,294	11,278,376	21,875,670	
JPN02	6A	7,570,717	7,949,090	15,519,807	83,790,681
Lower Leaf	6B	15,757,443	16,754,196	32,511,639	
	6C	8,341,297	8,628,045	16,969,342	
	6D	9,074,779	9,715,114	18,789,893	
JPN02	7A	15,145,250	16,073,522	31,218,772	100,692,927
Petiole	7B	11,317,371	12,034,239	23,351,610	
	7C	12,710,000	13,530,865	26,240,865	
	7D	9,595,536	10,286,144	19,881,680	

B.16   Putative cluster containing *MaOSC1*
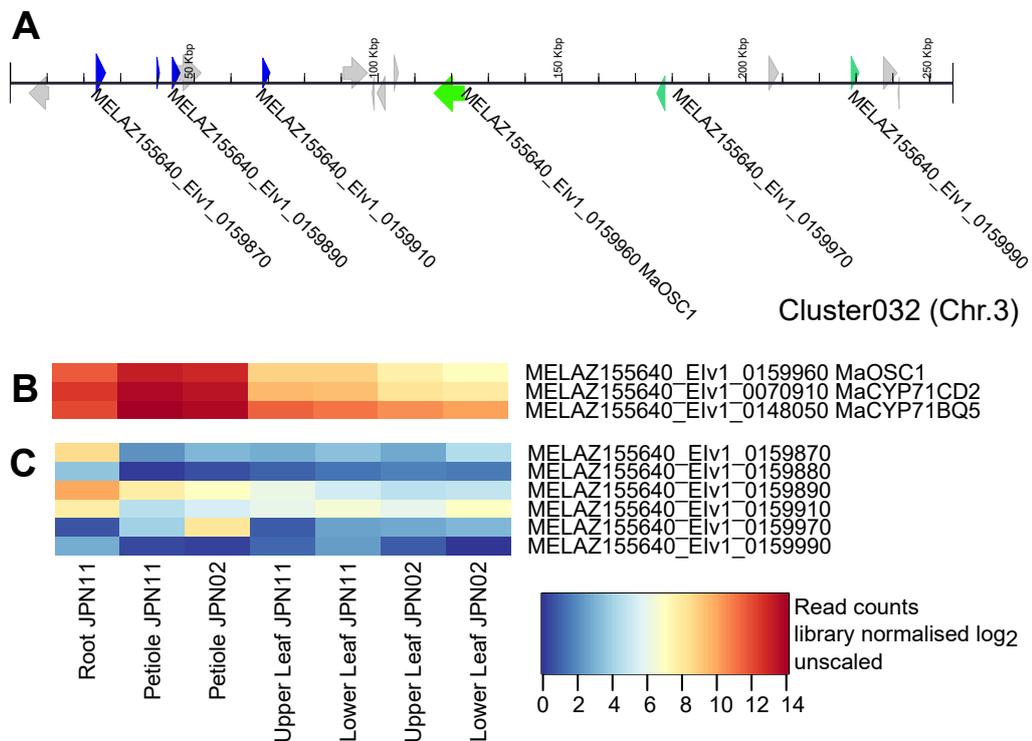
Figure B.16.1: **Putative plantiSMASH biosynthetic gene cluster 032.**
(A) Structure of putative biosynthetic gene cluster containing *MaOSC1* (predicted by plantiSMASH (229)). Heatmap displaying expression patterns of (B) functional melianol biosynthetic genes (*MaOSC1, MaCYP71CD2* and *MaCYP71BQ5*) and (C) genes occurring within putative cluster 032. Heatmap was constructed by Heatmap3 V1.1.1 (271) using read counts normalised by library size and $\log_2$ transformed. Values in rows have not been scaled.

B.17   Full list of differentially expressed genes

Table B.17.1: **Genes identified as co-expressed with melianol biosynthetic genes.**

EIv1 ID	Human Readable Description	InterPro domains
0000150.1	Leucine-rich repeat receptor kinase	IPR003591,IPR011009, IPR013210
0003510.1	Calcium-binding EF-hand	None
0004080.1	EXORDIUM	IPR006766
0004300.1	DUF668/DUF3475 domain	IPR007700,IPR021864
0006360.1	Membrane-associated kinase regulator, putative	None
0008560.1	Disease resistance	IPR001611,IPR002182, IPR027417
0009090.1	Heparanase	IPR005199
0009440.1	Chaperone	IPR001270,IPR004176, IPR019489,IPR027417
0010720.1	MIZU-KUSSEI 1	IPR006460
0011260.1	Heavy metal-associated isoprenylated	IPR006121
0011420.1	Ribosomal L34Ae	IPR008195, IPR012870
0015190.1	2-oxoglutarate and Fe(II)-dependent oxygenase	IPR005123,IPR026992, IPR027443
0015210.1	2-oxoglutarate and Fe(II)-dependent oxygenase	IPR005123,IPR026992, IPR027443
0015320.1	NADPH-dependent codeinone reductase	IPR020471, IPR023210
0015350.1	Aldo/keto reductase	IPR020471, IPR023210
0015800.1	GDSL esterase/lipase	IPR013830
0015820.1	NRT1/PTR 4.4	IPR000109, IPR020846
0018850.1	Tetraspanin	IPR018499
0019990.1	cyclin-dependent kinase inhibitor SMR1	None
0020520.1	IQ domain	IPR000048
0021010.1	Ubiquitin carboxyl-terminal hydrolase 12	IPR008974
0023520.1	EFFECTOR OF TRANSCRIPTION 2	None
0027140.1	Patellin-6, putative	IPR001251, IPR009038,IPR011074
0030030.1	Myb transcription factor	IPR009057,IPR031055
0030250.1	Dof zinc finger	IPR003851
0031400.1	Kinase	IPR011009,IPR013320, IPR024788,IPR032675
0033220.1	Homeobox-leucine zipper	IPR003106,IPR009057
0033630.1	NRT1/PTR 2.2	IPR000109,IPR020846
0034340.1	Nucleobase-ascorbate transporter 6	IPR006043
0034410.1	WAT1-related	IPR000620, IPR030184
0036000.1	SIEVE ELEMENT OCCLUSION B	IPR027942,IPR027944
0036170.1	Methyltransferase type	IPR029063
0036220.1	regulator of G- signaling 1	IPR016137
0036270.1	Mitogen-activated kinase kinase kinase	IPR011009,IPR013320, IPR024788,IPR032675
0037240.1	of unknown function (DUF 3339)	IPR021775
0038550.1	Chaperonin 60 subunit beta	None

Table B.17.1: (continued)

