Ludwig-Maximilians-Universität München
Fakultat Biologie
Bereich Genetik

Orthologous gene networks in defensive diterpenoid production

Reshi Shanmuganathan

München 2024

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades „Doctor rerum naturalium" (Dr. rer. nat.) an der Fakultät für Biologie der Ludwig-Maximilians-Universität München

Erster Gutachter: Prof. Dr. Claude Becker
Biozentrum der Ludwig-Maximilians-Universität München
Bereich Genetik

Zweiter Gutachter: Prof. Dr Klaus Mayer
Helmholtz Zentrum München – Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)

Datum der Abgabe: 25.09.2024
Datum der mündlichen Prüfung: 04.12.2024

Erklärung
Declaration

Hiermit erkläre ich, *
Hereby I declare

☒ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.
that this work, complete or in parts, has not yet been submitted to another examination institution

☒ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.
*that I did **not** undergo another doctoral examination without success*

☐ dass ich mich mit Erfolg der Doktorprüfung im Hauptfach
that I successfully completed a doctoral examination in the main subject

und in den Nebenfächern
and in the minor subjects

bei der Fakultät für
at the faculty of _____

der
at _____
(Hochschule/*University*)
unterzogen habe.

☐ dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.
that I submitted a thesis or did undergo a doctoral examination without success

München, 19.09.2024
Ort, Datum/*place, date*

Reshi Shanmuganathan
Unterschrift/*signature*

*) Nichtzutreffendes streichen/
delete where not applicable

Eigenständigkeitserklärung

Hiermit versichere ich an Eides statt, dass die vorliegende Dissertation mit dem Titel

Orthologous gene networks in defensive diterpenoid production

von mir selbstständig verfasst wurde und dass keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, wurden in jedem Fall unter Angabe der Quellen (einschließlich des World Wide Web und anderer elektronischer Text- und Datensammlungen) kenntlich gemacht. Weiterhin wurden alle Teile der Arbeit, die mit Hilfe von Werkzeugen der künstlichen Intelligenz de novo generiert wurden, durch Fußnote/Anmerkung an den entsprechenden Stellen kenntlich gemacht und die verwendeten Werkzeuge der künstlichen Intelligenz gelistet. Die genutzten Prompts befinden sich im Anhang. Diese Erklärung gilt für alle in der Arbeit enthaltenen Texte, Graphiken, Zeichnungen, Kartenskizzen und bildliche Darstellungen.

München, 19.09.2024

(Ort / Datum)

Reshi Shanmuganathan

(Vor und Nachname in Druckbuchstaben)

(Unterschrift)

Affidavit

Herewith I certify under oath that I wrote the accompanying Dissertation myself.

Title: Orthologous gene networks in defensive diterpenoid production

In the thesis no other sources and aids have been used than those indicated. The passages of the thesis that are taken in wording or meaning from other sources have been marked with an indication of the sources (including the World Wide Web and other electronic text and data collections). Furthermore, all parts of the thesis that were de novo generated with the help of artificial intelligence tools were identified by footnotes/annotations at the appropriate places and the artificial intelligence tools used were listed. The prompts used were listed in the appendix. This statement applies to all text, graphics, drawings, sketch maps, and pictorial representations contained in the Work.

München, 19.09.2024

(Location/date)

Reshi Shanmuganathan

(First and last name in block letters)

(Signature)

Table of Contents

I. Abbreviations and Definitions

AAP	Amino Acid Permeases	MEP	2-C-Methyl-D-Erythritol-4-Phosphate
BAS	B-Amyrin Synthase	MS	Murashige Skoog
BGC	Biosynthesis Gene Clusters	MVA	Mevalonate
bHLH	Basic Helic-Loop-Helix	NB-ARC	Nucleotide-Binding Adaptor Shared By Apaf-1, R Proteins, And Ced-4
BXs	Benzoxazinoids	NLR	Nucleotide-Binding Oligomerisation Domain-Like Receptor
CAS	Cycloartenol Synthase	OETGN	Orthologous Edge Terpenoid Gene Networks
CPQ	Cucurbitadienol Synthase	OSC	Oxidosqualene Cyclase
CPS	Copalyl Synthase	OTGN	Orthologous Terpenoid Gene Networks
CRRSP	Cysteine Rich Secreted Proteins	PCA	Principle Coordinate Analysis
CYP450	Cytochrome P450	PDR	Pleiotropic Drug Receptor
DMAPP	Dimethylallyl Diphopshate	PPPS	Polyprenyl Diphosphate Synthase
eccDNA	Extra Chromosomal Dna	PR	Pathogenesis Related
EMS	Ethyl Methanosulfate	PRL	Primary Root Length
FIMO	Find Individual Motif Occurrences	RAPDB	Rice Annotation Project Database
FPP	Farnesyl Diphosphate	SDR	Short Chain Dehydrogenase
FPPS	Farnesyl Diphosphate Synthase	SDRLK	S-Domain Receptor Like Kinase
GA	Gibberellic Acid	SPPS	Solanesyl Diphosphate Synthase
GGPP	Geranyl Geranyl Diphosphate	SPS	Solanesyl Diphosphate Synthase
GPP	Geranyl Diphosphate	SSD	Single Seed Descent
HEM	Homozygous EMS Mutants	TF	Transcription Factor

HPPD	4-Hydroxyphenylpyruvate Dioxygenase	TOM	Topological Overlap Matrix
IPI	Isopentenyl Diphosphate Isomerase	TPS	Terpene Synthase
KS	Kaurene Synthase	WAK	Wall Associated Kinase
LRK	Lectin Receptor Kinase	WGCNA	Weight Gene Correlation Network Analysis
LRR	Leucine Rich Repeat	WLE	W-Box Like Element
MATE	Multidrug And Toxic Compound Extrusion	Zx	Zealexin

II. Abstract

As sessile organisms, plants respond to biotic stresses by producing a variety of defensive metabolites, such as complex terpenoids catalysed in modular biosynthesis steps. Terpenoid biosynthesis is mediated by orthologous terpene synthases, resulting in more than 40 defensive diterpenoids and sesquiterpenoids having been reported in rice and maize. While much is known about the similarities between these two species in producing such compounds, which ancillary genes govern the transcriptional regulation, transport, and deployment of defensive terpenoids in plant immune responses are less well understood. Here, I set out to identify novel genetic components that drive the terpenoid biosynthesis machinery. Recent bioinformatic approaches have utilised gene co-expression networks to identify ancillary genes which drive the production of defensive metabolites in plants. Therefore, using publicly available transcriptomes, I built gene co-expression networks in both rice and maize, extracted the terpenoid gene networks, and identified orthologous genes present in the two networks. This yielded a number of candidate ancillary genes, which I further investigated via bioinformatic approaches, and to which I assigned putative functions within the terpenoid biosynthesis machinery. Furthermore, analysis of the known terpenoid biosynthesis enzymes shed light on specific families of enzymes which had undergone positive selection. Lastly, I investigated the mode of action of the rice potent diterpenoid allelochemical, momilactone B, by performing forward genetic screens. Such screens are powerful tools in identifying mechanisms of resistance to herbicides and allelochemicals, allowing for combating the incessant increase in herbicide resistance, a major global agronomic problem.

1. Introduction

1.1 Defensive specialised metabolites in grasses

As sessile organisms, plants have evolved various methods of responding to attacks by pathogens, herbivores, or parasites, commonly grouped under the term 'biotic stress'. One critical component is the production of specialised defensive metabolites against these stresses. The biosynthesis of such defensive metabolites often branches out from primary metabolism and involves a wide array of enzymes, leading to the production of highly diverse defensive molecules. Such compounds are a key research focus especially in agronomically relevant crops such as rice, maize and wheat, with the goal of boosting agricultural yields by improving endogenous defensive abilities.

One of the most researched classes of compounds in crop grasses are benzoxazinoids (BXs). BXs are defensive indole-derived compounds found in many grasses, including major crops such as rye, maize and wheat (Bakera et al., 2015; Makowska et al., 2015; Nomura et al., 2002; Nomura et al., 2003; Rakoczy-Trojanowska et al., 2017). BXs have been reported to mediate a wide variety of roles in plants. Firstly, they attract beneficial bacterial species such as Pseudomonas putida (Neal et al., 2012; Neal & Ton, 2013), which prime plant defence in maize. Second, they have growth-inhibitory properties: APO, one of the final conversion products derived from BXs once they enter the soil, is a potent antibiotic (Fritz & Braun, 2006). APO also inhibits the growth of roots of some non-BX-producing species (Venturelli et al., 2015), thereby acting as an allelochemical. Lastly, BXs also act as siderophores, binding extracellular iron and forming complexes which are taken up by rice and oat roots (Hu et al., 2021). In sum, BXs play diverse roles in plant defence and act against a wide variety of organisms.

Another major class of defensive compounds in grasses are terpenes, well known for their incredible structural diversity. Terpenes are divided into subtypes, categorised by the number of carbons: monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), triterpenes (C_{30}) and sesterpenes (C_{25}) represent the best -characterised subtypes (Rudolf & Chang, 2020). Over the last few decades, increasing interest in plant defence has led to the identification of such terpenes in key crop species such as rice, maize and wheat, as well as in other grasses. The availability of transcriptomic and genomic data has led to the identification of a number of unique biosynthesis gene clusters (BGCs) underlying the production of these terpenes.

1.1.1 Biosynthesis of terpenes in grasses

Terpenoid production in grasses, as in all plants, begins with the condensation of two 5-carbon compounds; isopentenyl diphosphate (IPP) and dimethylallyl diphopshate (DMAPP). Both of these compounds are generated via two cellular pathways, the plastidic 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway and the cytosolic mevalonate (MVA) pathway. Prenyltransferases catalyse the condensation of IPPs and DMAPPs into precursors such as C_{20} geranyl geranyl diphosphate (GGPP), C_{10} geranyl diphosphate (GPP), and C_{20} farnesyl diphosphate (FPP) (Zhou & Pichersky, 2020). These three compounds are then used to produce diterpenes, monoterpenes and sesquiterpenes & triterpenes respectively (Jia & Chen, 2016), which were the main focus of my thesis. There are also prenyltransferases such as solanesyl diphosphate synthases (SPPS) and polyprenyl diphosphate synthases (PPPS) which catalyse the production of solanesyl diphosphate (SPP) and polyprenyl diphosphate, respectively, the precursors to (Jia & Chen, 2016) solanesol and polyprenyls such as plastoquinones and ubiquinones (Liu et al., 2019). Lastly, non-canonical terpene synthases (TPS) have been reported in dicots to catalyse

terpene biosynthesis (Rivera et al., 2001), further highlighting that terpene metabolism is highly complex and flexible. Canonical terpene biosynthesis in cereals has been well understood, and some examples are elaborated upon in this section.

1.1.1a Triterpene biosynthesis in cereals

Sesquiterpenes and triterpenes are both catalysed within the cytosol by cytosolic sesquiterpene and triterpene synthases respectively. Triterpene synthases in plants are also known as squalene synthases (SQS), as squalenes are the products of SQS acting on FPP (Phillips et al., 2006). Squalenes are then oxidised into 2,3-oxidosqualenes, which are cyclicised by oxidosqualene cyclases (OSCs) into i) cycloartenols by cycloartenol synthases (CAS) (Phillips et al., 2006) ii) cucurbitadienol by cucurbitadienol synthases (CPQ) (Shibuya et al., 2004) iii) β-amyrins by β-amyrin synthases (BAS) (Abe et al., 2004). Respectively, each of these products are then utilised in three different biosynthesis pathways to produce i) stigmasterols and brassinosteroids, ii) cucurbitacins, and iii) triterpene glycosides such as avenacins, potent anti-microbial metabolites found in the roots of oat plants (Thimmappa et al., 2014). These three biosynthesis pathways utilise varying cytochrome P450 enzymes (CYP450s), sugar transferases, oxidases and glycosyltransferases to generate end products such as avenacin (Leveau et al., 2019; Yan Li et al., 2021; Louveau et al., 2018; Mylona et al., 2008; Orme et al., 2019; Owatworakit et al., 2013).

1.1.1b Sesquiterpene biosynthesis in cereals

The other main class of cytosolic terpenes considered within the scope of this project, are the sesquiterpenes. Sesquiterpenes can be acyclic, such as (E)-β-farnesene, or cyclic, such as β-bisabolene. Sesquiterpene synthases catalyse the production of acyclic, monocyclic, bicyclic (Block et al., 2019), tricyclic (Garcia et al., 2019), and tetracyclic sesquiterpenes. Most sesquiterpene synthases, or at least those characterised in grasses so far, catalyse the production of multiple products at differing ratios. For example, ZmTPS1 converts FPP into (E)-β-farnesene as the major product, and (E,E)-farnesol and (3R)-(E)-nerolidol as minor products (Schnee et al., 2002). This diversity arises from the many conformations in which FPP can be condensed, generating diverse end-products or intermediates upon which cytosolic enzymes such as CYP450s can act to further amplify terpenoid diversity. Such metabolic diversity in grasses has been more extensively studied in grasses recently, with switchgrass (Muchlinski et al., 2019), maize (Block et al., 2019), sorghum (Zhuang et al., 2012) and centipedegrass (Lee et al., 2019) all having been shown to produce a variety of volatile and non-volatile sesquiterpenes in leaves and roots.

1.1.1c Monoterpene biosynthesis in cereals

The other major class of volatile terpenes in grasses are the monoterpenes, produced by plastidic monoterpene synthases. Effectively, they catalyse very similar reactions to sesquiterpene synthases. In fact, in vitro characterisations of sesquiterpene and monoterpene synthases show affinities for both GPP and FPP, albeit to differing degrees (Schnee et al., 2002). Monoterpene synthases can also produce acyclic monoterpenes such as geraniol, or cyclic monoterpenes such as limonenes (Block et al., 2019). Similar to sesquiterpene synthases, monoterpene synthases can produce major and minor products. A perfect example would be TPS6 and TPS11 in maize, both of which produce acyclic monoterpenes (linalool and β-myrcene) as major products and cyclic monoterpenes (limonene) as minor products (Block et al., 2019; Huffaker et al., 2011). When either of these enzymes were incubated with FPP as a precursor, instead of GPP, both enzymes produced the monocyclic sesquiterpene, β-bisabolene. As low molecular weight compounds, most reported monoterpenes and sesquiterpenes in grasses have been shown to be volatile compounds emitted upon herbivory by insects such as the fall armyworm (Block et al., 2019; Muchlinski et al., 2019; Zhuang et al., 2012). However, there are also non-volatile sesquiterpenes and monoterpenes, which will be elaborated upon in subsequent sections.

1.1.1d Diterpene biosynthesis in cereals

Diterpenes are the final type of terpenes within the scope of this project and its primary focus. Diterpene biosynthesis in dicots has been shown to involve singular bifunctional terpene synthases which can catalyse two cyclisation events of GGPP into diterpene scaffolds. However, within the Poaceae family (i.e., grasses), these C_{20} compounds are formed from plastidic GGPP by plastidic terpene synthases such as copalyl synthases (CPS) and kaurene synthases (KS). CPS are known as Type II terpene synthases which catalyse the initial cyclisation of GGPP into a prenyl diphosphate intermediate. Subsequently, Type I terpene synthases such as KS catalyse a secondary cyclisation of these intermediates, leading to the production of diterpene scaffolds such as syn- / ent-pimaradienes (Christianson, 2017; Zhou & Pichersky, 2020). Type I TPS contain a DDxx(D,E) motif at their C-terminus, while Type II TPS contain a DxDD motif at their N-terminus. The aforementioned triterpene, sesquiterpene and monoterpene synthases are all Type I TPS, similar to KS. Once the various scaffolds such as kaurenes, pimaradienes, sandropimaradienes, etc. are formed, an array of CYP450s, belonging to various families, as well as cytosolic dehydrogenases and 2-oxoglutarate-dependent dioxygenases (2ODDs), amongst other enzymes, perform additional modifications (Bathe & Tissier, 2019). The result of such multi-factorial biosynthesis pathways is a variety of diterpenes which can be secreted into the environment or accumulate in various tissues, acting as anti-fungal compounds as well as allelochemicals (Pelot et al., 2018; Peters, 2006; Wyatt et al., 2024).

1.1.2 Examples of potent defensive terpenes in grasses

To emphasise the functions of the various triterpenes, sesquiterpenes, monoterpenes and diterpenes thus far characterised in grasses, I discuss notable examples of each terpene type in the Poaceae in this section.

1.1.2a Avenacins: a classic anti-fungal triterpene in oat

Avenacins are a class of anti-microbial triterpenes, first discovered in the roots of oat plants (Turner, 1953). Subsequently, follow-up efforts isolated avenacin A1, the major product, and showed that it was highly toxic to a variety of fungal pathogens (Maizel et al., 1964). Avenacins have been shown to act by permeabilising fungal membranes in a sterol-dependent manner, leading to pore formation, membrane destabilisation, and accumulation of sterol-saponin complexes in the membrane (Armah et al., 1999). Recent efforts in the assembly of the oat genome also led to the full elucidation of the avenacin biosynthesis pathway (Yan Li et al., 2021), with heterologous expression in tobacco producing avenacins. The same group also showed that two more oat species, *A. eriantha* and *A. atlantica*, both contain BGCs similar to that of the avenacin BGC in *A. strigosa*. However, while *A. atlantica* produces avenacins in roots, *A. eriantha* produces avenacins in leaves, presumably against foliar pathogens (Yan Li et al., 2021). The same class of defensive triterpenes are produced in closely related oat species, utilising similar genes located in syntenic genomic regions, yet in different tissues.

1.1.2b Volatile monoterpenes and sesquiterpenes act as chemoattractants to predators of herbivorous insects

Recent efforts to understand the defensive terpenes produced by switchgrass, Panicum virgatum, an important biofuel crop, found that production of volatile sesquiterpenes and monoterpenes was induced by fall armyworm larvae (*Spodoptera frugiperda*) (Muchlinski et al., 2019). Volatile sesquiterpenes and monoterpenes can act as anti-feedants on insects or as inducers of plant defence, thereby priming surrounding plants for predation, as well as for fungal or bacterial infections (Chen et al., 2018; Taniguchi, Hosokawa-Shinonaga, et al., 2014). This mechanism of action holds true for below-ground interactions as well. Maize roots damaged by insects release the sesquiterpene (E)-β-caryophyllene, which attracts

nematodes to prey on the invading beetle larvae (Rasmann et al., 2005). ZmTPS23 catalyses the production of (E)-β-caryophyllene in the roots to attract nematodes, and in leaves to attract parasitic wasps which prey on herbivorous leafworm larvae (Degenhardt, 2009; Köllner et al., 2008). The volatile monoterpene (S)-limonene has been shown to be inducibly produced by OsTPS19, leading to the inhibition of spore germination in *Magnaporthe oryza*, a notorious fungal pathogen of rice (Chen et al., 2018).

1.1.2c Defensive diterpenes are a vital component of anti-fungal plant defence

Lastly, the defensive diterpenes within the Poaceae family are best characterised in rice and maize and will be elaborated upon in the next section. Therefore, this section will focus on the recently reported class of anti-fungal defensive diterpenes in barley (Hordeum vulgare) (Yaming Liu et al., 2024). A BGC on chromosome 2 of barley was found to contain 9 genes encoding for CYP450s, 1 KS gene and 1 CPS gene. Heterologous expression of *HvCYP89E31*, *HvCYP99A66*, *HvCYP99A67* and *HvCYP99A68* with *HvCPS2* and *HvKSL4* in yeast and tobacco led to the production of at least 9 different hordedanes, a subset of the 22 found in fungal-infected barley roots. Furthermore, barley mutants deficient in hordedane biosynthesis had higher colonisation by the generalistic pathogen *Fusarium graminearum*, whereas the same mutants were less infected by *Bipolaris sorokiniana*. The authors of that study suggested that it could be due to *B. sorokiniana* having co-evolved with barley as it was domesticated, therefore relying on hordedanes as chemoattractants. *F. graminearum*, on the other hand, was adapted to North American grasses, as evidenced by the higher tolerance of the latter to Fusarium mycotoxins compared to wheat and barley (Lofgren et al., 2018). Nevertheless, it is a fascinating insight into the ancient role that diterpenes play in pathogen defence. Hordedanes are merely the first defensive diterpenes reported in barley. As research progresses, it is a foregone conclusion that defensive terpenes of varying structures will be discovered in most grasses.

The incredible diversity of terpenes as well as the modularity of their biosynthesis pathways has selected for the evolution of terpenes as key defensive compounds in all plants, not only grasses. Amongst the domesticated crop grasses, maize and rice are some of the most widely and intensively grown crops globally. As of 2021, 1.2 billion tonnes of maize and ~800 million tonnes of rice were harvested (*Agricultural production statistics 2000–2022*, 2023). As highly vital crops, they have been studied extensively to engineer higher productivity and faster growth rates, while also making them more resilient to abiotic and biotic stresses. Both rice and maize produce defensive terpenes upon biotic and abiotic stress, leading to resistance to fungal infections or drought tolerance respectively (Quan & Xuan, 2018; Umemura et al., 2003; Vaughan et al., 2015).

1.2 Defensive terpenes produced by rice and maize

Defensive terpenoids have been very well studied in rice and maize, with major classes having been identified (Figure 1). Both species produce a variety of monoterpenes, diterpenes and sesquiterpenes in response to biotic stresses such as fungal or bacterial infections, as well as in response to abiotic stresses such as drought (Vaughan et al., 2015; Xuan et al., 2016).

1.2.1 Little is known about maize triterpenes

Triterpenes in maize have only recently been researched, with a single publication that discovered *ZmOSC1*, encoding for an enzyme that produced the triterpenes hop-17(21)-en-3-ol, hopenol B and simiarenol when heterologously expressed in yeast (Fan et al., 2022). It is assumed that maize produces triterpenes via otherwise undiscovered triterpene synthases, as such compounds are required for sterol biosynthesis and cuticular wax formation (Matschi et al., 2020).

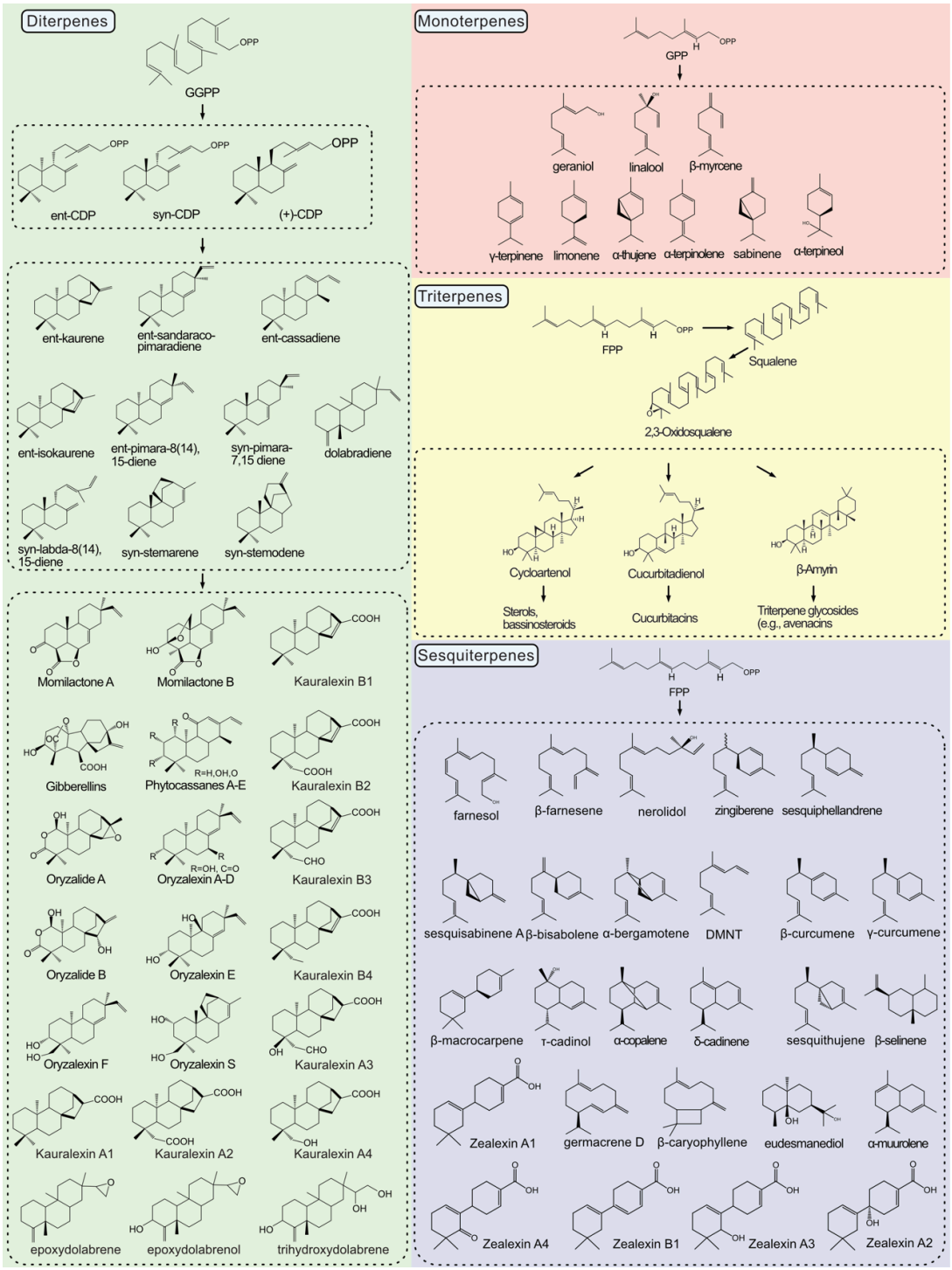


Figure 1: The collection of known terpenes produced in maize and rice, segregated as mono-, di-, tri- and sesquiterpenes. The diterpene and triterpene pathways show the intermediate compounds in the sequential biosynthesis. The sesquiterpene and monoterpene panels only show the structural diversity of compounds as both classes are produced by single-step reactions. Adapted from (Murphy et al, 2020, Block et al, 2019)

1.2.2 Maize produces a variety of sesquiterpenes in response to fungal and insect triggers

Maize sesquiterpenes and monoterpenes have been more extensively studied than triterpenes. As mentioned in the previous section, ZmTPS23 catalyses the production of volatile sesquiterpenes in roots and leaves, which attract parasitic nematodes and wasps, respectively, to combat herbivorous insects. Maize has been shown to produce at least 30 different volatile sesquiterpenes and monoterpenes in response to insect predation as well as fungal infections (Figure 1) (Becker et al., 2014). A previous attempt to identify all known sesquiterpenes in maize proposed that they could be separated into 5 groups (A-E), with each group containing sesquiterpenes that accumulated to similar levels and in the same tissues (Köllner et al., 2004b). Group A comprises (E)-β-farnesene and α-bergamotene, Group B α-copaene, germacrene D and δ-cadinene, Group C β-bisabolene, Group D bisabolane-, sesquithujane- and bergamotane-type sesquiterpenes, and lastly Group E γ-cadinene, δ-cadinene and α-cadinene (Figure 1) (Saldivar et al., 2023). The only sesquiterpenes found in maize roots belong to group C, whereas husks accumulate a combination of groups A, B and D. Lastly, leaves accumulated a mix of compounds from groups A, D and E. While this experiment was by no means extensive, it serves to highlight the variation of sesquiterpenes produced in maize and the differential accumulation in tissues.

A major class of non-volatile defensive maize sesquiterpenes are zealexins (Huffaker et al., 2011). Reported first in 2011, zealexins are β-macrocarpene related acidic sesquiterpenes, produced in response to fungal infections, insect feeding and treatment with jasmonic acid and ethylene. Although 14 acidic sesquiterpenes were detected in infected maize stems, only 5, Zealexins A1-4 and B1, have been identified to date (Christensen et al., 2018; Huffaker et al., 2011). Zealexins A1-3 and B1 accumulated in stem tissue in response to *C. heterostrophus*, *R. microsporus*, *C. sublineolum*, and *A. flavus*, but not in response to *C. graminicola*, a specialist maize pathogen. Zealexins have variable effects on fungal growth; 100 µg / ml of zealexin A1 having had the strongest inhibitory activity against *R. microsporus* (45%), *A. flavus* (80%), and *F. graminearum* (37%) (Huffaker et al., 2011). Zealexin A2 had no inhibitory activity against these 3 pathogens at concentrations up to 100 µg / ml, whereas zealexin A3 at the same concentration inhibited growth of *A. flavus* and *F. graminearum* by 32% and 20% in liquid culture. Therefore, zealexins are speculated to have differential activities against differing fungal pathogens, collectively providing broad anti-fungal resistance to maize.

1.2.3 Monoterpenes in maize play a key role in volatile defence against insects

There are 9 known monoterpenes in maize, all acting as volatile defence compounds against insects (Figure 1). Linalools and myrcene were shown to repel aphids but act as attractants to larvae of army worms (Yactayo-Chang et al., 2024). β-myrcene, limonene, γ-terpinene, terpinolene, 4-terpineol, and α-terpineol are all produced in response to damage from insects feeding on tissues (Lin et al., 2008; Shen et al., 2000). It seems that the main role for monoterpenes in maize is to act as volatile defence signals against herbivory.

1.2.4 Maize diterpenes are potent antibiotics against pathogenic fungi

There are two main classes of defensive diterpenes in maize which have been uncovered recently; kauralexins and dolabralexins (Figure 1). Dolabralexins are produced in the roots at a higher concentration than kauralexins, with concentrations of trihydroxydolabrene reaching 225 µg/g of fresh weight in maize roots elicited by Fusarium spores (Mafu et al., 2018). Kauralexins were measured to accumulate to 9 µg/g of root tissue in the same experiment, suggesting that dolabralexins are the primary anti-fungal diterpenes in maize roots.

Furthermore, epoxydolabranol reduced hyphal growth of *F. graminearum* and *F. verticillioides* by 87% and 71%, respectively, at concentrations of 50 µg/ml (Mafu et al., 2018). This suggested that even if such concentrations were not observed in planta, a combination of dolabralexins acting in concert could act as effective fungicides.

First observed in 2011, kauralexins were inducibly produced in stems subject to insect and fungal infections. Amongst the 6 characterised kauralexins, 10 µg/ml of kauralexin B3 significantly inhibited the growth of *R. microsporus* and *Colletotrichum graminicola* (Schmelz et al., 2011). Kauralexins were also shown to be produced in the roots and vital for resistance to *F. verticillioides* infection (Veenstra et al., 2019).

That summarises the effects of maize terpenes as well as their inducibility. Similar to maize, a number of defensive terpenes have also been characterised in rice thus far. This includes sesquiterpenes, monoterpenes and diterpenes. These are summarised in this following section.

1.2.5 Rice does not produce any known defensive triterpenes

Despite efforts to characterise the production of defensive triterpenes in rice, the only progress made thus far has been to identify 12 genes encoding OSCs which might produce triterpene precursors such as beta-amyrins (Inagaki et al., 2011). Five of these genes have been characterised in vitro, producing isoarborinol, cycloartenol, parkeol, achilleol B and orysatinol (Ito et al., 2011; Xue et al., 2012, 2018). Furthermore, none of these genes occur in a BGC akin to the avenacin BGC in oat. Therefore, rice is thought to produce no defensive triterpenes, but instead produce cuticular wax sterols or other forms of triterpene derivatives (Inagaki et al., 2011).

1.2.6 Rice produces anti-bacterial sesquiterpenes and monoterpenes

Like other cereals, rice also produces volatile sesquiterpenes, especially in response to biotic stresses (Figure 1). Bacterial blight (*Xanthamonas oryza*) infection of rice led to the production of volatile (E)-nerolidol, which also inhibited the growth of Xanthamonas at 100 µM (Kiryu et al., 2018).

Xanthamonas oryza infection in rice was found to induce the production of a variety of monoterpenes, including (S)-limonene, which inhibited the growth of the bacteria at 5 mM (Lee et al., 2016). Furthermore, pre-treatment of rice leaves with 1 - 5 mM of (S)-limonene confers noticeable resistance to *X. oryzae* infection symptoms, suggesting that (S)-limonene acted as a primer of plant defence, as well as inhibitor of bacterial growth. Monoterpene production in rice is also induced by jasmonic acid, a common plant defence signal. γ-terpinene was found in rice plants induced with jasmonic acid, which inhibited growth of *X. oryzae* as well (Yoshitomi et al., 2016). Lastly, jasmonic acid treatment also induced the production of linalool in rice, however linalool did not directly inhibit growth of *X. oryzae* (Taniguchi, Hosokawa-Shinonaga, et al., 2014). Instead, plants overexpressing linalool biosynthesis genes were more resistant to *X. oryzae* infection, implying that linalool derived monoterpenes were anti-bacterial compounds, akin to previously mentioned rice monoterpenes.

1.2.7 Rice produces five major classes of diterpenoids with varied roles in plant defence

Amongst all grasses, rice defensive diterpenes have been best characterised and understood. There are five major classes of diterpenes: phytocassanes, oryzalexins, oryzalides, momilactones, and casbenes (Figure 1).

Phytocassanes A-E were first isolated from leaves of rice infected with *Magnaporthe grisea* (rice blast) (Koga et al., 1997, 1995). The 5 compounds accumulated at edges of necrotic lesions, suggesting a role in preventing fungal spread to surrounding tissue (Umemura et al., 2003). Respectively, phytocassanes A-E inhibited spore germination of *M. grisea* at 20 µg/ml, 4 µg/ml, 7µg/ml, 25 µg/ml, 6 µg/ml. Phytocassane F was reported recently in rice leaves and was shown to have comparable inhibition of spore germination against *M. grisea* as phytocassane A (Horie et al., 2015). Phytocassanes were also found in root exudates of rice, suggesting that they are anti-fungal defence compounds for above-ground and below-ground pathogens (Toyomasu et al., 2008).

Oryzalexin A - D were isolated in 1984 from *M. grisea*-infected rice leaves, similar to phytocassanes (Akatsuka et al., 1983; Kono et al., 1984; 治知 et al., 1986). Oryzalexin E, F and S were isolated from UV-irradiated rice leaves (H. Kato et al., 1993, 1994; O. Kodama et al., 1992). Oryzalexin F had an ED_{50} of 0.3 mM for spore germination of *M. grisea*. Oryzalexin D had an IC_{50} of 0.7 mM against *M. grisea* when measuring mycelial growth and is speculated to disrupt fungal membranes (Sekido & Akatsuka, 1987). Overall, oryzalexins seem to be less well understood than phytocassanes, leading to the conclusion that they may be secondary anti-fungal metabolites.

The third major class of rice diterpenoids are oryzalides. First discovered in rice leaves infected with the bacterial pathogen *X. oryzae*, oryzalide A inhibits colony formation of *X. oryzae* at 0.468 mM (Minoru Watanabe et al., 1990). Oryzalide B and oryzalic acid A were identified later from healthy rice leaves, marking a change from prior studies (Kono et al., 1991). Oryzalide B and and oryzalic A inhibits colony formation of *X. oryzae* at 0.6 mM and 0.2 mM respectively. Varying forms of oryzalides have been characterised but remain unnamed or uninvestigated (Manabu Watanabe et al., 1992). Due to the presence of oryzalides in *X. oryzae* infected leaves as well as their accumulation in healthy leaves, they are thought to be to prohibitins acting as the first line of defence against bacterial pathogens.

Momilactones are likely the most studied rice diterpenoids. They were first isolated in 1973 from rice husks (known as ‘momi’ in Japanese, thus the name momilactones) (Kato et al., 1973). They were then extracted from UV irradiated rice leaves as well as *M. oryzae* infected leaves (Cartwright et al., 1981). Rice produces momilactones A, B, C, D and E, with momilactones A and B being the most potent contributors to plant defence (Cartwright et al., 1981; J.-G. Cho et al., 2015; T. Kato et al., 1973; Tsunakawa et al., 2006; Zhao et al., 2018). Momilactone B in particular was shown to be actively exuded into the rhizosphere and acted as an allelochemical, inhibiting growth of other plants (Kato-Noguchi & Ino, 2003), including *Echinochloa crus-galli* (Kato-Noguchi et al., 2010), *Arabidopsis thaliana* (Kato-Noguchi et al., 2012), and *Echinochloa colonum* (Kato-Noguchi, 2011), to name a few. In general, the IC_{50} of momilactone B ranged from 1 µM to 20 µM in inhibiting hypocotyl extension or root extension, whereas momilactone A had much higher IC_{50} of 28 µM to 240 µM (Kato-Noguchi, 2011). Uniquely within grasses, momilactones are allelopathic diterpenoids, warranting much more research interest than the other rice diterpenoids.

The final major rice diterpene is 5,10-diketo-casbene, also known as ent-10-oxodepressin, and was first reported to be produced in rice in 2013 (Inoue et al., 2013). 5,10-diketo-casbene was shown to accumulate in the leaves of rice that had been irradiated with UV as well as in leaves inoculated with *M. oryzae* spores, albeit at lower concentrations than momilactones A and B. 5,10-diketo-casbene was shown to inhibit germ tube elongation of *M. oryzae* at an IC_{50} of 33 µM and spore germination at 100 µM (Inoue et al., 2013). Mutants deficient in producing 5,10-diketo-casbene exhibit stronger symptoms of *M. oryzae* infection (Liang et al., 2021). 5,10-diketo-casbene is produced by a recently assembled BGC, which will be discussed in the next section, and is therefore still undergoing selection in Oryza species, domesticated and wild (Zhan et al., 2020). Furthermore, considering that it

has a high IC_{50} and accumulates in tissues at lower concentrations than other anti-fungal compounds, it is not as significant in anti-fungal responses as the other rice diterpenes.

Recent reports have proposed a new class of diterpenes which still need better characterisation. These were termed oryzalactones and were speculated to originate from a novel KSLX-OL enzyme acting in concert with OsCPS4 to produce syn-abieta-7,12-diene, which is then converted to oryzalactones via unknown enzymes (Kariya et al., 2024). Furthermore, the gene encoding *OsKSLX-OL* was found to exist in a subset of rice cultivars and arose from fusion of *OsKSL8* and *OsKSL9* on chromosome 11. Clearly, terpenoid biosynthesis is still evolving in rice cultivars and varieties in order to generate a sufficiently large array of defensive compounds to combat rapidly evolving fungal and bacterial pathogens.

In summary, the array of terpenes produced by maize and rice contribute to plant defence against agronomically relevant fungal, bacterial and weed species. By accumulating in tissues before damage, after damage and actively being secreted into the environment, terpenes play a key role as the first line of plant defence in limiting fungal and bacterial damage. Allelochemicals confer competitive advantages to rice, in allowing seedlings to establish themselves while reducing competition. All of these traits are vital points of improvement in the agricultural chase for higher yields. Therefore, it is crucial to identify the biosynthesis of defensive terpenes in rice and maize.

1.3 Defensive terpene biosynthesis genes in rice and maize

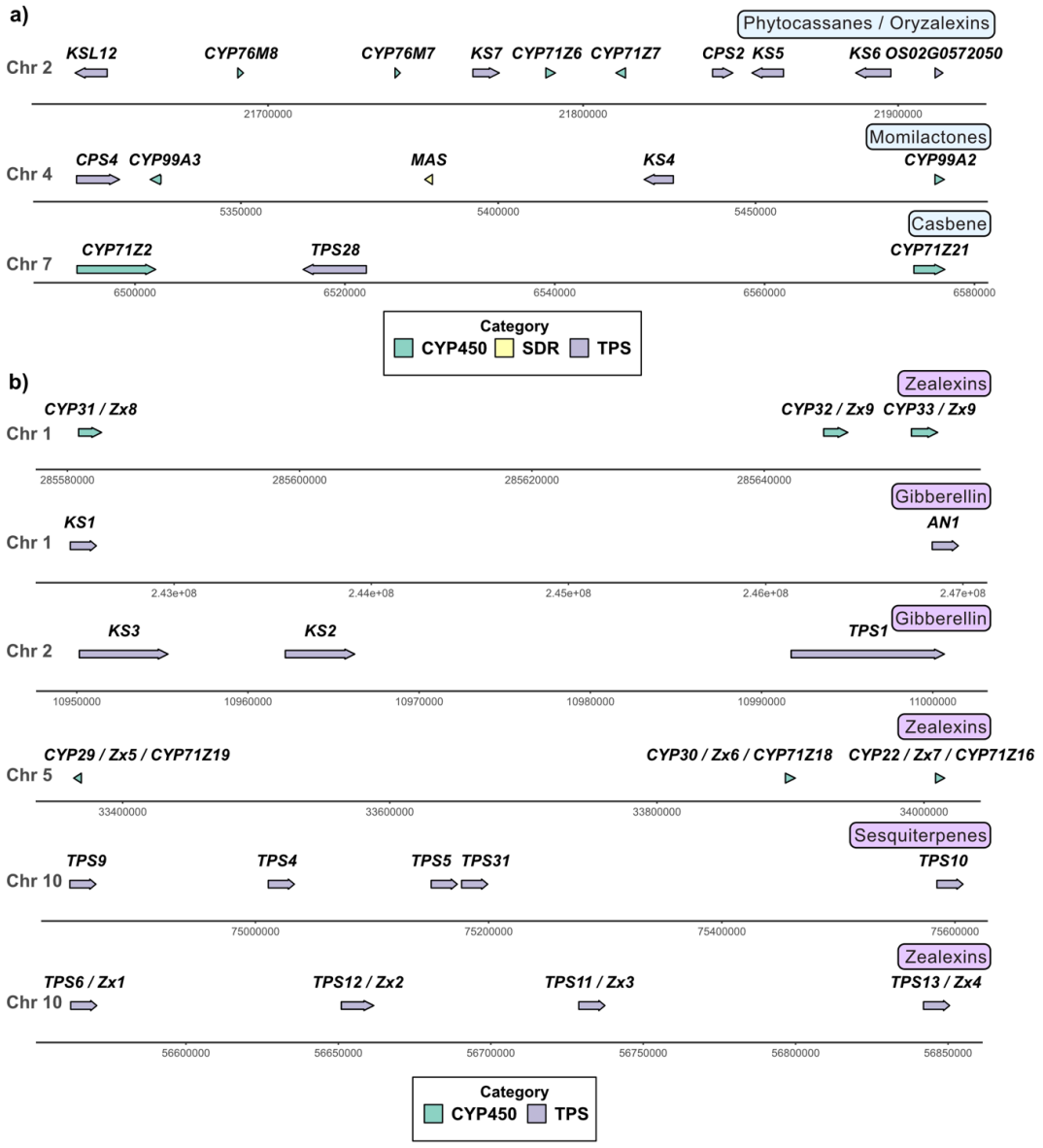


Figure 2: The relevant, known BGCs present in a) rice and b) maize. The genes are coloured by their function and the gene length corresponds to the genomic distances shown in the y-axis only for rice. Due to the large size of the maize genome, the gene lengths in maize have been scaled up in order to better visualise the presence of gene clusters.

The terpene biosynthesis genes in rice and maize are largely organised in clusters of genes, known as biosynthetic gene clusters (BGCs). These clusters are shown in Figure 2 above. This section covers the discovery of these genes and their roles within each terpene biosynthesis pathway in rice and maize.

1.3.1 Maize sesquiterpene and monoterpene synthases catalyse a variety of major and minor products

The main sesquiterpene biosynthesis genes appear to be *ZmTPS8*, *ZmTPS10* and *ZmTPS23* (Köllner et al., 2013; Saldivar et al., 2023). Together, these 3 genes cyclicise FPP into 11 of the 21 known sesquiterpenes in maize (Block et al., 2019). The remaining ten are produced by sesquiterpene synthases encoded by *ZmTPS1 / 4 / 5 / 6 / 7 / 8 / 10 / 11 / 21*. *ZmTPS26* encodes for a monoterpene synthase which solely reacts with only GPP and yields β-myrcene, γ-terpinene, α-terpinolene, and geraniol (Lin et al., 2008). *ZmTPS1* is another monoterpene synthase which can produce linalool and geraniol (Schnee et al., 2002).

1.3.2 Maize diterpene production is intricately linked to zealexin production

Zealexin, kauralexin and dolabralexin biosynthesis was recently reported to be intertwined, incorporating promiscuous CYP450s which act on acyclic FPP as well as cyclical diterpenes (Ding et al., 2020). The first step in zealexin production is the conversion of FPP to β-macrocarpene and β-bisabolene via *ZmTPS6* (*Zx1*), *ZmTPS12* (*Zx2*), *ZmTPS11* (*Zx3*), and *ZmTPS13* (*Zx4*) (Christensen et al., 2018; Huffaker et al., 2011). Then, *ZmCYP71Z19* (*Zx5*), *ZmCYP71Z18* (*Zx6*) and *ZmCYP71Z16* (*Zx7*) produce zealexin D1/2 from β-bisabolene, and zealexin A1 from β-macrocarpene. Zealexin A1 is metabolised into at least 12 other zealexin A, B and C products via *ZmCYP81A37* (*Zx8*), *ZmCYP81A38* (*Zx9*), *ZmCYP81A39* (*Zx10*) (Ding et al., 2020). GGPP is cyclicised into ent-CDP by the CPS *ZmAN2*. Subsequently, *ZmKSL2* and *ZmKSL4* produce ent-isokaurene and dolabradiene respectively. *ZmCYP71Z19* (*Zx5*), *ZmCYP71Z18* (*Zx6*), *ZmCYP71Z16* (*Zx7*), *ZmKR2* and *ZmKO2* then produce kauralexins from ent-isokaurene and dolabralexins from dolabradiene (Ding et al., 2019, 2020; Mafu et al., 2018). The respective compounds are shown in Figure 1 and the BGCs of zealexins in Figure 2.

1.3.3 Rice triterpene synthases

As mentioned before, although 12 genes encoding OSCs have been identified and tentative products characterised, the final active triterpenes are unknown and have no clear roles (Inagaki et al., 2011; Xue et al., 2012).

1.3.4 Rice sesquiterpene and monoterpene synthases

Sesquiterpenes and monoterpenes are generally catalysed by single terpene synthases in rice, unlike the diterpene biosynthesis pathways which incorporate multiple TPS, KS and CYP450s. OsTPS3, OsTPS19, OsTPS20, OsTPS24 produce the bulk of volatile sesquiterpenes and monoterpenes in rice (Cheng et al., 2007; Taniguchi, Miyoshi, et al., 2014; Yoshitomi et al., 2016). Geraniol, a monoterpene produced by rice in response to *X. oryza* infection inhibits growth of said pathogen by suppressing expression of cell division genes (Kiyama et al., 2021).

1.3.5 Diterpene biosynthesis in rice is highly modular and involves promiscuous enzymes

Diterpenes in rice are catalysed by cyclicisation of GGPP into ent-copalyl diphosphate (ent-CDP) or syn-CDP via OsCPS4 or OsCPS1/2 respectively. OsKSL4 / 8 / 10 and 11 can cyclicise syn-CDP into syn-pimara-7,15-diene, syn-stemarene, syn-labda-8(14)15-diene and syn-stemodene, respectively. Syn-pimara-7,15-diene is converted into momilactones A and B by OsCYP99A3, OsCYP76M8, OsCYP701A8, OsCYP76M14, and momilactone A synthase (OsMAS) (De La Peña & Sattely, 2020; Kitaoka et al., 2016; Q. Wang et al., 2012,

2011, 2012-3). OsDTC1 is necessary for oryzalexin S production (Nemoto et al., 2004). Syn-stemarene leads to the production of oryzalexin S (Nemoto et al., 2004) whereas syn-stemodene has an unknown function as well as undiscovered biosynthesis pathway. ent-CDP can be cyclicised into gibberellins by OsKS1, ent-sandaracopimaradiene by OsKSL10, ent-cassadiene by OsKSL7, ent-isokaurene by OsKSL6 or ent-pimara-8(14),15-diene by OsKSL5. Ent-sandaracopimaradiene is converted to oryzalexin A-F by OsCYP701A8, and OsCYP76M6/8 and OsSDR110c-MS3, a short chain dehydrogenase (Kitaoka et al., 2016; Q. Wang et al., 2012-3). Phytocassanes arise from ent-cassadienes via oxidations catalysed by CYP76M7/8 and CYP71Z7 (Swaminathan et al., 2009; Wu et al., 2011). Finally, ent-isokaurenes are converted to oryzalides via OsCYP71Z6 (Wu et al., 2011). Amongst the terpenes in rice and maize, the modular organisation of diterpene biosynthesis in rice, involving various CPS and KS yielding different chiralities, an array of CYP450s and SDRs is most impressive. It is evident that such modularity and promiscuity confers high diversity in terpenes, thus providing a higher chance of anti-fungal, anti-bacterial and allelopathic defences. The phytocassane, oryzalexin and momilactone biosynthesis genes are located in BGCs in chromosomes 2 and 4, respectively (Figure 2). There is a 3rd BGC on chromosome 7, which contains 3 genes involved in casbene production (Figure 2).

The identification of biosynthesis genes and pathways contribute to plant breeding for more resistant cultivars, especially with climate change forecasted to greatly impact agriculture. Known biosynthesis pathways and genes can be transferred to high yielding varieties of crops, supplementing endogenous resistance mechanisms. However, other approaches to engineer more stress-tolerant cultivars rely on the identification of regulatory mechanisms governing the production of defensive compounds.

1.4 Transcriptional regulation of terpene biosynthesis genes in rice and maize

In parallel to the discovery of defensive terpenes in rice and maize, regulatory mechanisms driving the production of such compounds have been partially identified in rice and maize. Seven transcription factors (TFs) in rice have been identified that affect the production of defensive terpenes in rice. OsDPF, OsTGAP1, OsWRKY45 and OsWRKY10 were all found to increase momilactone, phytocassane and oryzalexin production, as well as upregulate transcription of the respective biosynthesis genes (Akagi et al., 2014; Okada et al., 2009; Wang et al., 2023; Yamamura et al., 2015). OsDPF was found to bind to N-boxes (5'-CACGAG-3'), sequences in the promoter regions of OsCPS2 and OsCYP99A2, thereby activating their expression. OsWRKY10 binds to W-boxes (5'-TTGACC-3') and W-box like element (WLE) (5'-TGACA-3') in the promoter regions of *OsKSL7*, *OsKSL4*, *OsCYP99A3* and *OsKSL10*, thereby activating their transcription. OsWRKY62, OsWRKY76 and OsBZIP79 repress diterpenoid accumulation when transgenically overexpressed, perhaps playing a role in shunting metabolic flux towards primary metabolism of gibberellic acid (Liu et al., 2016; Miyamoto et al., 2015).

The regulatory mechanism of maize terpenoid production has not been as well understood as in rice. Only three TFs have been identified thus far. First, *ZmTPS10* is regulated by ZmEREB58, a jasmonic acid regulated TF (Li et al., 2015), thereby controlling sesquiterpene production. Second, ZmWRKY79 regulates *ZmAN2* and *ZmTPS6* via W-box and WLEs in the promoter regions of the two terpene synthases (Fu et al., 2018). Furthermore, WLEs in the promoter regions of *ZmKSL5* were also found to regulate its expression, suggesting that WRKYs play a larger role in regulating the production of zealexins and kauralexins (Yang et al., 2020). Finally, gene co-expression network analyses identified ZmNACTF7 as a key regulator of a module of genes which contained zealexin, kauralexin and diterpene biosynthesis genes (Ma et al., 2017). Such gene co-expression

networks are highly useful tools for fully elucidating biosynthesis pathways as well as identifying regulatory genes.

1.5 Gene co-expression networks shed light on specialised metabolism in plants

Gene co-expression network tools have been useful tools in deciphering biology since gene expression data was available (Butte & Kohane, 1999). With the introduction of microarray and RNA-Seq, a great variety of tools were designed to build co-expression networks from single and multiple projects. In the latter case, tools could utilise data from multiple projects to identify networks of genes that commonly impact disease symptoms (Soh et al., 2011). The gradual development of these tools as well as increasing interest from traditional plant biologists led to projects identifying Fusarium-responsive BGCs in wheat and TFs regulating anthocyanin production in eggplant, amongst many other publications (He et al., 2021; Perochon et al., 2021).

1.5.1 Examples of use of WGCNA in studying plant specialised metabolism

The identification of SmWRY44 in eggplant which positively regulates the accumulation of anthocyanins (Yongjun He et al., 2021), is highly relevant to this thesis. In this example, transcriptomic data from 33 time point samples were used in a gene co-expression network building tool known as Weighted Gene Correlation Network Analysis (WGCNA). This generated modules of genes which were highly correlated with each other. The main core of the known anthocyanin biosynthesis genes was present in module "tan", and analysis of the module showed that a single gene, encoding a TF, was highly correlated to the rest of the genes in the module, and hence termed a hub gene. Further investigations of this TF SmWRKY44 showed that it was induced by light, activated the promoters of known anthocyanin biosynthesis genes, and - when overexpressed - caused the accumulation of anthocyanins in stems. Thereby, a gene co-expression network successfully identified a TF regulating a core metabolic process in plants.