EIv1 ID	Human Readable Description	InterPro domains
0038650.1	kinase domain	IPR001611,IPR011009, IPR013210,IPR013320, IPR032675
0043260.1	F-box/kelch-repeat	IPR001810,IPR011043, IPR015915,IPR017451
0044980.1	GLABRA2 expression modulator	IPR004182
0045130.1	Haloacid dehalogenase hydrolase (HAD)	None
0045140.1	Haloacid dehalogenase hydrolase (HAD)	IPR023214
0047010.1	Ethylene-responsive transcription factor	IPR016177
0047610.1	xyloglucan 6-xylosyltransferase 2	IPR008630
0047710.1	Heavy metal-associated isoprenylated	IPR006121
0049630.1	Ankyrin repeat	IPR020683,IPR026961
0052040.1	Cytochrome b5	IPR001199
0052810.1	Unknown	None
0052900.1	RBR-type E3 ubiquitin transferase	IPR002867,IPR013083, IPR031127
0052970.1	Disease resistance	IPR002182,IPR027417, IPR032675
0052990.1	Cytochrome P450	IPR001128
0053980.1	Ubiquitin-conjugating enzyme, E2	IPR016135
0054280.1	MYB transcription factor	IPR009057
0054560.1	zinc finger CONSTANS	IPR010402
0055440.1	RING/U-box	IPR013083, IPR033275
0055610.1	Homeobox-leucine zipper	IPR009057,IPR013978, IPR023393
0055820.1	Actin cytoskeleton-regulatory complex pan	None
0056290.1	Early nodulin	IPR008972
0058770.1	DJ-1	IPR006287, IPR029062
0059930.1	Callose synthase	IPR003440,IPR023175, IPR026899
0061220.1	Cytochrome b5	IPR001199
0061380.1	Small heat shock	IPR008978,IPR031107
0061950.1	Cytochrome P450 ent-kaurenoic acid oxidase	IPR001128
0061960.1	Cytochrome P450 ent-kaurenoic acid oxidase	IPR001128
0062310.1	deoxycytidylate deaminase	IPR015517
0062520.1	F-box/LRR-repeat	IPR001810,IPR019494
0063430.1	Unknown	None
0064420.1	Callose synthase	IPR003440,IPR023175, IPR026899
0064770.1	HVA22	IPR004345
0065650.1	WAT1-related	IPR000620,IPR030184
0065680.1	Inactive receptor kinase	IPR011009,IPR013210, IPR032675
0065860.1	Bidirectional sugar transporter SWEET	IPR004316,IPR020846
0067010.1	SIEVE ELEMENT OCCLUSION B	IPR027942,IPR027944
0067100.1	SIEVE ELEMENT OCCLUSION B	IPR027942,IPR027944
0067140.1	SIEVE ELEMENT OCCLUSION B	IPR027942,IPR027944

Table B.17.1: (continued)

EIv1 ID	Human Readable Description	InterPro domains
0067860.1	UPF0481 At3g47200	IPR004158
0069730.1	LIGHT-DEPENDENT SHORT HYPOCOTYLS 10	IPR006936
0070020.1	RING-H2 finger	IPR013083
0070190.1	Cytochrome P450 ent-kaurenoic acid oxidase	IPR001128
0070780.1	pollen-specific leucine-rich repeat extensin	IPR006121
0070910.1	Cytochrome P450	IPR001128
0071880.1	Subtilisin protease	IPR015500
0072390.1	WEB	None
0073200.1	Jacalin-related lectin	IPR001229
0073520.1	Mannan endo-1,4-beta-mannosidase	IPR017853
0073910.1	Glucan endo-1,3-beta-glucosidase, putative	IPR000490,IPR012946, IPR017853
0074200.1	deoxyuridine 5'-triphosphate nucleotidohydrolase	IPR008180,IPR029054
0074660.1	B3 domain	IPR015300
0074770.1	Nucleobase-ascorbate transporter	IPR006043
0074880.1	Phosphoglycerate mutase 4	IPR029033
0076210.1	Subtilisin protease	IPR015500
0077260.1	Auxin-responsive	IPR000270,IPR003311, IPR033389
0077330.1	Octicosapeptide/Phox/Bem1p domain	IPR000270
0078440.1	O-fucosyltransferase	IPR019378
0078490.1	Alcohol dehydrogenase	IPR011032,IPR013149, IPR016040
0078600.1	F-box FBW2	IPR001810,IPR032675
0080030.1	WRKY transcription factor	IPR003657,IPR018872
0081740.1	BTB/POZ domain	None
0082630.1	E3 ubiquitin ligase DRIP2	None
0083530.1	inactive RESTRICTED TEV MOVEMENT 2	IPR008978
0089350.1	Calcineurin subunit B	IPR011992
0089450.1	Class I heat shock	IPR008978,IPR031107
0090170.1	Transmembrane	None
0091390.1	Unknown	None
0092190.1	Unknown	None
0094380.1	Universal stress A	IPR006015
0094830.1	major allergen Pru ar 1	IPR023393,IPR024949
0094840.1	major allergen Pru ar 1	IPR023393,IPR024949
0095290.1	bifunctional epoxide hydrolase 2	IPR000639,IPR029058
0095420.1	Late embryogenesis abundant Lea5	IPR004926
0096390.1	Unknown	None
0096710.1	phosphatase 2c, putative	IPR001932,IPR015655
0097190.1	RING zinc finger	None
0098480.1	BI1	IPR006214
0098680.1	Argonaute	IPR003100,IPR012337, IPR014811,IPR032472, IPR032473,IPR032474
0100910.1	Histidine phosphotransfer , putative	IPR008207
0103360.1	B-cell receptor-associated 31	IPR008417

Table B.17.1: (continued)

EIv1 ID	Human Readable Description	InterPro domains
0104100.1	Heavy metal-associated isoprenylated	IPR006121
0104590.1	Lysosomal Pro-X carboxypeptidase	IPR008758,IPR029058
0104940.1	Stress responsive A/B Barrel Domain	IPR011008
0105850.1	DUF724	IPR007930
0108070.1	Embryo-specific	IPR001024,IPR010417
0108370.1	LIGHT-DEPENDENT SHORT HYPOCOTYLS	IPR006936
0110320.1	Heavy metal-associated domain	None
0110950.1	heavy metal-associated isoprenylated 3	IPR006121
0111290.1	Pectinesterase	IPR006501,IPR011050
0111360.1	phosphatase 2C	IPR001932,IPR015655
0112790.1	transport Sec24	IPR002035,IPR006895, IPR006896,IPR006900, IPR012990,IPR029006
0115670.1	Cysteine/Histidine-rich C1 domain	IPR002219,IPR004146
0117220.1	3-phosphatidyltransferase	IPR000462
0118130.1	Xyloglucan endotransglucosylase/hydrolase	IPR013320,IPR016455
0119690.1	Kynurenine formamidase	IPR007325
0120380.1	2-hydroxyisoflavanone dehydratase	IPR029058
0121420.1	Phospholipid/glycerol acyltransferase	IPR002123,IPR032098
0123750.1	TOPLESS	IPR006594, IPR006595,IPR015943
0123880.1	synaptotagmin-5	IPR000008,IPR031468
0124880.1	TIFY 8	IPR010399
0125840.1	Homeobox-leucine zipper	IPR009057, IPR013978,IPR023393t
0126870.1	MYB transcription factor	IPR009057
0128140.1	Class I glutamine amidotransferase	IPR011697,IPR029062
0128310.1	plastidic glucose transporter 4	IPR005828,IPR020846
0128570.1	Cytochrome P450, putative	IPR001128
0129090.1	DUF1262	IPR010683
0132430.1	ABC transporter B	IPR003439,IPR011527, IPR027417
0134260.1	Sentrin/sumo-specific protease, putative	IPR003653
0134290.1	Chaperone DnaJ-domain	IPR001623
0134700.1	Unknown	None
0135060.1	tobamovirus multiplication 1	IPR009457
0135650.1	carboxylesterase 1	IPR029058
0136100.1	LRR receptor kinase	IPR001611,IPR003591, IPR013210,IPR032675
0136380.1	Dof zinc finger	IPR003851
0138230.1	Glutathione peroxidase	IPR000889,IPR012336
0139550.1	Ubiquitin-conjugating enzyme, E2	IPR016135
0140540.1	Dof zinc finger	IPR003851
0141150.1	Heavy metal-associated isoprenylated	IPR006121
0141190.1	oxidoreductases, acting on NADH or NADPH	None
0141380.1	Plastid movement impaired	IPR025322
0142020.1	myosin-binding 3	IPR007656

Table B.17.1: (continued)

EIv1 ID	Human Readable Description	InterPro domains
0142070.1	vinorine synthase	IPR003480,IPR023213
0142870.1	F-box	IPR001810,
		IPR006566,IPR032675
0143460.1	Phosphate transporter PHO1	IPR004331,IPR004342
0147870.1	Histidine kinase	IPR003594,IPR003661,
		IPR004358,IPR005467,
		IPR006189,IPR011006
0148050.1	Cytochrome P450	IPR001128
0148250.1	Methyltransferase	IPR005299,IPR029063
0148480.1	zinc finger CONSTANS 9	IPR000315,IPR010402
0152400.1	LIGHT-DEPENDENT SHORT HYPOCOTYLS 10	IPR006936
0153350.1	AWPM-19	IPR008390
0154150.1	Transmembrane	IPR011990
0154500.1	Ankyrin repeat domain	IPR020683,IPR021832
0155320.1	Auxin-responsive	IPR000270,IPR003311,
		IPR033389
0155970.1	Mitochondrial carrier	IPR002067,IPR023395
0156270.1	F-box/LRR-repeat	IPR001810,IPR032675
0159840.1	F17F16.3	None
0159960.1	Terpene cyclase/mutase member	IPR018333,IPR032696,
		IPR032697
0161110.1	Ubiquitin carboxyl-terminal hydrolase 12	IPR008974
0161120.1	Ubiquitin carboxyl-terminal hydrolase 12	IPR008974
0161130.1	Ubiquitin carboxyl-terminal hydrolase 12	IPR008974
0162330.1	Calcium-binding EF hand	IPR011992
0162590.1	NUCLEAR FUSION DEFECTIVE 4	IPR010658,IPR020846
0163280.1	Ankyrin repeat	IPR020683, IPR026961
0163600.1	inactive RESTRICTED TEV MOVEMENT 2	IPR008978
0163610.1	inactive RESTRICTED TEV MOVEMENT 2	IPR008978
0164450.1	vinorine synthase	IPR003480,IPR023213
0165010.1	Defensin	IPR008176
0165910.1	transmembrane ascorbate ferrireductase 1	IPR006593
0169240.1	AAA-type ATPase	IPR003959,
		IPR007330,IPR015415,
		IPR027417
0170810.1	WRKY transcription factor	IPR003657
0170970.1	Cytochrome P450	IPR001128
0173120.1	26S proteasome non-ATPase regulatory subunit 4	IPR027040
0173620.1	Peptide methionine sulfoxide reductase MrsB	IPR002579
0174510.1	RNA-binding	IPR012677
0174570.1	Auxin efflux carrier , putative	IPR004776
0175400.1	Unknown	None
0175520.1	Acyl-[acyl-carrier-] desaturase	IPR005067,IPR009078
0175570.1	Acyl-[acyl-carrier-] desaturase	IPR005067,IPR009078
0177570.1	DUF674	IPR007750
0177590.1	DUF674	IPR007750
0179950.1	Bifunctional inhibitor/lipid transfer /seed storage	IPR016140

Table B.17.1: (continued)

EIv1 ID	Human Readable Description	InterPro domains
0180860.1	VAN3-binding	IPR008546,IPR013666
0181090.1	Phosphatidylinositol/phosphatidylcholine transfer	IPR001071,IPR011074
0181910.1	Transmembrane	None
0182890.1	21 kDa	IPR006501
0185570.1	AP2 ethylene-responsive TF	IPR016177,IPR031112
0186060.1	Auxin-responsive	IPR000270,IPR003311, IPR033389
0187600.1	F-box	IPR001810,IPR017451
0188240.1	RING-H2 finger	IPR013083
0188740.1	Egg cell-secreted 1.1,	IPR008502
0188750.1	Egg cell-secreted 1.1,	IPR008502
0188900.1	S-acyltransferase	IPR001594
0189050.1	WRKY transcription factor	IPR003657
0190240.1	Alanine-glyoxylate aminotransferase 2, putative	IPR005814,IPR015424
0190600.1	of unknown function	IPR007650
0190640.1	Glucan 1,3-beta-glucosidase	IPR008999,IPR010431, IPR017853
0191770.1	DETOXIFICATION	IPR002528
0192180.1	Hydroxyacylglutathione hydrolase	IPR017782,IPR032282
0192980.1	Sterol-8,7-isomerase	IPR007905,IPR033118
0193320.1	Trihelix transcription factor	None
0193860.1	Subtilisin protease	IPR015500
0194630.1	SMAX1 3	IPR004176,IPR027417
0198190.1	Short-chain dehydrogenase/reductase	IPR002347
0198240.1	secoisolariciresinol dehydrogenase	IPR002347
0198810.1	Phosphate transporter PHO1	IPR004331,IPR004342
0198830.1	peroxisomal membrane PMP22	IPR007248
0202990.1	Beta-glucosidase	IPR001360,IPR017853
0204630.1	E3 ubiquitin ligase drip2	None
0205830.1	Ribonuclease	IPR001568
0206230.1	Trehalose-6-phosphate synthase, putative	IPR006379,IPR023214
0206240.1	Trehalose-6-phosphate synthase	IPR001830
0206340.1	Inositol-tetrakisphosphate 1-kinase	IPR008656
0209740.1	Aspartic ase Asp1	IPR001461, IPR021109,IPR032799, IPR032861
0209850.1	Ubiquitin carboxyl-terminal hydrolase 12	IPR008974
0211970.1	SMAX1 4	IPR004176,IPR027417
0212340.1	Cadmium/zinc-transporting ATPase 3	IPR001757,IPR023214
0212860.1	Syntaxin/t-SNARE	IPR010989
0213750.1	Beta-1,6-N-acetylglucosaminyltransferase	IPR003406
0214480.1	UBX domain 1	IPR009060,IPR018997
0214610.1	Multidrug resistance 3	None
0214640.1	Leucine-rich receptor kinase , putative	IPR001611,IPR003591, IPR011009,IPR013210, IPR032675
0215380.1	Zinc transporter	IPR003689