Gene co-expression networks are also useful for identifying novel BGCs which produce defensive compounds. WGCNA was used to build gene networks from fungus-inoculated wheat tissue, eventually leading to the identification of 6 BGCs. These 6 BGCs were predicted to produce diterpenes, triterpenes and flavonoids (Polturak et al., 2022). Heterologous expression of the triterpene and flavonoid BGCs in tobacco leaves confirmed the production of triterpenes and flavonoids, albeit not necessarily the final end products. It was shown that upon fungal infection, these BGCs were activated, producing defensive compounds in leaves and roots. While this has been the most recent outstanding example of gene network analysis in pathway discovery, it is by no means the last.

Clearly, gene co-expression networks are useful tools for identifying novel BGCs and regulatory genes in specialised metabolism. It is crucial to understand the key, relevant concepts of gene networks as they have been used within the scope of this project.

1.5.2 What is a WGCNA derived gene network?

Gene networks can be built from a variety of methods, including traditional Pearson’s correlation and Spearman’s correlation. These methods compare the expression levels of all genes across all samples. Genes which are similarly differentially expressed across similar samples and conditions have a high Pearson’s correlation score and are considered co-expressed. These genes can then be ranked by the correlation score and be considered

as strong candidates for missing biosynthesis genes (De La Peña & Sattely, 2020). The correlation score derived from Pearson's correlation can be used to build gene networks, which can be informative as well.

Modern methods such as WGCNA are built from Pearson's correlation analysis. After obtaining a correlation matrix of all genes, the next step in WGCNA is to perform a topological similarity analysis, producing a topological overlap matrix (TOM). The neighbourhood of a gene is defined as the genes most associated with it in the correlation matrix. The topological similarity analysis compares the neighbourhoods of all genes and scores them for similarity, thereby producing a TOM. Hierarchical clustering of the TOM then leads to genes being grouped into modules which can be associated with biological and cellular functions. For example, the 6 wheat BGCs in the previous subsection contained genes largely within module 25 of the wheat WGCNA network (Polturak et al., 2022). This module contained genes consistently upregulated in treatment conditions (fungal inoculations) when compared to control samples, aligning with the upregulation of the 6 BGCs upon fungal stress.

Once the gene network is separated into modules, the average expression of each module can be correlated to the "Treatment" of each sample. Again, relating back to the wheat BGC example in the previous subsection - 5 modules of genes, out of 69 modules, were consistently upregulated in treatment conditions compared to control conditions. One of those 5 modules contained the majority of the genes in the 6 BGCs. By correlating module expression to treatments, most projects can identify a few modules or a single module which contributes the most to a treatment and possibly a phenotype.

At this stage, the genes in the module are extensively investigated by mining available literature or performing GO Enrichment analysis to identify cellular pathways which are represented in the module etc. Another approach is to dissect the topology of a module. Within each module, there are hub genes, which are defined as genes having a high TOM score to most of the other genes in the module. For example, *SmWRKY44* was a hub gene which most strongly correlated with many of the other genes in the "tan" module, and therefore was picked as a candidate for anthocyanin biosynthesis regulation. By mining such topologies, hub genes can be identified and can be characterised better in experiments. Most publications involved in gene network analysis take these approaches and are successfully published and well received. However, this project leverages one key concept of gene networks, which will be elaborated upon in the next subsection.

1.5.3 Orthologous gene networks accelerate our understanding of core cellular processes

A core concept in gene networks are the neighbourhoods of a gene, which refer to the most strongly correlated genes of any single gene, i.e., it's closest neighbours. By exploring the neighbourhoods of genes known to perform particular roles, one can identify candidates contributing to metabolism, transcriptional regulation, cell immunity etc. For example, identification of *SmWRKY44* as an anthocyanin regulating TF was possible because it was in the neighbourhood of known anthocyanin biosynthesis genes.

Another approach, which partially inspired this doctoral thesis, studied gene neighbourhoods of known growth regulating genes in 3 different species, leading to the identification of 34 genes regulating growth in Arabidopsis (Curci et al., 2022). In Curci et al, 2020, a variety of gene network building algorithms, including TOM, were used to build meta-gene networks in three plant species (Curci et al., 2022). These networks were then filtered for orthologous genes present in all three species, and gene neighbourhoods of known growth regulators were analysed in each network. This yielded a substantial number of candidate genes in

each species which could contribute to cell growth and elongation. Curci et al showed that Arabidopsis plants lacking either of two growth regulating candidate genes had reduced rosette area. Therefore, by identifying conserved orthologous genes between 3 species and analysing their gene neighbourhoods, one can identify novel genes involved in core cellular pathways. This concept of orthologous gene networks has been a key component of this thesis project.

1.6 Momilactone B: A potent diterpenoid allelochemical

In section 1.2.7, one of the potent diterpenoids covered in rice was momilactone B. It is one of the unique terpenoids within the scope of this project as it has broad effects in inhibiting the growth of neighbouring competitor plants (Kato-Noguchi et al., 2010, 2012; Kato-Noguchi & Ino, 2003). Its potency against the agronomically relevant weed *E. crus-galli* provoked interest in developing it as a herbicide (Kato-Noguchi, 2011). There were early efforts to identify rice cultivars producing high amounts of allelochemicals and introducing them to farmers (MOlofsdotter et al., 1999; Olofsdotter et al., 1995; Olofsdotter et al., 1999), to alleviate the costs of herbicides as well as mitigate the impact of rising herbicide resistance (Powles & Yu, 2010). However, at that point, even though momilactone B had been discovered in 1964, its function as an allelochemical was overseen in these efforts; instead, focus was on phenolic acids, which were eventually concluded to be unlikely effective allelochemicals (Maria Olofsdotter et al., 2002). Momilactone B's allelopathic abilities were intensely investigated in the early 2000s and onwards, and its biosynthesis pathway was recently fully elucidated (De La Peña & Sattely, 2020; Shimura et al., 2007; Toyomasu et al., 2014; Wang et al., 2012-3).

1.6.1 Allelochemicals are strong starting points for herbicide development

As naturally occurring compounds that inhibit plant growth, allelochemicals are obvious candidates for herbicide development. The goal would be to identify the mode of action and test it against a range of agronomically relevant weeds. Leptospermone is the success story of exactly such a developmental method. While leptospermone itself required a high application rate to be a practical herbicide, its backbone was used to generate thousands of analogues. This led to the triketone class of 4-hydroxyphenylpyruvate dioxygenase inhibitor (HPPD), with mesotrione, also known as Callisto, as a well-known Syngenta product (Cornes, 2005).

Coincidentally, the discovery of potent allelopathy in rice led to intense focus on identifying the culprits. In the late 1990s as well as early 2000s, Maria Olofsdotter led the efforts in quantifying the allelopathic nature of a variety of rice cultivars against weeds as well as model organisms such as lettuce (Olofsdotter et al., 1999).

Multiple assays were conducted in a variety of plants to characterise the potency of momilactones. These assays varied greatly, as they were based on herbicide development assays where seeds were sown on filter paper imbibed with potential herbicides. These methods result in the crystallisation of chemicals, leading to unequal concentrations being encountered by test organisms. Thus, the IC_{50} was often inaccurate and imprecise, varying amongst published papers. These various IC_{50}s are included in Table 1 below as well as their dose response method.

Table 1: Summary of studies reporting momilactone B effective concentrations on target plants. Adapted from (Serra Serra et al., 2021).

Organism	Experimental method	Phenotype measured	Observation	Reference
Lepidum sativum L.	Co-cultivation on filter paper	Root and hypocotyl length	30% inhibition at 12 µM and 16µM	(Kato-Noguchi & Ino, 2003)
E. crus-galli	Co-cultivation on filter paper	Shoot and root growth	IC_{50} of 6.5 µM and 6.9 µM	(Kato-Noguchi et al., 2010)
A. thaliana	Co-cultivation on filter paper	Hypocotyl and root growth	IC_{50} of 6.5 µM and 12 µM	(Kato-Noguchi et al., 2012)
A. thaliana	Co-cultivation on filter paper	Hypocotyl and root growth	Internal IC_{50} of 0.09 µM and 0.95 µM	(Kato-Noguchi et al., 2012)
A. thaliana	Co-cultivation on filter paper	Germination rate	IC_{50} of 48.4 µM	(Chi et al., 2013)
E. crus-galli	Co-cultivation on filter paper	Root and shoot growth	IC_{50} of 6.31 µM and 6.11 µM	(Kato-Noguchi, 2011)
E. colona	Co-cultivation on filter paper	Root and shoot growth	IC_{50} of 12.5 µM and 5.04 µM	(Kato-Noguchi, 2011)

Various stresses were used to identify the mechanisms of induction of momilactone production in rice. Triggers such as copper chloride and UV treatment, fungal infections, or chitin elicitations drove the accumulation of momilactones in tissues as well as in root exudates. This was supposed by the increased expression of biosynthesis enzymes in transcriptomic analyses. This led directly to the elucidation of the key enzymes *CPS4*, *KSL4*, *CYP99A3* and *OsMAS* by analysing genes which were upregulated upon elicitation by one or several of the aforementioned triggers.

Through an act of convergent evolution, momilactone B is also produced in the moss *Calohypnum plumiformae*. The biosynthesis pathway is partially characterised with the enzymes *CpDTC1* (bifunctional diterpene synthase that combines activities of OsCPS4 and OsKSL4), CpMAS, CpCYP970A14 (CYP99A3 orthologue), and CpCYP964A1 (CYP701A8) (Mao et al., 2020; Okada et al., 2016).

Due to the ubiquitous occurrence in rice, a crucial crop, as well as its potency, it is vital that the mode of action of momilactone be identified. This will lead to a variety of applications including breeding for more allelopathic cultivars that utilise less herbicide as well as lead compounds in herbicide development. A prominent method of identifying modes of action of herbicides is through forward genetic screens.

1.6.2 Forward genetic screens in Arabidopsis

Forward genetic screens are extremely useful tools for deciphering cellular pathways (St Johnston, 2002). To date, the common fruit fly, *Drosophila melanogaster*, is known as the workhorse of genetics for two reasons; i) core processes were conserved between insects and animals, allowing for translational findings ii) fruit flies were highly amenable to forward genetic screens, allowing for quick deciphering of cellular pathways. Similarly, forward genetic screens in Arabidopsis have been used to understand core plant processes in plants.

Forward genetic screens or mutant screens in Arabidopsis have identified genes governing leaf and flower morphology (McKelvie, 1961; Röbbelen, 1957), flowering time (Koornneef et al., 1983), phytohormone biosynthesis and perception (Bleecker et al., 1988; Guzmán & Ecker, 1990; Koornneef et al., 1984; Koornneeff et al., 1982; Maher & Martindale, 1980), lipid biosynthesis (Lemieux et al., 1990) and plant-pathogen interactions (Glazebrook et al., 1997), to name a few. Forward genetic screens in Arabidopsis are effective at identifying mutants resistant to herbicides (Brotherton et al., 2007; Jander et al., 2003).

Herbicide resistance in plants can be classified into two categories - target site resistance and non-target site resistance (Délye et al., 2015). Target site resistance can arise from mutations in the target protein of the herbicide, therefore altering the affinity of a herbicide for a target protein (Beckie & Tardif, 2012). Target site resistance can also occur from mutations that increase the expression of a target protein either by increased transcription or target gene duplication, therefore surpassing a herbicide's effects (Baerson et al., 2002). Non-target site resistance occurs in 4 main mechanisms - altered transport, enzymatic modifications, sequestration and dosage compensation via extra-chromosomal DNA. Mutations in exporters / importers have been shown to reduce contact of herbicides with target proteins (Svyantek et al., 2016). Mutations in enzymes allow the detoxification or modification of herbicides into harmless metabolites, such as CYP450s in rice detoxifying the herbicide, bensulfuron-methyl, by O-demethylation (Deng & Hatzios, 2002). Mutations in metabolite sequestration mechanisms result in shuttling of herbicides into vacuolar spaces that prevent cellular damage, such as the shuttling of glyphosate to the edges of leaves (Feng et al., 2004; Gaines et al., 2019). Finally, extra-chromosomal DNA (eccDNA) or mini chromosomes can harbour copies of target genes of herbicides, thereby increasing expression to overcome the herbicidal effects. This mechanism was first reported in glyphosate resistant *Amaranthus palmeri* (Koo et al., 2018) and recent findings have shown that eccDNA is heritable across compatible species, further accelerating herbicide resistance in weeds (Koo et al., 2023).

Forward genetics screens in Arabidopsis are well established and fairly straight forward. Ethyl methanesulfonate (EMS) is used to mutate Arabidopsis seeds, which are grown on regular media, and selfed to generate an M2 population (Kim et al., 2006). The M2 population can have as many as 700 mutations, of which a significant percentage are homozygous mutations, therefore allowing for effective screening (Jander et al., 2003). This population is then screened for phenotypes; resistance to herbicides, morphological changes, etc. When screening for chemicals in particular, a dose response study must be conducted in order to choose an appropriate mutant screen concentration. Once herbicide resistant mutants are identified in a mutant screen, these mutants are usually crossed with wild-type Arabidopsis, producing an F1 generation (Page & Grossniklaus, 2002). This crossing step dilutes the large number of mutations in the M2 generation into individual plants, such that an F1 plant has much fewer mutated loci compared to M2. The F1 generation is self-fertilised, leading to an F2 generation which has homozygous mutations at the mutated loci (Schneeberger, 2014). The F2 generation is now screened for herbicide resistance; resistant individuals are sequenced in bulk to identify genomic loci associated with herbicide resistance. At this point, identification of genes at those loci or other genetic elements driving target site / non-target site resistance followed by characterisation of those genes allows for understanding how a herbicide binds to a target protein as well as the types of genes which could confer resistance.

In summary, as Arabidopsis has been extensively used in forward genetic screens and is susceptible to momilactone B, there was an opportunity to better understand this diterpene, within the context of this project.

1.7 Objectives

There were two primary objectives in this doctoral thesis. First, due to rice and maize utilising orthologous enzymes to produce similar defensive terpenoids, I hypothesised that there could be other orthologous genes encoding ancillary proteins involved in the terpenoid biosynthesis machinery. I aimed to build gene networks in both species and identify orthologous ancillary genes involved in transcriptional regulation, transport, and immune response, thereby collectively driving the production and deployment of defensive diterpenoids. I then aimed to verify the roles of these ancillary genes in the terpenoid biosynthesis machinery by using bioinformatic approaches. Collectively, this improves the current understanding of defensive terpenoid metabolism in two key crop species. Furthermore, it would validate using similar gene network approaches in other plant species to accurately identify the production of defensive metabolites.

Second, I hypothesised that the mode of action of momilactone B, a potent allelochemical in rice, could be elucidated using forward genetics approaches, akin to deciphering herbicide resistance in weeds. I aimed to use a forward genetic screen in Arabidopsis to identify resistant mutants, perform linkage mapping to identify resistant loci, and eventually identify genes conferring resistance to momilactone B in Arabidopsis. This would eventually lead to understanding of the mode of action of momilactone B, which would aid efforts to generate much needed herbicides with novel modes of action.

2. Materials & Methods

2.1 Mutant screens in Arabidopsis

2.1.1 Dose Response of Momilactone B in Arabidopsis

A. thaliana Col-0 seeds were sterilised with chlorine gas for 1 h and stratified for 6 d in the dark at 4˚C. 20 seeds per plate were sown on ½ MS media supplemented with various concentrations of momilactone B (0 µM, 0.1 µM, 1 µM, 4 µM, 10 µM and 20 µM) dissolved in DMSO, to a final DMSO concentration of 0.1% in all plates. Seedlings were grown in a 16 h / 8 h light/dark cycle chamber at 21˚C with a light intensity of 50 µM/m^2/sec. After 5 d of growth, the seedlings were imaged using a fixed camera and a ruler for scale. The primary root length was traced using ImageJ and primary root length was calculated using the scale as a reference. Primary root length was plotted as relative percentage of growth compared to the control sample. The drc package in RStudio was used to fit a dose–response model and to calculate the half-maximal-effect concentration; data was plotted using the ggplot2 package.

2.1.2 Mutant Screens

Two collections of Arabidopsis mutants were used in the course of this project. The first collection was an EMS mutagenised F2 generation of seeds obtained from collaborators at the Max Planck Institute for Plant Breeding Research, Cologne, Germany. The second collection was an F5 generation of single seed propagated EMS mutagenised plants, originally generated by the Mercier lab at the INRA, France (Capilla-Perez et al., 2018). Three different concentrations of momilactone B were used during the course of this project.

2.2 Building terpenoid gene networks

2.2.1 Curation of bait gene list in rice and maize

The bait gene lists in maize and rice were assembled by identifying known terpene synthases via annotation files available on Phytozome (https://phytozome-next.jgi.doe.gov/) for each species, as well as via literature review of terpene synthases and CYP450s which contribute to terpenoid production. Rice has two versions of gene identifiers, defined by the Rice Annotation Project (https://rapdb.dna.affrc.go.jp/) and the Rice Genome Annotation Project (http://rice.uga.edu/), respectively. Both versions of identifiers were used for literature search in order to exhaustively identify terpene synthases. Maize has a variety of gene identifiers available at the MaizeGDB database (https://www.maizegdb.org/). Most publications use identifiers from version 3 or 4 of the maize genome annotation project. Some publications refer to maize genes by their pan-gene identifiers, GRMZM2G132212 for example.

2.2.2 Identification of orthologous proteins in rice, maize, barley and sorghum

Orthofinder (Emms & Kelly, 2019) was used to identify orthologous proteins based on amino acid sequences of 27 species, consisting of 26 grasses and 1 outgroup species (*Ananas comosus*; pineapple). This generated an orthogroup file that listed orthologous genes in each species, classified into distinct orthogroups.

Table 2: Latin names of all species of monocots used in the OrthoFinder analysis to identify orthogroups and orthologous proteins. A. comosus is the only monocot that is not a grass species and was used as an outgroup. The common names for some popular species are in brackets.

Ananas comosus (Pineapple)	*Streptochaeta angustifolia*	*Pharus latifolius*
Zea mays (Maize)	*Sorghum bicolor (Sorghum)*	*Digitalis exilis*
Echinocloa crus-galli	*Echinocloa haploclada*	*Setaria italica (Foxtail millet)*
Panicum hallii	*Cleistogenes songorica*	*Eragrostis curvula*
Eragrostis tef (Teff)	*Brachypodium distachyon*	*Hordeum vulgare (Barley)*
Secale cereale (Rye)	*Triticum aestivum (Wheat)*	*Aegilops tauschii*
Olyra latifolia	*Phyllostachys edulis*	*Zizania palustris*
Leersia perrieri	*Oryza brachyantha*	*Oryza officinalis*
Oryza punctata	*Oryza sativa subsp. Japonica (Japonica cultivar of rice)*	*Oryza alta*

2.2.3 Curation of bait gene list in barley and sorghum

The bait genes in barley were identified by identifying orthogroups which contained known terpenoid biosynthesis genes in rice and extracting the barley genes in the same orthogroups. The bait genes in sorghum were identified in an identical manner but by using known maize biosynthesis genes.

2.2.4 Gene network construction

Dataset identification and mapping

Datasets for network building were identified on NCBI BioProjects via the criteria listed below. In general, recently published datasets were chosen for the analyses.

i) For comparability across studies, the tissue of origin must be from leaves treated with fungal pathogens or the respective control condition; plants must be of similar age 2 - 3 week.
ii) The cultivars used in the experiments must have a well-assembled and annotated genome. This is the case for cultivar ‘Morex’ for *H. vulgare*, cv. ‘Nipponbare’ for *O. sativa*, cv. ‘B73’ for *Z. mays*, cv. Btx623 for *S. bicolor*.
iii) The percentage of uniquely mapped reads in each sample must be > 80%, else the samples were omitted.

Reads were mapped and gene features counted using the nf-core pipeline RNAseq v3.9 with default parameters (Ewels et al., 2020). In brief, reads were mapped using STAR and quantified using salmon (Patro et al., 2017). The counts per gene were further analysed using R v4.4.1.

Batch effect correction

When performing meta-analyses that incorporate multiple transcriptomic datasets, a batch effect often arises. The count table was filtered such that genes with low or no reads across all samples were removed. Then, the count table was formatted into a DESeq object and PCAs were used to visualise variance among the top 500 genes. This showed that the variations amongst the samples in the datasets were driven to a significant extent by the

BioProjects. In order to minimise the impact of the differing BioProjects, batch effect correction was applied via the limma::RemoveBatchEffect function on a normalised count table (variance stabilising transformation, vst() function of the DESeq2 package). This generated batch corrected, normalised counts which were used for network construction.

Network construction

Batch-effect corrected datasets were then transposed and a soft power calculated for each species using WGCNA::pickSoftthreshold(). This function generates a curve to determine a power which is used as an exponent in transforming the adjacency values in the adjacency matrix. This is necessary for then determining the threshold values for defining modules in the topological overlap matrix. Networks for each species were built using the automatic network building function WGCNA::blockwisemodules(). Once modules were defined, all the edges in each species were extracted using WGCNA::ExportNetworkToCytoscape() and saved.

Neighbourhood identification

For each bait gene, the edges were ranked by weight and the strongest 2% of edges were extracted, as well as the corresponding nodes. This led to two files, a consolidated edge file and a node file that contained all nodes in the edge file, which together constituted a terpenoid gene network for each species.

Text mining for ancillary genes

The genome annotations available for *Oryza sativa* subsp. Japonica cv. Nipponbare, *Zea mays* subsp. Mays cv B73, *Sorghum bicolor* and *Hordeum vulgare* subsp. Vulgare cv Morex contained key terms that were attributed to gene types. For example, a WRKY TF would contain the term “WRKY” or IPR036576, which was the InterPro domain name for WRKY domains. By mining the annotation files for specific terms I was interested in, I assigned categories for each gene in the terpenoid gene networks. I have listed the text I mined for each category in the table below.

Table 3. Each category had a few key terms I could mine the annotation file for. I mined these terms in the general description column of the annotation file of each genome or in the InterPro columns.

Category	Mined Text
LRR / NB-ARC / WAK / LRK	Leucine-rich / NB-ARC / receptor kinase / WAK / RLCK / Lectin receptor / wall associated kinase
TPS Precursor	Solanesyl / Isopentenyl / 1-deoxy-D-xylulose / Geranylgeranyl / Isoprenoid biosynthesis / Farnesyl / prenyl transferase
Transporter	Transporter / pleiotropic drug / IPR002528
CYP450	P450, CYP450, cytochrome P450
TPS	Terpene synthase / sesquiterpene synthase / monoterpene synthase / squalene synthase
PR Protein	Pathogenesis-related / pathogenesis related / 0006952 / Allergen V5
Cytochrome B5	Cytochrome B5
SDR	Dehydrogenase
Methyltransferase	Methyltransferase
Transcription Factor	Transcription factor / IPR011598 / WRKY / IPR015495
TF Co-factor	IPR008889 / VQ

Orthologous network identification

The orthogroup table is the key output from the OrthoFinder run. It lists genes for each orthogroup, across 27 species. This list was used to filter the edge files for rice and maize, such that only orthologous genes which had copies in rice and maize were retained in the

edge file. The node tables were then filtered to contain nodes which were present in the edge files.

Network visualisation
Gene networks were visualised using the ggraph and igraph packages in R. Nodes / genes were coloured by either the module they were assigned into in the WGCNA network or by their categories (Table 3). Nodes were labelled by gene IDs and gene names where applicable.

Subnetwork visualisation
Subnetworks for each type of terpene were extracted from the overall terpenoid gene network by using the induced_subgraph package of igraph. This package extracts all associated genes / nodes from a graph when provided with a list of genes of interest. Then, ggraph was used to visualise the extracted nodes and edges.

2.2.5 Downstream analysis

TF binding site analysis
Regions of interest in each species were defined as 2kb upstream of the transcription start site. Bedtools was used to extract these regions of interest from genome files downloaded from Phytozome or Rice Annotation Project Database (RAP-DB). The MEME suite of tools, particularly the 'Find Individual Motif Occurrences' (FIMO) tool, was used to identify TF binding motifs in the regions of interest (2 kb upstream from start codon) (Bailey et al., 2015).

Subcellular localisation prediction
The protein sequences of genes of interest were obtained from the fasta files of each genome assembly. These were submitted to the online server of the DeepLoc v2.1 tool (https://services.healthtech.dtu.dk/services/DeepLoc-2.1/) (Ødum et al., 2024). The results are listed in Supplementary tables 2 and 3.

InterPro annotations of immune receptor proteins
The protein sequences of immune receptor genes were submitted to the web server of the InterPro website to annotate protein domains (Paysan-Lafosse et al., 2023). Protein domains for each gene were visualised using BioRender.

Synteny visualisation
The MCscan and jcvi tools were used with the fasta files of each genome and the bed files containing the positional information of each gene, to identify syntenic regions between two species (Tang et al., 2024; Wang et al., 2012). The default settings were used for synteny identification and for visualising macro and micro synteny.

Figure visualisation
Most figures were generated in R, linux or BioRender and then edited in Affinity Designer v1.10.5.

3. Results

3.1 Rationale and study design

Plant metabolic networks are complex; many of them combine promiscuous or specific enzymes which are spatially or temporally clustered in order to produce specific stress-induced defensive compounds. In the last decade, plants have been shown to produce defensive compounds of various types: indole alkaloids, steroidal glycoalkaloids, coumarins, hydroxamic acids, and terpenoids, to mention a few. As many of these compounds are induced by and act against agronomically relevant fungal pathogens or weeds, there is interest in elucidating their full biosynthetic and regulatory pathways. This would facilitate engineering additional resistance mechanisms into key crop species.

Many biosynthesis pathways of specialised defensive compounds branch out from primary metabolism. For example, the biosynthesis of hydroxamic acids such as benzoxazinoids branches out from tryptophan biosynthesis, whereas that of terpenoids often begins with terpene synthases acting on cytosolic FPP or plastidic GGPP. Monoterpenes, diterpenes and gibberellins are produced specifically via plastidic monoterpene synthases, CPS and KS, whereas production of other terpenes in plants occurs in cytosols via terpene synthases. In addition to terpene synthases, CYP450s and SDRs catalyse the production of various defensive terpenoids from GGPP. The momilactones, phytocassanes and oryzalexins in rice as well as the zealexins, kauralexins and dolabralexins that maize produces against biotic stresses have been well studied and their biosynthetic pathways elucidated to varying extents.

Previous work has shown that both rice and maize use orthologous CPS, KS and CYP450s in order to produce defensive diterpenes. From this derives the main hypothesis of my thesis, namely that the regulatory machinery driving the production of stress-induced terpenes is also orthologous between species, and that cross-species network conservation analysis can hence be exploited to identify yet unknown regulatory components of plant specialised metabolism.
Based on this hypothesis, I aimed to identify 3 components of the conserved regulatory machinery:

1. TFs driving terpenoid biosynthesis gene expression

2. Transporters which might transport intermediate and final compounds intra- and extracellularly

3. Immune receptors which mediate defensive signalling, leading to terpenoid production.

In order to capture such orthologous machineries, I used publicly available transcriptomic data from fungus-infected and mock-treated leaves to build gene networks, as shown in the workflow image below (Figure 3). I applied the following strict criteria for choosing the publicly available datasets used in this project.

i) The experimental design of the RNA-Seq studies includes a time course.

ii) Treatments need to be replicated.

iii) The plants used in the experiments are of a cultivar/genotype with a high-quality reference genome and genome annotation. Examples are cv. Nipponbare, one of the best assembled rice genomes, or cv. B73 for maize.

iv) The tissue used in the studies are leaves, as the compounds studied in this project are produced and accumulate in the leaves upon biotic stresses.

Table 4: A list of the publicly available datasets which were used in this study

Bioproject	Species	Tissue	Treatment
PRJNA727296	*Oryza sativa* cv Nipponbare	Leaves	Time course of leaf infection with *M. oryzae*
PRJNA352773	*Oryza sativa* cv Nipponbare	Leaves	Time course of leaf infection with *M. oryzae*
PRJNA739552	*Oryza sativa* cv Nipponbare	Leaves	Time course of leaf infection with *M. oryzae*
PRJNA868217	*Zea mays* cv B73	Leaves	Time course of leaf infection with *C. graminicola*
PRJEB10574	*Zea mays* cv B73	Leaves	Time course of leaf infection with *C. graminicola*

In total, I identified 3 studies in rice and 2 in maize which fulfilled the criteria above. The raw reads from these studies were downloaded and mapped through the nf-core/rnaseq pipeline. Samples with less than 70% of uniquely mapped reads were removed from the study, as these reads were mapped to multiple locations in the genome, disproportionately affecting gene counts. Subsequently, the per-gene read counts were corrected for batch effects using limma::removeBatchEffect. This step corrected the overall gene counts for differences which arise when using datasets from multiple studies. Thereby, when plotting the variability in gene counts in a Principal Coordinate Analysis (PCA), the key variable was the treatment of samples and not the source of the datasets. These batch effect-corrected counts were then subjected to variance stabilising transformation (VST) such that variance across samples was constant. Those counts were used to build gene networks using WGCNA.

3.2 Terpenoid gene networks in rice and maize segregate into discrete terpenoid types

The gene networks built by WGCNA rely on co-expression analyses performed with either Pearson's or Spearman's correlation of gene expression across the different treatments and time points. Both methods have been used extensively for gene co-expression analyses since microarrays and RNA-Seq were introduced, in order to identify genes which could be involved in biosynthetic pathways. For example, the gene encoding for the enzyme that catalyses the final step in the momilactone B biosynthesis pathway, *OsCYP76M14*, was identified via co-expression analysis in rice (De La Peña & Sattely, 2020). In that study, publicly available RNA-Seq data from rice blast-infected leaves were used to correlate expression of all CYP450s with that of known momilactone B biosynthesis pathway genes. The subsequent heterologous expression of the top candidate CYP450s in tobacco leaves, together with the already known biosynthetic enzymes, yielded substantial amounts of momilactone B. This example illustrates that co-expression analyses are a valuable tool for studying plant specialised metabolism. However, in contrast to co-expression analyses, gene network approaches such as WGCNA are able to identify not only novel enzymes but also TFs regulating plant specialised metabolism (He et al., 2021). This is due to TFs usually

acting as the hubs of gene networks, as they have a high correlation score with multiple genes involved in the same biosynthesis pathway.

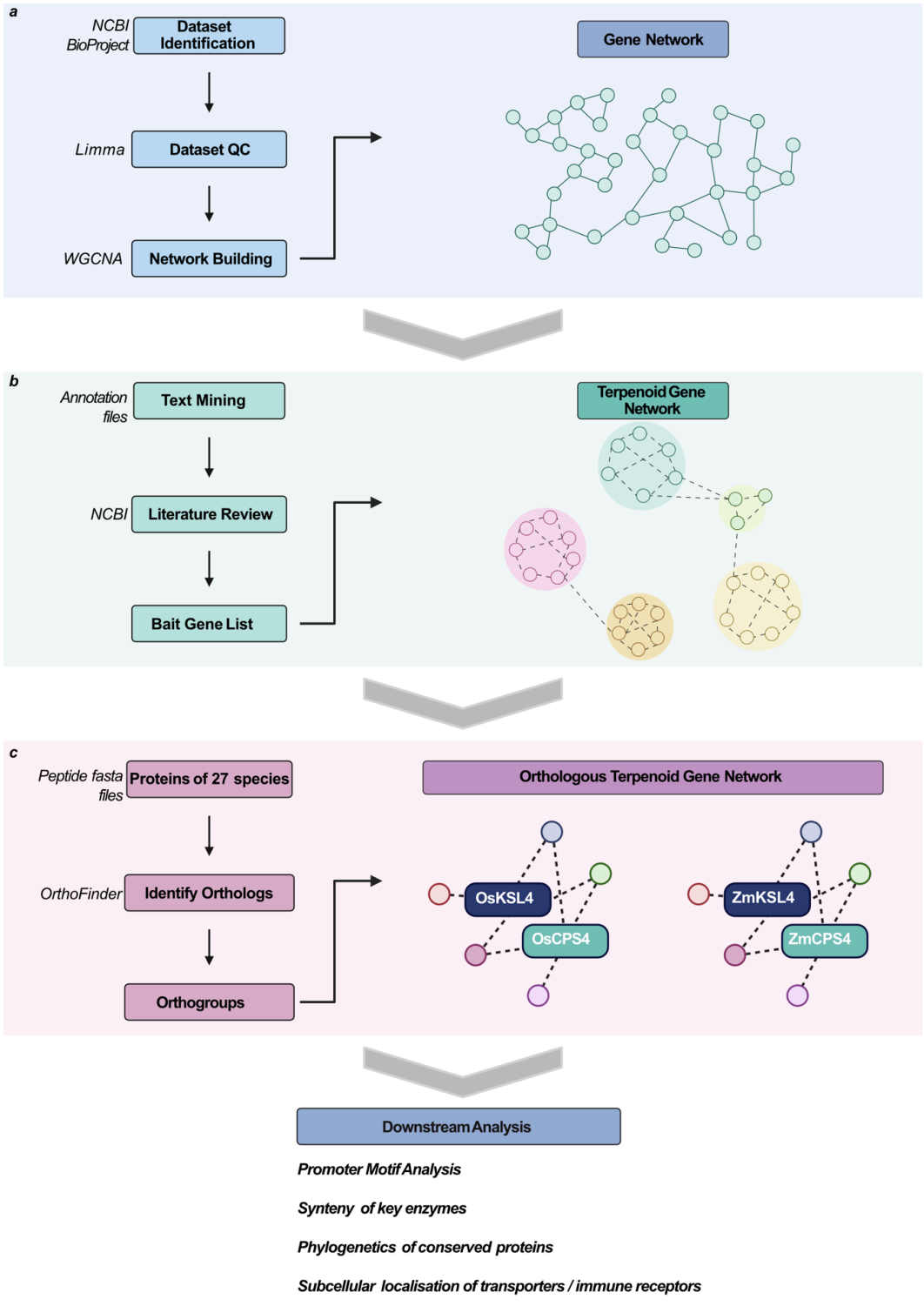


Figure 3: Overall workflow of analysis, beginning with construction of gene networks from publicly available transcriptome data and WGCNA. This is followed by curating a list of bait genes consisting of known terpene biosynthesis genes and extracting the neighbourhood of these genes from the gene network, resulting in a terpenoid gene network. Next, a phylogenetic orthology inference method, OrthoFinder, is used to identify orthologous proteins in rice and maize. The terpenoid gene networks are filtered such that only orthologous genes in rice and maize are retained, resulting in an orthologous terpenoid gene network (OTGN). The OTGNs are then analysed for genes known to encode regulatory, transport and defence proteins. Created with BioRender.com

WGCNA first correlates gene expression of all genes in all samples of an RNA-Seq dataset, yielding a matrix termed “adjacency matrix”. Next, a topological overlap matrix (TOM) is calculated from the adjacency matrix by comparing the adjacencies of every gene against all other genes. If genes A and B have similar adjacencies to the same group of genes, they end up with a high topological overlap, and correspondingly a high TOM score. The resulting TOM matrix is hierarchically clustered such that genes with high TOM scores are clustered together into modules. There are many variables in the network building process which can be altered in order to generate biologically relevant modules depending on the specific question that the study addresses. The output of the WGCNA analysis are modules of genes which behave similarly across all conditions. For example, a module of genes upregulated in disease conditions will likely contain plant defence genes encoding proteins such as NB-LRRs, PR proteins, chitinases or anti-fungal biosynthesis enzymes. The output of the gene network can be visualised as shown in Figure 4 below, linking genes within a module via edges that represent the TOM score.

Once a gene network has been built, the neighbourhood of each gene can be extracted by identifying the genes with the strongest TOM score for that gene. Since in my thesis project I was interested in identifying the gene network surrounding terpene biosynthesis, the genes of interest, or bait genes, were those encoding for enzymes known to be involved in terpene biosynthesis. Rice has a reported list of 62 terpene biosynthesis-related genes, whereas maize has 53 such genes. While most of these bait genes were terpene synthases, CPS, KS and CYP450s, there were also some SDRs and CYP450 reductases. Table 5 and 6 below show each bait gene used in this project, the associated biosynthetic pathway, their gene IDs, gene name, and references where available. The neighbourhood of each bait gene was extracted from the gene network, resulting in a terpenoid gene network, as shown in Figure 3.

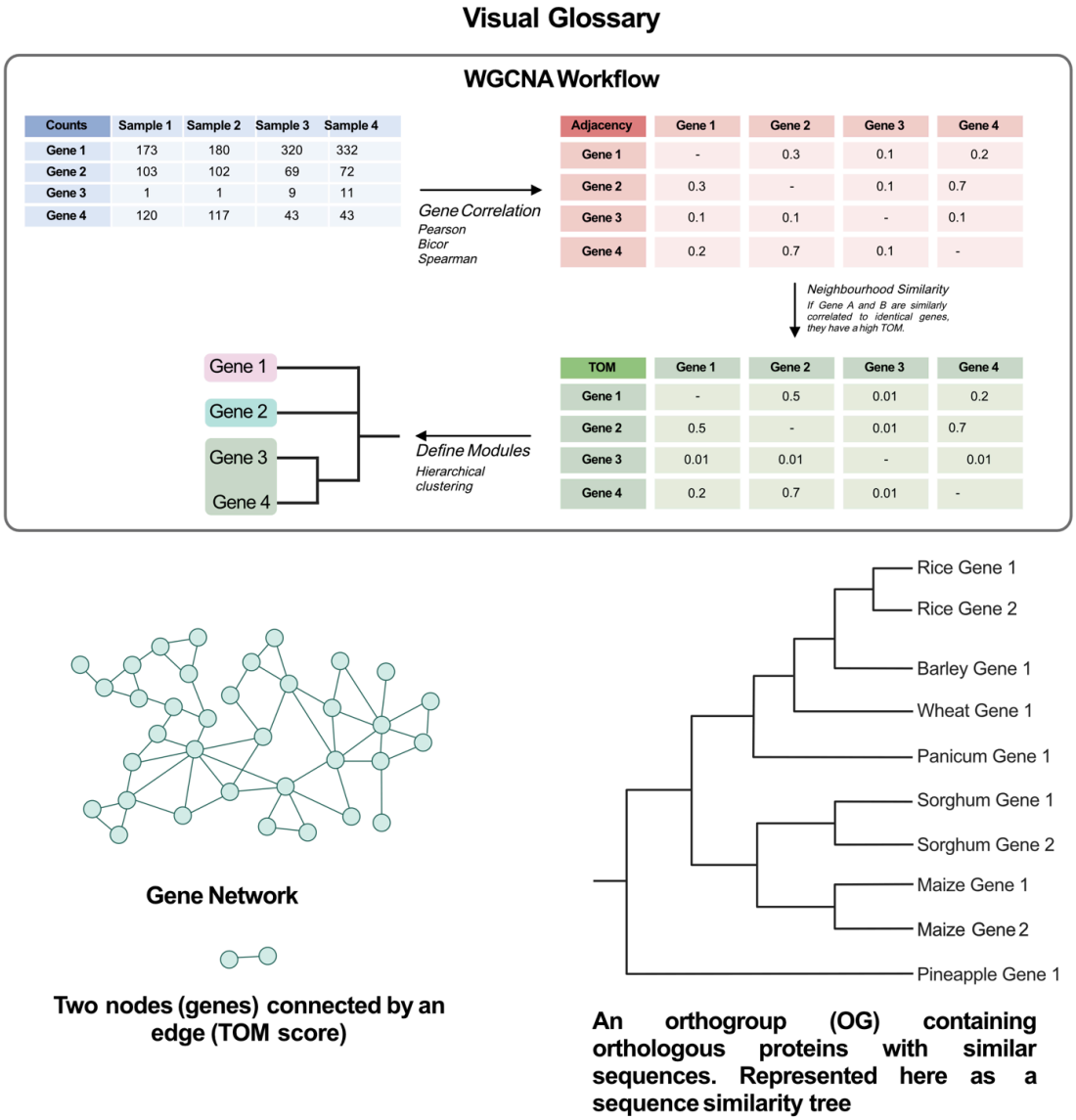


Figure 4. A visual glossary to understand key terms used in this paper. A simplified workflow of WGCNA to understand how modules of genes are defined in the approach. Key terms used in this work to interpret and describe networks and orthology are visualised and described also in the text. Created with BioRender.com

Table 5: These rice genes are known to contribute to terpenoid biosynthesis or act as terpene synthases according to literature or genome annotations. Where applicable, references are listed.

Gene ID	Gene Name	Role	Reference
OS01G0561600	CYP76M14	Diterpene	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7990393/
OS02G0568700	KSL12	Diterpene	https://academic.oup.com/bbb/article/85/9/1945/6318364?login=false
OS02G0569400	CYP76M8	Diterpene	https://www.jbc.org/article/S0021-9258(20)61105-X/fulltext
OS02G0569900	CYP76M7	Diterpene	https://www.jbc.org/article/S0021-9258(20)61105-X/fulltext
OS02G0570400	KSL7	Diterpene	https://www.tandfonline.com/doi/abs/10.1271/bbb.80044
OS02G0570500	CYP71Z6	Diterpene	https://www.sciencedirect.com/science/article/pii/S0014579311007174

OS02G0570700	CYP71Z7	Diterpene	https://www.sciencedirect.com/science/article/pii/S0014579311007174
OS02G0571100	CPS2	Diterpene	https://academic.oup.com/plphys/article/136/4/4228/6112544
OS02G0571300	KSL5	Diterpene	https://www.tandfonline.com/doi/abs/10.1271/bbb.80044
OS02G0571800	KSL6	Diterpene	https://www.tandfonline.com/doi/abs/10.1271/bbb.80044
OS02G0572050	KS7-Like	Diterpene	https://academic.oup.com/pcp/article/64/4/405/6874509?login=true
OS04G0178300	KSL4	Diterpene	https://pubmed.ncbi.nlm.nih.gov/22150231/
OS04G0178400	CYP99A3	Diterpene	https://www.jbc.org/article/S0021-9258(20)54453-0/fulltext https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-313X.2010.04408.x
OS04G0179200	OsMAS	Diterpene	https://www.jbc.org/article/S0021-9258(20)54453-0/fulltext
OS04G0179700	CPS4	Diterpene	https://onlinelibrary.wiley.com/doi/full/10.1111/ppl.12066?saml_referrer
OS04G0180400	CYP99A2	Diterpene	https://www.jbc.org/article/S0021-9258(20)54453-0/fulltext https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-313X.2010.04408.x
OS04G0611800	KS1	Diterpene	https://www.tandfonline.com/doi/epdf/10.1271/bbb.80044?needAccess=true
OS04G0612000	KS2	Diterpene	https://www.sciencedirect.com/science/article/pii/S0006291X15005604
OS06G0569500	CYP701A8	Diterpene	https://academic.oup.com/plphys/article/158/3/1418/6109264
OS07G0217600	CYP71Z2	Diterpene	https://www.sciencedirect.com/science/article/abs/pii/S0168945213000368
OS07G0218200	TPS28	Diterpene	https://www.nature.com/articles/s41477-020-00816-7
OS07G0218700	CYP71Z21	Diterpene	https://www.nature.com/articles/s41477-020-00816-7
OS07G0218900	CYP71Z30	Diterpene	https://link.springer.com/article/10.1007/s42994-022-00092-3
OS11G0474800	KSL8	Diterpene	https://www.tandfonline.com/doi/abs/10.1271/bbb.80044
OS12G0491800	KSL10	Diterpene	https://www.tandfonline.com/doi/abs/10.1271/bbb.80044
OS02G0278700	CPS1	Gibberellin	https://academic.oup.com/plphys/article/136/4/4228/6112544
OS03G0650200	CYP92C21	Homoterpene	https://onlinelibrary.wiley.com/doi/full/10.1111/pce.13924
OS02G0121700	TPS3	Linalool	https://www.sciencedirect.com/science/article/abs/pii/S0031942207002452
OS02G0458100	CAD	Monoterpene	
OS04G0344100	TPS20-1	Monoterpene	https://scholar.google.com/scholar_lookup?hl=en&volume=252&publication_year=2015&pages=997-1007&journal=Protoplasma&author=G.+Lee&author=S.+Lee&author=M.%E2%80%90S.+Chung&author=Y.+Jeong&author=B.+Chung&title=Rice+terpene+synthase20+%28OsTPS20%29+plays+an+important+role+in+producing+terpene+volatiles+in+response+to+abiotic+stresses
OS04G0345400	TPS24	Monoterpene	https://www.sciencedirect.com/science/article/pii/S0176161715002771#bib0090
OS04G0340300	TPS19	Monoterpene	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6131416/
OS01G0337100	TPS1	Sesquiterpene	https://link.springer.com/article/10.1007/s11427-022-2241-0
OS03G0348200	TPS10	Sesquiterpene	
OS03G0361700	TPS13	Sesquiterpene	https://www.sciencedirect.com/science/article/pii/S0031942207002452#bib29
OS04G0108600	TPS37	Sesquiterpene	
Os04g0339500	TPS18	Sesquiterpene	https://link.springer.com/article/10.1007/s10327-018-0774-7
OS04G0342100	STPS2	Sesquiterpene	https://thericejournal.springeropen.com/articles/10.1186/s12284-019-0274-1

OS08G0139700	TPS29	Sesquiterpene	https://onlinelibrary.wiley.com/doi/full/10.1111/j.1461-0248.2012.01835.x
OS08G0167800	TPS30	Sesquiterpene	
OS08G0168000	TPS46	Sesquiterpene	https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1365-313X.2008.03524.x?src=getftr
OS08G0168400	TPS31	Sesquiterpene	https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2008.03524.x
OS08G0223900	PTS1 / OSC12	Squalene	https://www.nature.com/articles/s41467-018-03048-8
OS09G0319800	CPS3	Squalene	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6868895/
OS11G0189600	OSC7	Squalene	https://pubs.acs.org/doi/10.1021/ol200777d
OS11G0285000	OSC8	Squalene	https://pubs.acs.org/doi/10.1021/ol200777d
OS11G0286800	OSC10	Squalene	https://nph.onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.2011.03712.x
OS11G0562100	OSC11	Squalene	https://nph.onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.2011.03712.x
OS02G0139700	OSC2	Triterpenoid	https://nph.onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.2011.03997.x
OS02G0140200	OSC3	Triterpenoid	https://nph.onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.2011.03997.x
OS02G0140400	OSC4	Triterpenoid	https://nph.onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.2011.03997.x
OS03G0347900	TPS4	Unknown	
OS03G0361100	TPS11	Unknown	
OS03G0361500	Unknown	Unknown	
OS03G0361600	TPS12	Unknown	
OS03G0362032	Unknown1	Unknown	
OS03G0362500	TPS9	Unknown	
OS03G0428200	TPS14	Unknown	
OS04G0341500	TPS20	Unknown	https://scholar.google.com/scholar_lookup?hl=en&volume=252&publication_year=2015&pages=997-1007&journal=Protoplasma&author=G.+Lee&author=S.+Lee&author=M.%E2%80%90S.+Chung&author=Y.+Jeong&author=B.+Chung&title=Rice+terpene+synthase20+%28OsTPS20%29+plays+an+important+role+in+producing+terpene+volatiles+in+response+to+abiotic+stresses
OS04G0344400	TPS23	Unknown	
OS04G0611700	KS3	Unknown	https://cir.nii.ac.jp/crid/1390001206479873152
OS10G0489500	TPS43	Unknown	

Table 6: These maize genes are known to contribute to terpenoid biosynthesis or act as terpene synthases according to literature or genome annotations. Where applicable, references are listed.

Gene ID	Gene Name	Role	Reference
ZM00001EB021200	ZmTPS38 / CPPS2 / AN2	Diterpene	https://link.springer.com/article/10.1007/s11103-005-1674-8
ZM00001EB047160	ZmTPS45 / KSL4 / KS1	Diterpene	https://academic.oup.com/plphys/article/176/4/2677/6116981?login=false
ZM00001EB133200	ZmTPS43 / KSL2 / KS4	Diterpene	https://link.springer.com/article/10.1007/s10658-018-1557-x

ZM00001EB167120	ZmTPS40/CPS4	Diterpene	https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.01542/full
ZM00001EB176190	ZmTPS42 / KSL1 / KS6	Diterpene	https://www.nature.com/articles/s41477-019-0509-6
ZM00001EB222680	ZmCYP71Z16 / Zx7 / CYP22	Diterpene	https://www.nature.com/articles/s41477-019-0509-6#Sec2
ZM00001EB300570	ZmKR2	Diterpene	https://www.nature.com/articles/s41477-019-0509-6#Sec2
ZM00001EB385070	ZmKO2	Diterpene	https://www.nature.com/articles/s41477-019-0509-6#Sec2
ZM00001EB415420	ZmTPS39/CPPS3	Diterpene	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6206430/
ZM00001EB415430	ZmTPS47 / KSL6 / KS5	Diterpene	https://www.nature.com/articles/s41477-019-0509-6
ZM00001EB432420	ZmCPR2	Diterpene	https://www.nature.com/articles/s41477-019-0509-6
ZM00001EB048020	ZmTPS37 / CPS1 / AN1	Gibberellin	https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.01542/full
ZM00001EB071070	ZmTPS44 / KS3	Gibberellin	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4734586/
ZM00001EB071080	ZmTPS46 / KSL5 / KS2	Gibberellin	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4734586/
ZM00001EB071090	ZmTPS1/KSL7	Gibberellin	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4734586/
ZM00001EB385100	ZmKO1	Gibberellin	https://www.nature.com/articles/s41477-019-0509-6#Sec2
ZM00001EB016730	ZmTPS27	Monoterpene	https://www.mdpi.com/2223-7747/12/5/1111
ZM00001EB208380	ZmTPS24	Monoterpene	
ZM00001EB208400	ZmTPS25	Monoterpene	
ZM00001EB230410	ZmTPS2	Monoterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB230440	ZmTPS3	Monoterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2

ZM00001EB267020	ZmTPS15	Monoterpene	https://link.springer.com/article/10.1007/s11103-020-01037-4
ZM00001EB278400	ZmTPS26	Monoterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB089360	ZmTPS17 / ZmEDS	Sesquiterpene	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6020683/
ZM00001EB017120	ZmTPS8	Sesquiterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB041770	ZmTPS7	Sesquiterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB374210	ZmTPS19/STC1	Monoterpene	https://academic.oup.com/plphys/article/146/3/940/6107317?login=false
ZM00001EB394330	ZmTPS21	Sesquiterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB413120	ZmTPS23	Sesquiterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB414190	ZmTPS22	Sesquiterpene	https://link.springer.com/article/10.1007/s11103-020-01037-4
ZM00001EB415070	ZmTPS9	Sesquiterpene	https://www.nature.com/articles/s41477-020-00787-9
ZM00001EB415080	ZmTPS4	Sesquiterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB415090	ZmTPS5	Sesquiterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB415100	ZmTPS31	Sesquiterpene	https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1162826/full
ZM00001EB415160	ZmTPS10	Sesquiterpene	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB416710	ZmTPS16	Sesquiterpene	https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1162826/full
ZM00001EB416720	ZmTPS20	Sesquiterpene	https://link.springer.com/article/10.1007/s11103-020-01037-4#Sec2
ZM00031AB020550	ZmTPS29	Sesquiterpene	https://link.springer.com/article/10.1186/s12864-023-09137-3

ZM00001EB061440	ZmSQSH1	Squalene	
ZM00001EB212630	ZmSQS1	Squalene	
ZM00001EB087570	ZmTPS18	Unknown	https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2023.1162826/full
ZM00001EB089110	ZmTPS14	Unknown	
ZM00001EB167050	ZmCYP71Z17	Unknown	https://www.sciencedirect.com/science/article/pii/S2095311918619215
ZM00001EB298110	ZmTPS28	Unknown	https://link.springer.com/article/10.1186/s12864-023-09137-3
ZM00001EB058050	ZmCYP31 / Zx8	Zealexins	https://www.nature.com/articles/s41477-020-00787-9
ZM00001EB058060	ZmCYP32 / Zx9	Zealexins	https://www.nature.com/articles/s41477-020-00787-9#Sec2
ZM00001EB058070	ZmCYP33 / Zx10	Zealexins	https://www.nature.com/articles/s41477-020-00787-9#Sec2
ZM00001EB222540	ZmCYP71Z19 / Zx5 / CYP29	Zealexins	https://www.mdpi.com/2223-7747/12/5/1111
ZM00001EB222660	ZmCYP71Z18 / Zx6 / CYP30	Zealexins	https://www.nature.com/articles/s41477-019-0509-6#Sec2
ZM00001EB412960	ZmTPS6 / Zx1	Zealexins	https://www.nature.com/articles/s41477-020-00787-9#Fig6
ZM00001EB412970	ZmTPS12 / Zx2	Zealexins	https://www.nature.com/articles/s41477-020-00787-9#Sec2
ZM00001EB412980	ZmTPS11 / Zx3	Zealexins	https://link.springer.com/article/10.1007/s00425-018-2999-2
ZM00001EB412990	ZmTPS13/Zx4	Zealexins	https://www.nature.com/articles/s41477-020-00787-9#Fig6

3.2.1a Rice terpenoid gene network

As shown in the workflow image above (Figure 3), I extracted the network around these bait genes (top 2 % of strongest edges) from the overall gene network in each species for further analysis. The rice terpenoid gene network segregated by terpenoid subtype as well as by biological function of the compounds (Figure 5). I observed that there were 4 major subnetworks: squalene, gibberellin, diterpene and sesquiterpene. Each subnetwork was centred around terpene synthases which are known to be tri-, di-, sesquiter- or monoterpene synthases. Squalenes are a class of plant triteprenes which are produced by triterpene synthases known as oxidosqualene cyclases (OSCs). Sesquiterpenes are known to accumulate in rice leaf tissues infected with bacterial or fungal pathogens (Cheng et al.,

2007; Taniguchi, Miyoshi, et al., 2014). Similarly, diterpenes in rice are known to accumulate at fungal sites of infection or be exuded into soil as allelochemicals (Imai et al., 2012; Kato-Noguchi & Ino, 2003). In some cases, I noticed that some subnetworks such as diterpene and sesquiterpene had overlapping neighbourhoods, i.e., the diterpene and sesquiterpene synthases were equally related to a common set of genes.

The squalene subnetwork was noticeably independent of other subnetworks, with the exception of *OsOSC10* and *OsOSC2*, which appeared to share neighbourhoods with gibberellin and sesquiterpene biosynthesis subnetworks (Figure 5). The squalene subnetwork was centred around *OsCPS3* (*OS09G0319800*), *OsOSC11* (*OS11G0562100*), *OsOSC8* (*OS11G0285000*), *OsOSC7* (*OS11G0189600*), *OsTPS19* (*OS04G0340300*), *OsTPS20* (*OS04G0341500*) and *OsTPS4* (*OS03G0347900*). Initially thought to be a pseudogene, *OsCPS3* has been implicated in the production of squalene (X.-Q. Wang et al., 2019), whereas *OsOSC11*, *OsOSC7* and *OsOSC8* have been implicated in the production of triterpenoids in in vitro systems and are only expressed in shoots (Inagaki et al., 2011; Ito et al., 2011; Xue et al., 2012). In vitro expression of *OsOSC11* in *E. coli* led to the accumulation of isoarborinol, a triterpene thought to be unique to higher plants (Lu et al., 2024) and a similar approach in yeast identified that *OsOSC7* catalysed the production of parkeol (Xue et al., 2012). Both *OsTPS19* and *OsTPS20* produce monoterpenes and sesquiterpenes, in particular (S)-limonene, with *OsTPS19* producing it in response to *M. grisea* infection, and *OsTPS20* producing it, along with a bouquet of volatile monoterpenes, in response to abiotic stresses (Chen et al., 2018; Lee et al., 2015). *OsCPS3*, *OsTPS19* and *OsTPS20* are predicted to localise to the chloroplast whereas *OsOSC11* and *OsOSC7* are predicted to localise to the peroxisome. Hence, the shared gene neighbourhoods between copalyl, monoterpene and triterpene synthases might contain common regulatory genes or subcellular trafficking genes. *OsOSC3* and *OsOSC4* have been shown to be pseudogenes (Inagaki et al., 2011), however this was performed in germinated seedlings. Therefore, considering that both genes are present in the rice gene network and that *OsOSC3* is present in the sesquiterpene subnetwork, it is highly likely that both genes lead to the production of novel defensive triterpenes upon biotic stress. Considering the clustering of squalene biosynthesis together with volatile monoterpene production, it is possible that *OsOSC7*, *OsOSC11* and *OsOSC8* are also involved in the production of volatile triterpenes. So far, only oat (*Avena sativa*) has been shown to produce defensive triterpenes in the form of avenacins (Burkhardt et al., 1964). So, the rice triterpene subnetwork may be of interest to future research focused on characterising novel defensive rice triterpenes.

The next subnetwork in the rice terpenoid gene network was the gibberellin subnetwork (Figure 5), which was largely constrained to the subnetworks around known gibberellin biosynthesis genes *OsCPS1* and *OsKS1*. Both gibberellin subnetworks are associated to a single uncharacterised sesquiterpene synthase (*OS03G0361500*) and do not share any other genes in their neighbourhoods. This is possibly due to *OsCPS1* being differentially regulated during defence responses, as plants produce less gibberellins during infection (Yimer et al., 2018). *OsKS1*, *OsKS3* and *OsKS5*, belonging to modules 0 and 3, clustered together with the triterpene synthases *OsOSC2* / *OsOSC3* / *OsCAD* and the sesquiterpene synthases *OsTPS29* and *OsCAD*. Considering the high degree of overlap in gene neighbourhood between the diterpene synthases *OsKS1*, *OsKS3*, *OsKS5*, the sesquiterpene synthase *OsCAD* and triterpene synthases *OsOSC2* and *OsOSC3*, it is likely that the genes in the shared neighbourhood between these 6 genes are involved in regulation, transport or biosynthesis of terpenes.

The genes involved in the diterpene subnetwork mostly belonged to module 6 and are annotated as CPS, KS and CYP450s known to be involved in diterpene biosynthesis. However, a single sesquiterpene synthase, *OsTPS46*, was centrally located within this subnetwork. *OsTPS46* catalyses the production of defensive sesquiterpenes in response to aphid infestation (Sun et al., 2017), therefore the shared neighbourhoods between *OsTPS46*

and the diterpene subnetwork could be due to a common set of genes encoding for biosynthesis enzymes or regulatory proteins. The genes in the diterpene subnetwork are otherwise known to be involved in biosynthesis of momilactones, phytocassanes, oryxalexins, oryzalides, and casbenes (Brown, 2016; De La Peña & Sattely, 2020; Kitaoka et al., 2015; J. Liang et al., 2021; Otomo et al., 2004; Shimura et al., 2007; Toyomasu et al., 2014; Wang et al., 2012-3; Wu et al., 2011; Xu et al., 2012; Zhan et al., 2021).

The sesquiterpene subnetwork (Figure 5) was further divided into two subnetworks, the first centred around *OsCYP92C21*, *OsTPS3* and *OsTPS30*. *OsCYP92C21* converts (E)-nerolidol and (E,E)-geranyllinalool into homoterpenes; however, the sesquiterpene synthases that produce (E)-nerolidol and (E,E)-geranyllinalool are unknown (Li et al., 2021). Since the sesquiterpene synthases *OsTPS3* and *OsTPS30* grouped together tightly with *OsCYP92C21*, it is likely that these 2 genes could produce the substrates needed for the production of volatile homoterpenes in rice. The second sesquiterpene subnetwork was centred around *OsTPS10*, *OsTPS29* and *OsOSC10*. Recombinant expression of *OsTPS29* alongside the substrate FPP resulted in the production of 5 sesquiterpenes with the major product, (E)-β-caryophyllene (Yuan et al., 2008). There is little information on the catalytic activity of *OsOSC10*, but based on phylogenetic trees, it is predicted to produce the triterpene friedelin (Yanlin Li et al., 2023; Liang et al., 2021). *OsTPS10* and *OsTPS29* were shown to produce sesquiterpenes upon fall armyworm infection and friedelins are known to function as antifeedants (Singh et al., 2023). One can therefore speculate that this second sesquiterpene subnetwork centred around *OsTPS10*, *OsTPS29* and *OsOSC10* contained genes which lead to the production and accumulation of anti-insect or more generally anti-herbivore compounds in plant tissues. Lastly, there are some sesquiterpene synthases which are unassociated with a specific sesquiterpene subnetwork. For example, *OsTPS13* encodes a sesquiterpene synthase which leads to the production of sesquiterpene alcohols from FPP (Cheng et al., 2007). Perhaps, these are sesquiterpenes which accumulate in tissues since the neighbourhood of *OsTPS13* does not overlap with either of the aforementioned sesquiterpene subnetworks.

Finally, genes such as *OsCYP71Z6*, *OsTPS13*, *OsTPS4*, *OsOSC12*, *OsKSL12*, *OS02G0572050* and *OS03G0361500* were loosely affiliated with diterpene subnetworks or appeared as independent subnetworks. This suggested that they play a minimal role in defensive terpenoid biosynthesis in response to fungal infections. In summary, the rice terpenoid gene network is segregated by terpenoid subtypes as well as volatile terpenes and terpenes reported to accumulate in tissues.

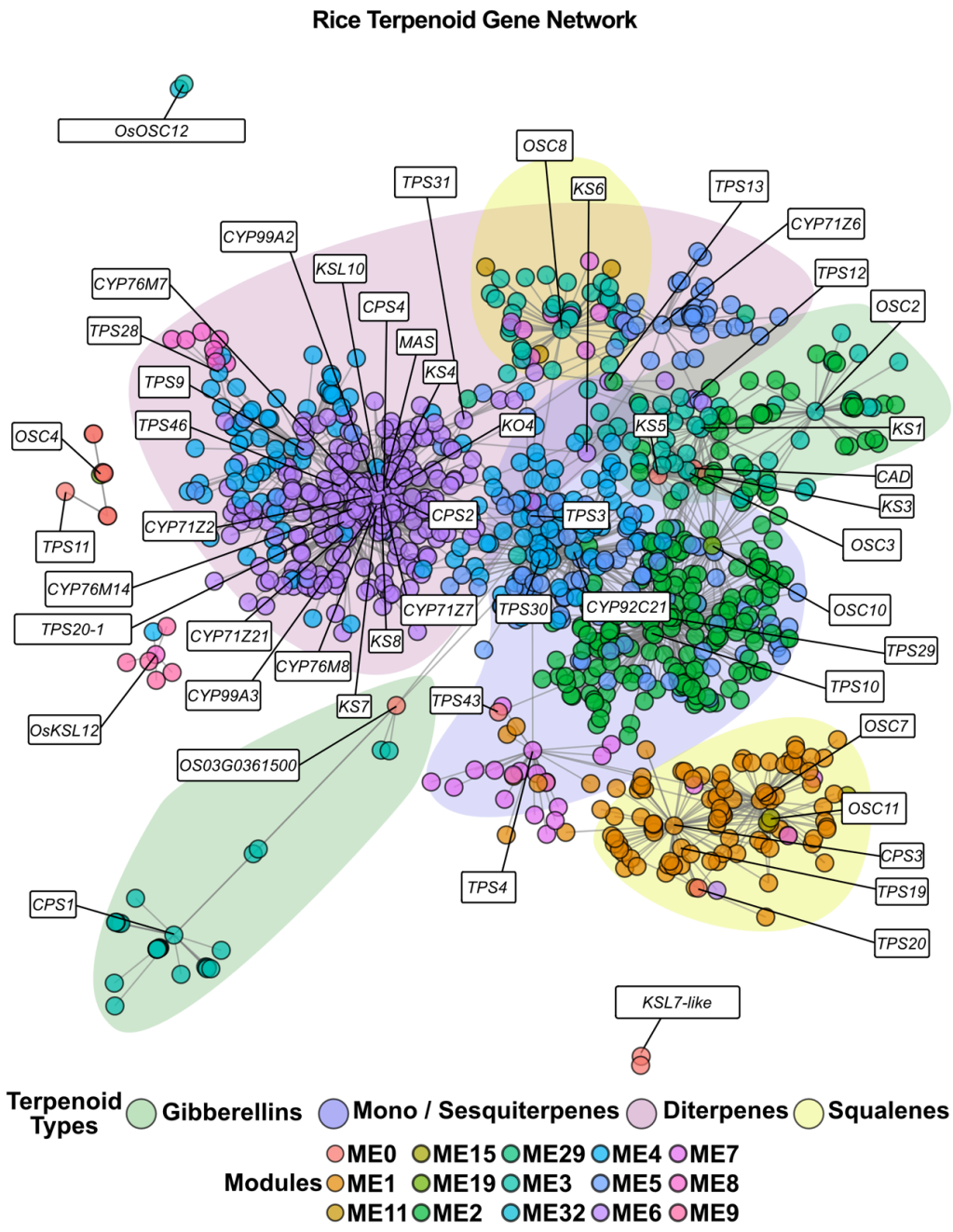


Figure 5: Top 2% of edges of each known diterpenoid biosynthesis enzyme and terpene synthase was extracted from the overall rice network and visualised. Colours of each node correspond to the modules they belong to in the gene network.

3.2.1b Maize terpenoid gene network

Similar to the rice terpenoid gene network, the maize terpenoid gene network (Figure 6) also formed distinctive subnetworks which can be associated with terpenoid subtypes. The first subnetwork contained genes known to encode for diterpene and zealexin synthases. The

second subnetwork consisted of three further subnetworks which contained genes encoding for sesquiterpene, monoterpene and gibberellin synthases (Block et al., 2019;Ding et al., 2019). Some orphan subnetworks centred around *ZmTPS13*, *ZmTPS14*, *ZmTPS20*, *ZmTPS23*, *ZmTPS28*, and *ZmTPS31*. While these genes have been annotated as sesquiterpene synthases, they belonged to a variety of modules and did not interact with any other subnetworks. This suggested that these genes likely play a minimal role in producing defensive terpenoids in maize.

The sesquiterpene subnetwork was centred around the sesquiterpene synthases ZmTPS5, 7, 17 and 20 (Figure 6). ZmTPS7 has been shown to produce the sesquiterpene τ-cadinol, and is inducibly expressed in leaves upon fungal spore inoculation and methyl jasmonate treatment (Ren et al., 2016). Although τ-cadinol was shown to significantly inhibit mycelial growth in *Cochliobolus heterostrophus* and *F. graminearum*, it is speculated that τ-cadinol is only the intermediate, and that other sesquiterpenes that *ZmTPS7* produces might be the primary ZmTPS7-derived anti-fungal compounds (Ren et al., 2016). *ZmTPS4* and *5* are closely related and produce the same types of sesquiterpenes: (Köllner et al., 2004a), 7-epi-sesquithujene, sesquithujene, (Z)-α-bergamotene, (E)-α-bergamotene, sesquisabinene B, sesquisabinene A, (E)-β-farnesene, (S)-β-bisabolene, β-curcumene, and γ-curcumene (Köllner et al., 2004a). *ZmTPS20* produces germacrene-A from farnesyl diphosphate (FPP) and *ZmTPS17* produces dihydroxylated eudesmane-2,11-diol from FPP (Liang et al., 2018; Luck et al., 2020).

The gibberellin and monoterpene subnetworks (Figure 6) were tightly associated to each other and centred around the plastidic diterpene synthases *ZmAN1, ZmKS2, ZmKS5, ZmTPS1, ZmKO1*, and *ZmKS3*, the plastidic monoterpene synthases *ZmTPS24, ZmTPS2, ZmTPS3, ZmTPS26*, and *ZmTPS15*, as well as the sesquiterpene synthases *ZmTPS8, ZmTPS18*, and *ZmTPS9*. The clustering of the diterpene synthases with the monoterpene synthases in the terpenoid gene network is likely due to both types of synthases being plastidic. Therefore, the genes in the shared neighbourhood between both types of synthases would correspond to chloroplast-localised proteins which facilitate plastidic terpene biosynthesis. The sesquiterpene synthase *ZmTPS18* is predicted to localise to the chloroplast, whereas *ZmTPS8* and *ZmTPS9* are predicted to be cytoplasmic. While this suggests that *ZmTPS18* might produce some unique plastidic sesquiterpenes, there is insufficient information about *ZmTPS18*, *ZmTPS8* and *ZmTPS9* to speculate on potential activities.

Lastly, the diterpene / zealexin subnetwork contained terpene synthases and CYP450s known to produce zealexins, a unique class of maize sesquiterpenoids, and defensive diterpenoids such as kauralexins and dolabralexins (Figure 6). The top of the diterpene / zealexin subnetwork contained the plastidic monoterpene synthase *ZmSTC1*, cytoplasmic sesquiterpene synthases, *ZmTPS22*, *ZmTPS27* and the triterpene synthases *ZmSQSH1* and *ZmSQS1* (Liu et al., 2021; Luck et al., 2020; Wan et al., 2023; Wang et al., 2023). Compared to the other genes in the diterpene subnetwork, which have been assigned to biosynthetic pathways, these five genes are less characterised and understood. Their clustering with the diterpene / zealexin genes suggested that the enzymes encoded by these five genes might be producing potent anti-fungal compounds also.

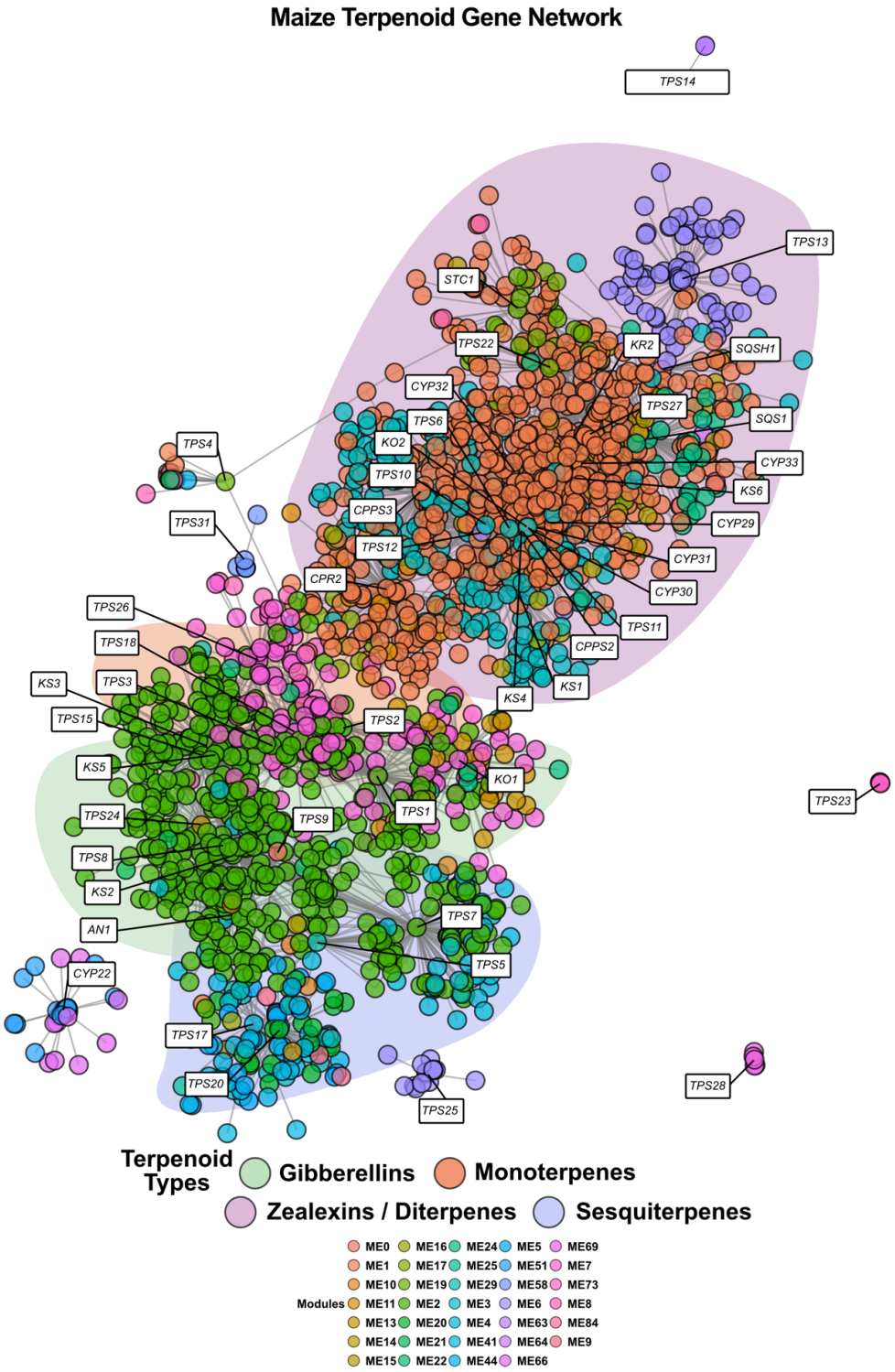


Figure 6: Top 2% of edges of each known diterpenoid biosynthesis enzyme and terpene synthase was extracted from the overall maize network and visualised. Colours of each node correspond to the modules they belong to in the gene network.

In summary, the overall terpenoid gene networks in both rice and maize segregated into subnetworks that correspond to terpenoid subtypes. In general, the volatile terpenes segregated independently from the defensive diterpenoids and zealexins in both species. Furthermore, the monoterpene and diterpene subnetworks often clustered together or overlapped to a certain extent, potentially due to a common relationship with genes associated with the chloroplast machinery.

3.2.2 Ancillary genes contribute to the biosynthesis of terpenes in rice and maize

Within the context of this project, I have defined ancillary genes as those involved in the following cellular processes; i) transcriptional regulation ii) immune perception and signalling iii) intra- and extracellular transport of metabolites. Together, ancillary genes and terpenoid biosynthesis genes are crucial for production of defensive terpenes in rice and maize. In order to identify ancillary genes, I mined the annotation files for the rice and maize genome, searching for terms such as TFs, WRKY, bHLH and assigning them as TFs. Within the scope of this project, I identified 11 categories of ancillary genes (CYP450s, LRR / NB-ARC / WAK / LRK, transporters, PR proteins, TFs, TPSes, SDRs, cytochrome B5s, methyltransferases, TF co-factors, TPS precursors) which have been associated with the 3 cellular processes I was interested in. The terpenoid gene networks in rice and maize clearly contained an abundant amount of information that could be useful to future work. Therefore, in order to derive meaningful results from the gene networks, I extracted the genes most closely associated with squalene, diterpene and sesquiterpene synthases in rice and visualised them in three separate networks (Figure 7). Similarly, I extracted the genes most closely associated with diterpene / zealexin, sesquiterpene, monoterpene, gibberellin synthases from the maize terpene gene network.

Squalene subnetwork in rice

Defensive triterpenes such as the anti-fungal compound avenacin, which is produced in the roots of Oat (*Avena sativa*), are as relevant as defensive diterpenes in the study of plant chemical defence. Thus far in rice, the products of a number of triterpene biosynthesis genes have been characterised (Inagaki et al., 2011; Xue et al., 2012) and these genes have been labelled as OSCs. So, I refer to triterpenes in rice as squalenes. The complete biosynthetic pathway for a finished triterpene product has not been identified, nor have any ancillary genes. In this section, I have briefly proposed some biosynthesis / ancillary genes that may contribute to triterpene biosynthesis in rice. The squalene subnetwork in rice (Figure 7) contained more methyltransferases than the diterpenoid subnetworks, presumably due to methyltransferases such as S-adenosyl methionine-dependent methyltransferases having been reported to be involved in monoterpene and triterpene metabolism (Lashley et al., 2022; Niehaus et al., 2012). *OS06G0315300* is a SAM dependent carboxyl methyltransferase, a category of enzymes known to mediate terpene biosynthesis. Its association with *OsOSC4* suggested that it might play a role in squalene biosynthesis. The subnetwork further included several TFs such as *OsMYB1* / Os2R_*MYB22*, which have been characterised as a phosphate starvation responsive regulator of GA biosynthesis (Gu et al., 2017). Those were not the only genes associated with phosphate starvation in the squalene subnetwork: *OsPT22* as well as *OsABCB26* are induced by phosphate starvation and are both plastidic transporters (Surhone et al., 2010). This might indicate that both of these transporters are involved in shuttling intermediates in the squalene biosynthesis pathway between organelles. *OsSWEET2a* encodes a sugar transporter which is closely related to *OsSWEET3a*, a known gibberellic acid transporter in rice (Morii et al., 2020). While this analysis highlighted several candidate genes which might play a role in squalene production, it is not exhaustive; future work is required to elucidate the role of these genes in squalene biosynthesis in rice.

Sesquiterpene production in rice

To date, sesquiterpenes in rice have been shown to be anti-insect volatiles or anti-bacterials which accumulate in tissues (Cheng et al., 2007; Kiryu et al., 2018; Xiao et al., 2012; Yuan et al., 2008; Zhan et al., 2023). In this section, I have described biosynthesis / ancillary genes which could contribute to sesquiterpene production in rice. As described in section 3.2.1a, I observed an overlap between the sesquiterpenoid and diterpenoid subnetworks (Figure 5). When visualising the rice sesquiterpene subnetwork (Figure 7), it became clear that it further separated into two subnetworks, the first centred around *OsTPS46* and *OsTPS31*, and the second *around OsTPS3, OsTPS30, OsTPS29* and *OsTPS10*.

The first sesquiterpene subnetwork shared genes with the diterpene subnetwork (Figure 7). *OsCYP76M5*/*6*/*8* and momilactone A synthase (*OsMAS*) are part of the phytocassane and momilactone biosynthesis pathway. *OsCYP716A16*, *OS03G0570100*, *OS06G0671300* and *OsSL* (*OsCYP71P1*) are novel CYP450s which have not been reported to be involved in diterpene biosynthesis. It is possible that these four CYP450s are involved in sesquiterpene biosynthesis downstream of *OsTPS31* and *OsTPS46*. Furthermore, OsDPF has been shown to directly upregulate the gene expression of *OsCPS2* and *OsCYP99A2* (Yamamura et al., 2015). Considering that *OsWRKY72*, *OsBHLH030* and *OsWRKY9* were also strongly associated with the diterpene subnetwork, it is highly likely that *OsTPS46* and *OsTPS31* share regulatory mechanisms with the diterpene subnetwork. *OsTPS31* and *OsTPS46* have been shown to produce sesquiterpenes in response to fall armyworm as well as aphids (Sun et al., 2017; Yuan et al., 2008). Overall, these findings suggest that diterpene and sesquiterpene biosynthesis in rice share a common set of TFs and enzymes, which might facilitate mounting a successful defence response against biotic stress.

The second sesquiterpene subnetwork contained 5 yet uncharacterised CYP450s potentially involved in sesquiterpene biosynthesis: *C4H2*, *Os09G0558900*, *Os10G0164500*, *Os01G0227400* and *Os08G0243500*. This subnetwork also contained genes encoding enzymes involved in the cytosolic MVA pathway that generates substrates for sesquiterpenes. This includes the genes farnesyl diphosphate synthase 1 (*FPPS1*), farnesyl diphosphate synthase 4 (*FPPS4*), isopentenyl diphosphate isomerase (IPI*),* solanesyl diphosphate synthase 2 (*SPS2*), and HMBPP synthase (*HDS*). Other biosynthesis enzymes found in the sesquiterpene subnetwork included the methyltransferases *OsCOMTL2*, *OsJMT1*, *OsSHMT2* and *OsASMT19* / *OsCOMT31*, which are all known to be variably induced by drought or salt stress and involved in lignin biosynthesis (Liang et al., 2022). Additionally, the subnetwork contained a number of short chain dehydrogenases (SDRs) such as *OsMDH5.1*, *OsMDH8.1*, *OsMAS*, *OsPDHE1A, OsCAD3*, hinting at their role in mediating sesquiterpenoid biosynthesis.

The second sesquiterpene subnetwork (Figure 7) further contained TFs such as *OsWRKY104*, which is upregulated upon rice blast infection, as well as *OsZDH1* which regulates floral transition (Cheng et al., 2015; Yoon et al., 2023). Noticeably, there were also a variety of transporters in the sesquiterpenoid subnetwork, including multi antimicrobial extrusion proteins (MATEs), pleiotropic drug resistance (PDRs), phosphate-, potassium- and boron-transporters, amino acid permeases, ABC transporters, and sugar transporters. PDRs have been shown to transport the sesquiterpene β-caroyphyllene in *Artemisinia annua* L (X. Fu et al., 2017) and sugar transporters such as *OsSWEET3a* have been shown to transport diterpenes (Morii et al., 2020). Therefore, these transporters might be involved in intracellular shuttling or extracellular secretion of sesquiterpenes and are therefore interesting candidates for future follow up experiments. Both the biotic stress-responsive *OsRLCK303* and the drought stress-responsive *OS07G0129800*, a lectin receptor kinase, suggested that sesquiterpene biosynthesis can be induced by a variety of environmental stresses (Sun et al., 2019).

Next, I more closely inspected the rice diterpene subnetwork (Figure 7), which had the highest density and contained the most interconnected genes. Since the orthologous

terpenoid gene network that is part of the diterpene subnetwork will be characterised in depth in subsequent sections, I will limit the discussion here to just the biosynthesis genes, CYP450s and SDRs. The immune receptors, PR proteins and TFs shown in the diterpene subnetwork below will be elaborated upon in the subsequent sections as they address the main objectives of my work. Among those were CYP450s known to catalyse the production of momilactones, phytocassanes and casbenes. However, I also identified several novel CYP450s such as *OsCYP716A16*, *OsCYP76M13*, *OsCYP71Z8*, *Os10g0513900* (*CYP89B9*) and *OS03G0594100* (*CYP71W3*), which are associated with *OsCYP71Z21*. Their close association with casbene biosynthesis suggests that these enzymes might catalyse the production of unknown diterpenes. Of particular interest was the subnetwork surrounding *OsCYP71Z6*. It has been shown that although both *OsCYP71Z6* and *OsCYP71Z7* are located in the phytocassane cluster on chromosome 2, only *OsCYP71Z7* is involved in phytocassane production (Wu et al., 2011). *OsCYP71Z6*, on the other hand, produces compounds which might lead to the production of oryzalides or oryzadione (Kitaoka et al., 2015; Wu et al., 2011). *OS02G0185300*, *OsCYP709C5* and *OsSDR7-6*, which were all associated with *OsCYP71Z6*, might therefore contribute to oryzalide and oryzadione biosynthesis.

In summary, in this section, I identify candidate genes which encode for biosynthesis or ancillary genes, capable of facilitating terpene biosynthesis for each terpene subtype. While the monoterpene and squalene subnetworks were significantly less complex when compared to the sesquiterpene and rice subnetworks, both subnetworks might still contain genes of interest for future research.

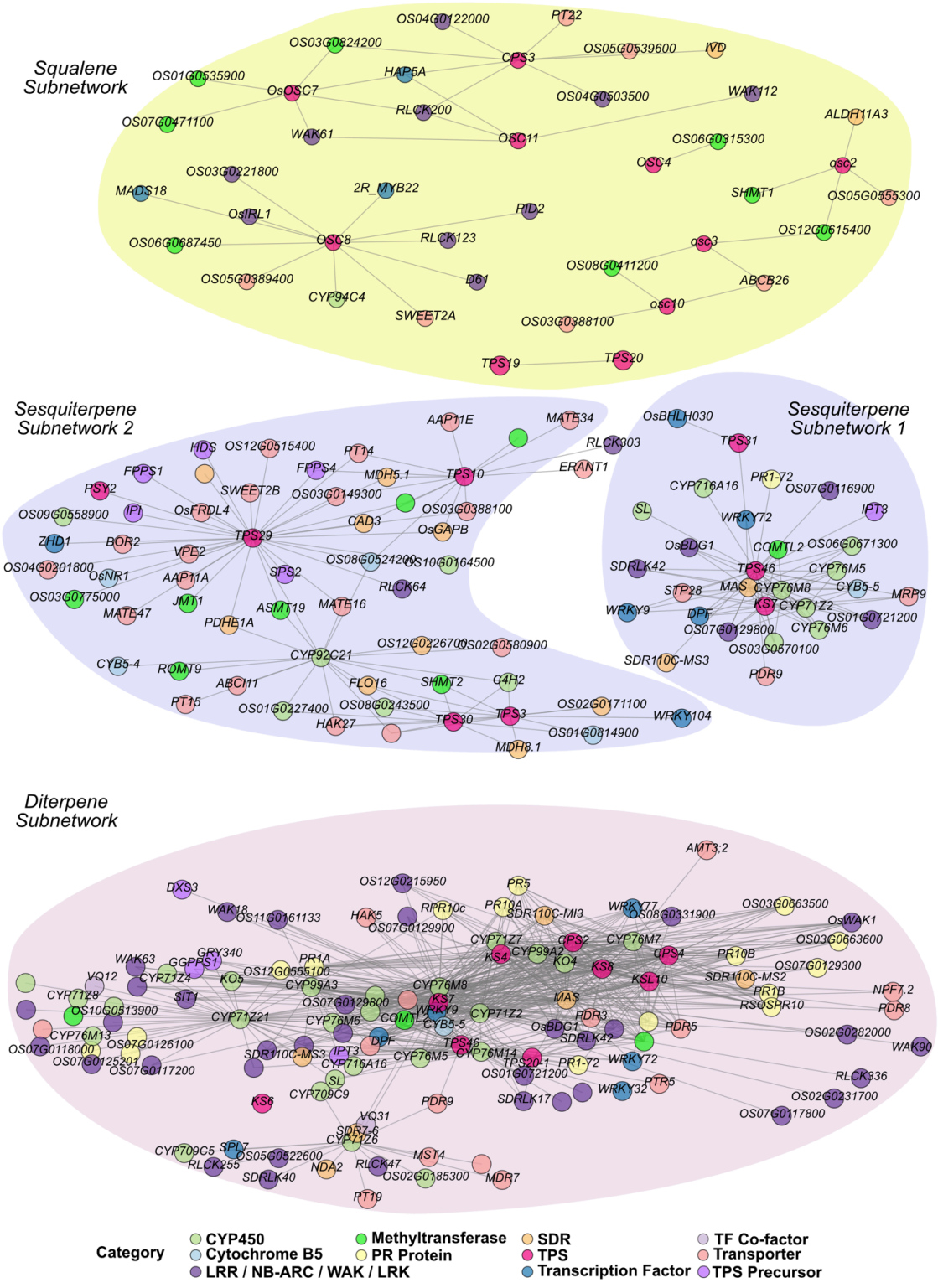


Figure 7: Top 2% of edges of each known diterpenoid biosynthesis enzyme and terpene synthase were extracted from the overall rice network; only ancillary genes were visualised. Colours of each node correspond to the categories they belong to in the gene network.

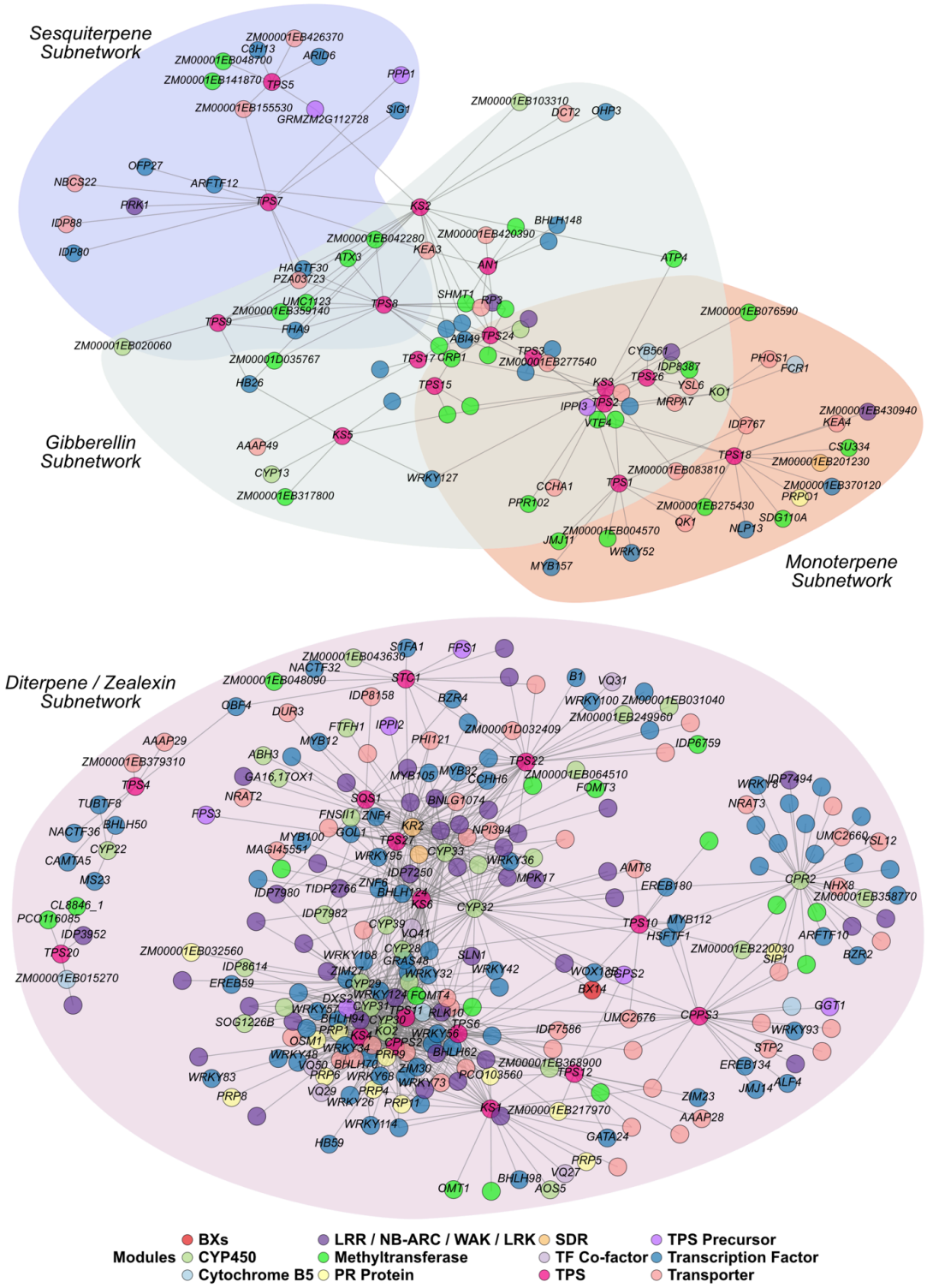


Figure 8: Top 2% of edges of each known diterpenoid biosynthesis enzyme and terpene synthase were extracted from the overall maize network; only ancillary genes were visualised. Colours of each node correspond to the categories they belong to in the gene network. The gene names of some nodes have been removed to increase readability.

Analogous to the approach applied to the rice network, I also filtered the maize terpenoid gene network to retain ancillary genes involved in the following cellular processes; i) transcriptional regulation ii) immune perception and signalling iii) intra- and extracellular transport of metabolites. The resulting monoterpene, gibberellin, sesquiterpene, and diterpene / zealexin subnetworks were thus extracted from the overall maize terpenoid gene network, filtered, and visualised (Figure 7).

Monoterpene Subnetwork in maize

ZmCYP709D1 (*IDP8387*) and *ZmCYP71C32* are the only CYP450s associated with the monoterpene synthases, suggesting that other types of enzymes might play a larger role in producing monoterpenes. *ZmIPPI3*, an isopentenyl pyrophosphate isomerase (IPPI) necessary for biosynthesis of GGPP precursors was also associated with *ZmTPS2* as well as *ZmKO1*, suggesting that of the 3 IPPI genes in maize, *ZmIPPI3* is the only one involved in monoterpene synthesis.

Methyltransferases, such as the S-adenosyl-L-methionine-dependent methyltransferases, are known to be involved in terpenoid production. There are many methyltransferases in the monoterpene subnetwork (Figure 7), including Vitamin E synthesis 4 (*ZmVTE4*), serine hydroxymethyltransferase 1 (*SHMT1*) and *ZM00001EB275150* an, S-adenosyl-L-methionine-dependent methyltransferase. As these enzymes are known to produce precursors to volatile sesquiterpenes or catalyse terpene production, future work in elucidating maize sesquiterpenes could include these as candidates (Köllner et al., 2010; Lanier et al., 2023; Lenk et al., 2012).

Of the 6 transporters in the monoterpene subnetwork, two are ABC transporters of class C and G, both known to transport defensive metabolites across the cell membrane. Lastly, the presence of a MYB TF (*ZM00001EB125300*) as well as the bHLH (*ZM00001EB202130*) suggest that these genes could be regulated in plant defence by these stress responsive TFs.

Gibberellin subnetwork in maize

The gibberellin and monoterpene subnetworks showed a high degree of overlap, probably due to both pathways originating in the plastids (Figure 7). The CYP450s *ZM00001EB103310* and *ZmCYP13* were related only to the gibberellin synthases *ZmKS5* and *ZmKS2*, respectively. QTL analysis associated *ZmCYP13* with zeaxanthin and carotenoid production (Venado et al., 2017) since both begin with plastidic GGPP, it is possible that *ZmCYP13* and *ZmKS2* are responsible for producing maize carotenoids. The gibberellin subnetwork in maize also contained methyltransferases, transporters and TFs. The majority of these genes were shared with the monoterpene and sesquiterpene subnetworks, suggesting a high degree of functional overlap between these three subnetworks. If other ancillary genes involved in plastidic processes were also studied, one might be able to identify interesting genes associated solely with gibberellin biosynthesis during biotic stress.

Sesquiterpenoid subnetwork in maize

The sesquiterpene subnetwork (Figure 7) in maize was centred around *ZmTPS7, 8, 9*, and *5* and contained a single novel CYP450 (*ZM00001EB020060*), annotated as a sesquiterpene producing enzyme (*MaizeMine: Gene Zm00001eb020060 Z. mays*, n.d.) as well as two S-adenosyl-L-methionine-dependent methyltransferases which might contribute to sesquiterpene biosynthesis. Volatiles can be sequestered in peroxisomes before emission into the environment (Shitan et al., 2023); therefore the ABC type D transporter *ZM00001EB338660* / *IDP88*, which showed association with *ZmTPS7*, is a promising candidate for acting as the facilitating transporter (Shitan et al., 2023).

Based on available literature on TFs that regulate plant specialised metabolism, the MYB, bHLH, WRKY and VQ classes of TFs are the most likely to regulate sesquiterpenoid biosynthesis (Li et al., 2024; Schluttenhofer & Yuan, 2015). In particular *ZmbZIP84*, which has been shown to act as a hub of gene regulatory networks inferred from transcriptomic data of maize leaves infected with *Puccinia sorghii* (Kim et al., 2021) that could be upregulating the production of anti-fungal sesquiterpenoids in response to infection, to be investigated in future functional studies.

Due to lack of sufficient literature, the 19 transporters in the sesquiterpenoid subnetwork could not be assigned putative functions in sesquiterpenoid production. However, based on available genome annotations, there are at least 4 ABC transporters which might mediate the extracellular accumulation of sesquiterpenes or intracellular transport of sesquiterpenes.

Diterpene / Zealexin subnetwork

The maize diterpene / zealexin subnetwork (Figure 7) could be further divided into two parts; the first was enriched for genes encoding for enzymes involved in the early stages of the diterpene / zealexin pathway. The second contained genes encoding for enzymes involved in the late stage of that pathway, which generates the highly diverse types of diterpenes and zealexins. The early-stage subnetwork was centred around *ZmTPS6/11/12 or Zx1/2/3* and *ZmCYP29/30/31* or *Zx5/6/8*. These genes are known to produce enzymes which catalyse the production of kauralexins, dolabralexins and zealexins D1, D2 and A1 (Ding et al., 2020). Subsequently, these compounds are converted into further varieties of kauralexins, dolabralexins and zealexin A and B via Zx8/9/10. The late stage diterpene / zealexin subnetwork was centred around *Zx9/10* or *ZmCYP32/ZmCYP33* as well as *ZmTPS27*, *ZmKS6*, *ZmTPS22* and *ZmKR2*. The latter encodes a reductase which catalyses the production of kauralexin A1-A4 from kauralexin B (Murphy & Zerbe, 2020), further supporting the notion that this part of the subnetwork revolves around late-stage biosynthesis. *ZmTPS27* encodes a monoterpene synthase which is known to produce geraniol in response to fungal infections (Jiang et al., 2023); one might therefore hypothesise that it utilises *ZmCYP32/33* to produce further monoterpenes, sesquiterpenes or zealexin subtypes (Jiang et al., 2023).

While many of the CYP450s that catalyse the production of zealexins are well known, 12 uncharacterised CYP450s were strongly associated with the zealexin subnetwork and could play a role in producing the diverse subtypes of zealexins or uncharacterised related metabolites. They could also be facilitating the production of geraniol via ZmTPS27, although *ZmFNSII* has been shown to produce flavone, specifically apigenin (Righini et al., 2019), whereas *ZmFTFH1* is a flavonoid 3',5'-hydroxylase. The occurrence of genes encoding for the biosynthesis of defensive metabolites such as flavonoids in the diterpene / zealexin subnetwork is probably due to upregulation of flavonoid biosynthesis upon biotic stress (Förster et al., 2022). ZmABH1 degrades abscisic acid, while ZmGA16 is a gibberellin 16,17 epoxidase which inactivates gibberellic acid (Hedden, 2020; Krochko et al., 1998). Their occurrence in this subnetwork suggested that biosynthesis of defensive terpenes co-occurs with degradation of the growth promoting gibberellins and the stress hormone abscisic acid.

The extracellular transport of zealexin could be mediated by PDRs or MATEs, both of them classes of transporters known to transport specialised metabolites across the cell membrane. All 5 PDRs in the zealexin subnetwork associated with the early-stage pathway, whereas a single MATE gene clustered with the late-stage pathway. Although this could facilitate the exudation or accumulation of specific zealexins in tissues, there is insufficient knowledge of these genes to make founded conclusions.

Whereas the monoterpene / sesquiterpene / gibberellin subnetworks displayed many methyltransferases (see above), the zealexin subnetwork contained only 7, some of which

have been characterised. Specifically, ZmFOMT3 and ZmFOMT4 are involved in the production of O-methylated flavonoids (Förster et al., 2022). ZmBX14 is an O-methyltransferase that converts DIM2BOA-Glc to HDM2BOA-Glc, the endpoint in the maize benzoxazinoid biosynthesis pathway. The presence of *ZmBX14* in the zealexin subnetwork and its close association with key diterpene / zealexin biosynthesis genes *ZmCYP32*, *ZmCPPS3*, and *ZmTPS6* suggested a common regulatory mechanism that leads to the production of defensive diterpene, zealexins, and benzoxazinoids in response to biotic stress.

Finally, the zealexin subnetwork (Figure 7) featured 9 PR proteins belonging to the PR1, PR5 and PR10 families. All were associated with the early-stage pathway subnetwork. The same 3 classes of PR proteins were also found in the rice diterpene subnetwork, suggesting that these proteins are crucial for mediating the production of defensive diterpenes and zealexins in both species upon fungal infection.

In summary, analysis of the subnetworks surrounding the genes associated with producing diterpenes, gibberellins, monoterpenes, sesquiterpenes and zealexins, revealed abundant information on candidate genes involved in biosynthesis and regulation of terpenes. Hence, this analysis lays the ground for future research that aims at identifying novel biosynthesis genes.

3.3 Terpenoid metabolism in rice and maize is driven by a conserved machinery

The main goal of this project was to identify orthologous genes involved in terpenoid production in both rice and maize which encode for proteins involved in i) immune signalling ii) transcriptional regulation iii) intracellular transport of biosynthesis intermediates and extracellular transport of defensive diterpenes and zealexins. In order to achieve that, I built orthologous terpenoid gene networks (OTGNs) and orthologous edge terpenoid gene networks (OETGNs) (illustrated below in Figure 9) from the overall terpenoid gene networks in section 3.2. Both of these networks only contained orthologous genes between rice and maize.

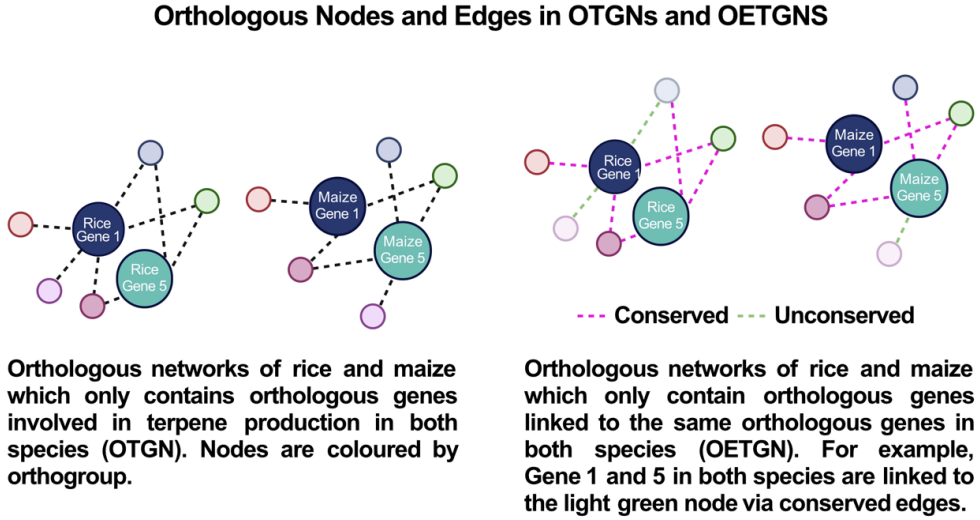


Figure 9: A representation of OTGNs and OETGNs, where nodes are genes and edges between nodes represent the strength or weight of an interaction. Genes are assigned to orthogroups so an edge which exists between the same orthogroups of genes / nodes in both rice and maize is a conserved edge.

I used OrthoFinder to compare protein sequences from 27 species of monocots, which includes 26 grasses and *Ananas comosus* (pineapple) as an outgroup. Orthofinder defines orthogroups as similar protein sequences (orthologs) across all 27 species which are descended from a last common ancestor. The pineapple genome serves as an outgroup genome, which is a reference point used to determine evolutionary relationships between the other 26 genomes. Once orthogroups were defined, I filtered the maize and rice overall terpenoid gene networks (section 3.2) such that only genes belonging to the same orthogroup were retained, yielding OTGNs.

OETGNs are an extension of OTGNs in that one can identify edges between two orthologous genes which are conserved in both OTGNs. For example, if Genes OsAA/ZmAA and OsBB/ZmBB in rice and maize belong to Orthogoups 1 and 2 respectively and are associated to each in the rice and maize OTGNs, that constitutes an orthologous edge. Therefore, by filtering the OTGNs such that only orthologous edges between orthologs were retained, I obtained OETGNs which highlight relationships conserved since rice and maize last shared a common ancestor.

The second key step in visualising the OTGNs and OETGNs was to identify genes involved in immune perception / signalling, transcriptional regulation and transport of chemicals. I used gene annotation files available for the rice and maize genome to assign genes into 11 categories of genes: CYP450s, LRR / NB-ARC / WAK / LRK, transporters, PR proteins, TFs, TPSes, SDRs, cytochrome B5s, methyltransferases, TF co-factors, TPS precursors. Genes encoding TPS Precursors, CYP450s, TPSes, SDRs, cytochrome B5s and methyltransferases would be involved in biosynthesis of terpenoids and genes encoding transporters would be involved in transporting metabolites within or out of the cell. Finally, genes encoding for LRR / NB-ARC / WAK / LRK, PR Proteins are immune receptors. Similar to the last subsection of results, the genes in these 11 categories will be referred to as ancillary genes.

Table 7: The number of nodes present in each type of orthologous network and in each species is summarised. Each node / gene is also assigned to orthogroups, which are summarised in the table below. In brackets are the number of genes or orthogroups which have been assigned to ancillary genes.

	Rice	**Rice**	**Maize**	**Maize**
	Orthologous terpenoid gene network (OTGN)	Orthologous edge terpenoid gene network (OETGN)	Orthologous terpenoid gene network (OTGN)	Orthologous edge terpenoid gene network (OETGN)
Nodes / Genes	260 (113)	129 (74)	273 (113)	119 (59)
Orthogroups	194 (61)	88 (37)	204 (65)	88 (35)

The differences in the number of categorised orthogroups between rice and maize are due to inconsistencies in annotations between rice and maize. In order to limit the impact of unavailable annotation on downstream analysis, I further defined ancillary genes as those being well annotated in both rice and maize publicly available annotation files. For example, a gene annotated as "hypothetical gene" with available InterPro domains, will not be considered an ancillary gene. This filtering step removes such noise from downstream analysis.

The table above summarises the number of genes and orthogroups found in the rice and maize OTGNs & OETGNs. When analysing the genes in the OTGNs, the ancillary genes belong to 61 orthogroups in rice and 65 in maize. This is ~30% of the orthogroups in both OTGNs, suggesting that the 147 rice and 160 maize genes belonging to the remaining 70% of orthogroups in either OTGNs could play a key role in terpenoid production. This highlights the need for accurate and better gene annotations in publicly available genome assemblies.