Table B.17.1: (continued)

EIv1 ID	Human Readable Description	InterPro domains
0215940.1	CLAVATA3/ESR-related TDIF	None
0217180.1	beta-amyrin 28-oxidase	None
0218850.1	Class I heat shock	IPR008978,IPR031107
0218870.1	Class I heat shock	IPR008978,IPR031107
0220510.1	Ankyrin repeat , putative	IPR020683,IPR026961
0221370.1	dnaJ homolog	IPR012724
0221500.1	Thermospermine synthase	IPR001045,IPR029063
0221670.1	17.6 kDa class II heat shock	IPR008978,IPR031107
0221830.1	Ubiquitin carboxyl-terminal hydrolase 12	IPR008974
0221850.1	17.9 kDa class II heat shock	IPR008978,IPR031107
0221860.1	Ubiquitin carboxyl-terminal hydrolase 12	IPR008974
0221880.1	17.9 kDa class II heat shock	IPR008978,IPR031107
0222160.1	kinase	IPR011009
0223340.1	Disease resistance	IPR002182,IPR027417
0223450.1	Disease resistance	IPR032675
0223780.1	NADPH-dependent codeinone reductase	IPR020471,IPR023210
0224550.1	NAC domain	IPR003441
0225020.1	Myosin-binding	IPR007656
0225170.1	Unknown	None
0227040.1	Glucan endo-1,3-beta-glucosidase 1	IPR012946,IPR013781
0227990.1	Endoribonuclease Dicer	IPR000999,IPR003100, IPR014720
0228000.1	Endoribonuclease Dicer	IPR000999,IPR003100, IPR014720
0228670.1	Zinc finger	IPR000315,IPR010402
0228690.1	amino-acid permease BAT1 homolog	IPR002293
0228850.1	Reticulon	IPR003388
0229060.1	LRR receptor kinase	IPR011009,IPR013210, IPR013320,IPR032675
0230050.1	Myosin-binding	IPR007656
0230180.1	CADMIUM RESISTANCE 2	IPR006461
0232400.1	Pectin lyase	IPR000743,IPR011050
0233960.1	Disease resistance	IPR002182,IPR027417, IPR032675
0235320.1	NAC domain ,	IPR003441
0235630.1	vinorine synthase	IPR003480,IPR023213
0236020.1	of unknown function	IPR006502
0236040.1	Ankyrin repeat	IPR020683,IPR026961
0236560.1	Glucose-6-phosphate/phosphate translocator 2	IPR004696,IPR004853
0237210.1	Ankyrin repeat , putative	IPR020683,IPR026961
0238770.1	vinorine synthase	IPR003480,IPR023213
0242600.1	NAD(P)-linked oxidoreductase	IPR023210
0242690.1	Glucose-1-phosphate adenylyltransferase	IPR011831,IPR029044
0245390.1	SCARECROW	IPR005202
0245950.1	Leukocyte immunoglobulin receptor A	None
0246010.1	Disease resistance	IPR002182,IPR027417, IPR032675

Table B.17.1: (continued)

EIv1 ID	Human Readable Description	InterPro domains
0248570.1	MLO	IPR004326
0248950.1	Ankyrin repeat , putative	IPR020683,IPR026961
0251930.1	FAD-binding Berberine	IPR012951,IPR016166
0253000.1	3-isopropylmalate dehydratase subunit	IPR001030

Candidates selected for further analysis are detailed in Table 6.2.1 and Figure 6.2.1. EIv1 identifier (MELAZ155640_EIv1_), truncated human readable annotation and InterPro domains are given.

B.18   Metabolite analysis of leaf material from *M. sexta* assay

Table B.18.1: **Metabolite analysis of *N. benthamiana*-leaf material from *M. sexta* assays.**

Sample (infiltration)	AZ	T1	T2	DI	ME
negative (empty vector)	0.00	0.00	0.00	0.00	0.00
azadirachtin (NeemAzalT/S)	0.25	0.00	0.00	0.00	0.00
negative (tHMGR)	0.00	0.00	0.00	0.00	0.00
tirucalla-7,24-dien-3β-ol (tHMGR, MaOSC1)	0.00	2.15	0.00	0.00	0.00
tirucalla-7,24-dien-3β,21-diol (tHMGR,MaOSC1, MaCYP71BQ5)	0.00	0.71	0.15	0.00	0.00
dihydroniloticin (tHMGR, MaOSC1, MaCYP71CD2)	0.00	0.23	0.00	0.25	0.00
melianol (tHMGR, MaOSC1, MaCYP71CD2, MaCYP71BQ5)	0.00	0.24	0.04	0.01	0.26

Mean relative values (n=3, mg/g DW) for all metabolites quantified from *M. sexta* assay.   Rows are labelled by metabolite of interest and combination of genes expressed within this leaf tissue.   Columns are labelled by metabolite being measured azadirachtin (AZA), tirucalla-7,24-dien-3β-ol (T1), tirucalla-7,24-dien-3β,21-diol (T2), dihydroniloticin (DI), melianol (ME).