Key functions of genes in the rice OTGN and OETGN

Both monoterpene and gibberellin subnetworks (Figure 10) in the orthologous rice terpenoid network contained conserved edges / relationships between TPSes and genes encoding for methyltransferases (*OsSHMT1*), SDRs (*OsFLO16*), CYP450s (*OsC4H2*, *Os08g0243500*, *OsCYP71Z6*) and transporters (*OsABCB26* / *OsPTI19*). The unconserved edges in the monoterpene and gibberellin subnetworks indicate that there has been selection for new genes encoding for transporters (*OsMATE47*), SDRs (*OsMDH3.1 / OsERD1*) and cytochrome B5 reductases (*Os01g0814900*), to name a few ancillary genes. While analysing the non-diterpenoid subnetworks may lead to new findings, I here only touch on them tangentially, since the focus of this project was on the diterpenoid subnetworks.

The diterpenoid subnetwork within the rice OTGN (Figure 10) contains CYP450s belonging to 8 orthogroups and to the families CYP89B, CYP76C, CYP76M, CYP99A, CYP71Z, CYP701A and CYP73A. While most of the CYPs belonging to CYP76M, CYP99, CYP71Z and CYP701 have been well studied and assigned to diterpene biosynthesis pathways in rice, there are also some unassigned CYP450s such as *OS10G0513900* (*OsCYP89B9*), *OS10G0164500*, *OsCYP76M13*, *OsCYP71Z8*, *OsCYP71Z4*, *OS08G0243500* and OS09G0558900. These 7 genes are only mentioned in available literature as differentially regulated upon biotic or abiotic stresses. Nevertheless, perhaps their associations with some TPSes could guide future studies investigating their biological functions.

When studying the OETGN (Figure 10), the only conserved CYPs belonged to the CYP71Z, CYP89B, CYP701, and CYP73 families and were linked to other CYP450s, TPSes, transporters, NB-LRRs, PR genes, cytochrome B5, and VQ genes. This suggests that the recruitment of CYP76 and CYP99 and the subsequent gene expansion of both families in diterpene biosynthesis occurred after rice and maize diverged. This will be explored in greater detail in subsequent sections.

OsJMT1 encodes a carboxyl methyltransferase which catalyses the conversion of jasmonic acid to methyl jasmonate, a volatile defence hormone (Qi et al., 2016; Jia Wang et al., 2020). *OsSHMT1* has been implicated in lowering ROS in chloroplast processes (T. Pan et al., 2024). Finally, the maize ortholog of *OsCOMTL2*, *ZmFOMT4*, catalyses the production of various flavonoids upon fungal infection (Förster et al., 2022). In the OETGN (Figure 10), the methyltransferases *OsSHMT1* and *OsCOMTL2* were associated with kaurene synthases, CYP701, CYP71Z and sesquiterpene synthases. This suggests that both rice and maize produce defensive flavonoids alongside defensive diterpenes, sesquiterpenes and zealexins. *OsJMT1* was associated with the sesquiterpene synthase *OsTPS29*, which together with *OsTPS3* modulates herbivory behaviour on rice (Xiao et al., 2012).

Orthologous diterpenoid network in rice

The diterpenoid subnetwork in the rice OTGN (Figure 10) contained 4 TFs: *OsDPF*, *OsWRKY72, OsWRKY77* and *OsSPL7*. *OsSPL7* encodes a heat stress responsive TF which has been reported to regulate rice defences against fungal pathogens as well as heat stress and cold stress (Hoang et al., 2019). *OsDPF* encodes a basic helix loop helix TF (bHLH) which has been reported to bind directly to bHLH binding motifs upstream of *OsCYP99A2* and *OsCPS2* to drive their transcription (Yamamura et al., 2015). *OsWRKY72* and *OsWRKY77* encode WRKY TFs and have been reported to potentially negatively regulate chilling responses in rice (Viana et al., 2021) positively regulate expression of defence genes in Arabidopsis (Lan et al., 2013). VQ proteins are co-factors of WRKY proteins and aid in activating them as well as specifying their binding to DNA sequences. The rice OTGN contained two VQ genes, *OsVQ12* and *OsVQ35*. *OsVQ35* has been shown to be upregulated in rice shoots upon chilling stress and *M. oryzae* infection (Viana et al.,

2021). Altogether, these findings suggest that there is a conserved transcriptional mechanism between rice and maize that requires further investigation.

Both the OTGN and the OETGN in rice (Figure 10) contained 11 genes encoding PR proteins, belonging to 4 orthogroups and 3 PR families (PR1, PR5 and PR10). PR1 proteins have been shown to be involved in abiotic and biotic stress response, possibly by sequestering sterols or directly binding to pathogen proteins (Akbudak et al., 2020; Breen et al., 2017). Prunus domestica PR5 proteins have been shown to drive phytoalexin production in Arabidopsis (El-kereamy et al., 2011), implying a conserved role of PR5 in pathogen response. Lastly, the PR10 family of proteins in rice have been well studied in response to abiotic as well as biotic stress conditions. Some PR10 proteins in plants have been shown to have RNAse activity and some PR10 proteins have been shown to contain a hydrophobic cavity which binds ligands such as flavonoids (Morris et al., 2021). The 5 rice PR10 genes in the rice OTGNs and OETGNs belong to the latter category of ligand binding PR10 proteins, according to available genome annotation. *OsPR10a* and *OsPR10b* were shown to be transcriptionally upregulated in response to *M. grisea* infection, while *OsPR10c* did not respond and was assumed to be a pseudogene (McGee et al., 2001). *RSOsPR10* has been characterised in great detail in 2011 when Takeuchi and colleagues showed that it is specifically expressed in the roots in response to jasmonic acid, ethylene, and 1-aminocyclopropane-1-carboxylic acid, and that it accumulates in the cortex cells surrounding root vasculature. Furthermore, they showed that RSOsPR10 accumulation is suppressed by salicylic acid treatment, suggesting a complex regulation of *RSOsPR10* that is dependent on abiotic and biotic stresses (Hashimoto et al., 2004; Takeuchi et al., 2011). It seems therefore evident that the PR proteins in general, but specifically members of the PR10 family, play a key role in mediating the production of diterpenoids in response to abiotic and biotic stress, and that this role is likely conserved between rice and maize.

When analysing the genes encoding for immune receptors that could mediate the production of defensive terpenes, I consolidated all genes containing the following domains into a single category: leucine-rich repeat (LRR), nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4 (NB-ARC), wall-associated kinases (WAK), leucine-rich kinases (LRK), and S-domain subfamily of receptor-like kinases (SDRLKs). This was done because these proteins are all crucial for mounting successful immune responses against fungal pathogens (Stephens et al., 2022; Tang et al., 2017; van der Biezen & Jones, 1998).

Both the rice OTGN and OETGN (Figure 10) contained 5 orthogroups of genes belonging to this LRR / NB-ARC / WAK / LRK category. This suggests that before rice and maize diverged, these immune-response-related genes had existing relationships with the terpene biosynthesis genes. *OsSDRLK42* and *OS07G0117900* are the only two immune receptors in the rice OTGN which do not exist in the OETGN, with *OsSDRLK42* associated with 9 TPSes and CYP450s, while *OS07G0117900* was associated with *OsCYP76M14* in the OTGN.

Among the 15 LRR / NB-ARC / WAK / LRK genes in the OTGN, there were 2 SDRLKs, *OsSDRLK42* and *OsSDRLK54*. *OsSDRLK42* is associated with resistance to *B. glumae*, whereas *OsSDRLK54* is associated with cold tolerance (Naithani et al., 2021). Of the 4 remaining orthogroups of genes, *OsSIT1* is the best characterised gene; it encodes a salt-stress-responsive lectin receptor-like kinase which phosphorylates mitogen-activated protein kinases (OsMPK3 / OsMPK6) and promotes reactive oxygen species (ROS) accumulation (C.-H. Li et al., 2014). The genes in the other 3 orthogroups are annotated as lectin receptor-like kinases (*OS07G0129800*, *OS07G0129900*), NB-ARCs (*OS01G0721200, OS01G0721300, OS01G0721400, OS07G0116900, OS07G0117200, OS07G0117800, OS07G0117900, OS07G0118000*), and LRRs (*OS05G0522600, OsBDG1, OS11G0514500*). All of these immune-response-related genes were equally associated with the diterpene biosynthesis genes in the network; therefore, all of these genes and their products are

interesting targets for future studies on immune responses that trigger the production of defensive diterpenes in rice.

The diterpenoid subnetwork in the OTEGN contained 7 genes encoding transporters, distributed across 4 orthogroups, whereas the diterpenoid subnetwork in the OTGN contained 8 such genes in 5 orthogroups. There were 5 genes annotated as PDRs in the OETGN that were strongly associated with most diterpene biosynthesis genes (*OsPDR3 / 5 I 8 / 9*, and *OS01G0342750*). PDRs have been shown to be stress-responsive and to transport specialised metabolites. As such, these 5 genes are the strongest candidates for future work on understanding the transport of defensive diterpenes during stress response.

In summary, the rice OETGN and OTGN shed light on key genes which might regulate, transport and contribute to the biosynthesis of diterpenes in rice.

To dissect the roles played by ancillary genes in the OTGNs and OETGNs, there are few bioinformatic approaches available. For example, the extent to which one could validate a transcription factor in silico would be to identify the presence of its binding motifs in promoter regions of genes and to ensure that it is in open chromatin regions. Molecular biology methods on the other hand would be able to verify the importance of such motifs via luciferase assays for example, proving that the motif and transcription factor regulate the expression of any target gene. Due to such limitations of bioinformatic analyses, only some of the ancillary genes in the OTGNs and OETGNs can be explored in greater detail in subsequent sections. Transporters, immune-response-related genes, and genes encoding for enzymes can only be fully characterised via in vitro experiments, which was beyond the scope of this dissertation project. Therefore, I have utilised available bioinformatic tools to ascertain putative roles for some key ancillary genes in the OTGNs.

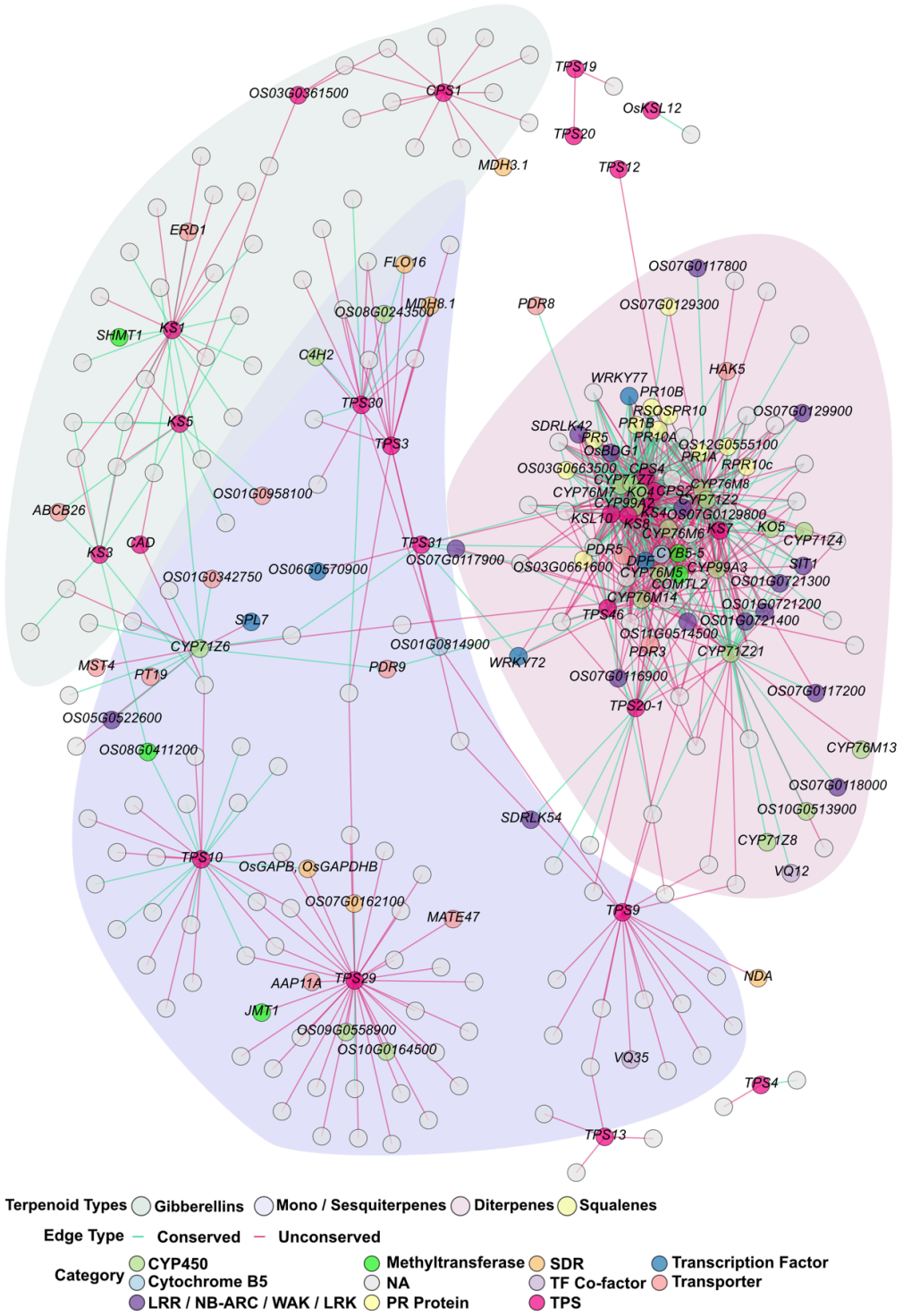


Figure 10: Top 2% of edges of each known diterpenoid biosynthesis enzyme and terpene synthase were extracted from the overall rice network and only orthologous genes were retained. Ancillary genes were visualised and colours of each node correspond to the categories they belong to.

Key functions of genes in the maize OTGN and OETGN

Figure 11 below shows the maize OTGN and OETGN. Most of the conserved edges were in the diterpenoid and gibberellin subnetworks. There were a few conserved edges linking *ZmTPS7* with genes encoding proteins involved in cellular respiration and photosynthesis. The gibberellin subnetwork in the maize OTGN was conserved for genes encoding the methyltransferases *ZmSHMT1, ZM00001EB04228, ZM00001D035767* and CYP99 (*ZM00001EB020060*). The TFs *ZmHAGTF41*, *ZmHB26* and *ZmOHP3* have not been characterised although they are all annotated as chloroplastic or thylakoidal transcriptional processes. *ZmPZA03723* is an ABC subfamily B protein, which has been shown in multiple studies to transport auxin (M. Cho & Cho, 2013) it was associated with the gibberellin, monoterpene and sesquiterpene synthases, suggesting promiscuous transport of common intermediates or substrates in the respective pathways. *ZmPZA03723* is orthologous to *OsABCB26*, which was also associated with a kaurene synthase in the rice OETGN (Figure 11). Thus, it is likely that *OsABCB26* and *ZmPZA03723* are transporting substrates or intermediates in the gibberellin biosynthesis pathway and are strong candidates for in vitro validation experiments.

The sesquiterpene subnetwork surrounding *ZmTPS5* and *ZmTPS7* was associated only with the protein kinase ZmPRK1 and the chloroplast-localised TF *ZmSIG1*. It is possible that, due to limitations in the maize genome annotation, other genes that might be crucial for sesquiterpene biosynthesis could not be identified. The sesquiterpene subnetwork around *ZmTPS9* and *ZmTPS8* was closely linked to the gibberellin subnetwork, sharing several genes with it. *ZmTPS9* was associated with the CYP99 *ZM00001EB020060*, whose ortholog is involved in momilactone biosynthesis in rice (Wang et al., 2011). I thus speculate that ZM00001EB020060 may catalyse similar oxidations of methyl groups in sesquiterpenes (Wang et al., 2011). *ZmTPS8* had 6 conserved relationships with genes which are also in the gibberellin subnetwork. This strongly suggests that *ZmTPS8* may act downstream of gibberellin biosynthesis in maize to produce novel, uncharacterised metabolites.

The monoterpene subnetwork in the maize OTGN (Figure 11) contained only unconserved edges, suggesting substantial differences in monoterpene production between rice and maize. The maize monoterpene subnetwork contained 3 genes encoding transporters (*ZmQK1*, *ZmMRPA7*, *ZM00001EB147760*). *ZM00001EB147760* is an uncharacterised phosphate transporter. *ZmMRPA7* belongs to the ABC subfamily C, a subfamily of exporters known to transport specialised metabolites and other chemicals (Pan et al., 2021). *ZmQK1* is a metal transfer protein required for iron transport in maize (Nie et al., 2021). The monoterpene subnetwork contained *ZmKO1* and *ZmFCR1*, which encode a CYP701A and a cytochrome B5 reductase respectively. Both of these proteins may facilitate monoterpene biosynthesis, although *ZmKO1* has been shown to produce gibberellins constitutively (Mao et al., 2017). There was a single SDLRK (*ZM00001EB220500*) in the monoterpene subnetwork, orthologous to *OsSDLRK42* / *54*. Overall, ~75 uncategorised genes were associated with sesquiterpenoid biosynthesis, belonging to a variety of cellular processes such as plastid biosynthesis, sugar metabolism and transcription.

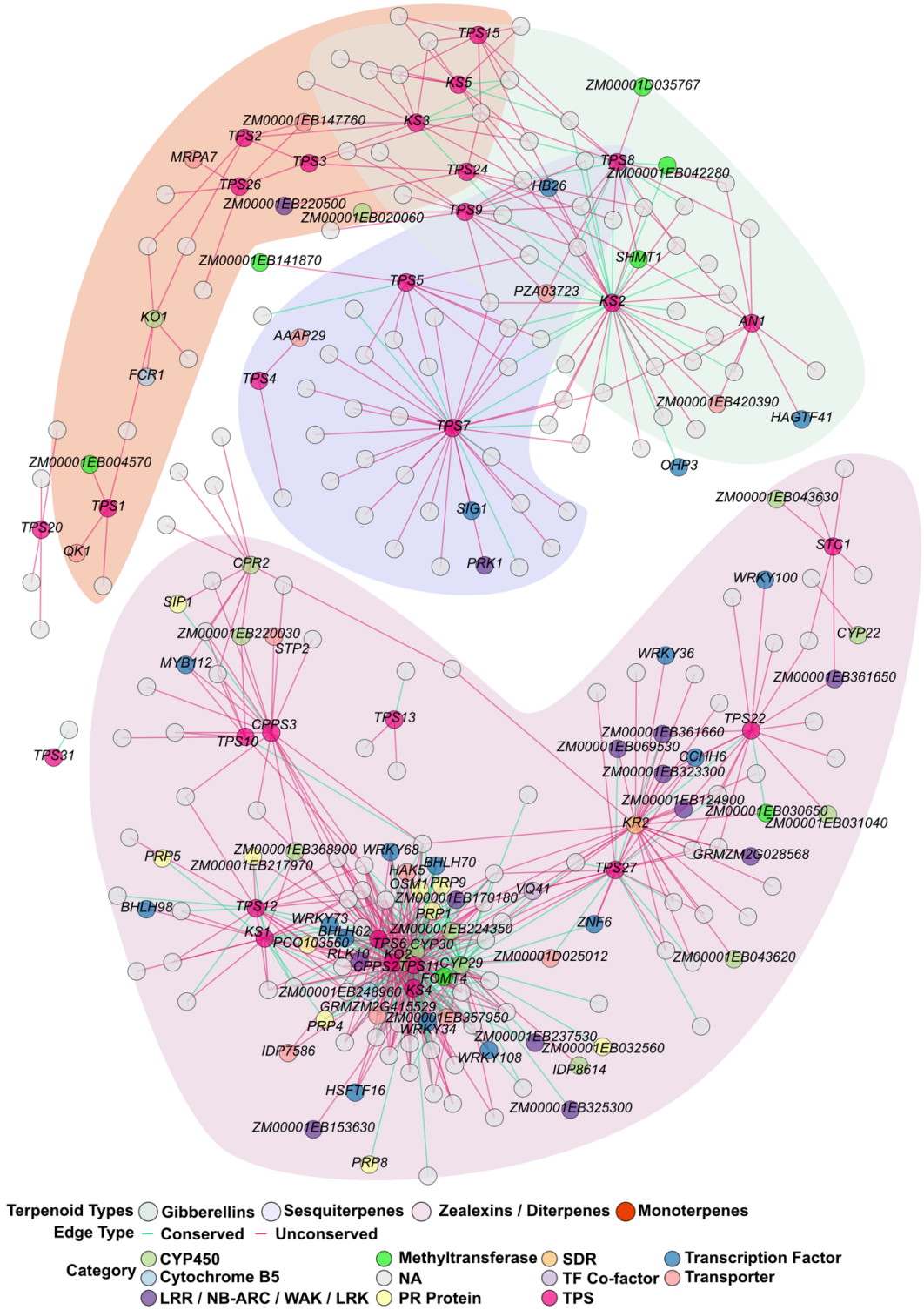


Figure 11: Orthologous terpenoid network in maize. Top 2% of edges of each known diterpenoid biosynthesis enzyme and terpene synthase were extracted from the overall maize network and only orthologous genes were retained. Ancillary genes were visualised and colours of each node correspond to the categories they belong to.

Orthologous diterpenoid network in maize

The diterpenoid subnetwork in the maize OTGN contained 12 CYP450s, of which 5 (*ZmKO2, ZmCYP29, ZmCYP30, ZM00001EB224350, ZM00001EB368900*) were in the OETGN as well. *ZM00001EB368900* (CYP73A) and *ZM00001EB224350* (CYP89B) have no reported functions in maize specialised metabolism so far. *ZM00001EB368900* was only linked with *ZmTPS12* and *ZmTPS6*, therefore the CYP73A encoded by it could be involved in sesquiterpene biosynthesis. *ZM00001EB224350* was associated only with the zealexin synthases, *ZmCYP29* and *ZmCYP30*, therefore it too is most likely involved in catalysing the various subtypes of zealexins.

In the maize OTGN (Figure 11), *ZM00001EB220030* and *IDP8614* are CYP450s which have not been associated with diterpenoid biosynthesis before. These are novel findings and makes these genes good candidates for in vitro validation for activity in dolabrelexin, kauralexin or other diterpenoid biosynthesis, especially downstream of *ZmCPPS3* and *ZmCPS4*. The presence of the known CYP450s is expected, as the CYP71Z / Zx family plays a key role in zealexin biosynthesis and are conserved in rice to produce casbene via a cluster on chromosome 7. Cytochrome B5 proteins are known cofactors of CY450s and an annotated Cytochrome B5, *ZM00001EB248960,* was strongly associated with other diterpene synthases in the orthologous network. Experimental reconstitutions of maize diterpene biosynthesis could include such cytochrome B5s to fully elucidate crucial proteins for diterpene biosynthesis. Lastly, *ZmFOMT4* is a flavonoid O-methyltransferase which methylates anti-fungal flavonoids at position 7 (Förster et al., 2022). Its occurrence in this network is probably due to diterpenes, zealexins and anti-fungal flavonoids being produced in response to fungal infections. Hence flavonoids and diterpenes may share some common regulatory elements.

10 PR genes encoding PR proteins were found in the maize OTGN and OETGN, spanning 4 orthogroups and 3 PR families, PR1, PR5 and PR10. The PR1 genes, *ZmPRP1/4/8*, are associated with *ZmKO2, ZmKS2, ZmKS4*, and *ZmKS1*, which are all diterpenoid synthases. Two PR5 proteins belonging to the orthogroup OG0000630 (*PCO103560, ZM00001EB21797*) were closely associated to *ZmCPPS2*, *ZmCYP30, ZmTPS6/11/12* and *ZmKS1*. The PR5 protein, ZmSIP1, also belonging to OG0000630 was closely associated with *ZmCPPS3* and *ZmCPR2*. *ZmCPPS3* is constitutively expressed and has been shown to produce pimara-8,14-diene together with *ZmKSL4* (Murphy et al., 2018). OG0000856 also contains 3 PR5 proteins, *ZM00001EB032560, ZmOSM1* and *ZmPRP5*. *ZmOSM1* is associated with *ZmKS4*, *ZmCPPS2, KO2, ZmCYP30, ZmTPS11* and *ZmCYP29*, whereas *ZM00001EB032560* was associated solely with *ZmCYP29* and *ZmPRP5* with *ZmKS1*. Lastly, *ZmPRP9*, belonging to the PR10 subfamily, was associated with a multitude of diterpene synthases and CYP450s such as *ZmKO2*. Overall, while the OsPR10 family has undergone expansion while being strongly associated with diterpenoid biosynthesis, maize diterpenoid biosynthesis incorporates PR5 mediated cellular interactions to a higher degree.

There were two main classes of transcription factors, bHLHs and WRKYs, which were found to be strongly associated with the diterpenoid subnetwork; the WRKYs: *ZmWRKY34 / 68 / 73 / 108* and the bHLHs: *ZmBHLH62 / 70 / 98*. *ZmBHLH70* and *ZmBHLH62* are strongly associated with diterpene biosynthesis and are orthologous to *OsDPF*, a well characterised regulator of rice diterpene biosynthesis (Yamamura et al., 2015). *ZmBHLH62* has been shown to be expressed in differentiating tissues, whereas *ZmBHLH70* has been shown to be upregulated in biotic stress (Hayford et al., 2023; Zhang et al., 2018). They are key candidate genes for downstream experiments directed towards dissecting the transcriptional regulation of diterpene / zealexin biosynthesis in maize. The 4 WRKYs associated with diterpene biosynthesis belonged to two orthogroups; *ZmWRKY108* (belonging to OG0000320) was associated with the zealexin biosynthesis genes *ZmTPS11, ZmCYP29*, *ZmCYP30* and *ZmKO2*. It has been implicated in defence response to *F. verticillioides*

(Wang et al., 2022). *ZmWRKYs 34 / 68 / 73* belong to the same orthogroup and are associated with *ZmCPPS2, ZmTPS6, ZmTPS11, ZmKS4, ZmCYP30*. *ZmWRKY73* was the only WRKY associated with *ZmKS1*. Together, these WRKYs suggest a regulatory hierarchy that might overlap due to induction by various stresses. For example *ZmWRKY34 / 68 / 73* have been shown to act in a regulatory network when maize plants were infected by *Puccinia sorghi* (Kim et al., 2021). Moreover, *ZmWRKY34* was associated with driving *ZmTPS6* and *ZmTPS11* in a meta-regulatory network (Zhou et al., 2020). The final TF worth discussing in the maize OTGN is *ZmHSFTF16*, which is orthologous to *OsSPL7* and a heat stress TF belonging to class A4. Presumably, these two genes are the abiotic stress responsive TFs which regulate diterpenoid biosynthesis in both species.

ZM00001EB357950 is a PDR (ABC-G subtype) transporter, orthologous to *OsPDR8/9* and has been found to be responsive to biotic stress as well as transcriptionally upregulated in both abiotic and biotic stress (Hayford et al., 2023; Wang et al., 2022). *GRMZM2G415529* and *ZM00001D025012* belong to the same orthogroup as *OsPDR5* and *OsPDR3*, and are closely associated with diterpenoid biosynthesis. They have both been identified as differentially expressed in the aforementioned meta-analysis of abiotic and biotic stress treated maize tissues (Hayford et al., 2023). PDRs have been shown to export defensive metabolites into the extracellular environment in Arabidopsis, and these 3 PDRs would be the strongest candidates for downstream validation for the export of defensive metabolites in maize (He et al., 2019). IDP7586 is a MATE protein, which are transporters associated with transport of metabolites within cells, across tissues, and into extracellular environments in plants (Upadhyay et al., 2019).

The final category of genes in the diterpenoid subnetwork belong to immune receptors such as LRRs, WAKs, LRKs, RLKs and NB-ARCs. *ZmRLK10* and *ZM00001EB153630* are RLKs which have been shown in rice (*OsBDR1*) to bind to mitogen activated protein kinase 3 (*OsMAPK3*) after rice blast infection and drive the production of terpenoids. *ZM00001EB069530* belongs to a class of SDRLK and the orthologs in rice have been shown to be responsive to cold stress and *B. glumae* infection (Naithani et al., 2021). *ZM00001EB069530* might therefore play a similar role here in perceiving either stress and driving the production of diterpenoids. *ZM00001EB124900* and *ZM00001EB156250* are WAK-LRKs, which have been increasingly associated with fungal response in cereals (Fan et al., 2024). *ZmRLK10* and *ZM00001EB153630* are LRR-RLKS and have been recently shown to be upregulated in fungal infection of maize (Hayford et al., 2023; Wang et al., 2022; Yan et al., 2023). While *ZmRLK10* was strongly associated with the diterpene biosynthesis genes, *ZM00001EB153630* was associated only with *ZmKS4* in the maize orthologous network. *ZM00001EB237530* and *ZM00001EB325300* are RLKs whose orthologs in rice is *OsSIT1*, a salt responsive immune receptor. Furthermore, both of these genes were peripherally associated with *ZmKS4* and *ZmCYP29* and were found to be differentially expressed in biotic as well as abiotic / biotic conditions in recently published literature (Hayford et al., 2023). The final category of genes belong to the NB-ARC, of which there are 8 genes in the rice as well as maize orthologous terpenoid subnetwork. Within the maize network, 4 genes *(ZM00001EB154670, ZM00001EB298840, ZM00001EB298800, ZM00001EB398950*) were strongly associated with diterpene biosynthesis whereas the other 4 genes (*ZM00001EB298830, ZM00001EB361660, ZM00001EB298890, ZM00001EB361650*) were associated with *ZmTPS27, ZmKR2, ZM00001EB414190* and *ZmSTC1*, which produce sesquiterpenes. Therefore, perhaps these 8 NB-ARCs could represent novel switches which regulate the production of defensive diterpenoids and zealexins in maize as well as rice. A number of the immune receptor genes listed above are also found to be upregulated in maize transcriptomes from varying fungal infections (Wang et al., 2022). In summary, these immune receptors respond to biotic or abiotic and can also respond to particular fungal pathogens in order to potentially drive the production of defensive diterpenoids / sesquiterpenoids in maize. Downstream characterisation of these proteins may be challenging if they perform redundant roles. Recent literature has shown

that NB-ARCs can bind to plasma membrane localised cysteine rich secreted proteins (CRRSPs) in order to trigger immune responses mediated by PR proteins and thereby trigger cell death (Wang et al., 2023).

The categories of genes identified and described so far in the analysis are crucial candidates that address the aim of this project: to identify genes which regulate, transport, facilitate and drive the production of defensive diterpenoids in maize and rice. In order to substantiate my findings thus far, further analyses were conducted on candidate genes.

3.4 A conserved transcriptional mechanism drives the production of defensive terpenoids

The OTGNs and OETGNs in both rice and maize contain bHLHs, WRKYs and the WRKY cofactors, VQ proteins which are known to regulate cellular responses to abiotic and biotic stresses. The bHLHs belong to two orthogroups: OG0008510, which consists of *ZmbHLH70*, *ZmbHLH62* and *OsDPF*, and OG0003325, which consists of *ZmBHLH98* and *OsBHLH30*. The WRKY cofactors, the VQs also belong to two orthogroups, OG0004644 (*ZmVQ41*, *OsVQ12*) and OG0004352 (*ZM00001EB084880*, *OsVQ35*).

OsDPF is known to directly bind to the N-boxes in the promoter regions of *OsCYP99A2* and *OsCPS2* and to activate their transcription (Yamamura et al., 2015). ZmWRKY79 is a transcriptional activator binding to W-boxes in the promoter regions of *ZmAN2* and *ZmTPS6* (Fu et al., 2018). Due to these previous reports of bHLHs and WRKYs being involved in diterpene biosynthesis in both species, I chose to focus on those two classes of TFs to understand how they might be further regulating diterpene / zealexin biosynthesis in rice and maize. I analysed the promoter regions (designated as 2 kb upstream of the transcription start site of a gene) for presence of N-boxes (5'-CACGAG-3') and G-boxes (5'-CACGTG-3') for putative bHLH binding, and W-boxes (5'-TTGACC-3') and W-box like elements (WLEs)(5'-TGACA-3') as putative WRKY binding sites.

Table 8. Distribution of bHLH motifs in promoter regions of genes in orthologous networks. Rice has 260 genes in the network, maize has 273.

Species	N-box	G-box	N-box & G-box	Neither
Rice	49	62	33	116
Maize	56	73	37	106

3.4.1 bHLH TFs regulate defensive terpenoid production via N-boxes and G-boxes

Analysis of the promoter regions of the genes in the OTGNs showed that both N-boxes and G-boxes were present in promoter regions of 82/260 genes and 95/260 genes, respectively (Table 8). Similarly, 96/273 genes in maize had at least 1 N-box in the promoter region and 111/275 genes in maize had at least 1 G-box (Table 8).

bHLHs such as OsDPF bind to their motifs via the residues His/Lys290, Glu294 and Arg298 in the basic region of the protein (Yamamura et al., 2015). An amino acid alignment of all 5 bHLHs showed that the N-box binding residues were conserved in all of them (Figure 12). Furthermore, recent research from our collaborator Kazunori Okada had also shown that *OsDPF* as well as N-boxes in momilactone biosynthesis genes had been conserved acros 5 wild rice species, *O. rufipogon, O. punctata, O. officinalis, O. brachyantha* and *L. perrieri* (Liu et al., 2024).

I identified N-box motifs in the promoter regions of *OsCPS2, OsKSL4, OsKSL7, OsKS5* and *OsKSL10*, which catalyse the initial conversion of GGPP into intermediates of the momilactone, phytocassane and oryzalexin biosynthesis. My analysis further showed that promoter regions of CYP450s known to be involved in the casbene, momilactone and phytocassane pathways, such as OsCYP71Z21, OsCYP99A3, OsCYP99A2, OsCYP76M14, and OsCYP76M7 also contained N-boxes, suggesting that momilactone and phytocassane biosynthesis is directly regulated by OsDPF.

The promoter regions of the diterpene synthase genes *OsKSL12, OsKS7, OsCPS2* and *OsKS5* contained G-box motifs. I detected G-boxes also in the promoter regions of genes encoding for known diterpene biosynthesis CYP450s: *OsCYP99A2, OsCYP76M6/7/8, OsKO5* and *OsKO4*, as well as uncharacterised enzymes such as *OsCYP71Z2/4/8* and *OsCYP89B9* (*Os10g0513900*). N-boxes and G-boxes were also found in promoter regions of genes involved in monoterpene, linalool, sesquiterpene biosynthesis.

Aside from biosynthesis genes, I identified N-boxes and G-boxes in promoter regions of conserved ancillary genes which are elaborated upon in the following subsections. The promoter regions of genes encoding the PR proteins, *OsPR1A, RSOSPR10, PR10b, RPR10c*, *PR10a* contained N-boxes and G-boxes. The immune receptor genes *OsSDRLK54, OsSDRLK42, OsBDG1, OsSIT1, OS01G0721200, OS01G0721300* and *OS01G0721400* had either G-boxes or N-boxes, same as the TFs *OsDPF, OsBHLH30, OsWRKY77* and *OsWRKY72* as well as transporters such *as OsPDR3 / 5 / 8 / 9*, *OsABCB26* and *OsMATE47*. While OsDPF has been well characterised to activate the transcription of diterpene biosynthesis genes, the data presented here strongly suggests that it may as well be regulating the transcription of ancillary genes in the diterpene biosynthesis network. I also propose that *OsBHLH30* is a strong candidate that may be binding to G-boxes, which are abundant in essential genes found in the orthologous terpenoid gene network in rice. In summary, I find that the OTGNs in rice and maize contain 2 BHLHs which can bind to motifs abundant in promoter regions of biosynthesis genes as well as genes encoding regulatory, transport and immune-related proteins. A future in-depth characterisation of both of these genes may yield novel findings about the regulatory network of diterpene biosynthesis.

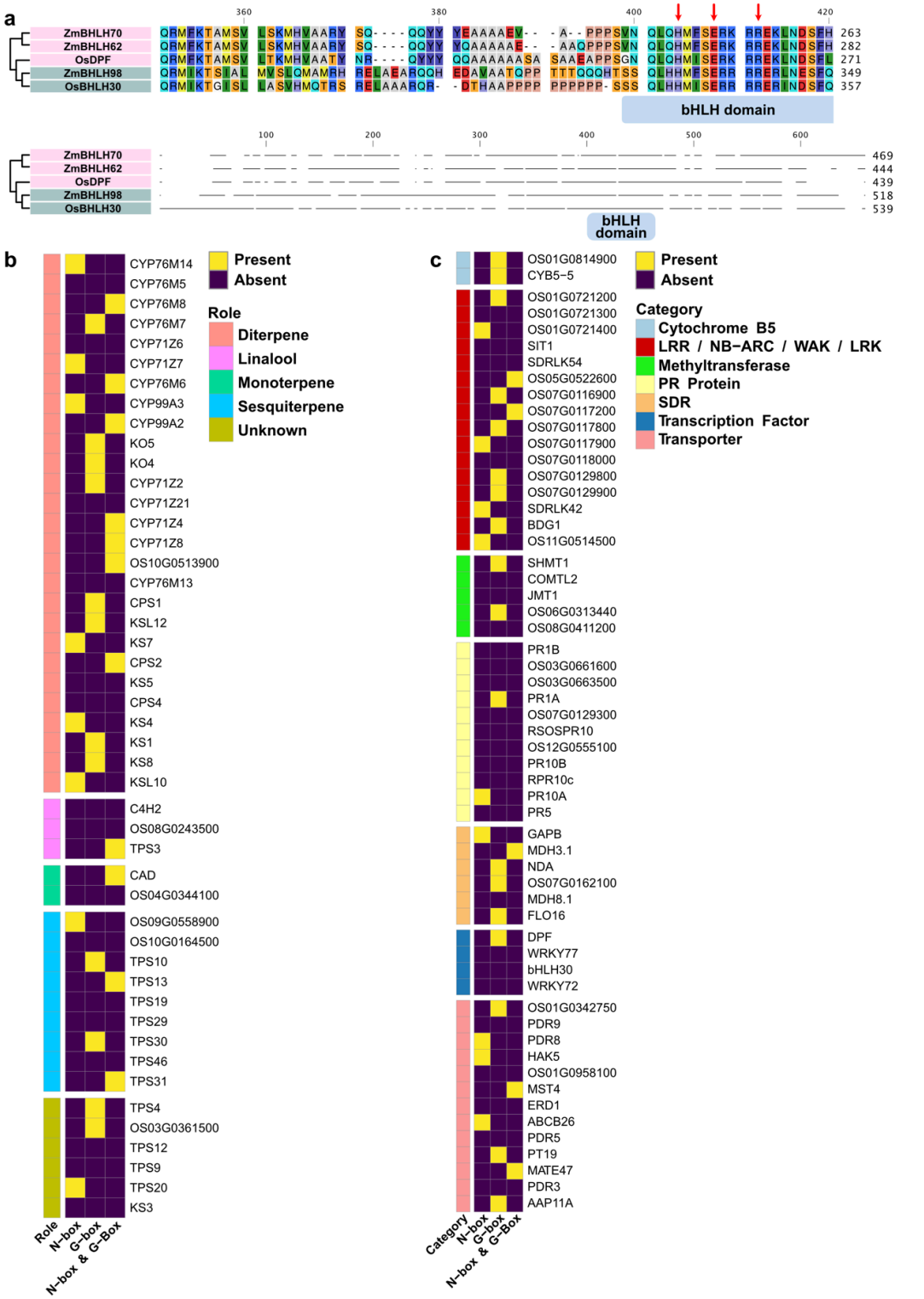


Figure 12: A) An amino acid alignment of 5 conserved bHLHs in both networks. Proteins in OG0003451 are shaded in light green whereas proteins in OG0008510 are shaded in light pink. A neighbour-joining tree shows the phylogenetic relationship between all 5 proteins. The red arrows highlight the conserved His, Glu and Arg residues which are necessary for binding to N-boxes as well as G-boxes. b) A heatmap showing the presence of N-box and G-box motifs in promoter regions of genes sorted by their roles

according to published research and annotation files. c) A heatmap showing the presence of N-box and G-box motifs in promoter regions of genes encoding proteins which might regulate, transport and facilitate defensive terpene production.

The maize OTGN contained 3 genes encoding bHLHs and two of them were in the same orthogroup as *OsDPF* and one in the same orthogroup as *OsbHLH30*. As mentioned above, the 3 key amino acid residues for binding to N-boxes and G-boxes were conserved across all 5 bHLH proteins. As such, I assumed that *ZmbHLH98*, *ZmbHLH70 and ZmbHLH62* were also binding to N-boxes and G-boxes in promoter regions of genes. When the promoter regions of the genes encoding diterpene biosynthesis enzymes in the maize OTGN were scanned for N-boxes and G-boxes, I found that N-boxes were only present in 2 genes encoding CYP450s (CYP76M and an CYP450 reductase), and *ZmKS5*. Similarly, G-boxes were found in promoter regions of 2 CYP450s (CYP76M and CYP89B), *ZmKS1* and *ZmKS4*. The rice orthologs of these 5 maize genes also had either G-boxes or N-boxes.

Of the genes encoding gibberellin biosynthesis enzymes, only *ZmKO1* contained a G-box motif. Since gibberellin biosynthesis is downregulated upon biotic stress, it is logical that few of the gibberellin biosynthesis genes had bHLH binding motifs.

The promoter regions of 4 genes encoding monoterpene biosynthesis enzymes contained either N-boxes or G-boxes, which suggests regulation via bHLHs as well. Although the maize OTGN did not contain any bHLHs associated with monoterpene biosynthesis genes, the monoterpene subnetwork in the overall terpenoid network in maize contained *ZmbHLH148*, a potential regulator.

Of the genes assigned as sesquiterpene synthases in the figure above, only *ZmTPS4 / 5 / 7 / 8 / 9* were found in the sesquiterpene subnetwork of the maize OTGN. All genes except *ZmTPS9* contained either a G-box or an N-box, suggesting that an unknown bHLH might be regulating expression of these sesquiterpene synthases. Otherwise, 5 of the sesquiterpene synthases had a G-box or both motifs. Furthermore, 4 out of 6 zealexin biosynthesis genes had only G-box motifs. This suggests that defensive diterpene and zealexin biosynthesis in maize is regulated by bHLH proteins binding to G-box motifs.

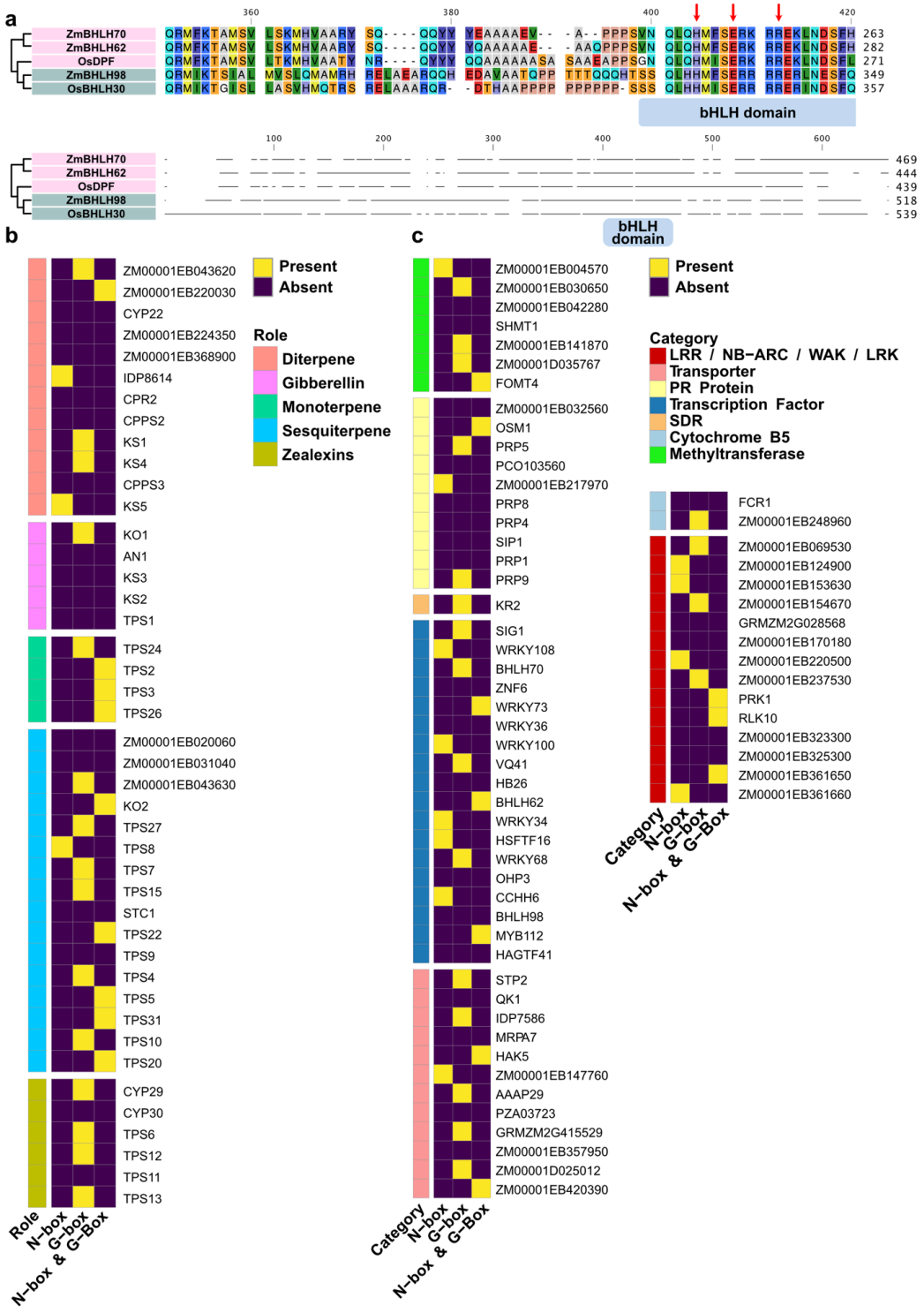


Figure 13: a) An amino acid alignment of 5 conserved bHLHs in both networks. Proteins in OG0003451 are shaded in light green whereas proteins in OG0008510 are shaded in yellow. A neighbour-joining tree shows the phylogenetic relationship between all 5 proteins. The red arrows highlight the conserved His, Glu and Arg residues which are responsible for binding to N-boxes as well as G-boxes. b) A heatmap showing the presence of N-box and G-box motifs in promoter regions of genes sorted by their roles according to published research and annotation files. c) A heatmap showing the presence of N-box and

G-box motifs in promoter regions of genes encoding proteins which might regulate, transport and facilitate defensive terpene production.

I detected G and N-box motifs also in the promoter regions of flavonoid biosynthesis genes (*ZmFOMT4*), PR5 (*ZmOSM1, ZmPRP5, ZM00001EB217970*) and PR10 (*ZmPRP9*) genes. These genes were also tightly associated with diterpene / zealexin biosynthesis in maize, hence the presence of these motifs in their promoter regions increases the likelihood of a master bHLH regulating diterpene / zealexin biosynthesis in maize. The TFs *ZmbHLH62* and *ZmbHLHh70* both had at least a single G-box motif in their promoter region. This could be indicative of a positive feedback loop or an uncharacterised master bHLH regulator that was not captured in this network. Finally, two PDR transporters (*GRMZM2G415529*, *ZM00001D025012*) and a MATE transporter (*IDP7586*) in the diterpene subnetwork of the maize OGTN had G-box motifs in their promoter regions. All of this points towards bHLHs regulating defensive diterpene / zealexin production in maize by binding to G-box motifs. I'd like to point out that while the maize OTGN only contained 3 bHLHs, the overall maize terpenoid network (Section 3.2.1b) contained 8 genes in total encoding for bHLHs. Therefore, while *ZmbHLH62 / 70 / 98* are excellent candidates for studying the regulation of diterpene / zealexin biosynthesis in maize, the remaining 5 bHLH genes in the overall maize terpenoid network should also be considered as strong candidates.

3.4.2 WRKYs and VQ proteins regulate defensive terpenoid production

Table 9. Distribution of WRKY motifs in promoter regions of genes in orthologous networks. Rice has 260 genes in the network, maize has 272.

Species	WLE	W-box	WLE & W-box	Neither
Rice	75	50	39	96
Maize	78	51	41	102

WRKYs have been shown to bind to W-boxes as well as W-box like elements (WLE) via the conserved WRKYGQK motif (Choi et al., 2015). Analysis of the promoter regions of the genes in the rice OTGN showed that both W-boxes and WLEs were present in promoter regions of 50/260 genes and 75/260 genes, respectively (Table 8). Similarly, 51/272 maize genes had at least 1 W-box in the promoter region and 78/272 maize genes had at least 1 WLE (Table 9).

WRKYs also require VQ proteins, containing a conserved FxxxVQxhTG motif that binds to the C terminal WRKY domains of group I and group IIc WRKYs in order to promote or repress DNA transcription (Cheng et al., 2012). The orthologous networks of rice and maize contained 2 and 6 WRKYs, respectively, as well as 2 VQ proteins, each classified into 2 orthogroups. *OsWRKY77* was orthologous to *ZmWRKY34*, *ZmWRKY68* and *ZmWRKY73*, whereas *OsWRKY72* was orthologous to *ZmWRKY36, ZmWRKY100* and *ZmWRKY108*. These 8 WRKYs were classified as group IIc WRKYs, containing a single C-terminal WRKY DNA-binding domain followed by a C2H2 zinc finger domain (Tang et al., 2021; Zhang et al., 2017). According to available literature spanning 3 maize reference genomes, *ZmWRKY108*, *ZmWRKY36* and *ZmWRKY68* have been shown to be drought-responsive (Wang et al., 2018; Zhang et al., 2017) whereas *OsWRKY77* and *OsWRKY72* have been previously characterised to be involved in the fungal infections and abiotic stress responses, respectively (Lan et al., 2013; Xie et al., 2005; Yu et al., 2010).

My analysis of the promoter regions of genes (Figure 14) in the rice OTGN for presence of W-boxes and WLEs showed that only 2 characterised CYP450s (*OsCYP76M5*, *OsCYP99A3*) do not contain either motif. Furthermore, of all diterpene synthases, only the

gibberellin synthase *OsKS1* and the oryzalexin S synthase *OsKS8* did not contain WRKY motifs in their promoter regions. The prevalence of WRKY binding motifs in the promoter regions of diterpene biosynthesis genes suggests that WRKYs are crucial for regulating diterpene biosynthesis in rice. Considering that both *OsWRKY72* and *OsWRKY77* have been reported to be responsive to abiotic and biotic stress respectively, and trigger the production of phytoalexins in Arabidopsis, they are excellent candidates for further validation (Lan et al., 2013; Yu et al., 2010).

W-boxes and WLEs were present in the promoter regions (Figure 14) of biosynthesis genes known to produce sesquiterpenes (*OsTPS10 / 30 / 31*). Since the overall rice terpenoid network contained 3 other WRKYs (*OsWRKY9 / 32 / 104*), it is likely that these WRKYs might regulate sesquiterpene production in rice via binding to W-boxes and WLEs in the promoter regions of *OsTPS10 / 30 / 31*.

I identified WRKY binding motifs in the promoter regions of 11 immune receptor genes. *OsSDRLK54, OsSDRLK42, OS01G0721400* and O*S05G0522600* only contained W-boxes in their promoter regions, whereas the remaining seven genes contained both motifs or only WLEs. There are 8 NB-ARCs in the orthogroup OG0000059, located in two clusters on chromosomes 1 and 7. All 3 genes in chromosome 1 (*OS01G0721200, OS01G0721300, O01G0721400*) had at least 1 WRKY binding motif, whereas only 1 of 5 genes on chromosome 7 (*OS07G0116900, OS07G0117200, OS07G0117800, OS07G0117900, OS07G0118000*) had a WRKY binding motif. Four out of five genes in the chromosome 7 cluster contained bHLH binding motifs, and 2 of 3 genes on chromosome 1 contained bHLH binding motifs. This suggests that the NB-ARCs on chromosome 1 might be regulated by both bHLHs and WRKYs, but that the NB-ARCs on chromosome 7 might be regulated by only bHLHs.

The promoter regions of *OsPR1a* as well as 4 out of 5 PR10 genes contained WRKY binding motifs (Figure 14), in contrast to only *OsPR1a* and *OsPR10a* having bHLH binding motifs (Figure 13). Similar to the NB-ARCs above, this suggests that these PR proteins in the rice OTGN are more likely to be regulated by WRKYs than by bHLHs. Furthermore, the WLE motif in the promoter region of *OsPR10a* was found to bind to salicylic induced *OsWRKY6* (Choi et al., 2015). All 4 TFs in the rice OTGN contained WRKY binding motifs, which suggests that WRKYs promote the transcription of *OsDPF* and *OsbHLH30*, thereby increasing the expression of diterpene biosynthesis genes. The key transporter-encoding genes, PDRs, *OsMATE47*, and *OsABCB26* all contain at least a single WRKY binding motif, further emphasising the potential role of WRKYs in affecting all stages of terpenoid biosynthesis in rice: perception of immune signals, transcriptional regulation, biosynthesis, and transport.

Taken together, the rice diterpene biosynthesis genes and the ancillary genes which encode for proteins mediating transcriptional regulation, transport and immune responses consistently contained WRKY binding motifs in their promoter regions, with WLE motifs being more prevalent. In some cases, there were genes which have either bHLH binding motifs or WRKY binding motifs. This might indicate a regulatory hierarchy, the further elucidation of which unfortunately has to remain beyond the scope of this project.

My confidence in WRKYs being crucial for regulating the diterpene biosynthesis machinery is also rooted in the finding of WRKY cofactors, the VQ proteins (*OsVQ12, OsVQ35, ZmVQ41, ZM00001EB084880*), in the OTGNs. There were 3 genes encoding for VQ proteins in the rice terpenoid gene networks (Section 3.2.1a) and 2 in the rice OTGN. AlphaFold3 predictions (in Supplementary Table 1) did not suggest a strong binding affinity between the predicted VQ structures and the predicted OsWRKY72 / 77 structures. However, due to a lack of deep understanding of interactions between VQs and WRKYs, as well as of the overall WRKY machinery that drives transcription, AlphaFold predictions for

interactions cannot be relied upon at this stage. *OsVQ12* was shown to be biotic stress responsive, like *OsWRKY77*, whereas *OsVQ35* was shown to be drought responsive, like *OsWRKY72* (Kim et al., 2013). Of the 39 known VQ genes in rice, only *OsVQ12* and *OsVQ35* were present in the rice OTGN, implying a conservation of these two VQs alongside the WRKYs, lending confidence to my hypothesis that the WRKY transcriptional machinery regulating terpene biosynthesis requires *OsVQ12* and *OsVQ35*.

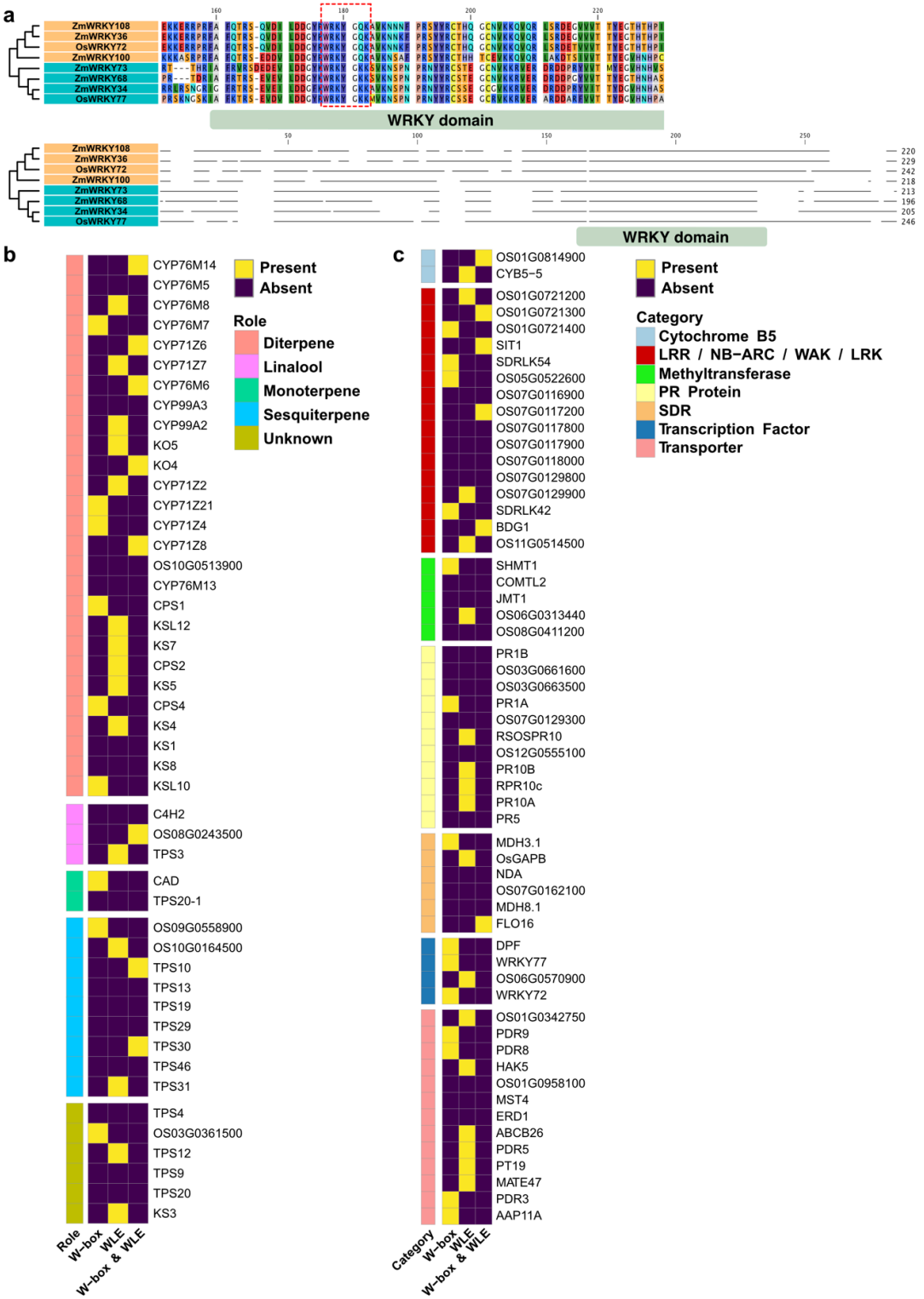


Figure 14: a) Alignment of WRKY protein sequences in the orthologous terpenoid network. WRKYs in orthogroup OG0000320 are in light orange and WRKYs in orthogroup OG0000481 are in turquoise. The yellow box highlights the conserved WRKYG(Q/K)K sequence that is necessary for binding to W-box or WLE motifs. The WRKY domain is illustrated for both the zoomed in sequence alignment and the overall protein alignment. b) A heatmap showing the presence of W-box and WLE motifs in promoter regions of

genes sorted by their roles according to published research and annotation files. c) A heatmap showing the presence of W-box and WLE motifs in promoter regions of genes encoding proteins which might regulate, transport and facilitate defensive terpene production.

When analysing the promoter regions of the maize diterpene biosynthesis genes (Figure 15), 8 of the CYP450s and TPSes contained at least a single WRKY binding motif. However, *ZmCPPS2* did not have a bHLH or WRKY binding motif in its promoter region. *ZmCPPS2* is crucial for producing dolabralexins and kauralexins (Harris et al., 2005) and might thus be regulated by other stress responsive TFs in the diterpene / zealexin subnetwork, such as basic leucine zippers (BZIPs), EREBs or MYBs, as has been shown for *ZmTPS10* (Li et al., 2015). WRKY binding motifs were found in promoters of 9 sesquiterpene biosynthesis genes in the diterpene / zealexin subnetwork, and 5 of the 6 zealexin biosynthesis genes. Overall, this suggests that WRKYs in the maize OTGN could be regulating the expression of diterpene / zealexin biosynthesis genes. WRKY motifs were also found in genes associated with biosynthesis of gibberellins, sesquiterpenes and monoterpenes.

Four out of five gibberellin biosynthesis genes contained WRKY binding motifs in their promoter regions. *ZmAN1* did not contain WRKY motifs or bHLH motifs in its promoter region, similar to *ZmCPPS2*. Promoters of 3 genes encoding monoterpene synthases contained WRKY motifs, with *ZmTPS24* and *ZmTPS2* containing both W-boxes and WLE motifs. The sesquiterpene synthases ZmTPS8 / 7 / 9 also contained at least a single WRKY binding motif. Aside from *ZmWRKY34 / 36 / 68 / 73 / 100 / 108*, the monoterpene / sesquiterpene / gibberellin subnetworks in maize contained 2 WRKYs, *ZmWRKY52* and *ZmWRKY127*, which might be involved in maize terpene biosynthesis regulation.

I identified WLE motifs in promoter regions of 4 out of 7 genes encoding methyltransferases, and a W-box in the promoter region of the flavonoid biosynthesis gene *ZmFOMT4*. W-boxes were found in the promoter regions of all PR5 genes except *ZmOSM1*, and all PR1 but not PR10 genes. This suggests that while the PR10 proteins are important immune signalling components in rice, the PR5 and PR1 proteins might be more relevant in maize. W-boxes were also present in promoter regions of 7 genes belonging to the "LRR / NB-ARC / WAK / LRK" category, and WLE motifs in promoters of 5 such genes. There were no discernable relationships between the orthogroups that these immune genes belonged to and the types of motifs found in their promoter regions. The promoter regions of the PDRs, *ZM00001EB357950*, *GRMZM2G415529, ZM00001D025012*, contained W-boxes and WLEs. 5 out of 6 WRKY genes and 2 out of 3 bHLH genes in the maize OTGN contained WRKY binding motifs.

In summary, genes encoding terpene biosynthesis enzymes as well as ancillary genes involved in cellular signalling, regulation and transport of terpenes contained WRKY motifs in their promoter regions. WLE motifs were more frequent than W-boxes in the promoter regions of genes in the maize OTGN, although the relevance of this discrepancy remains unknown.

The conservation of bHLHs and WRKYs in the OETGNs and OTGNs of both rice and maize as well as their strong relationships to known diterpene and zealexin biosynthesis genes leads me to conclude that these TFs are the most likely candidates to directly regulate diterpene and zealexin biosynthesis in rice and maize. Furthermore, the finding of WRKY and bHLH binding motifs in promoter regions of ancillary genes that are known to mediate immune signalling, regulate biosynthesis, or transport metabolites, suggests that the TFs covered in this section may directly bind and regulate expression of ancillary genes, and hence regulate diterpene-based defence systems beyond the core biosynthetic component. Finally, the conservation of genes encoding VQ proteins in both OTGNs positions them as strong candidates for in vitro binding assays to the WRKYs in the OTGNs. In summary, I find evidence for a conserved regulatory network that governs the transcriptional regulation,

biosynthesis, transport and immune signalling leading to the production of defensive diterpenes and zealexins in rice and maize.

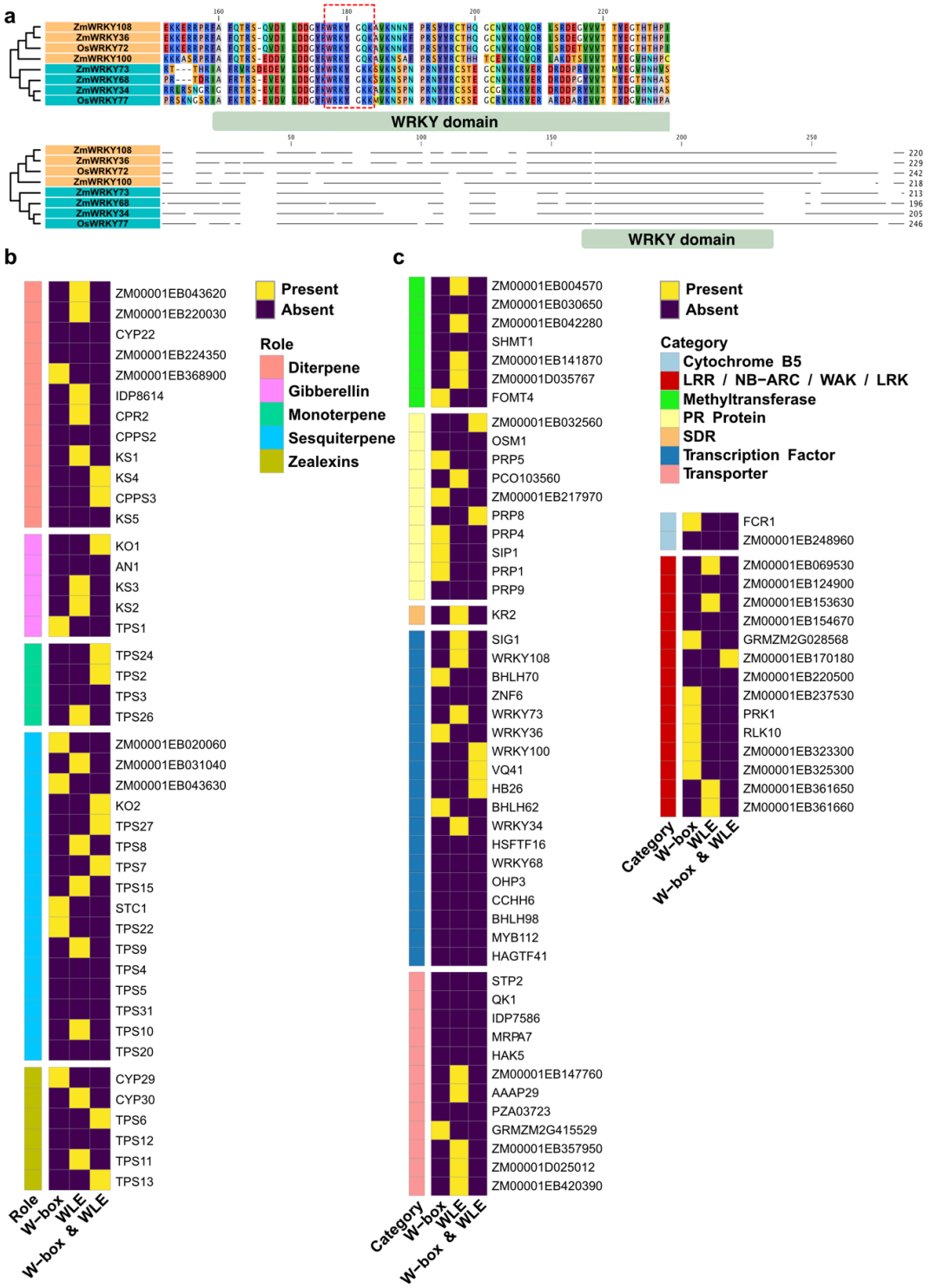


Figure 15: a) Alignment of WRKY protein sequences in the orthologous terpenoid network. WRKYs in orthogroup OG0000320 are in light orange and WRKYs in orthogroup OG0000481 are in turquoise. The yellow box highlights the conserved WRKYG(Q/K)K sequence that is necessary for binding to W-box or WLE motifs. The WRKY domain is illustrated for both the zoomed in sequence alignment and the overall

protein alignment. b) A heatmap showing the presence of W-box and WLE motifs in promoter regions of genes sorted by their roles according to published research and annotation files. c) A heatmap showing the presence of W-box and WLE motifs in promoter regions of genes encoding proteins which might regulate, transport and facilitate defensive terpene production.

3.5 Transporters in terpenoid network

Terpenoids in rice and maize have been shown to accumulate in leaves and roots during biotic stresses such as fungal infections as well as during abiotic stress such as drought (Hasegawa et al., 2010; Vaughan et al., 2015). In rice, momilactones are secreted into the surrounding environment to inhibit the growth of plants around them (Kato-Noguchi, 2004). In maize, (E)-β-caryophyllene is emitted into the rhizosphere upon root herbivory (Köllner et al., 2008) and kauralexins accumulate in roots upon seed infections with Fusarium (Veenstra et al., 2019). The active accumulation as well as exudation of defensive compounds require transporters belonging to specific classes, such as PDRs or ABC transporters (Fu et al., 2017; He et al., 2019; Pierman et al., 2017).Ten genes in the rice OTGN and 8 genes in the maize OTGN encoded for transport proteins, including PDRs and ABC transporters. In order to assign potential functions to the transporters found in the rice and maize OTGNs, I used DeepLoc2 to predict subcellular localisation of the proteins encoded by these genes (Ødum et al., 2024). These genes spanned 7 orthogroups and were the most promising candidates identified thus far in this project which could be involved in intracellular and extracellular transport of diterpenes and intermediate metabolites. Since bioinformatic approaches for characterising transporters are still developing, this section summarises available literature on each gene and assigns potential roles within terpene biosynthesis for each transporter gene.

Table 10: Subcellular localisation of transporters according to DeepLoc2 subcellular localisation predictions. Scores of predictions are listed for every protein. The orthogroup information for every gene is also listed.

Orthogroup	Rice Gene	DeepLoc Localisation / Score	Maize Gene	DeepLoc Localisation / Score
OG0000074	*OS01G0609300 (OsPDR9)*	Cell Membrane (0.7723)	*ZM00001EB357950 (ZmABCG40)*	Cell Membrane (0.7885)
OG0000074	*OS01G0609900 (OsPDR8)*	Cell Membrane (0.7386)	*ZM00001EB357950 (ZmABCG40)*	Cell Membrane (0.7885)
OG0000074	*OS01G0342750 (OsPDR16)*	Cell Membrane (0.7903)	*ZM00001EB357950 (ZmABCG40)*	Cell Membrane (0.7885)
OG0000896	*OS11G0587600 (OsPDR3)*	Cell Membrane (0.7488)	*ZM00001EB322880 (ZmABCG43)*	Cell Membrane (0.7560)
OG0000896	*OS11G0587600 (OsPDR3)*	Cell Membrane (0.7488)	*ZM00001EB419500 (ZmABCG34)*	Cell Membrane (0.7315)
OG0000896	*OS07G0522500 (OsPDR5)*	Cell Membrane (0.7360)	*ZM00001EB322880 (ZmABCG43)*	Cell Membrane (0.7560)
OG0000896	*OS07G0522500 (OsPDR5)*	Cell Membrane (0.7360)	*ZM00001EB419500 (ZmABCG34)*	Cell Membrane (0.7315)
OG0002387	*OS11G0126100 (OsMATE47)*	Lysosome / Vacuole (0.7577) & Cell Membrane (0.5991)	*ZM00001EB093580 (ZmMATE18)*	Cell Membrane (0.7806)
OG0005685	*OS03G0218400 (OsMST4)*	Cell Membrane (0.7995)	*ZM00001EB008810 (ZmSTP2)*	Cell Membrane (0.7752)

OG0008363	*OS01G0930400 (OsHAK5)*	Cell Membrane (0.7879) & Lysosome (0.6293)	*ZM00001EB142370 (ZmHAK5)*	Cell Membrane (0.7739) & Lysosome / (0.6404)
OG0012991	*OS07G0464600 (OsABCB26)*	Plastid (0.8202)	*ZM00001EB306760 (ZmABCB28)*	Plastid (0.8296)
OG0000392	*OS12G0181500 (OsAAP11A)*	Cell Membrane (0.7794)	*ZM00001EB178270 (ZmAAP14)*	Cell Membrane (0.6968)

The first orthogroup (OG0000074) contained PDRs *OsPDR9, OsPDR8, OsPDR16* and *ZM00001EB357950*. The latter has been found to be upregulated in maize leaves in response to fungal infections (Hayford et al., 2023; Lambarey et al., 2020) and no further information is currently available on this gene. *OsPDR9* expression is upregulated in rice roots for several hours in response to jasmonic acid treatment, and for a much shorter time period after salicylic acid treatment (Moons, 2008). Furthermore, expression of *OsPDR9* is induced upon PEG, cadmium, and zinc treatment, as well as under hypoxia and salt stress (Moons, 2003). Dithiothreitol, ascorbic acid and hydrogen peroxide also induced *OsPDR9* expression in rice roots. These findings led the authors to speculate that *OsPDR9* is involved in biotic and abiotic stress responses specifically in rice roots (Moons, 2003). *OsPDR8* expression in rice roots was upregulated in response to jasmonic acid treatment and rice blast infection, and in leaves infected with Rhizoctonia solani (Gupta et al., 2019; Moons, 2008). In summary, the PDRs in OG0000074 are transcriptionally upregulated in roots and leaves upon biotic and abiotic stress.

The second orthogroup (OG0000896) also consisted of genes encoding PDRs*; OsPDR3, OsPDR5, ZM00001EB322880* and *ZM00001EB419500*. *OsPDR3* and *OsDPR5* are upregulated in rice roots and shoots in response to abscisic acid, jasmonic acid and salicylic acid treatment (Gupta et al., 2019; Moons, 2008). *OsPDR5* is activated in roots upon salt stress and in leaf sheaths upon infection by *Rhizoctonia solani* (Gupta et al., 2019). There is little information on *ZM00001EB322880*. *ZM00001EB419500* is highly expressed under drought stress (Zhang et al., 2021). In summary, similar to the previous orthogroup, the PDRs in this one are also transcriptionally upregulated in roots and shoots upon biotic and abiotic stress. This is in line with available knowledge on PDRs as stress responsive transporters. All of the PDRs are predicted to localise to the plasma membrane, where they may actively transport defensive metabolites into extracellular environments.

The orthogroup OG0002387 contained *OsMATE47* and *ZmMATE18*. *OsMATE47* is upregulated in the root cortex upon iron deficiency and is also expressed constitutively in shoot tissue (Du et al., 2021; Ogo et al., 2014). There is no information available on *ZmMATE18*. MATE proteins are known to be generally stress-responsive (Upadhyay et al., 2019). Based on limited available data, I can only speculate that *OsMATE47* and *ZmMATE18* might be abiotic stress responsive, plasma membrane-localised transporters of defensive terpenes.

OG0005685 consisted of *OsMST4* and *ZmSTP2*; both are known to encode sugar transporters. Firstly, *ZmSTP2* has been shown to be important in resistance against the agronomically relevant corn fungal pathogens *Cochliobolus heterostrophus*, *C. carbonum*, and *Setosphaeria turcical*. Mutants deficient in producing *ZmSTP2* had significantly enlarged lesion areas in leaves infected with the 3 pathogens (Ma et al., 2023). *OsMST4* is a sugar transporter that is expressed in all tissues to varying degrees and transports a variety of sugars, specifically having a high affinity for galactose and mannose. Furthermore, its expression is upregulated upon root exposure to air and PEG treatment (Deng et al., 2019; Wang et al., 2007). *OsMST4* and *ZmSTP2* are orthologous to *AtSTP13*, which has been

shown to sequester hexose sugars from pathogens by importing them constitutively from the extracellular environment. Upon phosphorylation by AtBAK1, AtSTP13 increases its efficiency of importing sugars, reducing pathogen growth (Yamada et al., 2016). Both the rice and maize OTGNs contain orthologues of *AtBAK1*, which suggests that *OsMST4* and *ZmSTP2* could play a similar role as *AtSTP13* in reducing pathogen growth by sequestering hexose sugars. If further investigated in future experiments, this could provide fascinating insights into the interplay between cellular immune responses and specialised metabolism in plants.

OG0008363 contained *OsHAK5* and *ZmHAK5*, both of which are well known potassium transporters known to mediate plant response to potassium deficiency (Qin et al., 2019; Yang et al., 2014). Both of these proteins are known to localise to the cell membrane to mediate potassium uptake upon potassium deficiency. As such, the presence of both of these genes in the rice and maize OTGNs is probably due to the strong association between potassium response and defensive terpene metabolism with drought response.

OsABCB26 and *ZmABCB28* were in OG0012991 and code for plastid-localised (predicted) ABC proteins. They are orthologous to *AtABCB28*, which encodes a plastidic transporter which transports auxin across the chloroplast envelope into the cytosol (Pang et al., 2013; Tamizhselvan et al., 2023). Little further information was available on the role of *OsABCB26* and *ZmABCB28*; I can therefore only speculate that these two ABCB proteins might transport intermediates in terpene biosynthesis from the chloroplast into the cytosol, where ER membrane-anchored CYP450s can complete the biosynthetic process.

The final orthogroup of transporter proteins I investigated was OG0000392, containing the amino acid permeases (AAPs) *ZmAAP14* and *OsAAP11A*. In a recent effort to characterise maize AAPs, ZmAAP14 was found to encode a plasma membrane localised AAP which was upregulated in roots during drought stress (Islam et al., 2024). *OsAAP11A* was upregulated by glutamate treatment and transports a wide variety of amino acids, except aspartate and ß-alanine (Taylor et al., 2015). AAPs in general are increasingly being characterised as being abiotic stress-responsive with some proteins having high affinity for particular amino acids and some proteins having low affinity for all amino acids (Wang et al., 2024; Zhou et al., 2020). As such, *OsAAP11A* and *ZmAAP14* in rice and maize most likely import amino acids upon abiotic stresses. Their association with terpenoid biosynthesis genes must be due to a shared regulatory mechanism which activates amino acid uptake as well as terpene biosynthesis during abiotic stress.

In summary, the transporters in the rice and maize OTGNs can be assigned as abiotic- or biotic-stress-responsive. Some, such as the PDRs, are strong candidates for directly facilitating transport of terpenoids, whereas other transporters, such as HAKs, AAPs and STPs, are involved in processes that share a regulatory machinery with terpene biosynthesis, hence their strong associations with terpene biosynthesis genes. All of these genes have been assigned roles within the abiotic / biotic stress model in Figure 17.

3.6 Conserved immune receptors perceive abiotic and biotic stresses in order to facilitate defensive terpenoid production

Both the rice and maize OTGNs contained 5 orthogroups of genes encoding proteins associated with immune response. These belong to categories such as lectin receptor like kinases (LRKs), S-domain receptor like kinases (SDRLKs), leucine rich repeats (LRRs) and nucleotide-binding/leucine-rich repeats (NLRs). Each of these proteins differ substantially in protein domains and the manner in which they bind ligands in order to mediate immune signalling. Four out of five orthogroups of proteins contained signal peptides (DeepLoc,

Supplementary Tables 1 and 2), which is a common finding in cell membrane bound immune receptors. Only a single orthogroup contained nuclear localisation signals (NLS), suggesting a cytoplasmic immune signalling role. As bioinformatic methods to better understand these immune receptors are still being developed, literature review of these proteins was the best method to decipher their roles in triggering defensive terpene production in rice and maize. Figure 16 below shows the protein domains, according to InterPro, of the five orthogroups of immune receptor proteins in the maize and rice OTGNs.

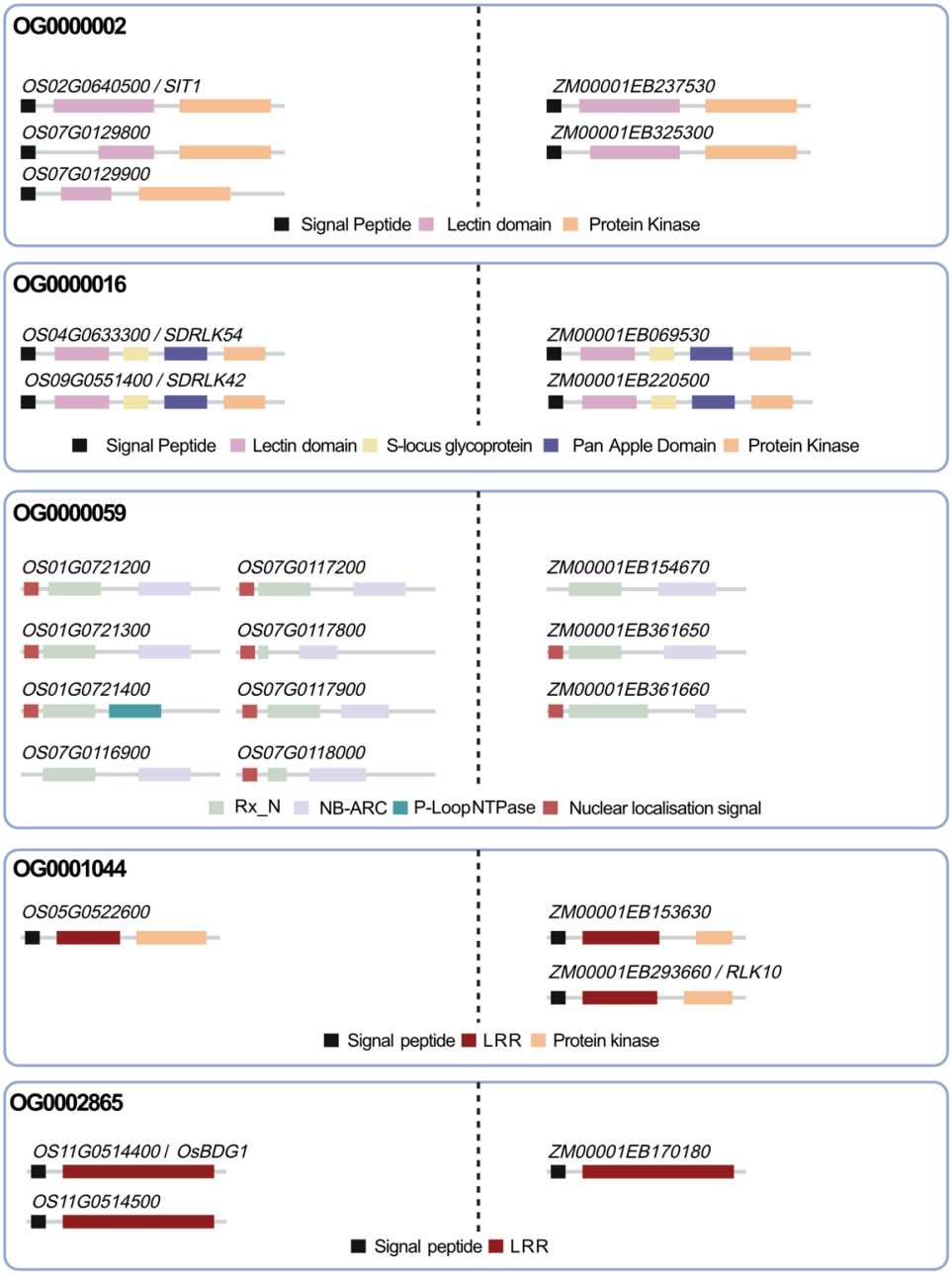


Figure 16: The predicted protein domains for each of the immune genes in the rice and maize OTGN, separated into the 5 orthogroups. Each orthogroup has a legend and genes are labelled with names

where available. The rice proteins are on the left and the maize proteins on the right. Subcellular localisation was predicted using DeepLoc v2.1. Created with BioRender.com

OG0000002, contained L-type LRLKs: *OsSIT1, OS07G0129800, OS07G0129900, OS10G0441900* (Wang et al., 2024), *ZM00001EB237530* and *ZM00001EB325300*. The expression of both maize genes is upregulated in the roots of maize grown in sandy soil to mimic drought stress (Ganther et al., 2022). The promoter region of *OS07G0129800* contains a number of abscisic acid responsive elements and expression of *OS07G0129900* is downregulated in rice roots inoculated with beneficial bacteria (Brusamarello-Santos et al., 2019; Passricha et al., 2017). *OsSIT1* is expressed in rice roots upon salt stress, leading to ROS accumulation and ethylene production, eventually causing cell death (Li et al., 2014). In summary, the LRLKs in OG000002 seem to mediate abiotic stress response in roots specifically, possibly leading to the production of defensive diterpenes in both rice and maize.

The SDRLKs *OsSDRLK42, OsSDRLK54, OS09G0551000, ZM00001EB069530* and *ZM00001EB220500* belonged to OG0000016. There is little information on the maize SD-RLKs, only that *ZM00001EB069530* has been shown to be downregulated in leaves upon drought stress (Wei et al., 2014). Expression profiles of *OsSDRLK42* and *OsSDRLK54* suggest that the former is responsive to bacterial infections and the latter is responsive to chilling stress (Naithani et al., 2021). Since that was the extent of available information for these 5 SDRLKs, I can only conclude that these SDRLKs might respond to both abiotic and biotic stresses.

The third orthogroup, OG0001044, contained RLKs *OS05G0522600, ZM00001EB153630* and *ZmRLK10*. *OS05G0522600* is differentially expressed in rice roots during nitrogen and phosphate stress as well as in leaves upon Xanthomonas infection (Cai et al., 2013; Wu et al., 2021). Both maize genes were recently shown to be hub genes in a co-expression gene network built in a meta-analysis of maize pathogen infections (Hayford et al., 2023). *ZmRLK10* is also upregulated in leaves infected with *C. heterostrophus* or *F. graminearum*. Furthermore, loss of function mutants of *ZmRLK10* accumulate fewer kauralexins and zealexins in leaves inoculated with *C. heterostrophus*. However, there were no differences in kauralexin and zealexin accumulation levels in leaves inoculated with *F. graminearum* (Block et al., 2021). In summary, this group of 3 RLKs mediate biotic stress-induced production of defensive diterpene in maize and possibly rice.

OG0002865, contained the LRRs *OS11G0514400, OS11G0514500* and *ZM00001EB170180*. Analysis of the protein sequences of all 3 proteins via InterPro showed that there is an N-terminal signal peptide followed by a single LRR domain in all 3 proteins. Furthermore, the LRR domains were annotated as extracellular, and subcellular localisation predictions via DeepLoc also suggested that these 3 genes are extracellular due to N-terminal signal peptides. *ZM00001EB170180* is part of a pan-gene set under the umbrella term *GRMZM2G145440*, as per MaizeGDB. It has also been annotated as a homolog ofi*AtBAK1* (BRI1 associated receptor kinase), which binds *AtBRI1* (Brassinosteroid insensitive 1) upon brassinosteroid binding and regulates cell growth (Li & Chory, 1997; Jia Li et al., 2002). ZM00001EB170180 / ZmBAK1 has been speculated to form a complex with ZmRLK10 based on co-expression data (Block et al., 2021). As such, it is possible that the RLKs in OG0001044 and the LRRs in OG0002865 could interact to form a receptor complex which responds to biotic stress signals in rice and maize.

The only orthogroup to contain NLRs was OG0000059, which contained 8 rice NLRs and 3 maize NLRs. NLRs contain 3 domains, a variable N-terminal domain, an NB-ARC domain which binds ADP / ATP and a C-terminal leucine rich repeat containing domain which perceives stress signals. The variable N-terminal domain can be a coiled-coil domain (CC) or a Toll/interleukin-1 receptor domain (TIR). Most of the NLRs in OG0000059 were

predicted to localise to the nucleus. The phylogenetic tree of OG0000059 showed that there were two clades containing the rice and maize NLRs (Supplementary Figure 1). The first clade contained the NLRs on chromosome 1 of rice and *ZM00001EB361650*, *ZM00001EB361650*. The second clade contained the NLRs on chromosome 7 of rice and *ZM00001EB154670*. According to Figures 13 and 14, the 5 NLRs in clade 1 all contained WRKY binding motifs and either G-boxes or N-boxes, whereas the 6 NLRs in clade 2 only contain N-boxes or G-boxes, suggesting differential transcriptional regulation of this orthogroup of NLRs.

According to InterPro, the rice NLRs on chromosome 1 have an N-terminal coiled-coil Rx_N domain and an NB-ARC domain (Figure 16). *ZM00001EB361650* and *ZM00001EB361660* only contain the NB-ARC domain. The rice NLRs on chromosome 1 exist in a locus containing 8 NLRs belonging to the same orthogroup. Clearly this is due to gene duplication which had occurred after rice and maize diverged, since multiple orthologs exist in the Oryza clade. The maize NLRs on the other hand have undergone gene duplication only in maize. Without empirical in vitro data, it is difficult to better understand these NLR proteins, as they could bind a variety of biotic or abiotic stress signals and signal immune responses.

According to InterPro, the clade 2 NLRs all contained an NB-ARC domain and some contained Rx_N (coiled-coil) domains / signal peptides / non-cytoplasmic domains. The clade 2 NLRs in rice were located in a locus containing 11 NLRs of the same orthogroup, whereas the maize NLR in clade 2 is located in a locus that contains 2 copies; *ZM00001EB154670* and *ZM00001EB154680*, which are both in OG0000059. *ZM00001EB154680* contains an Rx_N domain and is in a separate clade from *ZM00001EB154670*. In summary, the NLRs of OG0000059 have undergone gene duplication events in rice and maize, which might be indicative of an ongoing arms race typical to plant NLRs. The cellular function / role of any of the NLRs in OG0000059 is unknown, as NLR-mediated plant immunity is a complex field that is the focus of active research at the moment. The presence of bHLH and WRKY binding motifs, as well as the strong associations of these genes within the maize and rice OETGNs, suggest that these NLRs warrant in vitro investigations of their ability to regulate terpene biosynthesis.

The genes in this subsection of the results are well known to be cytosolic or membrane-bound immune receptors. Some of these genes, such as *OsSDRLK42* / *54*, or *ZmRLK10*, have been characterised to a certain extent, whereas others, such as those in OG0000059, have not been investigated to date. I believe that all of these genes play a role in intracellular immune signalling in response to abiotic or biotic stress and have played such a role before rice and maize diverged. Therefore, I proposed putative roles / functions for these genes in a summary model (section 3.7, Figure 17) for future experiments to refer to. I hope that future experiments can identify the exact interplay in immune signalling between LRRs, SDRLKs, RLKs and NLRs which lead to the production of defensive terpenes in both rice and maize.

3.7 A putative terpenoid regulatory model that responds to either biotic or abiotic stresses

The primary objective of this project was to analyse the orthologous proteins and genes involved in terpenoid production in rice and maize, in order to identify genes / proteins which act as regulators, transporters and immune signallers that collectively drive the production of defensive diterpenes in both species. The analyses thus far have identified strong candidates involved in each of these 3 categories; TFs, transporters and immune receptors, thereby satisfying the primary objective of this project.

Some genes in the OTGNs of both species are known to be differentially regulated upon biotic or abiotic stress. The EMBL expression atlas contains data from 13 experiments with differential gene expression analyses, involving biotic and abiotic stresses. This data and other publicly available literature were used to assign genes in OTGNs as responsive to biotic or abiotic stress. Incorporating these data as well as the functions of these genes allowed me to generate a putative model of defensive terpenoid production (Figure 17). These models should be considered as a summary of candidate genes involved in regulation, transport, and immune responses.

As shown in panel A of Figure 17 below, upon perception of either biotic or abiotic stress by SDRLKs, PR proteins could mediate cytosolic defence signalling leading to the expression of terpenoid biosynthesis genes. In biotic stress conditions, NLRs would play a key role in cytosolic defence signalling, as this is the hallmark role of NLRs in plant immunity. The transcriptions factors discussed in section 3.4 (WRKYs and bHLHs) would bind to their respective motifs in the promoter regions of terpene biosynthesis genes such as *OsCPS2*, *ZmTPS6*. Subsequently, the next steps of mRNA translation and trafficking of plastidic copalyl synthases and kaurene synthases are performed by known protein trafficking mechanisms such as the ALBINO or TIC / TOC translocons (Chuang et al., 2021; Thomson et al., 2020). Within the chloroplast, the terpene biosynthesis is well known; plastidic GGPP is converted into pimaradienes by CPSes and KSes, which would be exported from the chloroplast by chloroplast-localised ABCB transporters. The pimaradienes are then acted upon by ER membrane anchored CYP450s and cytosolic SDRs, producing momilactones, phytocassanes and oryzalexins in rice and kauralexins and dolabralexins in maize. These end products would then be transported out of the plant cells by plasma membrane localised transporters such as PDRs and MATEs. Defensive terpenoids would then accumulate in extracellular spaces at sites of fungal infections in leaves or exuded into the rhizosphere via the roots (Kodama et al., 1988; Veenstra et al., 2019).

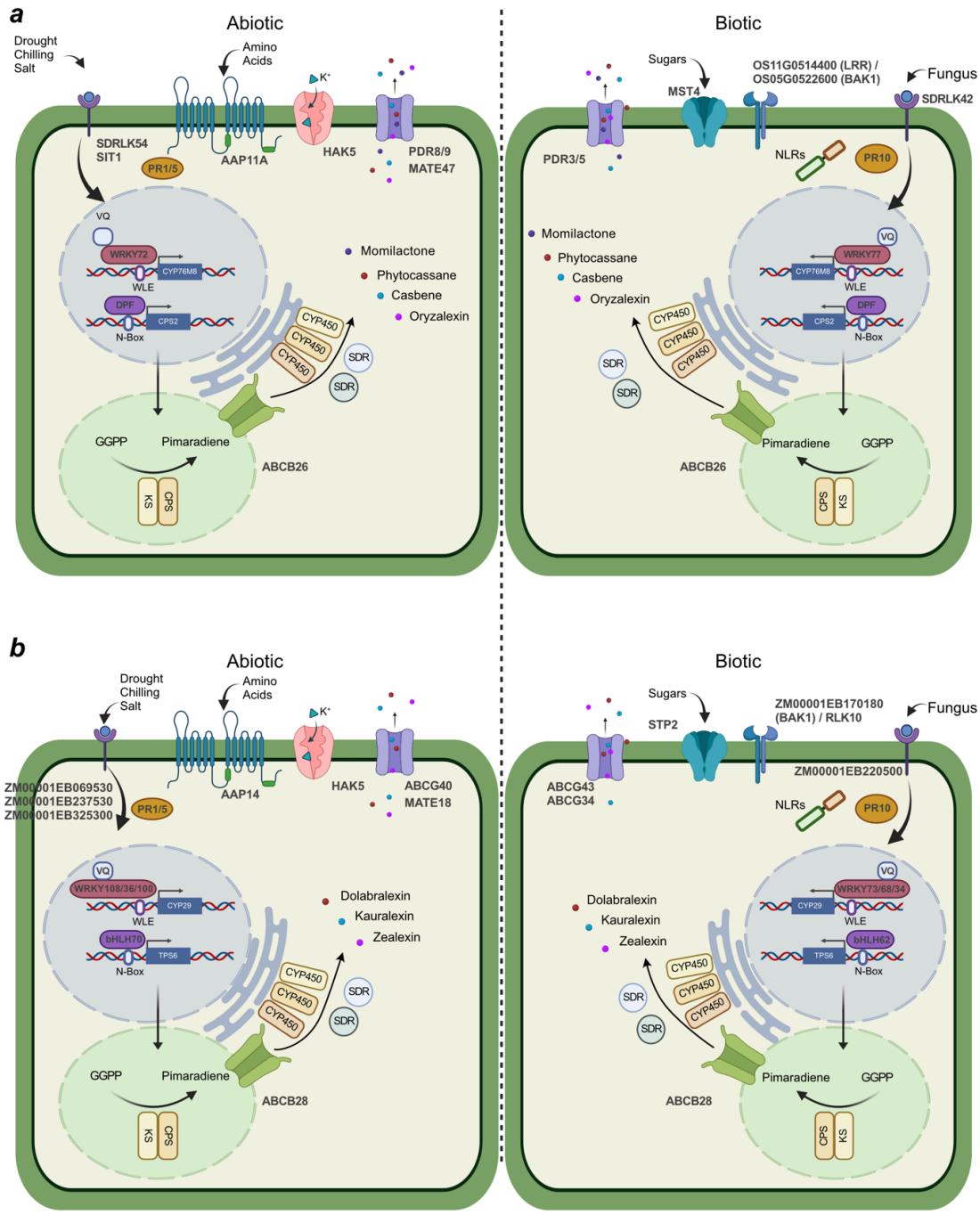


Figure 17: A model in a) rice and b) maize showing potential roles of candidate genes involved in i) transcriptional regulation of terpenoid biosynthesis gene ii) transporters which facilitate intra- and extracellular transport of intermediate compounds and complete compounds iii) immune responses which drive production of defensive terpenes. The candidate genes are shown in bold, grey font and known terpene biosynthesis genes / proteins are shown in black and white. Created with BioRender.com.

3.8 Syntenic relationships between rice and maize ancillary genes shed light on gene expansion and sub-functionalisation within the terpenoid gene network

Maize and rice diverged from a common ancestor approximately 70 million years ago. Since then, both species have undergone extensive evolution, including genome duplications and domestication, which selected for certain alleles (Ilic et al., 2003; Stein et al., 2018). Despite these substantial genomic changes, both species retained defensive terpenoid production catalysed by conserved CPSes, KSes and CYP450s. As per section 3.3, there were also conserved genes mediating immune responses, transport and transcriptional regulation of terpenoid biosynthesis. I noticed that rice had many members of the CYP76 family in the rice OTGN whereas the maize OTGN had many members of the CYP71 family. I was interested in analysing the gene family expansions of both of these types of CYP450s as it suggested that similar selection pressures in rice and maize had resulted in different CYP450s being used to produce defensive diterpenes. I was also interested in the PR10 family of proteins in the rice and maize OTGNs as rice had 5 members and maize had just 1. This suggested that there was a specific expansion of the PR10 family in rice associated with terpenoid biosynthesis.

Pathogenesis related proteins are a loose term used to describe a variety of proteins which accumulate in plant tissue upon abiotic and biotic stress. There are more than 17 families of PR proteins characterised across all plants so far, with each family having a specific biochemical function such as chitinase or peroxidase (Jain & Khurana, 2018; Morris et al., 2021). The PR10 family of proteins are generally known to be anti-fungal and anti-viral ribonucleases or protease inhibitors (Morris et al., 2021). The PR10 genes found in the rice and maize OTGNs are located in chromosome 12 of rice and in chromosome 10 of maize respectively. Both chromosomal regions are highly syntenic to each other, as shown in Figure 18a. When analysing the microsynteny between the PR10 loci in rice and maize, I observed that the 5 rice PR10 genes are syntenic to a single PR10 gene in maize, annotated as *ZmPRP9*. The other genes upstream and downstream of this PR10 locus were orthologous to each other as well, as expected in a syntenic block. This suggested that the PR10 gene family in rice had undergone a gene expansion after rice and maize diverged. In order to identify the earliest point of gene expansion, I analysed the phylogenetic tree (Figure 18c) of orthogroup OG0001341 containing the PR10 genes. This showed that *L. perrieri* and *O. brachyantha* both have multiple PR10 copies that are not closely related to *O. sativa* PR10s, indicating gene family expansion. Gene family expansions are usually associated with positive selection, especially within plant defence (Kahlon & Stam, 2021), suggesting that PR10s in *L. perrieri* and *O. brachyantha* already conferred advantages to both species. The PR10 genes in O. sativa are located in 5 separate clades which also contain *O. alta* and *O. officinalis* PR10s, indicating that the PR10 gene family involved in the rice OTGN expanded first in *O. alta* and *O. officinalis*. In summary, the PR10 family of proteins had undergone gene family expansion twice in the Oryza species analysed in this project, whereas no gene family expansion had been observed in maize. Since PR10 proteins have been associated with anti-pathogen ribonuclease activity, this gene family expansion might be specific to a particular pathogen. Future in vitro experiments can better dissect the exact role of PR10 proteins in diterpene biosynthesis in rice and maize.

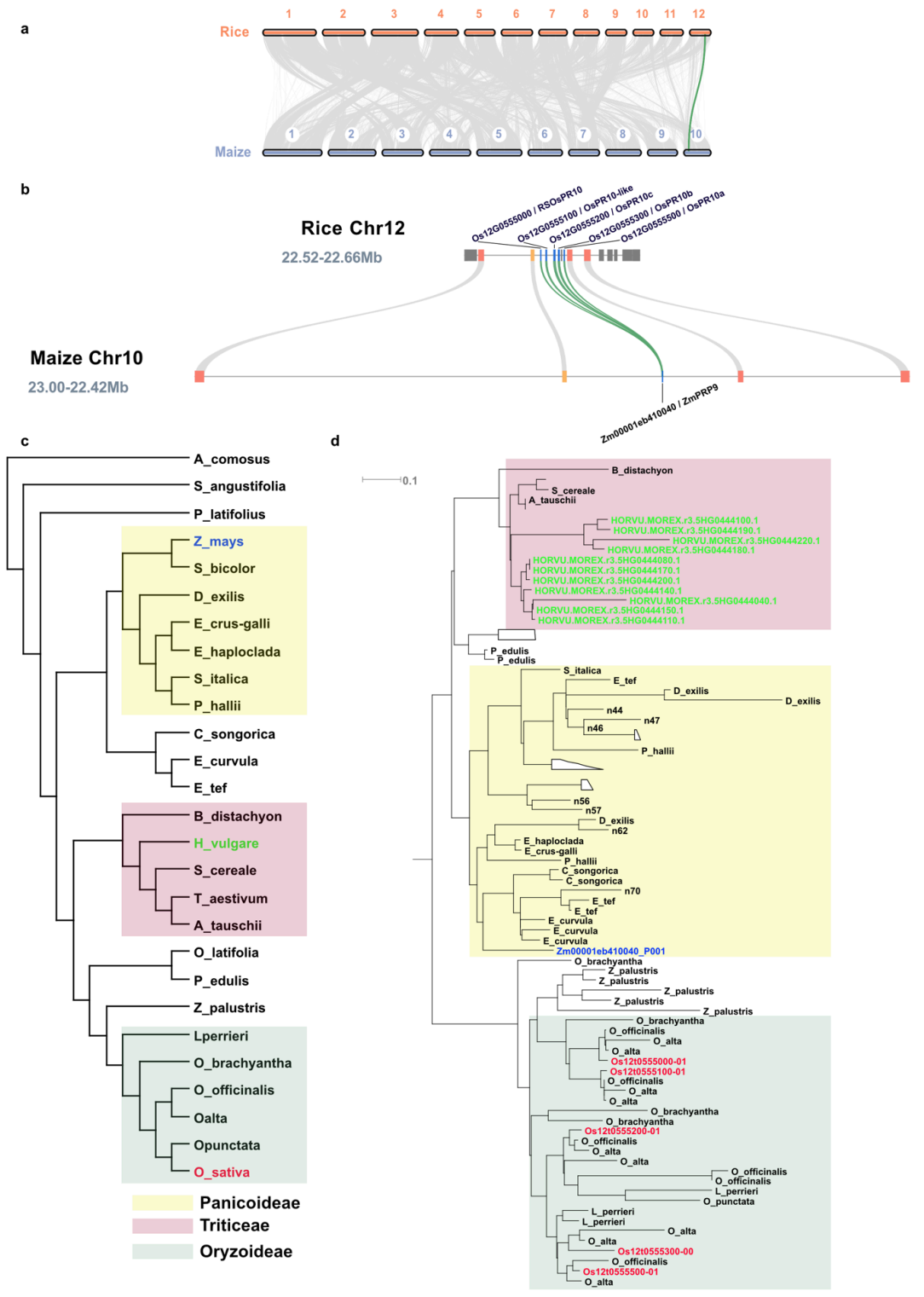


Figure 18: a) Macrosynteny between rice and maize where the green line indicates the syntenic block that contains OsPR10 and ZmPRP9. b) Microsynteny between rice and maize specific to the orthogroup containing the PR10 family of genes in rice and maize. c) The rooted species tree within the 27 species of plants used in the OrthoFinder analysis. d) Phylogenetic tree for the PR10 orthogroup.

I observed another example of gene family expansion and specialisation in the orthogroup OG0001170 containing *OsTPS46*. This gene is a sesquiterpenoid synthase and occurs in the diterpenoid subnetwork of the rice OTGN. It is known to produce farnesene (Yuan et al., 2008) and is in the same orthogroup that contained the maize zealexin biosynthesis genes (*ZmTPS6 / 12 / 11 / 13 / 9 / 4*). The phylogenetic tree of the orthogroup is shown below in Figure 19. It shows that there are 4 paralogs in rice, *OsTPS1 / 30 / 31 / 46*, which are in two separate clades. *ZmTPS31* was closely related to the clade containing *OsTPS1 / 30 / 31*, with only a few species in between maize and rice having any orthologs within this clade. Furthermore, in the maize OTGN (Figure 1 in section 3.3), *ZmTPS31* is only linked to a single other node. Whereas the other TPSes in the maize OTGN are connected to many more nodes. Together, this suggests that *ZmTPS31* has a minor role in maize terpene biosynthesis.

The OG0001170 tree showed that *OsTPS46* was in a clade which contained at least 1 protein from the 6 Oryzoideae species studied in this project. The closest maize ortholog of *OsTPS46* is *ZmTPS10*, which was in the diterpenoid subnetwork of the maize OTGN and was strongly associated with ZmCYP29. Considering that both *ZmTPS10* and *OsTPS46* were part of the diterpene subnetworks of the respective OTGNs, there are two possible scenarios. First, it is possible that *ZmTPS10* and *OsTPS46* are acting on products of plastidic diterpene synthases. Alternatively, the sesquiterpene products of *ZmTPS10* and *OsTPS46* are substrates for CYP450s associated with diterpene biosynthesis, thereby yielding novel sesquiterpenes. For example, the OETGNs showed that the relationships between genes encoding for CYP71Zs and *ZmTPS10* / *OsTPS46* were conserved, so it is possible that these sesquiterpene synthases are utilising CYP71Zs in their biosynthesis pathway.

When examining the remainder of the phylogenetic tree of OG0001170, I observed that there were 2 instances of gene duplications in maize which led to the formation of 2 BGCs. The first gene cluster contained the zealexin synthases *ZmTPS12 / 11 / 13 / 6* (*Zx2 / 3 / 4 / 1*), while the second gene cluster contained the sesquiterpene synthases *ZmTPS9 / 4 / 5* and a single pseudoegene, *ZM00001EB415130*. When comparing the presence of these two gene clusters in the maize OTGN, it was evident that the zealexin synthases occurred in the zealexin subnetwork, whereas *ZmTPS5 / 9* were present in the gibberellin / sesquiterpene subnetworks, and *ZmTPS4* was not associated with any subnetworks. It is likely that *ZmTPS5* and *ZmTPS9* produce uncharacterised metabolites in collaboration with the gibberellin biosynthesis enzymes. The evolution of the zealexin synthases in maize via gene duplication shows that zealexins are unique to maize and that it is unlikely that other closely related species such as Sorghum would produce zealexins.