C

Sequences

C.1 *M. azedarach* EIv1 candidate protein sequences

```
>MELAZ155640_EIv1_0061960.1_MaCYP88A108
MELNFLWLILAIFLGTYVVLFGFLRKVNDWYYVSRLGEKKKSLPPGDMGWPLLGNMLSFIRAFQSSDPDA
FVYNLVDRYGRTGVYKSHMFWSPSIVVTTPETCRRVLMDNEQFGRGNPESTKELLGKKTLALSNEEHKRL
RKLTTNPFRGDKALTMYVGYIEDIVVDLLDEWAGMKKPIVFLFEMRKLAFKVIGHIVFGTTSDHLLELME
QYYTDLLLGLRSPAINIPGFAFYRALKARKLLVKLLQDVLEERKKMVGIEQQKGNRGMIDLLIEAEDENG
KKLADENIIDLLIINLLAGHESSAHASMWAALYLYQHPEMLQKAKEEQEQILKRRPSTQKGLTLEEIKQM
DYLAKVIDETMRRTSLFIPIFREAKVDTDINGYTVPKGWQVLVWTRGVHMDPEVHPNPKEFDPSRWDNRA
KPGSYIPFGGGPWICPGADLTKLEIYIFLHYFLLYYKLELQNPDCPVAYLPVPRPSDNCIGKVIKVKNF
>MELAZ155640_EIv1_0015110.1_MaOGD5
MADHAAANGKESYVTAALNCIDLSNPDIHQSAALIKQACMDSGIFYVINHGISEELMAEVFAQTKKFFTL
PWEEKMKLRTGRNRGYDTRSKKIDSQTNKELNGEQFMMVRELPYSDKEAENPLTGPNVWPSADILPGWKE
TMLKYQQAVLNVGTAVARIIAVALDLKVDFFDQPAFLGNALPYVSLNHYGAEDVDPSKEYLLGTPAHCDP
PIITLLATDDVPGLQICKDKHAKAQIWQNVSPLKGAFIVNTGNTLERMSNCVFRSTMHRVLYRQDRYTIA
FLLYPSNDAVIECIPSCKSAENPPKFPPIKSDELLRVIANDIVAGRVTRDATDA
>MELAZ155640_EIv1_0015190.1_MaOGD1
MADHLTANGEEMGVYNLQCIDLSNPDIHESAALLRKACMESGIFYVINHGISQEFMDETFAQFKRFFDLP
IEEKMKLVWNKNLRGYRPPIQAVIDDKTKQKDFSEAFHIGVEVSEDDPNAHDFFYGPNLWPPADLLPGWR
KAMEKYHQEATNVARKIGRVIAVALDLNEDFFVQPELIGSANANYTNMFHYGVHGANLLKDVVGTTPHCD
LNLFTLLATDDIWGLQICREKNAESQVWEAIAPVKGAYIVNVGDMLEMLTNGIFRSILHRVVFNQERYST
GTFICPNHDYIIKCLPTCTSEENPPKYPTMRTGQYIYNRFNQLSASTVGKKA
>MELAZ155640_EIv1_0015210.1_MaOGD2
MAERIDGSLVLDCIDLSNPDIHQSASLMKKACMDSGIFYVTNHGIREEVMAEVFAQIKKLFSLPLKEKMK
LMIDNNRGYTVPKKNQFVDPETNLESQAEGYGIIQEVRDGETILEKPLGVPNVWPSNDTLPGWKEIMIKY
LQEVLDVGRAVSRIIAVALDLNVDFFDQPQFLGNALSYIVINHYKVGGVDPTKEPILGTPTHTDPCLITV
LASDGLPGLQICKDVNTKPRIWTNVAPLKGAFIVSVGDMLERMTNCVFKSPVHRVQHRQDRYSIPYFLFP
SRDAILECLPSCKSAENPAKFPPIESKELMSFIFGELRNGRVPVIDVTKY
>MELAZ155640_EIv1_0015210.1_MaOGD2.1
MAERIDGSLVLDCIDLSNPDIHQSASLMKKACMDSGIFYVTNHGIREEVMAEVFAQIKKLFSLPLKEKMK
LMIDNNRGYTVPKKNQFVDPETNLESQAEGYGIIQEVRDGETILEKPLGVPNVWPSNDTLPGWKEIMIKY
LQEVLDVGRAVSRVIAVALDLNVDFFDQPQFLGNALSYMNLNHYKVGGVDPTKEPILGTPTHTDPCLITI
LASDGLPGLQICKDVNTKPRIWTNVAPLKGAFIVSVGDMLERMTNCVFKSPVHRVQHRQDRYSIPYFLFP
```

SRDAILKCIPSCKSAENPAKFPPIESKEVLSIIFGELSNGRVPVIVEVKKY
>MELAZ155640_EIv1_0015210.1_MaOGD2.2
MAEQIEGSLVLDCIDLSNPDIHQSVSLIKQACMDSGIFYVINHGISEELMAEVFAQCKKFFSLPLKEKMK
LMIHKNRGYTVPKKNQFVDQETNLESQTEGYNIVQEVQNGEKIIEKPLGVPNIWPPNDTLPGWKELMLKY
VQEVLDVGRAVSRIIAVALDLNVDFFDQPQFLGNALSYIVINHYKVGGVDPTKEPILGTPTHTDPCLITV
LASDGLPGLQICKDVNTKPRIWTNVAPLKGAFIVSVGDMLERMTNCVFKSPVHRVQHRQDRYSIPYFLFP
SRDAILECLPSCKSAENPAKFPPIESKELMSFIFGELRNGRVPVIDVTKY
>MELAZ155640_EIv1_0015320.1_MaAKR1
MGAVPEVRISSCDKLMPIIGMGTGIYIQSKENDKEVMKAAFIEAIKVGYRHFDTSSDYPSEQPLGEVIME
ALRLGLIKSRDELFITTKLGPVYADKDLVLPAIKMSLRKLQLEYLDLYLIHVPIKLGEEAAVQWPNIAKD
MISPIDIKGVWEGMEECKNLGLTKAIGVSNFPRSMLEELLSVAKIPPAVNQVEMNPTWQQKKLREFCKEK
DIHVTAYSPLGAYKEMWGHKSIMESNVLEEIAKAKGKTKAQIALRWVYEQGASAVPKSLNKERMKENLQI
FDWCLTEKELNKINQLPQQKAVYFAKTIAAGHSFDFIDQLDAEL
>MELAZ155640_EIv1_0015350.1_MaAKR2
MEALHLGLIKSRDELFITTKLGPVFADKDLVLPAIKLSLRKLQLEYLDLYLIHVPIKLGEEAAVQWPNIA
KDMISPIDIKGVWEGMEECKNLGLTKAIGVSNFPRSKLEELLSVAKIPPAVNQVEMNPTWQQKKLREFCK
EKSIHVTAYSPLGAYKDVWGHKSIMESNVLEEIAKAKRKTKAQIALRWVYEQGVSIVPKSLNKDRMKENL
QIFDWCLTEEEVKRINQLPQQKAVYFAKTIAAGNSFDFIDQLDAEL
>MELAZ155640_EIv1_0015800.1_MaEst1
MDKRARMIRSGSKQQRLSLAELMVVLILSCWSGGVESQLVKGMFVFGDSLVDVGNNNFLYSFARSNYFPY
GIDFDSGPTGRFSNGKTFVDSLGRMLGLAYPPAFIDPCSMGPRLLGGVNYASAAAGILEETGRHYGERHS
LKQQVENFESTLMELRSIIDAKNLTYYLSKSMAVLVFGSNDYINNYLMPPDYSSSFTYTPRQFSKLLIGH
YAPQILALYNLGLRKFFICGIGPLGCIPNQRAKGQAAPGKCVDNVNKMVRMFNEELKSLVDLLNKRRLDS
GIFVYGNTFDAVMDIINKPEAFGFISADVGCCGMGKNKGEITCLPFGFPCLNRNKYVFWDAYHPTEAVND
ILAQRVVFGSTSYSYPMNVFNMTLH
>MELAZ155640_EIv1_0052990.1_MaCYP716-3
MELFLPSVLLILTVFCFYYLFKPKQTQNAPLPPGHVHWPLKIFETFDYMFKAKTNSIHKFIAERRNKYNT
KLFKTSHIGQNMVFLCSPEGNKFLFANDYKLVRSWWPVTFLRVFENAEEEITPEQVIRARKQFLSFFNEP
EALAKHVSITDEVVKDHVKLFWDGSDEVTVYPIARKLTFAISCRLLADIRDREILDELLPAMGDVVAAFF
ALPINLPGTKFNRAVKGSRKCRKIFVDIIKQRKIDLFEKGRKEANDVLSNILLENHRDGIEVNEVALAKN
LVSLLSAAFDNPSVTIVSIMKNLAENPEIYARVRSEQLEIAKGRAPGENLAMEDLKKMKFSMNVLSESLR
LEAPASGTFREALNDFTYEGYLIPKGWKVHWSTHATHRNPQYFKDPEKFDPARFERNDPIVPYSYVPFGG
GHHICPGKDFAKLQILIFIHHVVKKFNWEKVNPDEQMIRVPNLKAAKGLPVRLYPYNK
>MELAZ155640_EIv1_0061950.1_MaCYPU-1
MGSDLLWLILAIVVGTYVVLFGFLRRANEWYYSMKLGDKSRYLPPGDMGWPIIGNMIPYFKGFRSGEPES
FIFDLFEKYGRKGIYRNHIFGSPSIIVLAPEACRQVFLDDDNFKMGYPESTNKLTFRGSFNTASKEGQRR
IRKLATSPIRGHKAIAIYIDNIEDIVVKSMNEWASKDKPIEFLSEMRKATFKVIANIFLGSSSESVIGSV
EQYYVDYANGLISPLAINLPGFAFHKAMKARDMLGEILEPILRERRSMKEKDQLKGKRGLVDLLMEVEDE
NGEKLEDVDIVDMLIAFLSAGHESSAHIATWAIIHLHRHPEMLQKARKEQEEIVKKRPASQQGFSIEDFK
QMEYIAQVIDETLRITNLSSSSFREAEADVNLQGYIIPKGWKVLLYNRGVHRNPENYPNPKEFDPSRWDN
RANRPGYFIPFGGGPRICPGADLAKLEMSIFIHYFLLNYRLEPLNPECPTEYLPVPRPSDQCLARIVKLK
>MELAZ155640_EIv1_0078490.1_MaAD1
MENKSASPTTGKVIRCKAAVARNPGKPLVIEEIEVDPPKAWEVRIKILCTSLCRSDVTFWKSEPPIAIFP
RIFGHEAVGVVESVGEYVEEVKEGDLVLPVCKRQCGECRDCKSTKSNLCSKFGKDYRPNMLRDGTSRFRD
LKGEVIHNFIYVSSFTEYTVVDVTHVVKITADIPVDKACLLSCGVSTGVGAAWRVAGVEAGSTVAIFGLG
AVGLAVAEGARLRGASKIIGVDLNAEKFDLGKAFGITDFINPTNCGNKTVSQVIKEMTDGGADYCFECIG