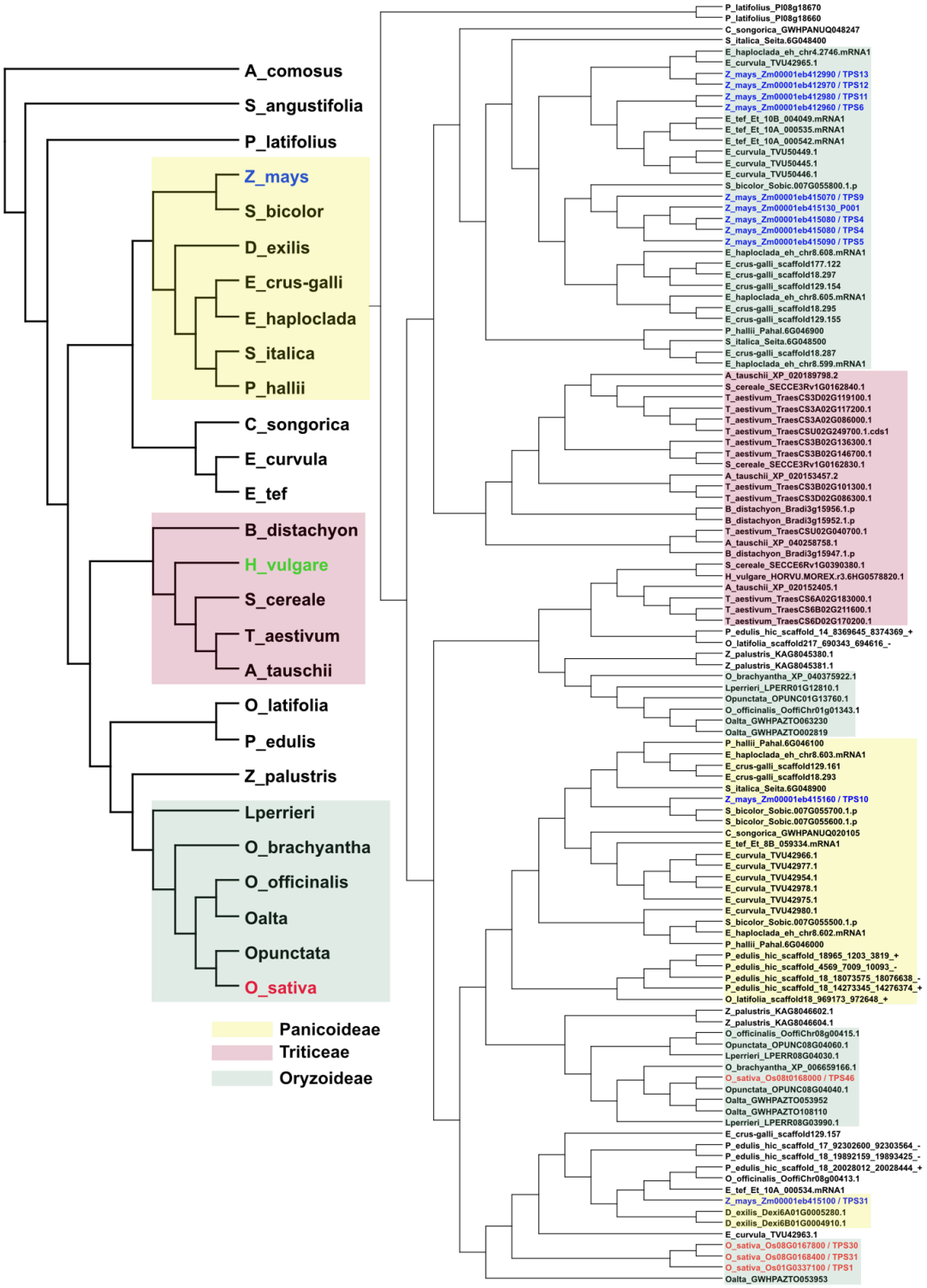


Figure 19. The phylogenetic tree of the orthogroup OG0001170 built from similarities of protein sequences amongst 27 species. The rice proteins are highlighted in red, maize proteins in blue. The species in the species tree are colour-coded by tribes of interest. Similarly, the phylogenetic tree highlights branches which contain members of the respective tribe.

When analysing the CYP450s involved in terpenoid biosynthesis in rice and maize (Figure 20), I observed that the CYP76M, CYP71Z, CYP701A8 and CYP99A families were always present, albeit to differing extent in both species. The Orthofinder analysis grouped proteins from each of these CYP450 families into 4 distinct orthogroups. When analysing the phylogenetic trees from each of these orthogroups, I noticed that each CYP450 family had undergone gene expansion to varying degrees in each species. There are 14 CYP76M proteins in rice (Figure 20a), located in chromosomes 1, 2, 6, 8 and 12, with two clusters of 4 genes each on chromosomes 2 and 8. The rice OTGN only contained the 4 genes from the chromosome 2 cluster as well as *OsCYP76M13 / 14* on chromosome 12 and 1, respectively. This suggested that the chromosome 8 cluster of CYP76M genes might be involved in non-terpene biosynthesis pathways or unknown terpenoid biosynthesis pathways. Maize had a single CYP76M member in the CYP76M orthogroup, one that was involved in the maize terpenoid network, associating with *ZmTPS22* and *ZmSTC1*, potentially producing sesquiterpenoids. This suggested that the CYP76M family in rice was selected over time and underwent multiple gene duplication events in order to produce diverse terpenes.

The CYP71Zs in both the rice and maize terpenoid networks were part of OG0001139 (Figure 20b). Rice CYP71Zs are located on chromosomes 2, 7 and 10, arising from the 3 ancestral copies in *L. perrieri*. Maize has 3 CYP71Zs on chromosome 5, of which 2 genes are duplicates. The chromosome 2 cluster of CYP71Zs in rice encodes CYP450s which produce oryzalides. The chromosome 7 cluster, on the other hand, produces casbenes, while the chromosome 10 cluster has not been associated with any known diterpenes so far. Furthermore, when analysing the phylogenetic tree of Orthogroup 1139, it was evident that there are no CYP71Zs of this orthogroup in *H. vulgare*, *T. aestivum*, *S. cereale* and *A. tauschii*, all members of the Triticeae tribe of grasses. This suggests that the CYP71Z family was only retained in Oryzoideae and Panicoideae, although it should be pointed out that the phylogenetic analyses conducted here are insufficient to confidently conclude this.

The CYP99 family belongs to the orthogroup 733 and contains 2 copies in rice, both in the terpenoid network, and 6 in maize, with only a single gene in the terpenoid network. Barley has 12 proteins in this orthogroup, suggesting a substantial expansion and recruitment of these genes into terpenoid metabolism in barley.

Overall, there were specific CYP450 gene family expansions observed in this analysis that could be associated with specific types of reactions selected over time (Brown, 2016). For example, CYP76M6 and CYP76M8 belong to the same orthogroup and are located in the chromosome 2 locus. CYP76M6 catalyses a C9 hydroxylation with the oryzalexin biosynthesis process, whereas CYP76M8 catalyses a C7 hydroxylation (Wu et al., 2013). The expansion of the CYP76M gene family within rice and sub functionalisation for specific hydroxylations of carbons likely enables generating specific compounds over time.

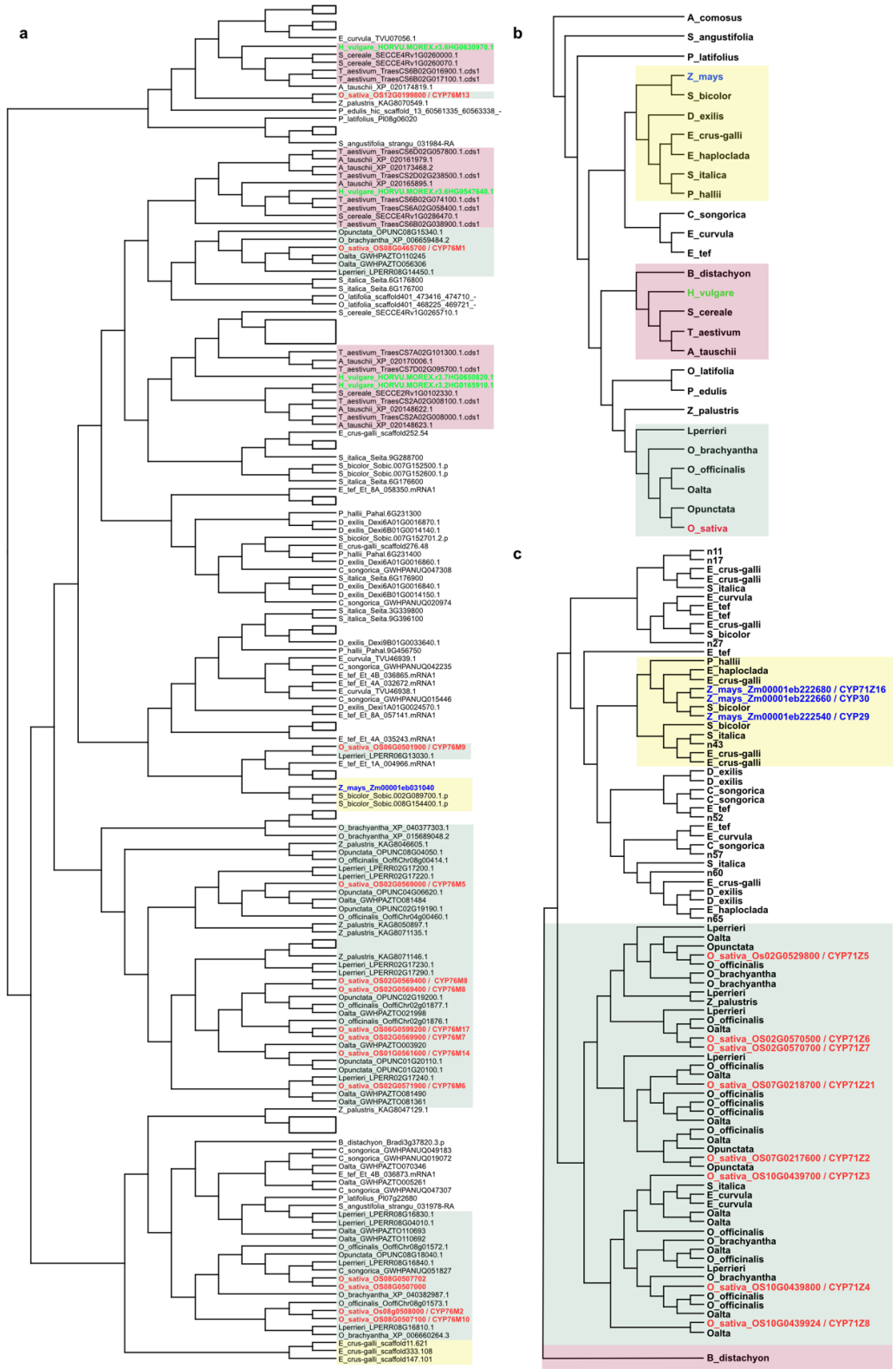


Figure 20: a) The phylogenetic tree of the orthogroup containing CYP76M genes. b) The species tree of the 27 species used in the OrthoFinder run. c) The phylogenetic tree of the orthogroup containing

CYP71Z genes. The species belonging to the Oryzoideae, Panicoideae and Triticeae tribes are highlighted.

3.9 Barley and Sorghum have extensive terpenoid networks based on homology to maize and rice respectively

Among crop species, defensive terpene biosynthesis has been best characterised in rice and maize, as shown in this project thus far. However, two vital crops which also have terpene synthases and the ability to produce defensive terpenes are sorghum and barley. Sorghum and maize are assigned to the Panicoideae clade of grasses, whereas rice and barley are in the BOP (Bambusoideae Oryzoideae Pharoideae) clade. A putative diterpenoid producing cluster on chromosome 2 has been identified in barley (Liu et al., 2023; Liu et al., 2021, 2024) and no terpene biosynthesis genes nor metabolites have been reported for sorghum. Since rice and maize share a number of conserved terpene biosynthesis genes, I hypothesised that the closer ancestors of barley and sorghum would also have conserved terpene biosynthesis genes. As such, I aimed to construct terpene gene networks in both barley and sorghum in order to identify novel defensive terpene biosynthesis subnetworks.

Table 8. A list of the publicly available datasets which were used in this study

Bioproject	Species	Tissue	Treatment
PRJNA431836	*Hordeum vulgare* cv Golden Promise	Leaves	Profiling of transgenic lines expressing the master regulator HvNPR1
PRJNA656491	*Sorghum bicolor* cv BTx623	Leaves	Time course of leaf infection with *Setosphaeria turcica*

A bait gene list was curated by identifying the orthologues of barley from the initial rice bait list (Table 1), likewise for sorghum and maize. I built a gene network in sorghum and barley using transcriptomic data from bacteria / fungus-infected leaves, and the gene neighbourhoods of the orthologous bait gene list were extracted (top 2% of edges). When the terpenoid gene network was visualised in both sorghum and maize, (Figures 21 and 22), it segregated by modules, similar to the rice and maize terpenoid networks. Furthermore, the barley terpenoid gene network segregated into 3 large subnetworks (Figure 21a), with one containing the chromosome 2 cluster of genes that has been reported to produce hordedanes, a type of diterpene (Liu et al., 2023). I decided to label this as putative diterpene subnetwork 1. The other two large subnetworks contained a number of sesquiterpene synthases and diterpene synthases respectively. Thus, I annotated them as putative sesquiterpene subnetwork and putative diterpene subnetwork 2, respectively.

The putative diterpene subnetwork 1 contained 9 genes, 2 terpene synthases (*2HG0099570*, *2HG0099360*) and 7 CYP450s (*2HG0099350, 2HG0099480, 2HG0099550, 2HG0099280, 2HG0099340, 2HG0099370, 2HG0099420*), which had been recently reported to form a BGC in chromosome 2 of barley (Liu et al., 2024). Six genes, *HvCPS2* (*2HG0099570*), *HvKSL4* (*2HG0099360*), *HvCYP89E31* (*2HG0099370*), *HvCYP99A66* (*2HG0099280*), *HvCYP99A67* (*2HG0099350*) and *HvCYP99A68* (*2HG0099550*) encode enzymes which catalyse the production of hordedanes, labdane related anti-microbial compounds (Liu et al., 2024). The remaining 11 genes in the putative diterpene subnetwork 1 may contribute to the production of more defensive diterpenes and should be the focus of future research.

The putative diterpene subnetwork 2 contained 2 CYP99As (*2HG0099350*, *2HG0191370*), a CPS (*1HG0028250*), a KSL (*5HG0499370*) and 2 monoterpene synthases (*3HG0219500*, *2HG0101990*). Together these genes might facilitate the production of defensive diterpenes although none of these genes are located in biosynthesis gene clusters.

The putative sesquiterpene subnetwork contained 24 genes orthologous to known rice terpene biosynthesis genes. Of these 24 genes, 3 monoterpene synthases (*6HG0549330*, *6HG0549080*, *5HG0422720*) clustered tightly together. These 3 monoterpene synthases and the 7 genes of putative diterpene subnetwork 2 shared a number of genes enclosed in a circle in Figure 21a. As speculated in the previous section, the shared genes might be associated with chloroplast processes as both monoterpene and diterpene biosynthesis originates in plastids.

Of the remaining genes in the putative sesquiterpene subnetwork, 4 genes (*2HG0191330*, *2HG0191390, 2HG0191410, 2HG0191420*) encoded kaurene synthases and were tandemly arranged in the barley genome. These genes were strongly associated with other genes in the sesquiterpene subnetwork, suggesting that diterpene and sesquiterpene biosynthesis may be intertwined in barley. The remaining genes in the putative sesquiterpene subnetwork were orthologous to sesquiterpene and triterpene synthases in rice. These genes are also strong candidates for future work aiming to identify novel terpene metabolites in barley.

Considering that rice and maize shared ~ 200 orthologous genes including TFs, immune receptors and transporters, I hypothesised that rice and barley as well as maize and sorghum would share a number of such regulatory genes as well. I filtered the barley terpenoid gene network for ancillary genes such as TFs, immune receptors and transporters and visualised in Figure 21b. This highlighted a few key findings. Firstly, there are 3 TFs in the putative diterpene subnetwork 1 which encoded for two WRKYs and one GRAS family TF. One of the WRKYs (*2HG0096750*) is orthologous to *OsWRKY72*, which was also closely associated with diterpenoid production in rice in response to biotic stresses (Figure 14). Secondly, *5HG0538180* is a PDR transporter, orthologous to *OsPDR3* and *OsPDR5*, which were both involved in the rice diterpenoid subnetwork (Figure 7, Table 10). Lastly, there were 3 PR proteins, *5HG0444080*, *5HG0444170* and *5HG0444200*, orthologous to *OsPR10a*, which was also strongly associated with rice diterpenoid biosynthesis and occurred in the rice diterpenoid subnetwork (Figure 7). Overall, this analysis suggests that some core genes such as WRKYs, PDRs and PR proteins, which potentially regulate and transport diterpenoid production, are conserved between rice and barley. These are strong candidates for downstream experiments that focus on identifying how diterpenoid production is facilitated and how they accumulate in extracellular spaces and / or are secreted into extracellular environments, as for example in the case of momilactones in rice. In summary, by building terpenoid gene networks in barley, I identified 3 major putative terpenoid subnetworks where 1 contained recently reported genes, validating my approach. Furthermore, I found evidence for conserved genes in the immune regulation, transcriptional regulation and transport of defensive diterpenes in rice and barley.

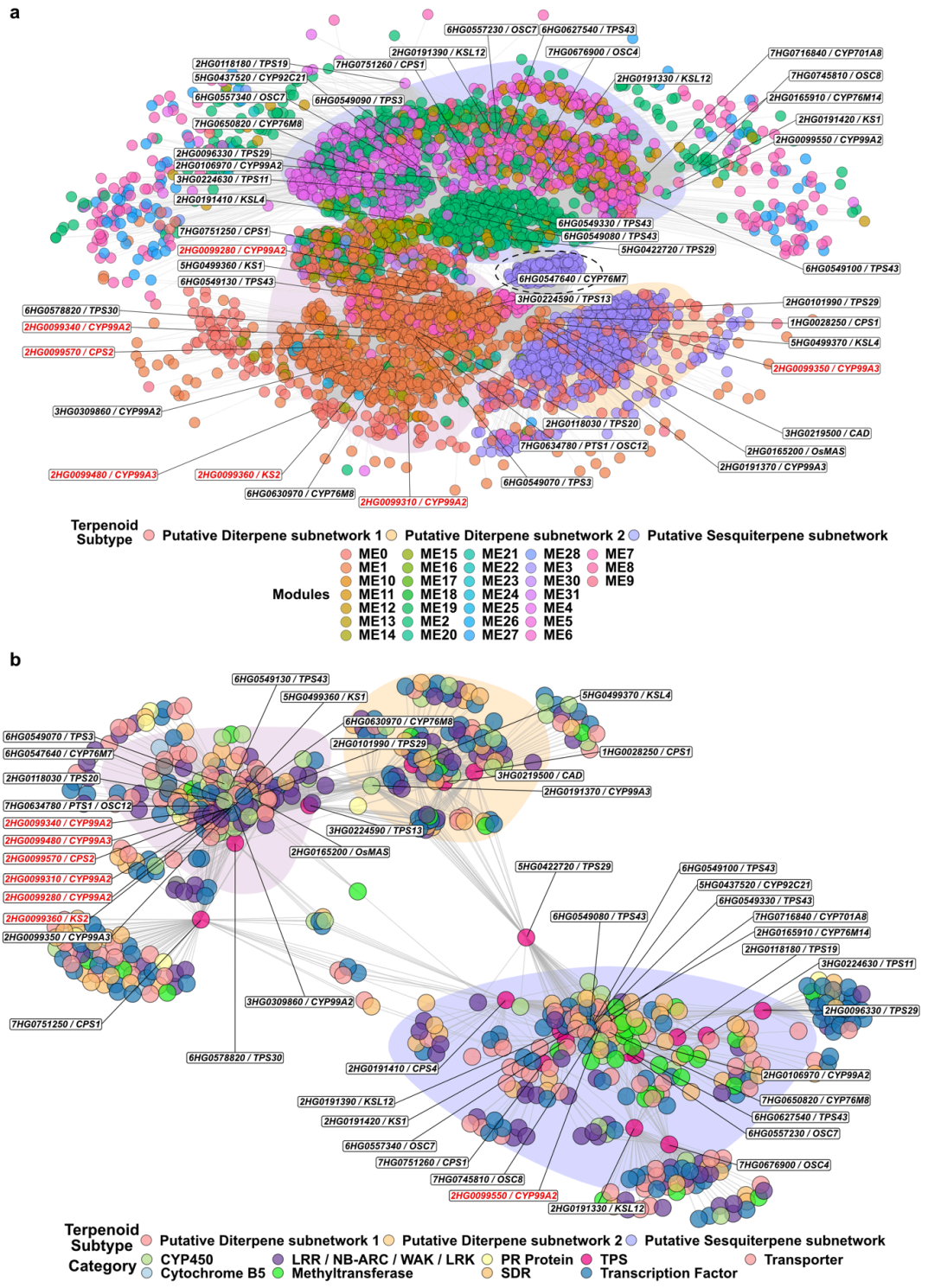


Figure 21. The putative barley terpenoid network was constructed by extracting gene neighbourhoods from barley genes known to be orthologs of rice terpene biosynthesis genes. a) The overall terpenoid network, with 2 putative diterpene and 1 putative sesquiterpene cluster being highlighted. The known barley diterpene biosynthesis genes are in red. b) The overall terpenoid network was filtered for genes corresponding to relevant categories. The known barley diterpene biosynthesis genes are in bold.

Similar to the previous section, a putative terpenoid gene network in sorghum was assembled based on orthologous sorghum genes of maize terpenoid gest edges (strongest 2%) of each bait gene were extracted from the sorghum network and visualised as a network (Figure 22) where overlapping neighbourhoods between bait genes might indicate a potential biosynthetic pathway / network. This resulted in a substantial terpenoid gene network containing 2200 genes. Where the maize terpenoid network segregated into two large subnetworks, the sorghum network consisted of 3 identifiable subnetworks: a putative monoterpene / sesquiterpene subnetwork, a putative sesquiterpene subnetwork and a putative diterpene / sesquiterpene subnetwork.

The putative monoterpene / sesquiterpene subnetwork was annotated as such because it contained 3 putative monoterpene synthases (*004G153900*, *004G019300*, *004G019400*) which were orthologous to known maize monoterpene synthases and 2 putative sesquiterpene synthases (*007G187100, 007G034700*) which were orthologous to maize sesquiterpene synthases. This subnetwork also contained the genes orthologous to the maize zealexin biosynthesis genes *ZmTPS10* and *ZmCYP31* (*007G055500, 001G082500*). It is feasible that in sorghum, there are promiscuous enzymes encoded by genes in the putative monoterpene / sesquiterpene subnetwork, acting in both monoterpene and sesquiterpene biosynthesis. As both of these compounds are known to act as anti-insect and anti-fungal metabolites, it is also feasible that this subnetwork contains genes encoding those particular defense responses.

The putative sesquiterpene subnetwork contained putative sesquiterpene synthases (*007G034700*, *007G055700*, *001G363400*, *005G130400*) orthologous to maize sesquiterpene synthases as well as CYP450s belonging to the CYP81 family (*002G189300, 001G082200, 002G065700, 001G082400, 003G360900*). This subnetwork shares a lot of genes with the diterpene / sesquiterpene subnetwork, which contained 3 diterpene synthases (*006G211500, 005G161200, 001G248600*) which were orthologous to *ZmKS3, ZmCPS3* and *ZmAN1* respectively. The diterpene / sesquiterpene subnetwork also contained some CYP450s (*001G369600, 001G082300, 010G172700*) which were orthologous to CYP450s involved in maize zealexin biosynthesis. Collectively, the sesquiterpene and diterpene / sesquiterpene subnetwork in Figure 22a resemble the diterpene / zealexin subnetwork in maize. Therefore, it is likely that the genes in both subnetworks encode enzymes and proteins which respond to fungal infection and drive the production of defensive terpenoids orthologous to those in maize.

The position of *ZmAN1* within the diterpene / sesquiterpene subnetwork is incongruent as it belongs to module 4 and most of the genes belonging to module 4 belong to the monoterpene / sesquiterpene subnetwork. However, this is similar to *OsCPS1* being distant from the rest of the network in the rice terpenoid gene network. So perhaps this suggests differential regulation of gibberellin biosynthesis upon biotic stress or novel diterpene metabolism in sorghum.

While it is reasonable to annotate these subnetworks by the types of metabolites they might produce, only in vitro experiments can accurately annotate these subnetworks. Nevertheless, in an attempt to identify subnetworks which might produce defensive terpenes, I filtered the sorghum terpenoid gene network for ancillary genes involved in immune regulation, transcriptional regulation and transport of terpenes.

This resulted in a less dense network shown in Figure 22b which was still segregated by potential terpenoid subtypes. PR proteins had been most strongly associated with subnetworks producing defensive terpenes such as momilactones in rice, zealexins in maize and hordedanes in barley. In the filtered sorghum terpenoid network, PR proteins were only strongly associated with terpene synthases in the putative sesquiterpene subnetwork. These PR proteins belonged to the PR5 and PR10 classes and were orthologous to the PR

proteins in the rice and maize OTGNs. As mentioned previously, PR5 and PR10 proteins are strongly upregulated under biotic stresses and in this project are always strongly associated with the production of defensive terpenes. Due to this finding, I believe that the subnetwork containing the genes (*001G363400, 002G065700, 002G065800, 001G082200, 001G082400, 007G055800*) is the strongest candidate for further research in sorghum terpene metabolism.

Since rice, maize, barley and sorghum have all shown the ability to produce defensive terpenoids, I analysed for the presence of common orthologous in the terpenoid gene networks of all 4 species. This identified 37 orthogroups of proteins being conserved across all 4 species. This includes 9 orthogroups spanning monoterpene, copalyl, kaurene and sesquiterpene synthases. Of the remaining 28 orthogroups of proteins, there were 7 orthogroups which contained proteins such as CYP450s, TFs, immune receptors, PR proteins and transporters. These are shown in Table 11 below. The remaining 21 orthogroups of proteins conserved in the terpenoid gene networks of rice, maize, barley and sorghum are involved in various cellular processes and may contribute to terpenoid production via currently unknown methods. In summary, I propose that an ancestral defensive terpenoid network utilised proteins represented in these 37 orthogroups in order to produce defensive terpenoids upon biotic stress and possibly abiotic stress.

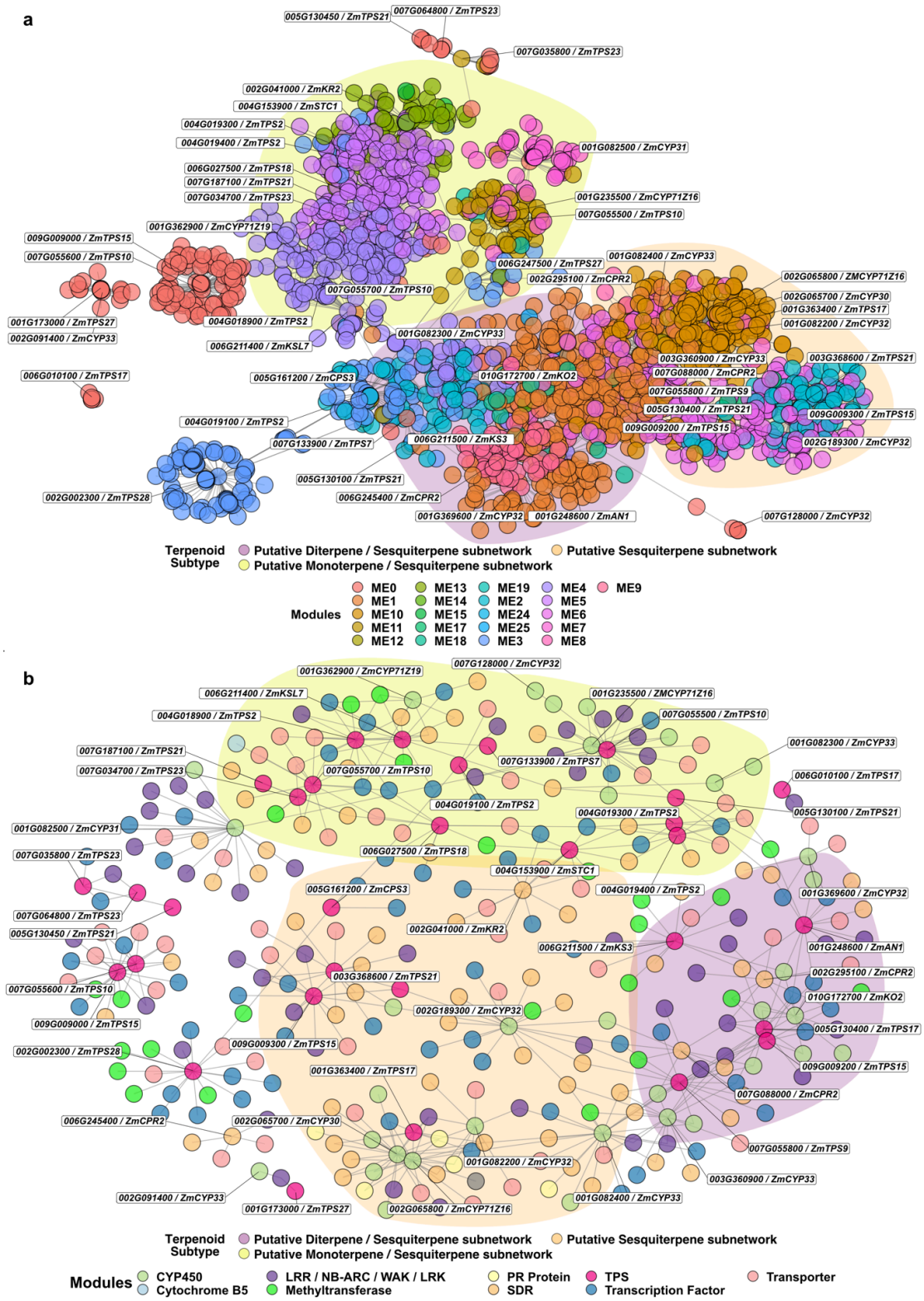


Figure 22: a) The overall terpenoid network in sorghum, based on sorghum orthologs of maize terpene biosynthesis genes. b) The terpenoid network with genes of interest filtered and highlighted in a

corresponding colour scheme. In both panels, the putative zealexin / diterpene subnetwork / cluster is highlighted in light red. Each sorghum gene is labelled with its gene ID and the closest maize ortholog.

Table 11. Orthogroups which have been conserved in terpenoid gene networks of rice, barley, maize and sorghum. The maize and rice genes IDs as well as gene names, if available, are listed. Cells are coloured by their categories, green for CYP450s, purple for immune receptors, yellow for PR proteins, blue for TFs and red for transporters.

Orthogroups	Functional Category	Maize Gene ID	Rice Gene ID
OG0000549	CYP76M family	*ZM00001EB031040*	*OS01G0561600* *OS02G0569000* *OS02G0569400* *OS02G0569900* *OS02G0571900* *OS12G0199800*
OG0000002	Lectin receptor like kinase	*ZM00001EB237530* *ZM00001EB325300*	*OS02G0640500 (OsSIT1)* *OS07G0129800* *OS07G0129900*
OG0000016	S-domain receptor like kinase	*ZM00001EB069530* *ZM00001EB220500*	*OS04G0633300 (OsSDRLK-54)* *OS09G0551400 (OsSDRLK-42)*
OG0001044	Leucine Rich Repeat protein	*ZM00001EB153630* *ZM00001EB293660 (ZmRLK10)*	*OS05G0522600*
OG0000630	PR Protein 5	*ZM00001EB052520* *ZM00001EB217970* *ZM00001EB325340*	*OS03G0661600* *OS03G0663500*
OG0000320	WRKY transcription factor	*ZM00001EB112840 (ZmWRKY108)* *ZM00001EB169460 (ZmWRKY36)* *ZM00001EB293420 (ZmWRKY100)*	*OS11G0490900 (OsWRKY72)*
OG0000074	PDR transporter	*ZM00001EB357950*	*OS01G0342750* *OS01G0609300 (OsPDR9)* *OS01G0609900 (OsPDR8)*

3.10 Traditional forward genetic screens in Arabidopsis to identify resistance or tolerance to momilactone B are ineffective

Forward genetic screens are powerful tools for identifying genes that underlie phenotypic traits, including the resistance or tolerance towards phytotoxic compounds. Particularly, forward genetic screens in Arabidopsis have been vital for identifying genes which confer sensitivity to chemicals such as sulfamethoxazole (Schreiber et al., 2012). As mentioned in the introduction, both target site and non-target site resistances can arise in forward genetic screens. Target site resistance is more robust as due to mutations which lead to structural changes in the target protein, the toxin is unable to bind to the target protein. Furthermore, screening for toxin resistance at a higher concentration often yields plants with target site resistance (Melero-Jiménez et al., 2021).

3.10.1 Dose response of momilactone B

Since Arabidopsis is sensitive to momilactone B, I decided to perform a forward genetics screen in Arabidopsis in order to identify genes which confer sensitivity to momilactone B or, in other words, the molecular mechanisms targeted or affected by momilactone B. In order to determine an effective screening concentration, I performed a dose response assay of momilactone B on Arabidopsis seedlings. Arabidopsis seeds were sown on plates containing ½ MS (½ Murashige and Skoogs) media supplemented with various concentrations of momilactone B and grown in a climate-controlled growth chamber in vertical orientation. After 7 days, plates were scanned and primary root length (PRL) was scored using ImageJ. The % of inhibition was calculated by comparing the PRL of seedlings at each concentration against seedlings grown on control media. Then, I used the drc package in R and the L.3 model for plotting a dose-response curve which determined a momilactone B half-maximal effect concentration (IC_{50}) of 2 µM (Figure 23a) (Ritz et al., 2015). Meaning that at 2 µM of momilactone B, a 50% reduction in PRL was observed in seedlings. At the highest concentrations used in the dose response assay, 20 µM, there was no radicle emergence observed after 7 days. Radicle emergence was observed in 10 µM but there was very little root growth observed. In order to increase the likelihood of identifying true momilactone B resistant mutants with target site resistance (i.e., to reduce the number of false positives), I chose to conduct the first momilactone B resistance screen at a concentration of 8 µM, corresponding approximately to the IC_{80}.

3.10.2 Momilactone B resistance screen in EMS mutagenised M2 generation of Arabidopsis

I had access to an M2 generation of EMS mutants of Arabidopsis, which was made available to us by Andreas Finke from the Max Planck Institute for Plant Breeding Research, Cologne. At the time this project was ongoing, it was highly challenging to obtain momilactone B from rice husks. As the extraction process was still being perfected, I had access to only limited amounts of momilactone B for my mutant screens. To minimise the need of the compound, mutant screens of momilactone B were conducted on horizontally grown plates of seedlings, unless otherwise mentioned. 400 M2 Arabidopsis seeds were scattered uniformly on each ½ MS plates supplemented with 8 µM momilactone B. After 6-8 days of growth, seedling fitness was scored. The dose response screens had shown that growth of seedlings after radicle emergence was stunted and root extension was arrested as concentrations increased. Therefore, in the M2 screen, seedlings with any form of radicle emergence and root extension that stood out in comparison to the rest of the seedlings on the plate were subject to a second round of screening on vertically oriented plates supplemented with momilactone B for 3-4 days. A group of negative control seedlings which had limited radicle emergence or no root extension were also transferred. The second round of screening

identified 5 seedlings which had longer roots or radicles compared to the negative control seedlings; those were also moved to momilactone B-containing media. If their growth was not significantly different, the seedlings were discarded. Five seedlings were scored as potentially resistant and moved to ½ MS so that they could be rescued and grown in soil in order to be crossed to Col-0 to progress further with the mutant analysis.

After these 5 seedlings were crossed with wild-type Arabidopsis to obtain an F1 generation, this F1 generation was selfed in order to generate an F2 population, in which resistant phenotypes would segregate. 10 seeds from each of the potentially resistant 5 lines were sown on ½ MS plates supplemented with 8 µM and grown for 7 days in regular growth conditions. Unfortunately, none of the seeds from any of the 5 lines showed any resistance (Figure 23b). At 8 µM, all seeds showed minimal radicle emergence and a lack of root extension.

Forward genetics screens in Arabidopsis can require screening a varied number of individuals, depending on the project. For example, resistance to sulfamethoxazole was identified in two screens of 12,000 and 16,000 seeds (Schreiber et al., 2012). Whereas other screens can utilise up to 50,000 seeds and be unsuccessful. To improve the chances of a successful forward genetics screen, I tried screening a new type of EMS mutant collection.

3.10.3 HEM Mutant Collection Screen

HEM mutant collections are the result of single seed descent (SSD) propagation of EMS-mutagenised Arabidopsis for 4 generations (Capilla-Perez et al., 2018). When a subset (25) of the ~ 700 HEM lines were sequenced and mutations analysed, there were on average 1003 mutations in each line with 70% of all mutations being homozygous. Of the 1003 mutations, 193 cause protein sequence changes and 15 may affect protein function. Across all ~700 lines of the HEM collections, effectively every gene should statistically be affected by at least one missense or nonsense mutation. Furthermore, since within each line there is a variable rate of homozygosity at mutated loci, I calculated that screening 4 seeds for each line should target all mutated loci within each line.

3.10.3a Momilactone B resistance screen in Arabidopsis HEM mutants

Hence, 4 seeds for each of the ~ 700 HEM lines were screened for resistance to momilactone B at 8 µM on horizontally oriented plates which were placed in the growth chamber for 5 - 7 days (Figure 23c). A control population of HEM seeds were grown on ½ MS media supplemented with DMSO alongside the mutant screen to ensure seed viability. Unfortunately, none of the screened seeds / seedlings exhibited any form of resistance according to root phenotypes. Even if root radicles had emerged, I did not observe any radicle extension or root elongation when the seeds were examined under the stereo microscope. Therefore, I terminated the mutant screen for resistance to momilactone B at this stage.

There are two major reasons which could explain the challenges of identifying momilactone B resistance in the EMS and HEM Arabidopsis mutant populations; lethal mutations in Arabidopsis and polygenic mutations for momilactone B resistance. These are elaborated upon further in the discussion. Since resistance to momilactone B had failed in both HEM and EMS populations, I chose to screen for tolerance to momilactone B, which might arise in the form of target site resistance or non-target resistance.

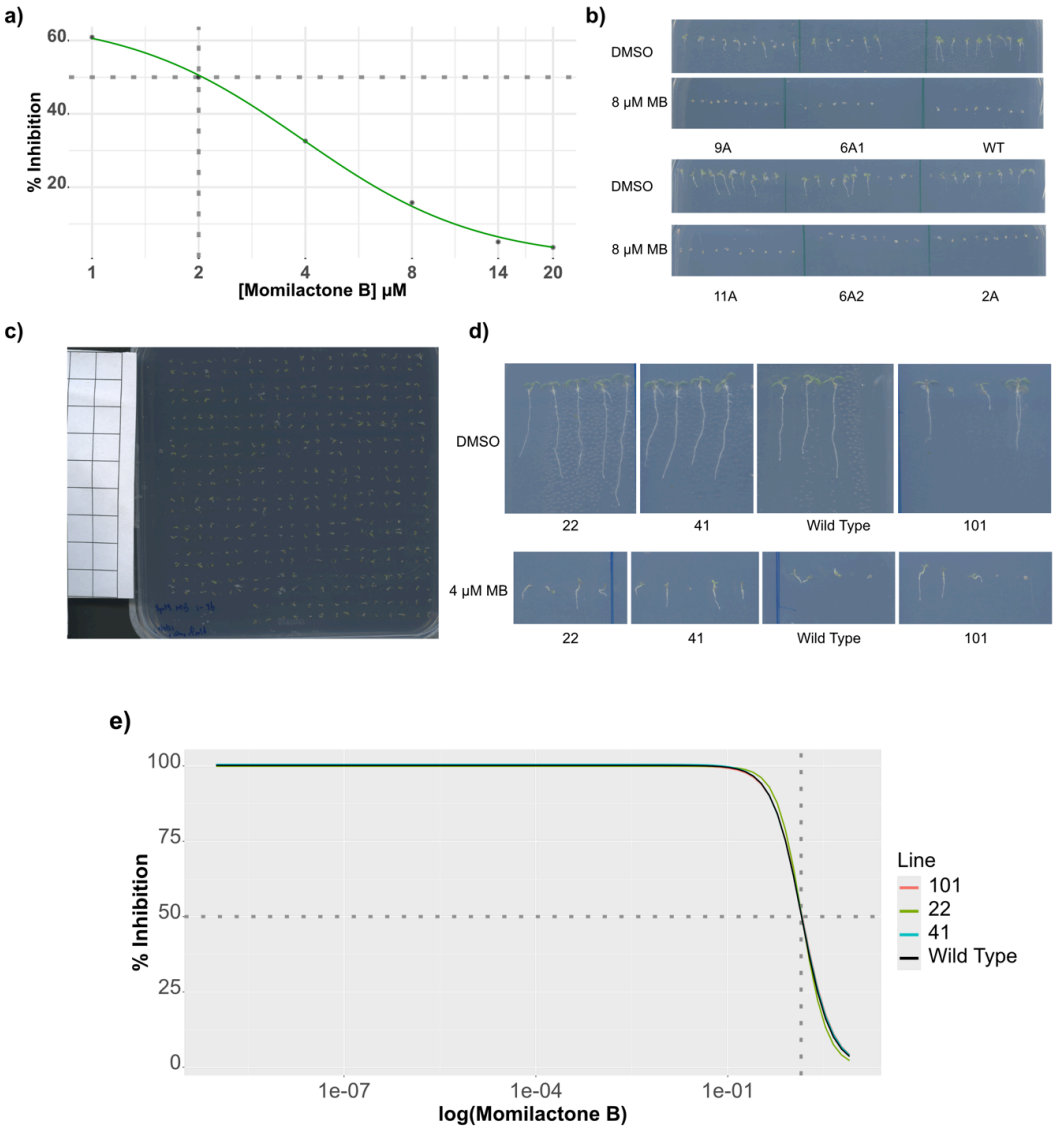


Figure 23: a) A dose response curve of inhibition of primary root length in Arabidopsis as concentration of momilactone B increases. The IC_{50} was 2 µM after fitting a model for the dose response curve using the drc package in R. b) Screening of backcrossed and selfed potentially resistant EMS M2 mutants at 8 µM momilactone B showed that resistance was not observed. c) A set up for screening ~ 696 Arabidopsis HEM mutant lines, with 4 seeds of each line plated in a square. d) Arabidopsis HEM mutants which were initially selected as potentially tolerant to momilactone B at 4 µM. e) A dose response curve of Arabidopsis and 3 lines of HEM mutants which exhibited strongest tolerance to Arabidopsis. The primary root length of each length at each concentration was compared to the root length of that line on control media (DMSO + ½ MS) and plotted as % inhibition. The IC_{50} was approximately 1.39 µM for all 3 backcrossed and selfed (BC1F2) HEM lines and Col-0.

3.10.3b Momilactone B tolerance screen in Arabidopsis HEM mutants

Similar to the previously conducted mutant screen for resistance, a screen for tolerance to momilactone B in Arabidopsis was conducted at 4 µM. At this concentration, Arabidopsis Col-0 seeds were observed to have higher rates of radicle emergence and root elongation. Therefore, a tolerant mutant would exhibit phenotypes similar to Col-0 growing on control media, with a substantially long primary root after 5 - 7 days of growth in light. The HEM mutant screen identified some potentially resistant plants at 4 µM momilactone B. Screening

of the F2 generation of these lines could not confirm these observations; none of the selected plants were tolerant or resistant (Figure 23d). A dose response curve of these plants showed that none of these mutants deviated from the wild-type Col-0 Arabidopsis (Figure 23e).

Resistance to a xenobiotic such as herbicides, or momilactone B in the context of our assays, can be a difficult trait to incur. As such, we screened for tolerance to momilactone B at 2 µM as well. Unfortunately, neither resistance nor tolerance to momilactone B was observed in the EMS mutagenised M2 population or the HEM population. This suggests that resistance to momilactone B in Arabidopsis could arise from two mechanisms. First, the momilactone B target gene could be a core gene which causes a lethal phenotype when mutated, hence it would not be detected in a forward genetic screen. Second, resistance to momilactone B could be a polygenic trait, which would be challenging to identify in a forward genetic screen.

The objective of this project was to utilise forward genetics approaches to identify the target protein of momilactone B in Arabidopsis. This would have provided insight into the cellular mechanisms which were disrupted by momilactone B, leading to lack of root elongation and eventually seedling death at higher concentrations. Although this objective was not fulfilled, the potency of momilactone B is still intriguing. It does not appear to have any degree of autotoxicity according to publicly available literature. Colleagues in our lab have shown that momilactone B treatment of rice leads to plant death at higher concentrations, suggesting that there is a cellular mechanism for momilactone B exudation into the environment or sequestration.

4. Discussion

The orthologous terpenoid gene network project was an ambitious attempt to leverage existing publicly available transcriptomic data in order to obtain insights into the evolutionary conservation of the terpenoid biosynthesis machinery in rice and maize, and to identify additional components of terpenoid biosynthesis, transport, modification, and regulation. Many of the individual results were discussed within the results section to ease the logical flow in the writing. However, several key insights warrant further discussion, especially in reference to future projects that might be inspired by or based on this work.

4.1 Diterpenes and sesquiterpenes are the primary anti-fungal metabolites

Out of the 4 major terpene subtypes which were within the scope of this thesis project (mono-, di-, tri-, sesqui-terpene), the diterpene and sesquiterpene networks in both rice and maize appear to be most involved in the response to biotic stress caused by fungal infections. Monoterpene synthases in both rice and maize were tightly related to gibberellin synthases, possibly due to gibberellin and monoterpene biosynthesis sharing components of the cellular machinery, e.g., chloroplast transport. Triterpene biosynthesis in both rice and maize was weakly associated with other terpene synthases or ancillary genes of interest, reflecting their lack of involvement in anti-fungal responses, at least at the transcriptional level.

It would be highly interesting to analyse a gene network built from transcriptomes of infected roots in rice and maize, especially from single cell data, as this has been used successfully to shed light on vinblastine biosynthesis in Catharanthus, a feat unachievable by bulk RNA-Seq (Li et al., 2023). A similar approach in rice and maize may identify novel triterpene biosynthesis, such as avenacin biosynthesis in roots of oats (Orme et al., 2019). In either scenario, it would be fascinating to analyse a terpenoid gene network of transcriptomes from other stresses such as insect, drought as well as tissues; root, stem, husk. Furthermore, advances in single cell transcriptomics have also made it possible to build single cell gene networks (Morabito et al., 2023), which can provide much information on cell type specific terpene biosynthesis.

4.1.1 Unexpected relationships exist amongst terpene subnetworks

The overall terpenoid networks in both species segregated into modules which also happened to be terpene subtype-specific. While this allowed me to neatly graphically represent functional subnetworks corresponding to terpene subtypes, this visualisation masks some of the underlying complexity.

For example, the maize network is divided into two clusters, one producing well-known anti-fungal diterpenoids and sesquiterpenoids (zealexins), and the other producing gibberellins, monoterpenes and other anti-insect compounds. In contrast, in rice, the overall terpenoid network is divided into volatile, anti-insect mono / tri / sesquiterpenes and defensive diterpenes. The co-occurrence of diterpene synthases with sesquiterpene synthases in maize is substantially different from the observations in rice. However, it fits to published findings that anti-fungal terpenoid metabolism in maize involves promiscuous sesquiterpene and diterpene synthases as well as CYP450s. Unexpectedly, a similar pattern is observed in rice to a lesser degree. The orthogroup of proteins that contains the zealexin synthases *Zx1 / 2 / 3 / 4* (*ZmTPS6 / 12 / 11 / 13*) also contained the rice genes *OsTPS31, OsTPS30* and *OsTPS46*. While *OsTPS30* and *OsTPS31* are located in the sesquiterpenoid networks, *OsTPS46* is integrated tightly into the diterpenoid network, suggesting that there

are as yet uncharacterised rice sesquiterpenoids which arise from interacting with promiscuous diterpene biosynthesis machinery. While the substrate and product of OsTPS46 have been identified (Yuan et al., 2008), future work could express *OsTPS46* together with the CYP450s in the diterpenoid subnetwork in order to identify novel defensive terpenoids.

A second example of complexity in gene networks are the close relationships between the sesquiterpene synthases *ZmTPS8 / 9* and the known gibberellin synthases *ZmAN1, ZmKS2 / 3 / 5*. According to DeepLoc's subcellular predictions, ZmTSP8 is plastid localised whereas ZmTPS9 is cytoplasmic. Both are sesquiterpene synthases, shown to produce sesquiterpenes (Block et al., 2019). Some monoterpene synthases have been shown to produce sesquiterpenes from FPP, while some sesquiterpene synthases have been shown to produce monoterpenes from GPP, implying substrate plasticity. In this case, ZmTPS8 is predicted to localise to the chloroplast and belongs to the same orthogroup as known maize and rice monoterpene synthases (*ZmTPS27, ZmTPS7, OsTPS19, OsTPS20*), yet has been shown to act as a sesquiterpene synthase. Therefore, it is possible that it is actually a plastid-localised monoterpene synthase which has substrate plasticity for FPP as well.

In summary, future efforts should build multiple terpenoid gene networks from varying datasets of diverse tissue types and stresses, to identify more regulatory mechanisms involved in terpene biosynthesis in grasses. It would also be fascinating to extend this analysis to grasses with well assembled genomes and transcriptomes as this would shed light on the extensive variety of terpene metabolism in plants. Considering recent reports of sesterterpenes and non-canonical terpenes in plants, there seem to be many more layers to uncover in plant specialised metabolism than meets the eye.

4.2 Phylogenetics of terpenoid production highlight conserved terpenoid biosynthesis machinery

4.2.1 Promiscuous ancient CYP450s drive terpenoid diversity

Rice and maize diverged approximately 50 - 70 million years ago. Even so, significant portions of the genome are syntenic and the corresponding genes as well as proteins retain significant homology. Amongst these orthologous proteins are CYP450s, which have undergone gene expansion and are involved to varying extents in terpenoid production. For example, the CYP71Z family in rice has been recently shown to be vital for casbene production (Liang et al., 2021), whereas the CYP71Z family in maize is crucial for zealexin production, the main maize antibiotic (Saldivar et al., 2023). To underpin this, members of the CYP71Z family in maize have undergone gene duplication events and are also functionally redundant to each other, in order to limit the negative effects of mutations on antibiotic production (Ding et al., 2020). The CYP76 and CYP701A families are both similar examples of CYP450s that were retained in both rice and maize for producing terpenoids (Mao et al., 2017; Wang et al., 2012, 2012-3). Both families are crucial and are also shown to be promiscuous, participating in a variety of biosynthesis pathways to produce differing terpenoids. CYP99s in monocots are grass-specific CYP450s which were reclassified into the CYP71 family later on (Nelson & Werck-Reichhart, 2011). This suggests that as the CYP71 superfamily of CYP450s diversified, the CYP99 family in rice was adopted into terpenoid production, forming two core enzymes for momilactone biosynthesis. Unlike the CYP71s, CYP76s and CYP701s discussed so far in rice and maize, CYP99s in rice are only involved in momilactone production and have not been shown to act in other diterpenoid biosynthesis pathways. Perhaps this is due to the recent recruitment of CYP99s into terpenoid production in Oryza. Overall, this paints an - admittedly incomplete - picture of the genomic dynamics at play in the evolution of terpene metabolism in rice and maize. If

OTGNs were built in many more grass species, it would be fascinating to study the various CYP450s facilitating terpene production as well as their gene family expansions and synteny.

4.2.2 OETGNs retain ancestral regulatory genes

The OETGNs contained a lower number of CYP450s (CYP71Z, CYP89, CYP701A, CYP73) than the orthologous node network (7 CYP450 families), firstly because it is common for new families of CYP450s to be introduced into or removed from a terpenoid network. A prominent example is CYP99, as discussed before. However, it is less likely for regulatory genes to be excluded from such a conserved network, as regulatory genes in organisms tend to be conserved over time. One key example of this is the conservation of the Hox genes, TFs which regulate body development in all animals (Maconochie et al., 1996).

4.2.2.1 Conserved transcription factors could regulate one half of functional roles of diterpenes

I identified TFs such *as OsDPF, ZmbHLH70 / 98* and WRKYs in both species, as well as the WRKY cofactors, VQ proteins. The conserved TFs also coincided with the corresponding TF binding motifs being conserved in the promoter regions of the genes in the terpenoid network. Noticeably, there were differing ratios of bHLH motifs and WRKY binding motifs between rice and maize, suggesting that there was further evolution of TF binding motifs after maize and rice diverged.

The conservation of the VQ proteins in the orthologous node and edge network is a surprising finding, as it is not common to identify the cofactors of proteins in a gene network. While the VQ-WRKY binding would have to be confirmed, the occurrence of VQs in an orthologous gene network specific for terpenoid production does substantially support their roles as cofactors for conserved WRKYs. Furthermore, AlphaFold prediction of WRKY-VQ binding for all available pairs in both networks suggests similar binding pockets centred around the VQ domains.

A comprehensive motif discovery and enrichment analysis via XSTREME (Meme-suite.org) of the putative promoter sequences in both rice and maize identified a number of common TF binding motifs. These motifs are known to be MYB, ERF, AP2/EREBP and BZIP binding sites. While some of these TFs are known to respond to abiotic stresses such as drought or zinc deficiencies, a significant portion (MYB, ERF, are known to drive leaf architecture as well as panicle development (Komatsu et al., 2003; Li et al., 2022). This could explain how momilactones, for example, accumulate in rice husks. Therefore, it is possible that while bHLHs and WRKYs drive the production of defensive terpenoids in response to biotic / abiotic stress, the bZIP, ERF and AP2 TFs facilitate the accumulation of terpenoids in rice husks in order to protect the developing seed and / or confer an allelopathic / microbiotic advantage to a developing seed.

4.2.2.2 Conserved immune receptors could signal towards co-evolution with pathogens

NLRs have been the subject of much research into plant immunity over the last two decades. They have been shown to be versatile, highly adaptable proteins with varied domains which co-evolve alongside pathogens such as *M. oryzae*. One example would be the evolution of the Pik1 NLR in rice to perceive Avr-PikD, an effector protein from *M. oryzae* (Białas et al., 2021). Paired with the sensor Pik1 NLR, the Pik2 helper NLR is crucial for mounting an effective immune response against *M. oryzae* (De la Concepcion et al., 2021).

The rice and maize OTGNs contained an orthogroup of NLRs with the coiled-coil domain, known as an N-terminal Rx domain (Hao et al., 2013). These NLRs are located in two

clusters on chromosomes one (Clade 1) and seven (Clade 2) (Supplementary Figure 1). Similarly in maize, the clade 1 NLRs are in chromosome eight and clade 2 in chromosome three. At both chromosome one and seven in rice, there are multiple copies of NLRs, of which only a subset is present in the rice OTGN. This gene duplication is only observed within the Oryzoideae, a common observation in NLR research (Borrelli et al., 2018). This suggests that both clade 1 and 2 NLRs are undergoing active selection and evolution in response to the rice specific pathogen, *M. oryzae*, akin to the Pik1/2 locus as well.

Lastly, clade 1 NLRs in both rice and maize contain bHLH binding motifs as well as WRKY binding motifs, whereas the clade 2 NLRs in both species only contained bHLH binding motifs. This suggests that there exists a difference in transcriptional regulation of each NLR. Perhaps this is indicative of each clade being regulated in response to differing pathogens or strains of *M. oryzae*. As NLRs with coiled-coil domains can perform many roles from signalling for cell death to forming pores as multimers (Wang et al., 2021), it was challenging to accurately surmise the function of these NLRs.

A final aspect worth noting on the NLRs in the rice and maize OTGNs is the lack of finding well/characterised NLRs, especially in rice, where key NLRs mediating resistance to M. oryzae have been found (Ding et al., 2020; Liu et al., 2020; Ortiz et al., 2017; Zdrzałek et al., 2020). Perhaps this is due to the NLRs in the OTGNs being helper NLRs which amplify immune responses, leading to production of defensive diterpenoids and zealexins. In any case, only future study can analyse the role these proteins play in driving terpenoid production.

4.2.2.3 Does differential PR10 gene family expansion and involvement in defensive terpene networks suggest sub-functionalisation?

The orthologous edge network also highlighted the conservation of a class of PR10 genes which had undergone gene expansion within the Oryza clade, whereas only a single PR10 gene in maize was strongly associated with anti-fungal terpenoids. First, this suggests that the PR10 gene family is crucial for defensive terpenoid production in both species, especially in rice. Second, the 5 PR10s within rice occur at the same locus that is syntenic to the maize locus containing a single PR10. Transcriptomics has shown that these 5 genes differ substantially in transcriptional response to biotic stresses. *RSOsPR10* is a root specific gene in rice which is induced by pathogen infection, yet it occurs in our leaf-oriented gene network. Of the remaining 4 PR10 genes, at least one is a pseudogene, whereas the other 3 are upregulated in different time frames in response to pathogen infection. While there is some literature that suggests putative roles for these cytoplasmic proteins as cell signalling proteins, there is insufficient evidence to assign specific functions for any of the PR10 genes in our networks. As such, future work on PR10 proteins can include their effects on driving defensive terpenoid production.

4.3 Building terpenoid gene networks from orthologous genes in grasses is a valid tool for predicting biosynthesis of novel metabolites

There have been some reports on a barley diterpene cluster on chromosome 2, involving KSL4, CPS4, CYP89 and CYP99 family members (Liu et al., 2024). In order to independently verify this cluster and identify orthologous regulatory elements in a putative barley diterpenoid network, I built a terpenoid network in barley based on networks extracted from bait genes orthologous to terpenoid genes in rice and maize. This showed that putative diterpene subnetwork 1 (Figure 21a & 21b) contained genes characterised in the hordedane biosynthesis pathway. The aligning of an independent transcriptomic approach with in vivo findings strongly suggests that the putative diterpenoid subnetwork 2 as well as the putative sesquiterpene subnetwork contain genes involved in novel terpene biosynthesis pathways. This is worth following up, and I am looking forward to future research using these data.

The sorghum terpenoid gene network contained many TPS from the OG0000434 orthogroup. The maize terpenoid gene network only contained 2 genes from this orthogroup, whereas sorghum had 7 genes. The opposing trend was observed for the orthogroup OG0001170, where the maize terpenoid gene network contained 9 genes of this orthogroup and sorghum only contained 4. It is clear that in maize, expansion of the genes in OG0001170 led to the production of zealexins, whereas there was a positive selection for the genes in OG0000434 in sorghum. In parallel, the sorghum terpenoid gene network also contained 9 CYP81s belonging to the orthogroup OG0000139, whereas maize only had 3, which catalyse zealexin production. It is possible that in sorghum, the TPS belonging to OG0000434 and the CYP81s in OG0000139 present in all three subnetworks of sorghum could collectively contribute to producing sorgolexins, analogous to the zealexins in maize.

A final observation made in the orthologous network analysis was that for each category of genes / proteins analysed, published literature assigned these genes as either biotic- or abiotic-stress-responsive. This suggests a dichotomy of a regulatory network that drives the production of these defensive terpenoids in response to biotic or abiotic stresses. It is possible that there are parallel regulatory elements which regulate identical steps in terpenoid production yet respond to various stresses. For example, *OsWRKY72* has been reported to be abiotic-stress-responsive, whereas *OsWRKY77* responds to biotic stress. Or, *OsSDRLK54* is responsive to cold stress whereas *OsSDRLK42* is responsive to bacterial infection. Diterpenes such as momilactones and phytocassanes in rice, dolabralexins and kauralexins in maize have been shown to accumulate in leaves during drought conditions as well as upon fungal infections. As such, the genes identified in our network could be responsive to biotic and / or abiotic stresses. Future work could consider these genes as strong candidates for in vitro experiments that focus on characterising this terpenoid network.

4.4 Resistance to momilactone B in Arabidopsis is a complex phenotype

Amongst all of the terpenoids discussed thus far, momilactones are the best studied class of diterpenes. Studies have focused on understanding the induction mechanism, the regulatory TFs and elucidating the complete biosynthetic network that spans 4 chromosomes and 2 known BGCs. In this project, we also attempted to identify the mode of action of these potent allelochemicals. However, multiple mutant screens of F2 generation of EMS mutants as well as HEM mutant collections to identify resistance or tolerance have failed. There are many possible reasons for this result or lack thereof.

It is possible that resistance or tolerance to momilactone B could be a pleiotropic effect. Replicating pleiotropic mutations in a mutant screen can be challenging, especially when considering the availability and cost of procurement of momilactone B. Alternatively, resistance to momilactone B in Arabidopsis could arise from mutations which are lethal in embryos. Therefore, these mutants would be unobtainable in EMS mutagenesis and instead would have to be conditionally knocked out or down, which are expensive and time-consuming projects. A final reason for true resistance or tolerance to be unobtainable was that momilactone B could have multiple sites of action on multiple proteins, thereby requiring the screenings of hundreds of thousands of mutants in order to identify true resistance.

It is also interesting to note that *Echinocloa crus-galli* contains some momilactone biosynthesis genes and has been shown to produce momilactone A, yet is highly sensitive to exogenously applied momilactone B (Kato-Noguchi et al., 2010; Van Quan et al., 2019). As a weed that is known to have been co-cultivated with rice by accident, one would expect

resistance to momilactone B to have evolved in *E. crus-galli*. Hence, if resistance has not emerged in *E. crus-galli*, it might come less as a surprise that a forward-genetic screen in A. thaliana was not able to return individuals resistant to momilactone B. A more informative approach here might be to study momilactone B autotoxicity amongst various rice cultivars and perform GWAS to identify causal regions / SNPs / genes.

In short, large-scale, high-throughput mutant screens as well as study of autotoxicity in rice cultivars could shed more light on the mode of action of momilactone B in plants. As a highly potent allelochemical which has the potential to be a lead compound in herbicide development, it is vital to understand the mode of action of momilactone B. The agronomical impact of momilactone B derived novel herbicides in an increasingly monoculture, highly industrial agricultural system is immense and therefore warrants more research into this topic.

5. Conclusion

Gene networks have increasingly been shown to be powerful tools in understanding regulatory mechanisms in cells in response to specific stresses. In this project, I showed that there are conserved ancillary genes which coordinate the production of defensive terpenoids in response to fungal infections of leaves. I identified key genes encoding for transporters, TFs and immune receptors which cumulatively might perceive biotic stress signals and then transcriptionally regulate the production and transport of defensive terpenoids. Therefore, these genes are strong candidates for future validation assays to completely characterise the production, transport and regulation of defensive terpenoids in rice and maize. I also showed that by comparing orthologous genes, it is possible to identify terpene gene networks in other related species as well. Collectively, the approaches in this thesis and the knowledge generated would substantially advance efforts to increase agricultural output by boosting endogenous defensive traits in crops.

The second goal of identifying the mode of action of momilactone B by identifying resistant mutants in Arabidopsis was unsuccessful, due to a variety of reasons laid out above. As a potent allelochemical, momilactone B's mode of action still warrants future research.

6. Acknowledgements

I would like to express my deepest, most sincere gratitude to Professor Claude Becker for accepting me as a doctoral candidate, encouraging my interests in a wide variety of projects ranging from moss and Chlamydomonas to co-expression gene network analysis. He has never failed to engender curiosity and independence in developing a research project to fruition and for that I thank him wholeheartedly. I would also like to thank Dr Niklas Schandry, an ever-helpful angel of bioinformatics who poked and prodded me into learning bioinformatics and applying it to my never-ending questions. I have had the luck to share my time in the Becker group with some excellent mentors; Jose, Isaac, Zane, Patrick and Eva who all contributed to my development as a young academic. They taught me to critique literature, design research questions more accurately and to harness my curiosity more effectively. Jinn, Liza, Kathi, Nuria, Santi, Duncan, Philipp and Peter were all excellent lab mates to have, bringing much cheer and joy alongside their intellectual excellence across a wide range of topics. They have inspired me during tough times and provided valuable feedback and criticism over the years.

Across my 6 years at the Vienna BioCenter, Gregor Mendel Institute and LMU, I have met some amazing friends who have collectively supported my doctoral journey. To Dejan, Maria, Mariya, Sean, Miguel, Paula, Natalie, Eliana, Jana, Filipe, Bernie, Constance, Alessa, Jessica, Isaac and Kate; you are better friends than I could have asked for and deserve. Thank you for creating such beautiful memories with me and being a part of my life.

I express my deep gratitude to the Genetics Department at the LMU, the Gregor Mendel Institute in Vienna, the LSM Graduate School, Nadine Hamze and the Vienna BioCenter PhD programme for playing a vital role in shaping my doctoral experience.

I deeply appreciate the continuous support from my parents and sister, my parents in law and the Ramans. Finally, the completion of my doctoral thesis would never have been possible without the never-ending, unyielding support of my wife, Sathya. You are the wind beneath my sails, the current of this river we call life and my forever joy. Your constant motivation, positive mindset and fiery spirit were invaluable over the last 6 years and I look forward to the rest of our lives together.

7. Appendix

Supplementary Table 1: Results of protein-protein docking attempted with ALPHAFOLD servers. Predicted template model (PTM) scores reflect the accuracy of the structures, with scores > 0.5 meaning that the predicted structures are similar to the true structures. The interface predicted template model (iPTM) scores reflect the accuracy of the predicted interaction area between the two proteins. Score > 0.8 are understood to be similar to the true structures.

Interaction Partners	VQ	WRKY	iPTM	PTM	Species
OsVQ12 x 2 OsWRKY72s	12	2 x WRKY72	0.27	0.25	Rice
OsVQ12 x 2 OsWRKY77s	12	2 x WRKY77	0.28	0.25	Rice
OsVQ12 x OsWRKY72	12	WRKY72	0.4	0.29	Rice
OsVQ12 x OsWRKY72 x OsWRKY77	12	WRKY72 x WRKY77	0.29	0.25	Rice
OsVQ12 x OsWRKY77	12	WRKY77	0.41	0.29	Rice
OsVQ35 x 2 OsWRKY72s	35	2 x WRKY72	0.28	0.22	Rice
OsVQ35 x 2 OsWRKY77s	35	2 x WRKY77	0.3	0.23	Rice
OsVQ35 x OsWRKY72	35	WRKY72	0.43	0.25	Rice
OsVQ35 x OsWRKY72 x OsWRKY77	35	WRKY72 x WRKY77	0.31	0.23	Rice
OsVQ35 x OsWRKY77	35	WRKY77	0.43	0.24	Rice
ZmVQ41 x ZmWRKY100	41	100	0.48	0.35	Maize
ZmVQ41 x ZmWRKY108	41	108	0.44	0.33	Maize
ZmVQ41 x ZmWRKY34	41	34	0.5	0.34	Maize
ZmVQ41 x ZmWRKY36	41	36	0.42	0.32	Maize
ZmVQ41 x ZmWRKY68	41	68	0.53	0.35	Maize
ZmVQ41 x ZmWRKY73	41	73	0.47	0.34	Maize
Zm00001eb084880 x ZmWRKY100	Zm00001eb084880	100	0.5	0.34	Maize
Zm00001eb084880 x ZmWRKY108	Zm00001eb084880	108	0.44	0.3	Maize
Zm00001eb084880 x ZmWRKY34	Zm00001eb084880	34	0.47	0.29	Maize
Zm00001eb084880 x ZmWRKY36	Zm00001eb084880	36	0.43	0.3	Maize
Zm00001eb084880 x ZmWRKY68	Zm00001eb084880	68	0.52	0.29	Maize
Zm00001eb084880 x ZmWRKY73	Zm00001eb084880	73	0.47	0.27	Maize

Supplementary Table 2: Subcellular localisation prediction of protein sequences of maize genes in OTGN. Prediction was performed by Deeploc.

Protein_ID	Localizations	Signals	Protein_ID	Localizations	Signals

ZM00001EB0 01950	Plastid	Chloroplast transit peptide	ZM00001EB2 22680	Endoplasmic reticulum	Signal peptide|Transme mbrane domain
ZM00001EB0 02270	Plastid	Chloroplast transit peptide	ZM00001EB2 23910	Cell membrane	Transmembrane domain
ZM00001EB0 04050	Plastid	Chloroplast transit peptide	ZM00001EB2 24350	Endoplasmic reticulum	Signal peptide|Transme mbrane domain
ZM00001EB0 04420	Cytoplasm		ZM00001EB2 30410	Plastid	Chloroplast transit peptide
ZM00001EB0 04570	Plastid		ZM00001EB2 30440	Plastid	Chloroplast transit peptide
ZM00001EB0 08810	Cell membrane	Transmembrane domain	ZM00001EB2 36900	Plastid	Chloroplast transit peptide
ZM00001EB0 10210	Cytoplasm|Nucleus	Nuclear localization signal	ZM00001EB2 37530	Cell membrane|Lysosome /Vacuole	Signal peptide|Transme mbrane domain
ZM00001EB0 12130	Plastid	Chloroplast transit peptide	ZM00001EB2 39740	Mitochondrion	
ZM00001EB0 14950	Plastid		ZM00001EB2 40970	Cell membrane|Lysosome /Vacuole	Signal peptide
ZM00001EB0 14960	Plastid		ZM00001EB2 42930	Nucleus	Nuclear localization signal
ZM00001EB0 15510	Cell membrane|Lysosome /Vacuole	Transmembrane domain	ZM00001EB2 43310	Cytoplasm|Nucleus	Nuclear localization signal
ZM00001EB0 15530	Cell membrane|Endoplas mic reticulum|Lysosome/ Vacuole	Signal peptide|Transme mbrane domain	ZM00001EB2 43590	Plastid	Chloroplast transit peptide
ZM00001EB0 16730	Cytoplasm		ZM00001EB2 45320	Plastid	
ZM00001EB0 17120	Plastid		ZM00001EB2 46090	Plastid	Chloroplast transit peptide
ZM00001EB0 17730	Cell membrane	Transmembrane domain	ZM00001EB2 46340	Plastid	Chloroplast transit peptide
ZM00001EB0 18860	Nucleus	Nuclear localization signal	ZM00001EB2 48540	Cytoplasm	
ZM00001EB0 19620	Cell membrane		ZM00001EB2 48960	Endoplasmic reticulum	Signal peptide|Transme mbrane domain|Peroxiso mal targeting signal
ZM00001EB0 20060	Endoplasmic reticulum	Signal peptide|Transme mbrane domain	ZM00001EB2 51520	Plastid	Chloroplast transit peptide
ZM00001EB0 21200	Plastid	Chloroplast transit peptide	ZM00001EB2 53760	Endoplasmic reticulum	Signal peptide
ZM00001EB0 21690	Cytoplasm		ZM00001EB2 55850	Plastid	Chloroplast transit peptide

ZM00001EB026100	Cytoplasm|Nucleus	Nuclear localization signal|Nuclear export signal	ZM00001EB258200	Plastid	Chloroplast transit peptide
ZM00001EB028930	Plastid	Chloroplast transit peptide	ZM00001EB258710	Mitochondrion	Mitochondrial transit peptide
ZM00001EB029480	Cytoplasm	Nuclear localization signal	ZM00001EB258750	Plastid	Chloroplast transit peptide
ZM00001EB030650	Cell membrane	Nuclear localization signal	ZM00001EB267020	Plastid	Chloroplast transit peptide
ZM00001EB031040	Endoplasmic reticulum	Signal peptide|Transmembrane domain	ZM00001EB267640	Cytoplasm	
ZM00001EB032520	Extracellular	Signal peptide	ZM00001EB268360	Plastid	Chloroplast transit peptide
ZM00001EB032560	Extracellular	Signal peptide	ZM00001EB271820	Plastid	Thylakoid luminal transit peptide
ZM00001EB032580	Extracellular	Signal peptide	ZM00001EB272110	Cytoplasm	
ZM00001EB032600	Extracellular	Signal peptide	ZM00001EB277710	Extracellular|Endoplasmic reticulum|Lysosome/Vacuole	Signal peptide
ZM00001EB039850	Plastid	Chloroplast transit peptide	ZM00001EB278070	Cytoplasm|Nucleus	Nuclear export signal
ZM00001EB041400	Plastid	Chloroplast transit peptide	ZM00001EB278400	Cytoplasm	
ZM00001EB041770	Cytoplasm		ZM00001EB280560	Plastid	Chloroplast transit peptide
ZM00001EB042200	Cell membrane	Peroxisomal targeting signal	ZM00001EB282710	Endoplasmic reticulum	Signal peptide
ZM00001EB042280	Plastid		ZM00001EB283060	Cytoplasm|Nucleus	
ZM00001EB042710	Cytoplasm|Nucleus	Nuclear localization signal	ZM00001EB287190	Plastid	Chloroplast transit peptide
ZM00001EB042940	Plastid	Chloroplast transit peptide	ZM00001EB288110	Mitochondrion|Plastid	Mitochondrial transit peptide
ZM00001EB043620	Endoplasmic reticulum	Signal peptide|Transmembrane domain	ZM00001EB289570	Cytoplasm|Plastid	Peroxisomal targeting signal
ZM00001EB043630	Endoplasmic reticulum	Signal peptide|Transmembrane domain	ZM00001EB290720	Plastid	
ZM00001EB047160	Plastid	Chloroplast transit peptide	ZM00001EB293420	Nucleus	Nuclear localization signal
ZM00001EB048020	Plastid	Chloroplast transit peptide	ZM00001EB293660	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
ZM00001EB050280	Cytoplasm|Nucleus	Nuclear localization signal|Nuclear export signal	ZM00001EB294780	Plastid	
ZM00001EB052520	Extracellular	Signal peptide	ZM00001EB296670	Cell membrane	

ZM00001EB055320	Plastid	Chloroplast transit peptide	ZM00001EB299340	Extracellular	Signal peptide
ZM00001EB056550	Mitochondrion	Mitochondrial transit peptide	ZM00001EB299370	Extracellular	Signal peptide
ZM00001EB058590	Plastid	Thylakoid luminal transit peptide	ZM00001EB300570	Endoplasmic reticulum	
ZM00001EB059070	Plastid		ZM00001EB301210	Plastid	Chloroplast transit peptide
ZM00001EB063160	Plastid		ZM00001EB302370	Plastid	Chloroplast transit peptide
ZM00001EB063710	Plastid	Chloroplast transit peptide	ZM00001EB306760	Plastid	
ZM00001EB069530	Cell membrane	Signal peptide|Transmembrane domain	ZM00001EB312350	Lysosome/Vacuole	Transmembrane domain
ZM00001EB071070	Plastid	Chloroplast transit peptide	ZM00001EB314770	Mitochondrion	Mitochondrial transit peptide
ZM00001EB071080	Plastid	Chloroplast transit peptide	ZM00001EB315490	Cytoplasm	
ZM00001EB071090	Plastid		ZM00001EB315600	Cell membrane	Signal peptide|Transmembrane domain
ZM00001EB071310	Cytoplasm		ZM00001EB319610	Plastid	Chloroplast transit peptide
ZM00001EB071870	Mitochondrion	Mitochondrial transit peptide	ZM00001EB319770	Plastid	Chloroplast transit peptide
ZM00001EB071890	Mitochondrion	Mitochondrial transit peptide	ZM00001EB322600	Plastid	
ZM00001EB072370	Plastid	Chloroplast transit peptide	ZM00001EB322880	Cell membrane	Transmembrane domain
ZM00001EB074700	Mitochondrion		ZM00001EB323300	Cell membrane	Signal peptide|Transmembrane domain
ZM00001EB075370	Plastid	Chloroplast transit peptide	ZM00001EB324080	Plastid	
ZM00001EB075660	Cytoplasm		ZM00001EB325300	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
ZM00001EB081570	Cytoplasm		ZM00001EB325340	Extracellular	Signal peptide
ZM00001EB084880	Nucleus	Nuclear localization signal	ZM00001EB325560	Plastid	Chloroplast transit peptide
ZM00001EB086340	Mitochondrion	Mitochondrial transit peptide	ZM00001EB325890	Endoplasmic reticulum	Transmembrane domain
ZM00001EB088390	Plastid		ZM00001EB327890	Nucleus	Nuclear localization signal
ZM00001EB089390	Plastid	Chloroplast transit peptide	ZM00001EB329630	Plastid	Chloroplast transit peptide
ZM00001EB093580	Cell membrane	Transmembrane domain	ZM00001EB329990	Cytoplasm|Nucleus	Nuclear localization signal|Nuclear export signal
ZM00001EB096060	Plastid	Chloroplast transit peptide	ZM00001EB330460	Plastid	Chloroplast transit peptide

ZM00001EB096890	Cell membrane	Nuclear localization signal	ZM00001EB335810	Plastid	Thylakoid luminal transit peptide
ZM00001EB101180	Cell membrane	Signal peptide|Transmembrane domain	ZM00001EB335940	Plastid	Chloroplast transit peptide
ZM00001EB108040	Endoplasmic reticulum		ZM00001EB336120	Cytoplasm	
ZM00001EB109480	Nucleus|Cell membrane		ZM00001EB337860	Nucleus	Nuclear localization signal
ZM00001EB112840	Nucleus	Nuclear localization signal	ZM00001EB341580	Extracellular	Signal peptide
ZM00001EB114130	Cytoplasm		ZM00001EB342550	Cytoplasm|Nucleus	Nuclear localization signal
ZM00001EB116470	Endoplasmic reticulum	Signal peptide	ZM00001EB346280	Plastid	Chloroplast transit peptide
ZM00001EB116560	Extracellular	Peroxisomal targeting signal	ZM00001EB346800	Plastid	
ZM00001EB117000	Extracellular|Cell membrane		ZM00001EB346880	Nucleus	Nuclear localization signal
ZM00001EB118450	Plastid	Chloroplast transit peptide	ZM00001EB349410	Plastid	Mitochondrial transit peptide
ZM00001EB120960	Nucleus	Nuclear localization signal	ZM00001EB353610	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
ZM00001EB122690	Cytoplasm		ZM00001EB353710	Nucleus	Nuclear localization signal|Nuclear export signal
ZM00001EB124300	Plastid	Chloroplast transit peptide	ZM00001EB354370	Nucleus	Nuclear localization signal
ZM00001EB124900	Cell membrane	Signal peptide|Transmembrane domain	ZM00001EB356430	Nucleus	Nuclear localization signal
ZM00001EB126080	Plastid	Chloroplast transit peptide	ZM00001EB357520	Plastid	Chloroplast transit peptide
ZM00001EB127150	Plastid		ZM00001EB357950	Cell membrane	Transmembrane domain
ZM00001EB127360	Extracellular	Signal peptide	ZM00001EB360740	Cell membrane|Endoplasmic reticulum	Signal peptide|Transmembrane domain
ZM00001EB127510	Plastid		ZM00001EB361650	Nucleus	
ZM00001EB128290	Cytoplasm		ZM00001EB361660	Nucleus	
ZM00001EB130870	Cytoplasm		ZM00001EB363910	Cytoplasm|Nucleus	Nuclear localization signal
ZM00001EB131480	Cell membrane	Nuclear localization signal	ZM00001EB365000	Plastid	Chloroplast transit peptide
ZM00001EB132050	Cytoplasm		ZM00001EB366810	Plastid	Chloroplast transit peptide
ZM00001EB133200	Plastid	Chloroplast transit peptide	ZM00001EB368070	Nucleus	Nuclear localization signal

ZM00001EB139760	Plastid		ZM00001EB368900	Endoplasmic reticulum	Transmembrane domain
ZM00001EB140490	Plastid	Chloroplast transit peptide	ZM00001EB368910	Cytoplasm	
ZM00001EB140540	Plastid	Chloroplast transit peptide	ZM00001EB370480	Plastid	Chloroplast transit peptide
ZM00001EB141870	Plastid	Chloroplast transit peptide	ZM00001EB371640	Plastid	Chloroplast transit peptide
ZM00001EB142370	Cell membrane|Lysosome/Vacuole	Transmembrane domain	ZM00001EB373350	Lysosome/Vacuole|Golgi apparatus	
ZM00001EB145860	Extracellular	Signal peptide	ZM00001EB374210	Plastid	Chloroplast transit peptide
ZM00001EB147750	Plastid	Chloroplast transit peptide	ZM00001EB377150	Plastid	
ZM00001EB147760	Plastid	Chloroplast transit peptide	ZM00001EB378020	Plastid	Chloroplast transit peptide
ZM00001EB147990	Cytoplasm	Nuclear export signal	ZM00001EB378030	Plastid	Chloroplast transit peptide
ZM00001EB148110	Cytoplasm|Peroxisome	Peroxisomal targeting signal	ZM00001EB379570	Plastid	Chloroplast transit peptide
ZM00001EB148580	Nucleus		ZM00001EB385070	Plastid	
ZM00001EB149860	Endoplasmic reticulum	Transmembrane domain	ZM00001EB385100	Plastid|Endoplasmic reticulum	
ZM00001EB153630	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain	ZM00001EB385110	Nucleus	
ZM00001EB154000	Cytoplasm		ZM00001EB386680	Plastid	Chloroplast transit peptide
ZM00001EB154670	Cytoplasm		ZM00001EB387370	Nucleus	
ZM00001EB154960	Nucleus	Nuclear localization signal	ZM00001EB387920	Cytoplasm	
ZM00001EB155590	Cytoplasm|Nucleus		ZM00001EB387930	Cytoplasm	
ZM00001EB156190	Cytoplasm		ZM00001EB393630	Lysosome/Vacuole	
ZM00001EB156250	Cell membrane	Signal peptide|Transmembrane domain	ZM00001EB400190	Cytoplasm	
ZM00001EB160050	Cytoplasm|Cell membrane	Nuclear localization signal	ZM00001EB400820	Cytoplasm	Nuclear localization signal
ZM00001EB169360	Plastid	Chloroplast transit peptide	ZM00001EB404590	Cytoplasm	
ZM00001EB169460	Nucleus	Nuclear localization signal	ZM00001EB406810	Plastid	Chloroplast transit peptide
ZM00001EB170180	Extracellular	Signal peptide	ZM00001EB409690	Endoplasmic reticulum	Signal peptide
ZM00001EB171030	Mitochondrion	Transmembrane domain	ZM00001EB409700	Endoplasmic reticulum	Signal peptide

ZM00001EB172970	Mitochondrion	Mitochondrial transit peptide|Peroxisomal targeting signal	ZM00001EB410040	Cytoplasm	Chloroplast transit peptide
ZM00001EB177670	Plastid	Chloroplast transit peptide	ZM00001EB411340	Cytoplasm	
ZM00001EB178270	Cell membrane	Transmembrane domain	ZM00001EB412960	Cytoplasm	
ZM00001EB180060	Cytoplasm		ZM00001EB412970	Cytoplasm	
ZM00001EB183770	Plastid	Chloroplast transit peptide	ZM00001EB412980	Cytoplasm	
ZM00001EB190340	Cytoplasm	Nuclear localization signal	ZM00001EB412990	Cytoplasm	
ZM00001EB190430	Extracellular	Signal peptide	ZM00001EB414190	Cytoplasm	
ZM00001EB198200	Peroxisome		ZM00001EB414630	Cytoplasm	Nuclear export signal
ZM00001EB198740	Lysosome/Vacuole	Transmembrane domain	ZM00001EB415070	Cytoplasm	
ZM00001EB200520	Cell membrane		ZM00001EB415080	Plastid	
ZM00001EB200910	Plastid	Chloroplast transit peptide	ZM00001EB415090	Cytoplasm	Nuclear export signal
ZM00001EB203770	Plastid	Thylakoid luminal transit peptide	ZM00001EB415100	Cytoplasm	
ZM00001EB204910	Cytoplasm	Nuclear localization signal	ZM00001EB415160	Cytoplasm	
ZM00001EB208380	Nucleus	Nuclear export signal	ZM00001EB415420	Plastid	Chloroplast transit peptide
ZM00001EB208940	Plastid	Chloroplast transit peptide	ZM00001EB415430	Plastid	Chloroplast transit peptide
ZM00001EB215460	Nucleus		ZM00001EB416690	Cytoplasm	Nuclear export signal
ZM00001EB217970	Extracellular	Signal peptide	ZM00001EB416700	Cytoplasm	
ZM00001EB220030	Endoplasmic reticulum	Signal peptide|Transmembrane domain	ZM00001EB416720	Cytoplasm	
ZM00001EB220500	Cell membrane	Signal peptide|Transmembrane domain	ZM00001EB418980	Plastid	Mitochondrial transit peptide
ZM00001EB222540	Endoplasmic reticulum	Signal peptide|Transmembrane domain	ZM00001EB419500	Cell membrane	Transmembrane domain
ZM00001EB222660	Endoplasmic reticulum		ZM00001EB419890	Cytoplasm	Nuclear export signal
ZM00001EB424460	Plastid	Chloroplast transit peptide	ZM00001EB420390	Mitochondrion	Mitochondrial transit peptide
ZM00001EB432420	Endoplasmic reticulum	Signal peptide|Transmembrane domain	ZM00001EB423470	Cytoplasm	

Supplementary Table 3: Subcellular localisation prediction of protein sequences of maize genes in OTGN. Prediction was performed by Deeploc.

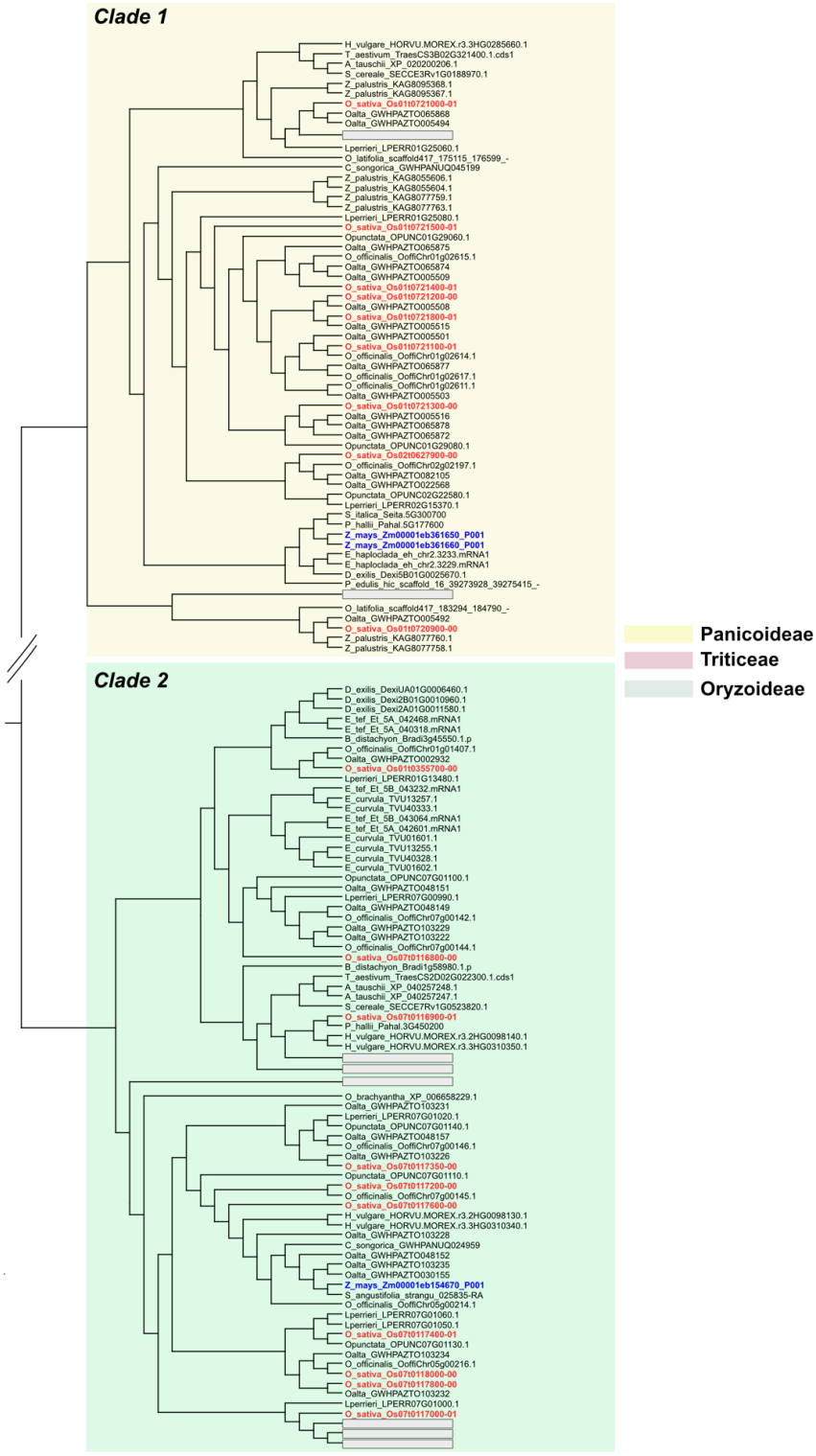
Protein_ID	Localizations	Signals	Protein_ID	Localizations	Signals
OS01G0101200	Plastid		OS04G0633300	Cell membrane	Signal peptide|Transmembrane domain
OS01G0118000	Plastid	Chloroplast transit peptide	OS04G0677300	Cell membrane	Transmembrane domain
OS01G0144100	Plastid	Thylakoid luminal transit peptide	OS05G0102000	Cytoplasm|Nucleus	Nuclear localization signal
OS01G0151200	Plastid	Chloroplast transit peptide	OS05G0141300	Cytoplasm	Nuclear localization signal
OS01G0151200	Endoplasmic reticulum		OS05G0156300	Endoplasmic reticulum	Signal peptide
OS01G0151200	Cytoplasm		OS05G0177500	Cytoplasm	
OS01G0158400	Nucleus		OS05G0291700	Plastid	Chloroplast transit peptide
OS01G0196300	Nucleus	Nuclear localization signal	OS05G0320700	Endoplasmic reticulum	Signal peptide|Transmembrane domain
OS01G0205500	Cytoplasm|Nucleus	Nuclear localization signal	OS05G0331200	Mitochondrion	Peroxisomal targeting signal
OS01G0252600	Plastid	Chloroplast transit peptide	OS05G0388600	Plastid	
OS01G0264700	Cytoplasm		OS05G0390100	Mitochondrion	Transmembrane domain|Peroxisomal targeting signal
OS01G0297200	Endoplasmic reticulum		OS05G0401100	Plastid	
OS01G0323600	Cytoplasm		OS05G0401200	Plastid	Thylakoid luminal transit peptide
OS01G0339900	Endoplasmic reticulum	Signal peptide	OS05G0480000	Plastid	
OS01G0342750	Cell membrane	Transmembrane domain	OS05G0480000	Cytoplasm|Plastid	Peroxisomal targeting signal
OS01G0376600	Nucleus	Nuclear localization signal	OS05G0522600	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
OS01G0382000	Extracellular	Signal peptide	OS05G0522600	Cytoplasm|Cell membrane	Nuclear localization signal
OS01G0561600	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS05G0526700	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
OS01G0584900	Nucleus	Nuclear localization signal	OS05G0530400	Nucleus	Nuclear localization signal
OS01G0589800	Plastid		OS05G0550800	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
OS01G0609300	Cell membrane	Transmembrane domain	OS05G0576600	Cell membrane	
OS01G0609900	Cell membrane		OS05G0582300	Plastid	Chloroplast transit peptide
OS01G0609900	Cell membrane	Transmembrane domain	OS06G0101600	Plastid	Chloroplast transit peptide
OS01G0612500	Cytoplasm|Nucleus	Nuclear localization signal	OS06G0133800	Plastid	
OS01G0658400	Cytoplasm|Nucleus	Nuclear export signal	OS06G0133800	Cytoplasm|Nucleus	Nuclear localization signal
OS01G0693300	Cell membrane|Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS06G0133900	Plastid	Chloroplast transit peptide
OS01G0693900	Plastid		OS06G0146300	Cytoplasm|Nucleus	Nuclear localization signal
OS01G0703400	Cytoplasm|Nucleus	Nuclear localization signal	OS06G0210900	Golgi apparatus	

OS01G0721200	Nucleus		OS06G0245800	Plastid	Chloroplast transit peptide
OS01G0721300	Nucleus		OS06G0254300	Endoplasmic reticulum	Signal peptide
OS01G0721400	Nucleus		OS06G0254300	Mitochondrion	
OS01G0734600	Cytoplasm		OS06G0264800	Plastid	Chloroplast transit peptide
OS01G0761000	Plastid	Chloroplast transit peptide	OS06G0313440	Cytoplasm	
OS01G0814900	Endoplasmic reticulum	Transmembrane domain	OS06G0354500	Cytoplasm|Peroxisome	Peroxisomal targeting signal
OS01G0838600	Nucleus	Nuclear localization signal	OS06G0549600	Extracellular	Signal peptide
OS01G0847700	Cytoplasm		OS06G0549900	Extracellular|Lysosome/Vacuole	Signal peptide
OS01G0862200	Plastid	Chloroplast transit peptide	OS06G0568600	Plastid|Endoplasmic reticulum	
OS01G0871300	Plastid	Chloroplast transit peptide	OS06G0569300	Cytoplasm	
OS01G0871300	Cytoplasm	Peroxisomal targeting signal	OS06G0569500	Plastid|Endoplasmic reticulum	
OS01G0879200	Extracellular|Lysosome/Vacuole	Signal peptide	OS06G0570900	Nucleus	
OS01G0897600	Cytoplasm		OS06G0647100	Plastid	Chloroplast transit peptide
OS01G0912700	Plastid	Chloroplast transit peptide	OS06G0729650	Plastid	
OS01G0925700	Cytoplasm|Nucleus	Nuclear localization signal	OS07G0116900	Golgi apparatus	
OS01G0930400	Cell membrane|Lysosome/Vacuole	Transmembrane domain	OS07G0117200	Nucleus	
OS01G0938100	Plastid	Chloroplast transit peptide	OS07G0117800	Nucleus	
OS01G0958100	Plastid	Chloroplast transit peptide	OS07G0117900	Nucleus	Nuclear localization signal
OS01G0958100	Nucleus	Nuclear localization signal	OS07G0118000	Nucleus	
OS01G0959100	Cell membrane	Nuclear localization signal	OS07G0129200	Extracellular	Signal peptide
OS02G0121700	Plastid	Chloroplast transit peptide	OS07G0129300	Extracellular	Signal peptide
OS02G0129900	Plastid	Chloroplast transit peptide	OS07G0129800	Cell membrane	Signal peptide
OS02G0151300	Extracellular	Signal peptide	OS07G0129900	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
OS02G0226000	Mitochondrion	Mitochondrial transit peptide	OS07G0162100	Endoplasmic reticulum	
OS02G0278700	Plastid		OS07G0175600	Cell membrane	
OS02G0285800	Plastid	Chloroplast transit peptide	OS07G0176900	Plastid	Chloroplast transit peptide
OS02G0285800	Cytoplasm|Nucleus	Nuclear localization signal	OS07G0190000	Plastid	Chloroplast transit peptide
OS02G0458100	Plastid	Chloroplast transit peptide	OS07G0217600	Endoplasmic reticulum	Signal peptide|Transmembrane domain
OS02G0536500	Nucleus	Nuclear localization signal	OS07G0218200	Cytoplasm	
OS02G0553200	Plastid	Chloroplast transit peptide	OS07G0218700	Plastid	
OS02G0568700	Plastid	Chloroplast transit peptide	OS07G0243150	Plastid	

OS02G0569000	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS07G0416700	Endoplasmic reticulum	Signal peptide
OS02G0569400	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS07G0416900	Endoplasmic reticulum	Transmembrane domain
OS02G0569400	Plastid		OS07G0417200	Endoplasmic reticulum	Signal peptide
OS02G0569900	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS07G0448100	Cell membrane	Transmembrane domain
OS02G0570400	Plastid	Chloroplast transit peptide	OS07G0448800	Cell membrane	Transmembrane domain
OS02G0570500	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS07G0449100	Mitochondrion	Transmembrane domain
OS02G0570700	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS07G0464600	Plastid	
OS02G0571100	Plastid	Chloroplast transit peptide	OS07G0513000	Plastid	
OS02G0571300	Plastid	Chloroplast transit peptide	OS07G0522500	Cell membrane	Transmembrane domain
OS02G0571900	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS07G0540600	Extracellular	Signal peptide
OS02G0581100	Plastid	Chloroplast transit peptide	OS07G0541000	Cell membrane	Signal peptide|Transmembrane domain
OS02G0595700	Plastid	Chloroplast transit peptide	OS07G0546000	Cytoplasm|Peroxisome	Peroxisomal targeting signal
OS02G0595700	Cytoplasm	Nuclear localization signal	OS07G0558200	Plastid	
OS02G0600200	Plastid	Chloroplast transit peptide	OS07G0558200	Cytoplasm|Nucleus	Nuclear export signal
OS02G0600200	Nucleus	Nuclear localization signal	OS07G0558300	Plastid	
OS02G0629200	Cell membrane	Transmembrane domain	OS07G0580900	Plastid	Chloroplast transit peptide
OS02G0640500	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain	OS07G0585900	Endoplasmic reticulum	Transmembrane domain
OS02G0649800	Endoplasmic reticulum	Peroxisomal targeting signal	OS07G0613500	Cell membrane	Nuclear localization signal
OS02G0705100	Plastid	Thylakoid luminal transit peptide	OS07G0619500	Cell membrane	
OS02G0718600	Cell membrane		OS07G0657900	Plastid	Chloroplast transit peptide
OS02G0735200	Cytoplasm		OS07G0663000	Nucleus	Nuclear localization signal
OS02G0759900	Plastid	Chloroplast transit peptide	OS08G0139700	Cytoplasm	
OS02G0804500	Plastid	Chloroplast transit peptide	OS08G0139700	Plastid	
OS02G0814400	Mitochondrion		OS08G0140300	Cytoplasm	Nuclear export signal
OS02G0815400	Plastid	Chloroplast transit peptide	OS08G0158200	Lysosome/Vacuole	Signal peptide
OS03G0129300	Plastid	Chloroplast transit peptide	OS08G0167800	Cytoplasm	
OS03G0129300	Cytoplasm		OS08G0168000	Cytoplasm	
OS03G0200500	Cytoplasm	Nuclear localization signal	OS08G0168400	Cytoplasm	
OS03G0218400	Cell membrane	Transmembrane domain	OS08G0243500	Cell membrane	

OS03G0240600	Cytoplasm|Nucleus	Nuclear localization signal	OS08G0243500	Cytoplasm	Nuclear export signal
OS03G0271100	Plastid	Chloroplast transit peptide	OS08G0243600	Cytoplasm	
OS03G0311300	Plastid	Chloroplast transit peptide	OS08G0288200	Plastid	
OS03G0311300	Mitochondrion|Plastid	Mitochondrial transit peptide|Peroxisomal targeting signal	OS08G0359000	Plastid	Chloroplast transit peptide
OS03G0311300	Plastid		OS08G0411200	Plastid	Chloroplast transit peptide
OS03G0326000	Cell membrane|Lysosome/Vacuole	Transmembrane domain	OS08G0434300	Plastid	Chloroplast transit peptide
OS03G0326200	Cell membrane|Lysosome/Vacuole	Transmembrane domain	OS08G0514033	Nucleus	Nuclear localization signal
OS03G0326200	Lysosome/Vacuole	Signal peptide|Transmembrane domain	OS08G0538200	Endoplasmic reticulum	
OS03G0343900	Plastid	Chloroplast transit peptide	OS08G0553800	Plastid	Chloroplast transit peptide
OS03G0347900	Cytoplasm		OS08G0566900	Mitochondrion|Peroxisome	Peroxisomal targeting signal
OS03G0348200	Cytoplasm		OS09G0364800	Cell membrane	Peroxisomal targeting signal
OS03G0361500	Cytoplasm		OS09G0370500	Nucleus	Nuclear localization signal
OS03G0361600	Cytoplasm		OS09G0394300	Extracellular	
OS03G0362500	Cytoplasm		OS09G0410400	Cytoplasm|Lysosome/Vacuole	Nuclear export signal
OS03G0363500	Cell membrane		OS09G0422000	Cytoplasm	Nuclear localization signal
OS03G0363500	Plastid	Chloroplast transit peptide	OS09G0454600	Mitochondrion	
OS03G0387300	Nucleus	Nuclear localization signal	OS09G0467200	Cytoplasm	
OS03G0432000	Cell membrane|Endoplasmic reticulum	Transmembrane domain	OS09G0468300	Cell membrane	Signal peptide|Transmembrane domain
OS03G0625300	Nucleus		OS09G0491820	Cytoplasm	
OS03G0661600	Extracellular	Signal peptide	OS09G0532700	Plastid	Chloroplast transit peptide
OS03G0663500	Extracellular	Signal peptide	OS09G0535000	Plastid	Chloroplast transit peptide
OS03G0718000	Plastid	Chloroplast transit peptide	OS09G0551000	Cell membrane	Signal peptide|Transmembrane domain
OS03G0738400	Mitochondrion	Mitochondrial transit peptide	OS09G0551400	Cell membrane	Signal peptide|Transmembrane domain
OS03G0738400	Mitochondrion		OS09G0558900	Endoplasmic reticulum	Transmembrane domain
OS03G0738400	Cytoplasm		OS09G0567366	Plastid	Chloroplast transit peptide
OS03G0765900	Plastid	Thylakoid luminal transit peptide	OS09G0567366	Cytoplasm|Plastid	Chloroplast transit peptide|Peroxisomal targeting signal
OS03G0773800	Cytoplasm		OS10G0100300	Plastid	Chloroplast transit peptide
OS03G0835900	Plastid	Chloroplast transit peptide	OS10G0439800	Endoplasmic reticulum	Signal peptide|Transmembrane domain

OS04G0102500	Cytoplasm|Nucleus		OS10G0439924	Endoplasmic reticulum	Signal peptide|Transmembrane domain
OS04G0104900	Cytoplasm		OS10G0441900	Cell membrane|Lysosome/Vacuole	Signal peptide|Transmembrane domain
OS04G0111200	Cytoplasm	Peroxisomal targeting signal	OS10G0478200	Cytoplasm	
OS04G0167800	Cytoplasm		OS10G0513900	Endoplasmic reticulum	Signal peptide|Transmembrane domain
OS04G0178300	Plastid		OS10G0530500	Cytoplasm|Plastid	Chloroplast transit peptide
OS04G0178300	Cytoplasm	Nuclear localization signal	OS10G0535800	Lysosome/Vacuole	Transmembrane domain
OS04G0178300	Plastid	Chloroplast transit peptide	OS11G0126100	Cell membrane|Lysosome/Vacuole	Transmembrane domain
OS04G0178400	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS11G0151300	Cytoplasm	
OS04G0179700	Plastid		OS11G0267000	Plastid	Chloroplast transit peptide
OS04G0180400	Endoplasmic reticulum	Signal peptide|Transmembrane domain	OS11G0474800	Plastid	Chloroplast transit peptide
OS04G0234600	Plastid	Chloroplast transit peptide	OS11G0490900	Nucleus	Nuclear localization signal
OS04G0298200	Mitochondrion	Mitochondrial transit peptide	OS11G0514400	Extracellular	Signal peptide
OS04G0304200	Cytoplasm	Nuclear localization signal|Nuclear export signal	OS11G0514500	Extracellular	Signal peptide
OS04G0304200	Cytoplasm|Nucleus	Nuclear export signal	OS11G0544500	Plastid	
OS04G0340300	Plastid	Chloroplast transit peptide	OS11G0582000	Cell membrane	
OS04G0341500	Cytoplasm|Endoplasmic reticulum	Signal peptide	OS11G0587600	Cell membrane	Transmembrane domain
OS04G0341500	Plastid	Chloroplast transit peptide	OS11G0641500	Extracellular	Signal peptide
OS04G0344100	Cytoplasm		OS12G0149900	Cytoplasm	Nuclear localization signal
OS04G0447700	Cytoplasm|Mitochondrion	Mitochondrial transit peptide	OS12G0181500	Cell membrane	Transmembrane domain
OS04G0465500	Plastid	Chloroplast transit peptide	OS12G0199800	Endoplasmic reticulum	Signal peptide|Transmembrane domain
OS04G0531750	Cytoplasm		OS12G0271700	Plastid	Chloroplast transit peptide
OS04G0539000	Plastid	Chloroplast transit peptide	OS12G0491800	Plastid	Chloroplast transit peptide
OS04G0591000	Plastid	Chloroplast transit peptide	OS12G0555000	Cytoplasm|Nucleus	Nuclear localization signal
OS04G0600300	Mitochondrion	Mitochondrial transit peptide	OS12G0555100	Cytoplasm|Nucleus	Nuclear localization signal
OS04G0608300	Cell membrane		OS12G0555200	Cytoplasm	
OS04G0611700	Cytoplasm		OS12G0555300	Cytoplasm	Nuclear localization signal
OS04G0611800	Cytoplasm		OS12G0555500	Cytoplasm|Nucleus	Nuclear localization signal



Supplementary Figure 1: A phylogenetic tree of the orthogroup OG0000059, which contains NLRs, split into two clades. Rice proteins are in red and bold, maize proteins in blue and bold. Some branches of the tree have been truncated as boxes in order to ease visualisation. The distance between the clades has also been truncated and is represented as two diagonal dashes.