LASAMEDAFSSSREGWGKTVILGAEMHGSPVSLKSFEILSGRTICGSLFGGMKPKSDIPILAQKYLDKEL
HLDEFITHEVSFQDINKAFDLLLEGKSLRTIIWMDNDKLN
>MELAZ155640_EIv1_0095290.1_MaEH1
MVSCHLPSTLISFCLIVLFLAFLANKQIQHRKPEIFKPEGNVILAHEEISNTEEKEIQHRTVYTNGIKMH
VAEKGCEGPVIIFIHGFPELWYSWRHQITALASLGYRAVAPDLRGYGDTDAPPEVRSYTCLDVVGDLIGL
IDAVAPNQEKVYVVGHDWGALIAWYLCLFRPDRVKALVNMSVVFNPRNPKRKPLENLKAFYGEDYYICRF
QEPGEIEAEFEQIGAETVLKEFFTYRNPGPLYLPKGKGFGHPPDAPIALPSWLSEEDIKYYTTKFEQKGF
TGGINYYRNIDLNWELTAAWTGVEIKVPVKFIVGDLDLTYNAPGTKEYIHNGGFKRDVPFLDEVVVMEGV
GHFINEEKPDEVNKHIFDFFQKHF
>MELAZ155640_EIv1_0120380.1_MaDT1
MASSGSDVAKEFRFFRVYKDGRVELFRPHWDKVPPSNHHTTGVRSKDVFISSEPPVSARVFIPKLTDPNQ
KIRLLLYVHGGGFSMQSAFAPYYHNFCSRVSAQANAIVVSVEYGLFPDRPIPACYDDSWVALKWAASHVS
RNGPDPWLNDHADFGKVFIGGDSAGGNISHTLAFRVGSTGLPGVKVVGVILLHPYFGGTEDDEMWLYMCP
NNGGLQDPRMNPPVEDLAKLGCERVLIFVAEKDHLNAVGKNYYEKLKRSGWKGTVELVENDGEDHCFLFQ
NLEYEKAVELTNKFVSFIKRES
>MELAZ155640_EIv1_0120440.1_MaHyd1
MDMDSSHNEIIHDFPPFFKIYKGGCVERYRKNEIADAGYDPSTGVESKDVVISPETGVRARIFIPKIKGP
DQKLPLLVHYHGGGFCIRSAFDVKTKSFLTSLVSKANIIAVSIDYRLAPEHLLPIAYHDSFAGLQWVATH
SSGLGPDPWLNAHADFGRVFLEGESAGGTIAHDVEIQAGITELAGLKIEGLLIVHSFFGSKKPHALYKFI
CPGSSGCDDDPRLNPEVDPNLKKMAGERVLACVAEKDWFIDRGMAYYETLAKYGWDGNVELWETLGEGHC
FHLFNPDSEKTAQLIEKMVHFINRD
>MELAZ155640_EIv1_0120480.1_MaHyd2
MDSTKIENEISHDFPPFFKVHKDGRVERYPFLESLSKCTPPGLDPKTAVQSKDVVISPKTGVKARIFVPK
VLDGQGQKLPLLVHYHGGGFCLGSPLDTITRDFLTYFVTQANIIAISIDYRLAPEHPLPIAYDDSWAGLQ
WVASHSNGTGPEPWLNEHADFGRVFLAGESAGANIAHYVAVQAGASRLAGLKIIGVLIVHPYFGCEKPDE
LYKYMCPTSSGCDDDPKLNPGVDPNLKSMGCSRVLVCVAEKDWLKNRGIGYYETLAKSEWKGKVEFYESL
GEDHCFHLIDSNGEKAAPLMKRMVDFIKQE
>MELAZ155640_EIv1_0128570.1_MaCYP735-1
MAMVVLATLFAIFISLLFRVAYDSISCYWLTPRRIKKIMNKQGVRGPKPRGLTGNILEMASLISQSTSKD
MGHIHHDIVGRLLPHYATWSEQYGKRFIYWHGTEPRMCISETELIKELLTKYSTKAGKSWLQQQGSKHFI
GKGLLMANGNDWYHQRHIVAPAFMGDRLKSYAGHMVECTKKMLQSLDNAVESGQTEFEIGEYMTRLTADI
ISRTEFDSSYEKGKQIFQLLTVLQHLCAQASRHLCFPGSRFFPSKYNREIKSLKMEVERLLMEIIQSRKD
CVEIGRSSSYGNDLLGMLLNEMQNNKRGNGFSLNLQLIMDECKTFFFAGHETTALLLTWTVMLLASNPSW
QEKIREEVKQVCNGETPSVDHLPKLTQLNMVINESLRLYPPATVLPRMAFEDIKLGDLFIPKGLSIWIPM
LAIHHSEELWGKDANEFKPDRFASRTFAPGRHFIPFAAGPRNCIGQSFAMMEAKIILAMLISRFSFNISD
NYRHAPVIVLTVKPKYGVQVYLKPINP
>MELAZ155640_EIv1_0135650.1_MaEst2
MSEESAPPHSTIDPYKFLQIIPNADGTLTRDCSNYPSSVATPDPDDHTLLVLSKDVHVNQLNNTWVRMFL
PRQALDSSFTTKLPLVVHFHGGGFIILSAATKVCHDFCSTIGAEIPAVVVSVNYRLAPEHRLPAAYNDAM
EVLHWIKTAHDEWIQKYVDFSCCFLTGESAGGNIAYHTGLRVAAEVDNLLPLKIKGFILNAPYFGGVNKT
ESELRLGTNPLSPLCANNLMWELSLPIGADRDHEYCNPTVGGGSKLLDKFKLLGWKVLVIGGYEDPLIDR
QIEFSKLIEQKGVKVVNRFDVGGHGSPKDIDKWIKDVVASTKSFISSSEMATGV
>MELAZ155640_EIv1_0141190.1_MaOR1
MFSETLKQSSHTHTPTEGDQKHKHKELDHEIRNMVNAITDHITDHHKPGSSLHHHEDDDDDRGVRIITLA
GTNSGATMRSELDDKQNPHGVSLGEPEPLSTYVNSNFQAVNNSIMLGSSYNTNDPGVHLDITDVFEQLGQ
KPEKQHAKKGKKKDKESFKSDQHSEKSD

>MELAZ155640_EIv1_0142070.1_MaBAHD1
MNLRITSSEIIKPSSPTPQNLKSYRLSIVDQLTPNVYFSIILLYTKTTENPTKTSDHLKKSLSETLTRYY
PLAGQLKYDQLIVDCNDQGVPFVEADVSNHMSELLKLPNIDVLEQLLPFKPHEGFNAERSNVTVQVNYFG
CGGMAIGLCFKHKVLDATTAAFFVKNWGVIARGAGEIKDVIYDQASLFPARDLSFLSKSVDEEFLKAESE
TKRFVFDGSAIASMREKFTHLGRRPTRFEVVSAVILGALISAAKESEEPPERLDTIISVNLRQRMVPPFP
EHCLGNIISGGLIYWPLEKKLDHGSLAEEIHQSIKKVDDQFARKFYGEAEFLNLPRLGANEVVKKREFWV
TSWCKTPLHQSDFGWGKPKWAGNSMRLNEITVLFDTSDGEGIEAWVGLPKKDMARFEQDATIVAYTSPNP
TIL
>MELAZ155640_EIv1_0148250.1_MaTF2
MEVEQVLRMKGGEGDNSYANNSFQQKAAILNSKPLLQQTMVEFYCNSFPNCIRFADMGCSSGPNSLLPTW
EAIDSLRKICRSLNRKPPIFQVFLNDLPGNDFNTVVKSLPSFYDRLKTEKGDEFGSCFVSATPGSFYGRL
FSPCFLHFVFSSYSLQWLSQVPDGLVGESGIPLNKDNIYPSKTSPESVHKAYSDQFQKDFTSFLRFRSEE
LKPGGRMILTLRGNQTKEESSTNYRPIFVEIIGTILHDMVSEGLVDESKLETFNMPMYAPRVEEVRRVIE
TEGSHNIDKLESYELRWDAGLENGGDSLEEKHLRGKHLSKLIRAVLEALLESHFGNAILDDLFQRFCIKA
TEHMEMGLGIQTNVIISLTKK
>MELAZ155640_EIv1_0164450.1_MaBAHD3
MELKIVSSEIIKPSSPTPQHLRTHKLSVLDQVAGDSLFPVILFYPQACSNDTTKTSDHLKKSLSETLSKY
HPFAGRFKDAFSIDCDDSGAVFVEACVAGEMSEILKQPKNDLLEELMPYKLNEKPSVPVNLAAKVTYFGC
GGMALCVCFSHVIADITTAANFIKSWVAISRGFSTSNDIPDTFFDCSTVFPPQDFSSFSRNAFADNHSQS
AEIITKRFTFDGVKLADLKEKIRKESSSGYQPTRVEAVSAVILGGIMSAEKEGEKFNHTNLVPSVSVNLR
NRTNPPLPELSIGNIIQAAITKLPIEKRINYSKLTEQLHGSIRSIDDEYLKKFHAVGEFMNNMQNALIAL
FDPNNRGFTISSWCRRPLYEADFGWGKPIWVSTAMKHKDVAVLLDSNDGKGIEALVALPKKEMDKFEQDP
GILAYASVDTSII
>MELAZ155640_EIv1_0170970.1_MaCYP714-1
MCTSLTLGKPTYLSKDRGPLLGQGILASSGPLWAHQRKVIAPQLYPERVKGMLKLMVDSTTSMVKSWESI
IESEGGIAEIRVDKYMQSLAADVISRACFGSNYSRGQEIFLKFESLVQLMSKGLVGVPGSRYLPTKNNRE
IWKLDKEISSMILQVVQERSQSSSHHKDLLQMVLEGAKTCGGYNGVGLNISGDKYIVDNCKNIYFAGHET
SAVTSSWALMLLAVHTEWQDRVRAEILRVCKDGCPDSDMLRSMKTLTMVIQETLRLYPPAGFVFREALDD
IKFRDIVIPKGMGIQVPIPILQQNPEVWGPDAHEFNPERFANGILAACKFPHAYMPFGAGPRACAGQHFA
MTELKVMLSLVLSKFSFSLSQKYQHCPVLRVNLIPEKGIILHVKRI
>MELAZ155640_EIv1_0192980.1_MaIsom1
MSDSSSVPVDFVLNFSTAALHAWNGLSLFLIVFISWFISGLTQAKTKMDRVVLCWWALTGLIHVFQEGYY
VFTPDLFKDDSPNFMAEIWKEYSKGDSRYATRHTSVLTIESMASVVLGPLSLLAAYALAKAKSYNYILQF
GVSIAQLYGACLYFLSAFLEGDNFASSPYFYWAYYVGQSSIWVIVPALIAIRCWKKINAICYLQDKKNKT
KVR
>MELAZ155640_EIv1_0192980.1.1_MaIsom1-I
MSDSSSVPVDFVLNFSTAALHAWNGLSLFLIVFISWFISGMSAY*SIKYTSYIILPQSYVVYCLTCISYI
YIYIRLHELIIACRVDTGENKNGQSGIMLVGSHWPYSCLSRGLLCFHSRFI*RRFS*FYG*NL*VQYTHM
CVYIHIYI*FTIFII*RNGIYIN*T*TCRERIQQR*FKICNKTHFSSYHRIDGFSCSGTS*PSSSVCFS*
SEVIQLHSSVWSLNCAAVWGLSIFPKCFPGGG*FCFFSVFLLGILRWTK*HLGYSTSTHSYTLLEKNQCY
LLSSRQEEQDQSSL
>MELAZ155640_EIv1_0192980.1.1_MaIsom1-I_ORF1
MSDSSSVPVDFVLNFSTAALHAWNGLSLFLIVFISWFISGMSAY*
>MELAZ155640_EIv1_0192980.1.1_MaIsom1-I_ORF2
MDRVVLCWWALTGLIHVFQEGYYVFTPDLFKDDSPNFMAEICKYNIHICVYIYIYIYDSQYLLSKEMGYI*
>MELAZ155640_EIv1_0192980.1.1_MaIsom1-I_ORF3

MSFKRVIMFSLQIYLKTILLILWLKFVSTIYTYVCIYTYIYMIHNIYYLKKWDIYKLNINLQGKNTAKVIQ
DMQQDTLQFLPSNRWLQLFWDLLAF*
>MELAZ155640_EIv1_0192980.1.1_MaIsom1-I_ORF4
MASVVLGPLSLLAAYALAKAKSYNYILQFGVSIAQLYGACLYFLSAFLEGDNFASSPYFYWAYYVGQSSIW
VIVPALIAIRCWKKINAICYLQDKKNKTKVR*
>MELAZ155640_EIv1_0198190.1_MaSDR1
MNSYSSAAPGKRLEGKVAIITGGASGIGATAVQIFHDNGAKVVISDVQDKLGQALADKLGEGVSYIHCDI
SNENDVINLVDTTVAKYGKLDIMYNNAGVIDRNFGSILDTPKSDLERLLSVNTIGGFLGAKHAARVMVPK
QKGCILFTASACTEIAGLGSPAYTVSKYGVVALVKSLAAELGQYGIRVNCVSPYGLATGMSTAGVDPALI
ESSLSEMGNLKGQVLKTDGIANAALYLACDEASYVSGQNLVVDGGFSILNPTIMKAYNLIN
>MELAZ155640_EIv1_0198240.1_MaSDR2
MATSSILSAVTRRLEGKVAVITGGASGIGACTAKVFAKQGAKVVIADIQDELGHSVVESIGKSNSSYVHC
DITNESHVKNVIDQAVATHGKLDIMFNNAGIGGENKPRIVDNGKDNFERVLKINVTGVFLGIKHAARVMI
PARSGSIISNASAASGISGVAPHAYTCSKHAVVGLTKNTAVELGQFGIRVNCLSPYAVATPATNEYSGMG
DEGFENFVNSIANLKGVTLKPEDVANSALYLASDEAKYVSGHNLFVDGGFTIVNPSFGMFKYTDS
>MELAZ155640_EIv1_0217180.1_MaCYPU-2
MEFFFLCGLSLFFLFISLSLLPFFFNHKSNASLPPPNHVWREHLRFPNWHVNIQRATPVEGPNGWKLYWS
TNSRHKFPGCFPEAAQNLDPSRFQGNCGKEYARLEMRVLMQNTLRISKYVDQTN
>MELAZ155640_EIv1_0223780.1_MaAKR3
MASNTTKFSMFNIPEVKLNHSSGHRKMPVIGLGSAVPKLDEDAVKFAVLEAIKLGYRHFDTAPIYRTEKA
LGEAIAEALRLDLLSSRKELFITSKLWCNDAHPDLVVPTLKKSLRTLQLEYLDLYLIHWPISATPGDIVI
PIPTENLLPMDYKGVWEAMEECQRLGLTKSIGVSNFSSKKIQTILSFATIPPSVNQVEMHPVWQQRKLRE
FCKARGIIVTAYSPLGAVGTIWGSDEVMENEVLKQIAKAHGKTVAQVSLRWIVEQGATVVVKSFNRERQK
QNLGIFDWALTDDDYNKINQIHQHKVGPSYIFLSPHGPFKTLEELWDE
>MELAZ155640_EIv1_0235630.1_MaBAHD2
MEPEIISRESIKPSSATPLHLKTHKLCLLDQYRHHAYFPIVLYYPFNQEPNISDPTQINHIVSERLQLLK
QSLPETLSRFYPFAGKIKDNLSVDCNDEGIYFVEARVKSSLSDYFNQPNFANANYKFIPFDVKELSGSIS
GLHVAKIQVTTFAYGGLVICACLSHLFGDGITLNSFLKSWVATACKNAEEAERPNNDASSLFPQQEIYPK
EATWTEMCKPFYRDGRFVSRRFLFDAKAIANLKDKVASSLVQNPSRVEAVSALLSRCIMTAFKGKFGSHR
PILLTHTVNMRRKAKPLMPEYSMGNIVWTANALCTNEEPELDGLVGKLREAIMEINGDFLKSLQGDEGFL
NLCEAAKNESALCSSAVERITFSSWCNFGLVDIDFGWGKPIWVSTIGIDGPVPCFSNTIILMDTRLKGEI
EAWVYLLEEDMNILELDKELIAFAKMDPSPM
>MELAZ155640_EIv1_0238770.1_MaBAHD4
MEVQIISKEDIKPSSPTPQHLTTFKLSFLDQFIPTPYAPVILFYPANDSSSNLDHINKKLALLKNSLSET
LALFYPLGGKFSDDLFIDCNDEGAYFVETRVNCFLDELLRQPDLLFLHRLLPRDFFPKESSEATYVSNIQ
VNVFQCGGIAIGVCISHKILDGAALSSFLKAWAAAARGDEEMICPNFAATSLFPANDLWLRDSSMVMWG
SMYKKGKSITKRFVFDASAIAEIKARAISSRVTVPTRVEAVSAFIWKCFMAVSKAKHGSQRPSLLTHLVN
LRRRMSPPLSKNSVGNFLWIAAAKCVSMQKVELPYLVNELRDAISKLDSDFVKKLRSDVGNSVMCESLKE
INEICSKDEVDYIGFTSWCKFGFYETDFGWGKPIWVSGFGLNCSPVMNLIILVDTRFGDGIEAWTTLEEE
EMAMLERNAEFIKFASMDPSPLLIGNPNPVS
>MELAZ155640_EIv1_0242600.1_MaOR1
MATVKRIKLGSQGLEVSAQGLGCMGMSAFYGAPKPEPEMIELIHHAIDSGITFLDTSDIYGPLTNEILLG
KALKGGYRERAEVATKFGINYADPNGEFRGDPAYVRASCEGSLKRLDVDCIDLYYQHRIDTRVPIEITMG
ELKKLVEEGKIKYIGLSEASASTIRRAHAVHPITAVQLEWSLWARDVEEEIIPTCRELGIGIVAYSPLGR
GFLSLGKKLTESFTKEDFRQHLPRFQPENLEHNLKLFERVKEIAARKGCTPSQLALAWVHHQGDDVCPIP
GTTKIANLNENIKALSVKLTPEEMGELESIASAEAVKGDRYPSVLFTYENADTPPLSSWKPS