8. Bibliography

Abe, I., Sakano, Y., Sodeyama, M., Tanaka, H., Noguchi, H., Shibuya, M., & Ebizuka, Y. (2004). Mechanism and stereochemistry of enzymatic cyclization of 24,30-Bisnor-2,3-oxidosqualene by recombinant beta-amyrin synthase. *Journal of the American Chemical Society*, *126*(22), 6880–6881.

Agricultural production statistics 2000–2022. (2023). FAO. https://doi.org/10.4060/cc9205en

Akagi, A., Fukushima, S., Okada, K., Jiang, C.-J., Yoshida, R., Nakayama, A., Shimono, M., Sugano, S., Yamane, H., & Takatsuji, H. (2014). WRKY45-dependent priming of diterpenoid phytoalexin biosynthesis in rice and the role of cytokinin in triggering the reaction. *Plant Molecular Biology*, *86*(1–2), 171–183.

Akatsuka, T., Kodama, O., Kato, H., Kono, Y., & Takeuchi, S. (1983). Short Communication 3-Hydroxy-7-oxo-sandaracopimaradiene (Oryzalexin A), a New Phytoalexin Isolated from Rice Blast Leaves. *Agricultural and Biological Chemistry*, *47*(2), 445–447.

Akbudak, M. A., Yildiz, S., & Filiz, E. (2020). Pathogenesis related protein-1 (PR-1) genes in tomato (Solanum lycopersicum L.): Bioinformatics analyses and expression profiles in response to drought stress. *Genomics*, *112*(6), 4089–4099.

Armah, C. N., Mackie, A. R., Roy, C., Price, K., Osbourn, A. E., Bowyer, P., & Ladha, S. (1999). The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. *Biophysical Journal*, *76*(1 Pt 1), 281–290.

Baerson, S. R., Rodriguez, D. J., Biest, N. A., Tran, M., You, J., Kreuger, R. W., Dill, G. M., Pratley, J. E., & Gruys, K. J. (2002). Investigating the mechanism of glyphosate resistance in rigid ryegrass (Lolium ridigum). *Weed Science*, *50*(6), 721–730.

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research*, *43*(W1), W39-49.

Bakera, B., Makowska, B., Groszyk, J., Niziołek, M., Orczyk, W., Bolibok-Brągoszewska, H., Hromada-Judycka, A., & Rakoczy-Trojanowska, M. (2015). Structural characteristics of ScBx genes controlling the biosynthesis of hydroxamic acids in rye (Secale cereale L.). *Journal of Applied Genetics*, *56*(3), 287–298.

Bathe, U., & Tissier, A. (2019). Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry*, *161*, 149–162.

Becker, E.-M., Herrfurth, C., Irmisch, S., Köllner, T. G., Feussner, I., Karlovsky, P., & Splivallo, R. (2014). Infection of corn ears by Fusarium spp. induces the emission of volatile sesquiterpenes. *Journal of Agricultural and Food Chemistry*, *62*(22), 5226–5236.

Beckie, H. J., & Tardif, F. J. (2012). Herbicide cross resistance in weeds. *Crop Protection* , *35*, 15–28.

Białas, A., Langner, T., Harant, A., Contreras, M. P., Stevenson, C. E., Lawson, D. M., Sklenar, J., Kellner, R., Moscou, M. J., Terauchi, R., Banfield, M. J., & Kamoun, S. (2021). Two NLR immune receptors acquired high-affinity binding to a fungal effector through convergent evolution of their integrated domain. *ELife*, *10*. https://doi.org/10.7554/eLife.66961

Bleecker, A. B., Estelle, M. A., Somerville, C., & Kende, H. (1988). Insensitivity to Ethylene Conferred by a Dominant Mutation in *Arabidopsis thaliana*. *Science*, *241*(4869), 1086–1089.

Block, A. K., Tang, H. V., Hopkins, D., Mendoza, J., Solemslie, R. K., du Toit, L. J., &

Christensen, S. A. (2021). A maize leucine-rich repeat receptor-like protein kinase mediates responses to fungal attack. *Planta*, *254*(4), 73.

Block, A. K., Vaughan, M. M., Schmelz, E. A., & Christensen, S. A. (2019). Biosynthesis and function of terpenoid defense compounds in maize (Zea mays). *Planta*, *249*(1), 21–30.

Borrelli, G. M., Mazzucotelli, E., Marone, D., Crosatti, C., Michelotti, V., Valè, G., & Mastrangelo, A. M. (2018). Regulation and evolution of NLR genes: A close interconnection for plant immunity. *International Journal of Molecular Sciences*, *19*(6). https://doi.org/10.3390/ijms19061662

Breen, S., Williams, S. J., Outram, M., Kobe, B., & Solomon, P. S. (2017). Emerging Insights into the Functions of Pathogenesis-Related Protein 1. *Trends in Plant Science*, *22*(10), 871–879.

Brotherton, J. E., Jeschke, M. R., Tranel, P. J., & Widholm, J. M. (2007). Identification of Arabidopsis thaliana variants with differential glyphosate responses. *Journal of Plant Physiology*, *164*(10), 1337–1345.

Brown, B. (2016). Biochemical characterization of new P450s in rice diterpenoid metabolism. *Honors Projects and Posters*. https://lib.dr.iastate.edu/honors_posters/201604/projects/4/

Brusamarello-Santos, L. C. C., Alberton, D., Valdameri, G., Camilios-Neto, D., Covre, R., Lopes, K. de P., Zibetti Tadra-Sfeir, M., Faoro, H., Adele Monteiro, R., Barbosa-Silva, A., John Broughton, W., Oliveira Pedrosa, F., Wassem, R., & Souza, E. M. de. (2019). Modulation of defence and iron homeostasis genes in rice roots by the diazotrophic endophyte Herbaspirillum seropedicae. *Scientific Reports*, *9*(1), 10573.

Burkhardt, H. J., Maizel, J. V., & Mitchell, H. K. (1964). AVENACIN, AN ANTIMICROBIAL SUBSTANCE ISOLATED FROM AVENA SATIVA. II. STRUCTURE. *Biochemistry*, *3*, 426–431.

Butte, A. J., & Kohane, I. S. (1999). MUTUAL INFORMATION RELEVANCE NETWORKS: FUNCTIONAL GENOMIC CLUSTERING USING PAIRWISE ENTROPY MEASUREMENTS. In *Biocomputing 2000* (pp. 418–429). WORLD SCIENTIFIC.

Cai, H., Xie, W., & Lian, X. (2013). Comparative Analysis of Differentially Expressed Genes in Rice Under Nitrogen and Phosphorus Starvation Stress Conditions. *Plant Molecular Biology Reporter / ISPMB*, *31*(1), 160–173.

Capilla-Perez, L., Solier, V., Portemer, V., Chambon, A., Hurel, A., Guillebaux, A., Vezon, D., Cromer, L., Grelon, M., & Mercier, R. (2018). The HEM Lines: A New Library of Homozygous Arabidopsis thaliana EMS Mutants and its Potential to Detect Meiotic Phenotypes. *Frontiers in Plant Science*, *9*, 1339.

Cartwright, D. W., Langcake, P., Pryce, R. J., Leworthy, D. P., & Ride, J. P. (1981). Isolation and characterization of two phytoalexins from rice as momilactones A and B. *Phytochemistry*, *20*(3), 535–537.

Chen, X., Chen, H., Yuan, J. S., Köllner, T. G., Chen, Y., Guo, Y., Zhuang, X., Chen, X., Zhang, Y.-J., Fu, J., Nebenführ, A., Guo, Z., & Chen, F. (2018). The rice terpene synthase gene OsTPS19 functions as an (S)-limonene synthase in planta, and its overexpression leads to enhanced resistance to the blast fungus Magnaporthe oryzae. *Plant Biotechnology Journal*, *16*(10), 1778–1787.

Cheng, A.-X., Xiang, C.-Y., Li, J.-X., Yang, C.-Q., Hu, W.-L., Wang, L.-J., Lou, Y.-G., & Chen, X.-Y. (2007). The rice (E)-beta-caryophyllene synthase (OsTPS3) accounts for the major inducible volatile sesquiterpenes. *Phytochemistry*, *68*(12), 1632–1641.

Cheng, H., Liu, H., Deng, Y., Xiao, J., Li, X., & Wang, S. (2015). The WRKY45-2 WRKY13

WRKY42 transcriptional regulatory cascade is required for rice resistance to fungal pathogen. *Plant Physiology*, *167*(3), 1087–1099.

Cheng, Y., Zhou, Y., Yang, Y., Chi, Y.-J., Zhou, J., Chen, J.-Y., Wang, F., Fan, B., Shi, K., Zhou, Y.-H., Yu, J.-Q., & Chen, Z. (2012). Structural and functional analysis of VQ motif-containing proteins in Arabidopsis as interacting proteins of WRKY transcription factors. *Plant Physiology*, *159*(2), 810–825.

Chi, W.-C., Chen, Y.-A., Hsiung, Y.-C., Fu, S.-F., Chou, C.-H., Trinh, N. N., Chen, Y.-C., & Huang, H.-J. (2013). Autotoxicity mechanism of Oryza sativa: transcriptome response in rice roots exposed to ferulic acid. *BMC Genomics*, *14*, 351.

Cho, J.-G., Cha, B.-J., Min Lee, S., Shrestha, S., Jeong, R.-H., Sung Lee, D., Kim, Y.-C., Lee, D.-G., Kang, H.-C., Kim, J., & Baek, N.-I. (2015). Diterpenes from the roots of Oryza sativa L. and their inhibition activity on NO production in LPS-stimulated RAW264.7 macrophages. *Chemistry & Biodiversity*, *12*(9), 1356–1364.

Cho, M., & Cho, H.-T. (2013). The function of ABCB transporters in auxin transport. *Plant Signaling & Behavior*, *8*(2), e22990.

Choi, C., Hwang, S.-H., Fang, I. R., Kwon, S. I., Park, S. R., Ahn, I., Kim, J. B., & Hwang, D.-J. (2015). Molecular characterization of Oryza sativa WRKY6, which binds to W-box-like element 1 of the Oryza sativa pathogenesis-related (PR) 10a promoter and confers reduced susceptibility to pathogens. *The New Phytologist*, *208*(3), 846–859.

Christensen, S. A., Huffaker, A., Sims, J., Hunter, C. T., Block, A., Vaughan, M. M., Willett, D., Romero, M., Mylroie, J. E., Williams, W. P., & Schmelz, E. A. (2018). Fungal and herbivore elicitation of the novel maize sesquiterpenoid, zealexin A4, is attenuated by elevated CO2. *Planta*, *247*(4), 863–873.

Christianson, D. W. (2017). Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews*, *117*(17), 11570–11648.

Chuang, M.-R., Chen, L.-J., & Li, H.-M. (2021). Chloroplast import of an intermembrane space protein is facilitated by translocon components Toc75 and Tic236. *Plant Direct*, *5*(11), e356.

Cornes, D. (2005). *Callisto: a very successful maize herbicide inspired by allelochemistry*. 569–572.

Curci, P. L., Zhang, J., Mähler, N., Seyfferth, C., Mannapperuma, C., Diels, T., Van Hautegem, T., Jonsen, D., Street, N., Hvidsten, T. R., Hertzberg, M., Nilsson, O., Inzé, D., Nelissen, H., & Vandepoele, K. (2022). Identification of growth regulators using cross-species network analysis in plants. *Plant Physiology*, *190*(4), 2350–2365.

De la Concepcion, J. C., Vega Benjumea, J., Bialas, A., Terauchi, R., Kamoun, S., & Banfield, M. J. (2021). Functional diversification gave rise to allelic specialization in a rice NLR immune receptor pair. *ELife*, *10*. https://doi.org/10.7554/eLife.71662

De La Peña, R., & Sattely, E. S. (2020). Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nature Chemical Biology*. https://doi.org/10.1038/s41589-020-00669-3

Degenhardt, J. (2009). Indirect defense responses to herbivory in grasses. *Plant Physiology*, *149*(1), 96–102.

Délye, C., Duhoux, A., Pernin, F., Riggins, C. W., & Tranel, P. J. (2015). Molecular Mechanisms of Herbicide Resistance. *Weed Science*, *63*(SP1), 91–115.

Deng, F., & Hatzios, K. K. (2002). Characterization of cytochrome P450-mediated bensulfuron-methyl O-demethylation in rice. *Pesticide Biochemistry and Physiology*, *74*(2), 102–115.

Deng, X., An, B., Zhong, H., Yang, J., Kong, W., & Li, Y. (2019). A Novel Insight into Functional Divergence of the MST Gene Family in Rice Based on Comprehensive Expression Patterns. *Genes*, *10*(3). https://doi.org/10.3390/genes10030239

Ding, L., Xu, X., Kong, W., Xia, X., Zhang, S., Liu, L.-W., Liu, A., & Zou, L. (2020). Genome-wide identification and expression analysis of rice NLR genes responsive to the infections of Xanthomonas oryzae pv. oryzae and Magnaporthe oryzae. *Physiological and Molecular Plant Pathology*, *111*(101488), 101488.

Ding, Y., Murphy, K. M., Poretsky, E., Mafu, S., Yang, B., Char, S. N., Christensen, S. A., Saldivar, E., Wu, M., Wang, Q., Ji, L., Schmitz, R. J., Kremling, K. A., Buckler, E. S., Shen, Z., Briggs, S. P., Bohlmann, J., Sher, A., Castro-Falcon, G., … Schmelz, E. A. (2019). Multiple genes recruited from hormone pathways partition maize diterpenoid defences. *Nature Plants*, *5*(10), 1043–1056.

Ding, Y., Weckwerth, P. R., Poretsky, E., Murphy, K. M., Sims, J., Saldivar, E., Christensen, S. A., Char, S. N., Yang, B., Tong, A.-D., Shen, Z., Kremling, K. A., Buckler, E. S., Kono, T., Nelson, D. R., Bohlmann, J., Bakker, M. G., Vaughan, M. M., Khalil, A. S., … Huffaker, A. (2020). Genetic elucidation of interconnected antibiotic pathways mediating maize innate immunity. *Nature Plants*, *6*(11), 1375–1388.

Du, Z., Su, Q., Wu, Z., Huang, Z., Bao, J., Li, J., Tu, H., Zeng, C., Fu, J., & He, H. (2021). Genome-wide characterization of MATE gene family and expression profiles in response to abiotic stresses in rice (Oryza sativa). *BMC Ecology and Evolution*, *21*(1), 141.

El-kereamy, A., El-sharkawy, I., Ramamoorthy, R., Taheri, A., Errampalli, D., Kumar, P., & Jayasankar, S. (2011). Prunus domestica pathogenesis-related protein-5 activates the defense response pathway and enhances the resistance to fungal infection. *PloS One*, *6*(3), e17973.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238.

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278.

Fan, Y., Ma, L., Pan, X., Tian, P., Wang, W., Liu, K., Xiong, Z., Li, C., Wang, Z., Wang, J., Zhang, H., & Bao, Y. (2024). Genome-Wide Association Study Identifies Rice Panicle Blast-Resistant Gene Pb4 Encoding a Wall-Associated Kinase. *International Journal of Molecular Sciences*, *25*(2). https://doi.org/10.3390/ijms25020830

Fan, Z., Wang, Y., Yang, C., Zhou, Z., Wang, P., & Yan, X. (2022). Identification of a novel multifunctional oxidosqualene cyclase from Zea mays sheds light on the biosynthetic pathway of three pentacyclic triterpenoids. *Synthetic and Systems Biotechnology*, *7*(4), 1167–1172.

Feng, P. C. C., Tran, M., Chiu, T., Douglas Sammons, R., Heck, G. R., & CaJacob, C. A. (2004). Investigations into glyphosate-resistant horseweed (Conyza canadensis): retention, uptake, translocation, and metabolism. *Weed Science*, *52*(4), 498–505.

Förster, C., Handrick, V., Ding, Y., Nakamura, Y., Paetz, C., Schneider, B., Castro-Falcón, G., Hughes, C. C., Luck, K., Poosapati, S., Kunert, G., Huffaker, A., Gershenzon, J., Schmelz, E. A., & Köllner, T. G. (2022). Biosynthesis and antifungal activity of fungus-induced O-methylated flavonoids in maize. *Plant Physiology*, *188*(1), 167–190.

Fritz, J. I., & Braun, R. (2006). Ecotoxicological effects of benzoxazinone allelochemicals and their metabolites on aquatic nontarget organisms. *Journal of Agricultural and*

Food Chemistry, *54*(4), 1105–1110.

Fu, J., Liu, Q., Wang, C., Liang, J., Liu, L., & Wang, Q. (2018). ZmWRKY79 positively regulates maize phytoalexin biosynthetic gene expression and is involved in stress response. *Journal of Experimental Botany*, *69*(3), 497–510.

Fu, X., Shi, P., He, Q., Shen, Q., Tang, Y., Pan, Q., Ma, Y., Yan, T., Chen, M., Hao, X., Liu, P., Li, L., Wang, Y., Sun, X., & Tang, K. (2017). AaPDR3, a PDR Transporter 3, Is Involved in Sesquiterpene β-Caryophyllene Transport in Artemisia annua. *Frontiers in Plant Science*, *8*, 723.

Gaines, T. A., Patterson, E. L., & Neve, P. (2019). Molecular mechanisms of adaptive evolution revealed by global selection for glyphosate resistance. *The New Phytologist*, *223*(4), 1770–1775.

Ganther, M., Vetterlein, D., Heintz-Buschart, A., & Tarkka, M. T. (2022). Transcriptome sequencing analysis of maize roots reveals the effects of substrate and root hair formation in a spatial context. *Plant and Soil*, *478*(1), 211–228.

Garcia, G. P., Sutour, S., Rabehaja, D., Tissandié, L., Filippi, J.-J., & Tomi, F. (2019). Essential oil of the malagasy grass Elionurus tristis Hack. contains several undescribed sesquiterpenoids. *Phytochemistry*, *162*, 29–38.

Glazebrook, J., Rogers, E. E., & Ausubel, F. M. (1997). Use of Arabidopsis for genetic dissection of plant defense responses. *Annual Review of Genetics*, *31*, 547–569.

Gu, M., Zhang, J., Li, H., Meng, D., Li, R., Dai, X., Wang, S., Liu, W., Qu, H., & Xu, G. (2017). Maintenance of phosphate homeostasis and root development are coordinately regulated by MYB1, an R2R3-type MYB transcription factor in rice. *Journal of Experimental Botany*, *68*(13), 3603–3615.

Gupta, B. B., Selter, L. L., Baranwal, V. K., Arora, D., Mishra, S. K., Sirohi, P., Poonia, A. K., Chaudhary, R., Kumar, R., Krattinger, S. G., & Chauhan, H. (2019). Updated inventory, evolutionary and expression analyses of G (PDR) type ABC transporter genes of rice. *Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale*, *142*, 429–439.

Guzmán, P., & Ecker, J. R. (1990). Exploiting the triple response of Arabidopsis to identify ethylene-related mutants. *The Plant Cell*, *2*(6), 513–523.

Hao, W., Collier, S. M., Moffett, P., & Chai, J. (2013). Structural basis for the interaction between the potato virus X resistance protein (Rx) and its cofactor Ran GTPase-activating protein 2 (RanGAP2). *The Journal of Biological Chemistry*, *288*(50), 35868–35876.

Harris, L. J., Saparno, A., Johnston, A., Prisic, S., Xu, M., Allard, S., Kathiresan, A., Ouellet, T., & Peters, R. J. (2005). The maize An2 gene is induced by Fusarium attack and encodes an ent-copalyl diphosphate synthase. *Plant Molecular Biology*, *59*(6), 881–894.

Hasegawa, M., Mitsuhara, I., Seo, S., Imai, T., Koga, J., Okada, K., Yamane, H., & Ohashi, Y. (2010). Phytoalexin accumulation in the interaction between rice and the blast fungus. *Molecular Plant-Microbe Interactions: MPMI*, *23*(8), 1000–1011.

Hashimoto, M., Kisseleva, L., Sawa, S., Furukawa, T., Komatsu, S., & Koshiba, T. (2004). A novel rice PR10 protein, RSOsPR10, specifically induced in roots by biotic and abiotic stresses, possibly via the jasmonic acid signaling pathway. *Plant & Cell Physiology*, *45*(5), 550–559.

Hayford, R. K., Haley, O. C., Cannon, E. K., Portwood, J. L., II, Gardiner, J. M., Andorf, C. M., & Woodhouse, M. R. (2023). *Functional Annotation and Meta-Analysis of Maize Transcriptomes Reveal Genes Involved in Biotic and Abiotic Stress*.

https://doi.org/10.21203/rs.3.rs-3684976/v1

He, Yongjun, Wang, Z., Ge, H., Liu, Y., & Chen, H. (2021). Weighted gene co-expression network analysis identifies genes related to anthocyanin biosynthesis and functional verification of hub gene SmWRKY44. *Plant Science: An International Journal of Experimental Plant Biology*, *309*, 110935.

He, Yunxia, Xu, J., Wang, X., He, X., Wang, Y., Zhou, J., Zhang, S., & Meng, X. (2019). The Arabidopsis Pleiotropic Drug Resistance Transporters PEN3 and PDR12 Mediate Camalexin Secretion for Resistance to Botrytis cinerea. *The Plant Cell*, *31*(9), 2206–2222.

Hedden, P. (2020). The Current Status of Research on Gibberellin Biosynthesis. *Plant & Cell Physiology*, *61*(11), 1832–1849.

Hoang, T. V., Vo, K. T. X., Rahman, M. M., Choi, S.-H., & Jeon, J.-S. (2019). Heat stress transcription factor OsSPL7 plays a critical role in reactive oxygen species balance and stress responses in rice. *Plant Science: An International Journal of Experimental Plant Biology*, *289*, 110273.

Horie, K., Inoue, Y., Sakai, M., Yao, Q., Tanimoto, Y., Koga, J., Toshima, H., & Hasegawa, M. (2015). Identification of UV-Induced Diterpenes Including a New Diterpene Phytoalexin, Phytocassane F, from Rice Leaves by Complementary GC/MS and LC/MS Approaches. *Journal of Agricultural and Food Chemistry*, *63*(16), 4050–4059.

Hu, L., Wu, Z., Robert, C. A. M., Ouyang, X., Züst, T., Mestrot, A., Xu, J., & Erb, M. (2021). Soil chemistry determines whether defensive plant secondary metabolites promote or suppress herbivore growth. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(43). https://doi.org/10.1073/pnas.2109602118

Huffaker, A., Kaplan, F., Vaughan, M. M., Dafoe, N. J., Ni, X., Rocca, J. R., Alborn, H. T., Teal, P. E. A., & Schmelz, E. A. (2011). Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. *Plant Physiology*, *156*(4), 2082–2097.

Ilic, K., SanMiguel, P. J., & Bennetzen, J. L. (2003). A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 12265–12270.

Imai, T., Ohashi, Y., Mitsuhara, I., Seo, S., Toshima, H., & Hasegawa, M. (2012). Identification of a degradation intermediate of the momilactone A rice phytoalexin by the rice blast fungus. *Bioscience, Biotechnology, and Biochemistry*, *76*(2), 414–416.

Inagaki, Y.-S., Etherington, G., Geisler, K., Field, B., Dokarry, M., Ikeda, K., Mutsukado, Y., Dicks, J., & Osbourn, A. (2011). Investigation of the potential for triterpene synthesis in rice through genome mining and metabolic engineering. *The New Phytologist*, *191*(2), 432–448.

Inoue, Y., Sakai, M., Yao, Q., Tanimoto, Y., Toshima, H., & Hasegawa, M. (2013). Identification of a novel casbane-type diterpene phytoalexin, ent-10-oxodepressin, from rice leaves. *Bioscience, Biotechnology, and Biochemistry*, *77*(4), 760–765.

Islam, M. N., Rabby, M. G., Hossen, M. M., Bonny, M., & Hasan, M. M. (2024). Genome-wide identification following functional analysis of amino acid permease and cationic amino acid transporter gene families in maize and their role in drought stress. *South African Journal of Botany: Official Journal of the South African Association of Botanists = Suid-Afrikaanse Tydskrif Vir Plantkunde: Amptelike Tydskrif van Die Suid-Afrikaanse Genootskap van Plantkundiges*, *168*, 360–371.

Ito, R., Mori, K., Hashimoto, I., Nakano, C., Sato, T., & Hoshino, T. (2011). Triterpene

cyclases from Oryza sativa L.: cycloartenol, parkeol and achilleol B synthases. *Organic Letters*, *13*(10), 2678–2681.

Jain, D., & Khurana, J. P. (2018). Role of Pathogenesis-Related (PR) Proteins in Plant Defense Mechanism. In A. Singh & I. K. Singh (Eds.), *Molecular Aspects of Plant-Pathogen Interaction* (pp. 265–281). Springer Singapore.

Jander, G., Baerson, S. R., Hudak, J. A., Gonzalez, K. A., Gruys, K. J., & Last, R. L. (2003). Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. *Plant Physiology*, *131*(1), 139–146.

Jia, Q., & Chen, F. (2016). Catalytic Functions of the Isoprenyl Diphosphate Synthase Superfamily in Plants: A Growing Repertoire. *Molecular Plant*, *9*(2), 189–191.

Jiang, H., Zhang, M., Yu, F., Li, X., Jin, J., Zhou, Y., Wang, Q., Jing, T., Wan, X., Schwab, W., & Song, C. (2023). A geraniol synthase regulates plant defense via alternative splicing in tea plants. *Horticulture Research*, *10*(10), uhad184.

Kahlon, P. S., & Stam, R. (2021). Polymorphisms in plants to restrict losses to pathogens: From gene family expansions to complex network evolution. *Current Opinion in Plant Biology*, *62*, 102040.

Kariya, K., Mori, H., Ueno, M., Yoshikawa, T., Teraishi, M., Yabuta, Y., Ueno, K., & Ishihara, A. (2024). Identification and evolution of a diterpenoid phytoalexin oryzalactone biosynthetic gene in the genus Oryza. *The Plant Journal: For Cell and Molecular Biology*, *118*(2), 358–372.

Kato, H., Kodama, O., & Akatsuka, T. (1993). Oryzalexin E, A diterpene phytoalexin from UV-irradiated rice leaves. *Phytochemistry*, *33*(1), 79–81.

Kato, H., Kodama, O., & Akatsuka, T. (1994). Oryzalexin F, a diterpene phytoalexin from UV-irradiated rice leaves. *Phytochemistry*, *36*(2), 299–301.

Kato, T., Kabuto, C., Sasaki, N., Tsunagawa, M., Aizawa, H., Fujita, K., Kato, Y., Kitahara, Y., & Takahashi, N. (1973). Momilactones, growth inhibitors from rice, oryza sativa L. *Tetrahedron Letters*, *14*(39), 3861–3864.

Kato-Noguchi, H. (2004). Allelopathic substance in rice root exudates: rediscovery of momilactone B as an allelochemical. *Journal of Plant Physiology*, *161*(3), 271–276.

Kato-Noguchi, H. (2011). Convergent or parallel molecular evolution of momilactone A and B: potent allelochemicals, momilactones have been found only in rice and the moss Hypnum plumaeforme. *Journal of Plant Physiology*, *168*(13), 1511–1516.

Kato-Noguchi, H., Hasegawa, M., Ino, T., Ota, K., & Kujime, H. (2010). Contribution of momilactone A and B to rice allelopathy. *Journal of Plant Physiology*, *167*(10), 787–791.

Kato-Noguchi, H., & Ino, T. (2003). Rice seedlings release momilactone B into the environment. *Phytochemistry*, *63*(5), 551–554.

Kato-Noguchi, H., Ota, K., & Kujime, H. (2012). Absorption of momilactone A and B by Arabidopsis thaliana L. and the growth inhibitory effects. *Journal of Plant Physiology*, *169*(15), 1471–1476.

Kim, D. Y., Kwon, S. I., Choi, C., Lee, H., Ahn, I., Park, S. R., Bae, S. C., Lee, S. C., & Hwang, D. J. (2013). Expression analysis of rice VQ genes in response to biotic and abiotic stresses. *Gene*, *529*(2), 208–214.

Kim, S.-B., Van den Broeck, L., Karre, S., Choi, H., Christensen, S. A., Wang, G.-F., Jo, Y., Cho, W. K., & Balint-Kurti, P. (2021). Analysis of the transcriptomic, metabolomic, and gene regulatory responses to Puccinia sorghi in maize. *Molecular Plant Pathology*, *22*(4), 465–479.

Kim, Y., Schumaker, K. S., & Zhu, J.-K. (2006). EMS Mutagenesis of Arabidopsis. In J.

Salinas & J. J. Sanchez-Serrano (Eds.), *Arabidopsis Protocols* (pp. 101–103). Humana Press.

Kiryu, M., Hamanaka, M., Yoshitomi, K., Mochizuki, S., Akimitsu, K., & Gomi, K. (2018). Rice terpene synthase 18 (OsTPS18) encodes a sesquiterpene synthase that produces an antibacterial (E)-nerolidol against a bacterial pathogen of rice. *Journal of General Plant Pathology: JGPP*, *84*(3), 221–229.

Kitaoka, N., Wu, Y., Xu, M., & Peters, R. J. (2015). Optimization of recombinant expression enables discovery of novel cytochrome P450 activity in rice diterpenoid biosynthesis. *Applied Microbiology and Biotechnology*, *99*(18), 7549–7558.

Kitaoka, N., Wu, Y., Zi, J., & Peters, R. J. (2016). Investigating inducible short-chain alcohol dehydrogenases/reductases clarifies rice oryzalexin biosynthesis. *The Plant Journal: For Cell and Molecular Biology*, *88*(2), 271–279.

Kiyama, H., Matsunaga, A., Suzuki, G., & Gomi, K. (2021). Monoterpene geraniol produced by rice terpene synthase 21 suppresses the expression of cell-division related genes in the rice bacterial pathogen, Xanthomonas oryzae pv. oryzae. *Physiological and Molecular Plant Pathology*, *115*, 101673.

Kodama, O., Xin Li, W., Tamogami, S., & Akatsuka, T. (1992). Oryzalexin S, a Novel Stemarane-type Diterpene Rice Phytoalexin. *Bioscience, Biotechnology, and Biochemistry*, *56*(6), 1002–1003.

Kodama, Osamu, Suzuki, T., Miyakawa, J., & Akatsuka, T. (1988). Ultraviolet-Induced Accumulation of Phytoalexins in Rice Leaves. *Agricultural and Biological Chemistry*, *52*(10), 2469–2473.

Koga, J., Ogawa, N., Yamauchi, T., Kikuchi, M., Ogasawara, N., & Shimura, M. (1997). Functional moiety for the antifungal activity of phytocassane E, a diterpene phytoalexin from rice. *Phytochemistry*, *44*(2), 249–253.

Koga, J., Shimura, M., Oshima, K., Ogawa, N., Yamauchi, T., & Ogasawara, N. (1995). Phytocassanes A, B, C and D, novel diterpene phytoalexins from rice, Oryza sativa L. *Tetrahedron*, *51*(29), 7907–7918.

Köllner, T. G., Held, M., Lenk, C., Hiltpold, I., Turlings, T. C. J., Gershenzon, J., & Degenhardt, J. (2008). A maize (E)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *The Plant Cell*, *20*(2), 482–494.

Köllner, T. G., Lenk, C., Schnee, C., Köpke, S., Lindemann, P., Gershenzon, J., & Degenhardt, J. (2013). Localization of sesquiterpene formation and emission in maize leaves after herbivore damage. *BMC Plant Biology*, *13*, 15.

Köllner, T. G., Lenk, C., Zhao, N., Seidl-Adams, I., Gershenzon, J., Chen, F., & Degenhardt, J. (2010). Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using s-adenosyl-L-methionine. *Plant Physiology*, *153*(4), 1795–1807.

Köllner, T. G., Schnee, C., Gershenzon, J., & Degenhardt, J. (2004a). The Variability of Sesquiterpenes Emitted from Two Zea mays Cultivars Is Controlled by Allelic Variation of Two Terpene Synthase Genes Encoding Stereoselective Multiple Product Enzymes. *The Plant Cell*, *16*(5), 1115–1131.

Köllner, T. G., Schnee, C., Gershenzon, J., & Degenhardt, J. (2004b). The sesquiterpene hydrocarbons of maize (Zea mays) form five groups with distinct developmental and organ-specific distributions. *Phytochemistry*, *65*(13), 1895–1902.

Komatsu, M., Chujo, A., Nagato, Y., Shimamoto, K., & Kyozuka, J. (2003). FRIZZY PANICLE is required to prevent the formation of axillary meristems and to establish floral

meristem identity in rice spikelets. *Development* , *130*(16), 3841–3850.

Kono, Y., Takeuchi, S., Kodama, O., & Akatsuka, T. (1984). Absolute Configuration of Oryzalexin A and Structures of Its Related Phytoalexins Isolated from Rice Blast Leaves Infected with *Pyricularia oryzae*. *Agricultural and Biological Chemistry*, *48*(1), 253–255.

Kono, Y., Uzawa, J., Kobayashi, K., Suzuki, Y., Uramoto, M., Sakurai, A., Watanabe, M., Teraoka, T., Hosokawa, D., Watanabe, M., & Kondo, H. (1991). Structures of Oryzalides A and B, and Oryzalic Acid A, a Group of Novel Antimicrobial Diterpenes, Isolated from Healthy Leaves of a Bacterial Leaf Blight-resistant Cultivar of Rice Plant. *Agricultural and Biological Chemistry*, *55*(3), 803–811.

Koo, D.-H., Molin, W. T., Saski, C. A., Jiang, J., Putta, K., Jugulam, M., Friebe, B., & Gill, B. S. (2018). Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed Amaranthus palmeri. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), 3332–3337.

Koo, D.-H., Sathishraj, R., Nakka, S., Ju, Y., Nandula, V. K., Jugulam, M., Friebe, B., & Gill, B. S. (2023). Extrachromosomal circular DNA-mediated spread of herbicide resistance in interspecific hybrids of pigweed. *Plant Physiology*, *193*(1), 229–233.

Koornneef, M., Reuling, G., & Karssen, C. M. (1984). The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. *Physiologia Plantarum*, *61*(3), 377–383.

Koornneef, M., van Eden, J., Hanhart, C. J., Stam, P., Braaksma, F. J., & Feenstra, W. J. (1983). Linkage map of Arabidopsis thaliana. *The Journal of Heredity*, *74*(4), 265–272.

Koornneeff, M., Dellaert, L. W. M., & van der Veen, J. H. (1982). EMS- and relation-induced mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *93*(1), 109–123.

Krochko, J. E., Abrams, G. D., Loewen, M. K., Abrams, S. R., & Cutler, A. J. (1998). (+)-Abscisic acid 8'-hydroxylase is a cytochrome P450 monooxygenase. *Plant Physiology*, *118*(3), 849–860.

Lambarey, H., Moola, N., Veenstra, A., Murray, S., & Suhail Rafudeen, M. (2020). Transcriptomic Analysis of a Susceptible African Maize Line to Fusarium verticillioides Infection. *Plants*, *9*(9). https://doi.org/10.3390/plants9091112

Lan, A., Huang, J., Zhao, W., Peng, Y., Chen, Z., & Kang, D. (2013). A salicylic acid-induced rice (Oryza sativa L.) transcription factor OsWRKY77 is involved in disease resistance of Arabidopsis thaliana. *Plant Biology* , *15*(3), 452–461.

Lanier, E. R., Andersen, T. B., & Hamberger, B. (2023). Plant terpene specialized metabolism: complex networks or simple linear pathways? *The Plant Journal: For Cell and Molecular Biology*, *114*(5), 1178–1201.

Lashley, A., Miller, R., Provenzano, S., Jarecki, S.-A., Erba, P., & Salim, V. (2022). Functional Diversification and Structural Origins of Plant Natural Product Methyltransferases. *Molecules* , *28*(1). https://doi.org/10.3390/molecules28010043

Lee, G. W., Chung, M.-S., Kang, M., Chung, B. Y., & Lee, S. (2016). Direct suppression of a rice bacterial blight (Xanthomonas oryzae pv. oryzae) by monoterpene (S)-limonene. *Protoplasma*, *253*(3), 683–690.

Lee, G. W., Chung, M.-S., Lee, S. S., Chung, B. Y., & Lee, S. (2019). Transcriptome-guided identification and functional characterization of key terpene synthases involved in constitutive and methyl jasmonate-inducible volatile terpene formation in Eremochloa ophiuroides (Munro) Hack. *Plant Physiology and Biochemistry: PPB / Societe*

Francaise de Physiologie Vegetale, *141*, 193–201.

Lee, G. W., Lee, S., Chung, M.-S., Jeong, Y. S., & Chung, B. Y. (2015). Rice terpene synthase 20 (OsTPS20) plays an important role in producing terpene volatiles in response to abiotic stresses. *Protoplasma*, *252*(4), 997–1007.

Lemieux, B., Miquel, M., Somerville, C., & Browse, J. (1990). Mutants of Arabidopsis with alterations in seed lipid fatty acid composition. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *80*(2), 234–240.

Lenk, C., Köllner, T. G., Erb, M., & Degenhardt, J. (2012). Two enzymes responsible for the formation of herbivore-induced volatiles of maize, the methyltransferase AAMT1 and the terpene synthase TPS23, are regulated by a similar signal transduction pathway. *Entomologia Experimentalis et Applicata*, *144*(1), 86–92.

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

Li, C., Wood, J. C., Vu, A. H., Hamilton, J. P., Rodriguez Lopez, C. E., Payne, R. M. E., Serna Guerrero, D. A., Gase, K., Yamamoto, K., Vaillancourt, B., Caputi, L., O'Connor, S. E., & Robin Buell, C. (2023). Single-cell multi-omics in the medicinal plant Catharanthus roseus. *Nature Chemical Biology*, *19*(8), 1031–1041.

Li, C.-H., Wang, G., Zhao, J.-L., Zhang, L.-Q., Ai, L.-F., Han, Y.-F., Sun, D.-Y., Zhang, S.-W., & Sun, Y. (2014). The Receptor-Like Kinase SIT1 Mediates Salt Sensitivity by Activating MAPK3/6 and Regulating Ethylene Homeostasis in Rice. *The Plant Cell*, *26*(6), 2538–2553.

Li, G., Xu, B., Zhang, Y., Xu, Y., Khan, N. U., Xie, J., Sun, X., Guo, H., Wu, Z., Wang, X., Zhang, H., Li, J., Xu, J., Wang, W., Zhang, Z., & Li, Z. (2022). RGN1 controls grain number and shapes panicle architecture in rice. *Plant Biotechnology Journal*, *20*(1), 158–167.

Li, J., & Chory, J. (1997). A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell*, *90*(5), 929–938.

Li, Jia, Wen, J., Lease, K. A., Doke, J. T., Tax, F. E., & Walker, J. C. (2002). BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell*, *110*(2), 213–222.

Li, Q., Duncan, S., Li, Y., Huang, S., & Luo, M. (2024). Decoding plant specialized metabolism: new mechanistic insights. *Trends in Plant Science*, *29*(5), 535–545.

Li, S., Wang, H., Li, F., Chen, Z., Li, X., Zhu, L., Wang, G., Yu, J., Huang, D., & Lang, Z. (2015). The maize transcription factor EREB58 mediates the jasmonate-induced production of sesquiterpene volatiles. *The Plant Journal: For Cell and Molecular Biology*, *84*(2), 296–308.

Li, W., Wang, L., Zhou, F., Li, C., Ma, W., Chen, H., Wang, G., Pickett, J. A., Zhou, J.-J., & Lin, Y. (2021). Overexpression of the homoterpene synthase gene, OsCYP92C21, increases emissions of volatiles mediating tritrophic interactions in rice. *Plant, Cell & Environment*, *44*(3), 948–963.

Li, Yan, Leveau, A., Zhao, Q., Feng, Q., Lu, H., Miao, J., Xue, Z., Martin, A. C., Wegel, E., Wang, J., Orme, A., Rey, M.-D., Karafiátová, M., Vrána, J., Steuernagel, B., Joynson, R., Owen, C., Reed, J., Louveau, T., … Osbourn, A. (2021). Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nature Communications*, *12*(1), 2563.

Li, Yanlin, Wang, J., Li, L., Song, W., Li, M., Hua, X., Wang, Y., Yuan, J., & Xue, Z. (2023).

Natural products of pentacyclic triterpenoids: from discovery to heterologous biosynthesis. *Natural Product Reports*, *40*(8), 1303–1353.

Liang, J., Liu, J., Brown, R., Jia, M., Zhou, K., Peters, R. J., & Wang, Q. (2018). Direct production of dihydroxylated sesquiterpenoids by a maize terpene synthase. *The Plant Journal: For Cell and Molecular Biology*, *94*(5), 847–856.

Liang, J., Shen, Q., Wang, L., Liu, J., Fu, J., Zhao, L., Xu, M., Peters, R. J., & Wang, Q. (2021). Rice contains a biosynthetic gene cluster associated with production of the casbane-type diterpenoid phytoalexin ent-10-oxodepressin. *The New Phytologist*, *231*(1), 85–93.

Liang, S., Xu, S., Qu, D., Yang, L., Wang, J., Liu, H., Xin, W., Zou, D., & Zheng, H. (2022). Identification and Functional Analysis of the Caffeic Acid O-Methyltransferase (COMT) Gene Family in Rice (Oryza sativa L.). *International Journal of Molecular Sciences*, *23*(15). https://doi.org/10.3390/ijms23158491

Lin, C., Shen, B., Xu, Z., Köllner, T. G., Degenhardt, J., & Dooner, H. K. (2008). Characterization of the monoterpene synthase gene tps26, the ortholog of a gene induced by insect herbivory in maize. *Plant Physiology*, *146*(3), 940–951.

Liu, H., Wu, H., Wang, Y., Wang, H., Chen, S., & Yin, Z. (2021). Comparative transcriptome profiling and co-expression network analysis uncover the key genes associated withearly-stage resistance to Aspergillus flavus in maize. *BMC Plant Biology*, *21*(1), 216.

Liu, J., Chen, X., Liang, X., Zhou, X., Yang, F., Liu, J., He, S. Y., & Guo, Z. (2016). Alternative Splicing of Rice WRKY62 and WRKY76 Transcription Factor Genes in Pathogen Defense. *Plant Physiology*, *171*(2), 1427–1442.

Liu, M., Ma, Y., Du, Q., Hou, X., Wang, M., & Lu, S. (2019). Functional Analysis of Polyprenyl Diphosphate Synthase Genes Involved in Plastoquinone and Ubiquinone Biosynthesis in Salvia miltiorrhiza. *Frontiers in Plant Science*, *10*, 893.

Liu, M.-H., Kang, H., Xu, Y., Peng, Y., Wang, D., Gao, L., Wang, X., Ning, Y., Wu, J., Liu, W., Li, C., Liu, B., & Wang, G.-L. (2020). Genome-wide association study identifies an NLR gene that confers partial resistance to Magnaporthe oryzae in rice. *Plant Biotechnology Journal*, *18*(6), 1376–1383.

Liu, Y., Mahdi, L. K., Porzel, A., Stark, P., Scherr-Henning, A., Bathe, U., Acosta, I. F., Zuccaro, A., Balcke, G. U., & Tissier, A. (2023). Metabolic arms race between a plant and a fungal pathogen. In *bioRxiv* (p. 2023.07.29.551000). https://doi.org/10.1101/2023.07.29.551000

Liu, Yaming, Balcke, G. U., Porzel, A., Mahdi, L., Scherr-Henning, A., Bathe, U., Zuccaro, A., & Tissier, A. (2021). A barley gene cluster for the biosynthesis of diterpenoid phytoalexins. In *bioRxiv* (p. 2021.05.21.445084). https://doi.org/10.1101/2021.05.21.445084

Liu, Yaming, Esposto, D., Mahdi, L. K., Porzel, A., Stark, P., Hussain, H., Scherr-Henning, A., Isfort, S., Bathe, U., Acosta, I. F., Zuccaro, A., Balcke, G. U., & Tissier, A. (2024). Hordedane diterpenoid phytoalexins restrict Fusarium graminearum infection but enhance Bipolaris sorokiniana colonization of barley roots. *Molecular Plant*. https://doi.org/10.1016/j.molp.2024.07.006

Liu, Youming, Tomiyama, S., Motegi, I., Yamamoto, N., Zheng, A., Mori, M., Kawahara, M., Tsujii, Y., Miyamoto, K., Furumi, H., Sato, Y., Nojiri, H., & Okada, K. (2024). Evolutionary conserved cis-trans regulation machinery for diterpenoid phytoalexin production in Poaceae. In *bioRxiv* (p. 2024.05.05.592300). https://doi.org/10.1101/2024.05.05.592300

Lofgren, L. A., LeBlanc, N. R., Certano, A. K., Nachtigall, J., LaBine, K. M., Riddle, J., Broz, K., Dong, Y., Bethan, B., Kafer, C. W., & Kistler, H. C. (2018). Fusarium graminearum: pathogen or endophyte of North American grasses? *The New Phytologist*, *217*(3), 1203–1212.

Louveau, T., Orme, A., Pfalzgraf, H., Stephenson, M. J., Melton, R., Saalbach, G., Hemmings, A. M., Leveau, A., Rejzek, M., Langdon, T., Field, R. A., & Osbourn, A. (2018). Analysis of Two New Arabinosyltransferases Belonging to the Carbohydrate-Active Enzyme (CAZY) Glycosyl Transferase Family1 Provides Insights into Disease Resistance and Sugar Donor Specificity. *The Plant Cell*, *30*(12), 3038–3057.

Lu, J., Yan, S., & Xue, Z. (2024). Biosynthesis and functions of triterpenoids in cereals. *Journal of Advertising Research*. https://doi.org/10.1016/j.jare.2024.05.021

Luck, K., Chen, X., Norris, A. M., Chen, F., Gershenzon, J., & Köllner, T. G. (2020). The reconstruction and biochemical characterization of ancestral genes furnish insights into the evolution of terpene synthase function in the Poaceae. *Plant Molecular Biology*, *104*(1–2), 203–215.

Ma, S., Ding, Z., & Li, P. (2017). Maize network analysis revealed gene modules involved in development, nutrients utilization, metabolism, and stress response. *BMC Plant Biology*, *17*(1), 131.

Ma, Y.-X., Zhou, Z.-J., Cao, H.-Z., Zhou, F., Si, H.-L., Zang, J.-P., Xing, J.-H., Zhang, K., & Dong, J.-G. (2023). Identification and expression analysis of sugar transporter family genes reveal the role of ZmSTP2 and ZmSTP20 in maize disease resistance. *Journal of Integrative Agriculture*, *22*(11), 3458–3473.

Maconochie, M., Nonchev, S., Morrison, A., & Krumlauf, R. (1996). Paralogous Hox genes: function and regulation. *Annual Review of Genetics*, *30*(1), 529–556.

Mafu, S., Ding, Y., Murphy, K. M., Yaacoobi, O., Addison, J. B., Wang, Q., Shen, Z., Briggs, S. P., Bohlmann, J., Castro-Falcon, G., Hughes, C. C., Betsiashvili, M., Huffaker, A., Schmelz, E. A., & Zerbe, P. (2018). Discovery, Biosynthesis and Stress-Related Accumulation of Dolabradiene-Derived Defenses in Maize. *Plant Physiology*, *176*(4), 2677–2690.

Maher, E. P., & Martindale, S. J. (1980). Mutants of Arabidopsis thaliana with altered responses to auxins and gravity. *Biochemical Genetics*, *18*(11–12), 1041–1053.

Maizel, J. V., Burkhardt, H. J., & Mitchell, H. K. (1964). AVENACIN, AN ANTIMICROBIAL SUBSTANCE ISOLATED FROM AVENA SATIVA. I. ISOLATION AND ANTIMICROBIAL ACTIVITY. *Biochemistry*, *3*, 424–426.

MaizeMine: Gene Zm00001eb020060 Z. mays. (n.d.). Retrieved 20 June 2024, from https://maizemine.rnet.missouri.edu/maizemine/report.do?id=216113065

Makowska, B., Bakera, B., & Rakoczy-Trojanowska, M. (2015). The genetic background of benzoxazinoid biosynthesis in cereals. *Acta Physiologiae Plantarum / Polish Academy of Sciences, Committee of Plant Physiology Genetics and Breeding*, *37*(9), 176.

Mao, H., Shen, Q., & Wang, Q. (2017). CYP701A26 is characterized as an ent-kaurene oxidase with putative involvement in maize gibberellin biosynthesis. *Biotechnology Letters*, *39*(11), 1709–1716.

Mao, L., Kawaide, H., Higuchi, T., Chen, M., Miyamoto, K., Hirata, Y., Kimura, H., Miyazaki, S., Teruya, M., Fujiwara, K., Tomita, K., Yamane, H., Hayashi, K.-I., Nojiri, H., Jia, L., Qiu, J., Ye, C., Timko, M. P., Fan, L., & Okada, K. (2020). Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants.

Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.1914373117

Matschi, S., Vasquez, M. F., Bourgault, R., Steinbach, P., Chamness, J., Kaczmar, N., Gore, M. A., Molina, I., & Smith, L. G. (2020). Structure-function analysis of the maize bulliform cell cuticle and its potential role in dehydration and leaf rolling. *Plant Direct*, *4*(10), e00282.

McGee, J. D., Hamer, J. E., & Hodges, T. K. (2001). Characterization of a PR-10 pathogenesis-related gene family induced in rice during infection with Magnaporthe grisea. *Molecular Plant-Microbe Interactions: MPMI*, *14*(7), 877–886.

McKelvie, A. D. (1961). A list of mutant genes in Arabidopsis thaliana (L.) Heynh. *Radiation Botany*, *1*, 233–241.

Melero-Jiménez, I. J., Bañares-España, E., Reul, A., Flores-Moya, A., & García-Sánchez, M. J. (2021). Detection of the maximum resistance to the herbicides diuron and glyphosate, and evaluation of its phenotypic cost, in freshwater phytoplankton. *Aquatic Toxicology (Amsterdam, Netherlands)*, *240*(105973), 105973.

Miyamoto, K., Nishizawa, Y., Minami, E., Nojiri, H., Yamane, H., & Okada, K. (2015). Overexpression of the bZIP transcription factor OsbZIP79 suppresses the production of diterpenoid phytoalexin in rice cells. *Journal of Plant Physiology*, *173*, 19–27.

Moons, A. (2003). Ospdr9, which encodes a PDR-type ABC transporter, is induced by heavy metals, hypoxic stress and redox perturbations in rice roots. *FEBS Letters*, *553*(3), 370–376.

Moons, A. (2008). Transcriptional profiling of the PDR gene family in rice roots in response to plant growth regulators, redox perturbations and weak organic acid stresses. *Planta*, *229*(1), 53–71.

Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E., & Swarup, V. (2023). hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Reports Methods*, *3*(6), 100498.

Morii, M., Sugihara, A., Takehara, S., Kanno, Y., Kawai, K., Hobo, T., Hattori, M., Yoshimura, H., Seo, M., & Ueguchi-Tanaka, M. (2020). The Dual Function of OsSWEET3a as a Gibberellin and Glucose Transporter Is Important for Young Shoot Development in Rice. *Plant & Cell Physiology*, *61*(11), 1935–1945.

Morris, J. S., Caldo, K. M. P., Liang, S., & Facchini, P. J. (2021). PR10/Bet v1-like Proteins as Novel Contributors to Plant Biochemical Diversity. *Chembiochem: A European Journal of Chemical Biology*, *22*(2), 264–287.

Muchlinski, A., Chen, X., Lovell, J. T., Köllner, T. G., Pelot, K. A., Zerbe, P., Ruggiero, M., Callaway, L., 3rd, Laliberte, S., Chen, F., & Tholl, D. (2019). Biosynthesis and Emission of Stress-Induced Volatile Terpenes in Roots and Leaves of Switchgrass (Panicum virgatum L.). *Frontiers in Plant Science*, *10*, 1144.

Murphy, K. M., Ma, L.-T., Ding, Y., Schmelz, E. A., & Zerbe, P. (2018). Functional Characterization of Two Class II Diterpene Synthases Indicates Additional Specialized Diterpenoid Pathways in Maize (Zea mays). *Frontiers in Plant Science*, *9*, 1542.

Murphy, K. M., & Zerbe, P. (2020). Specialized diterpenoid metabolism in monocot crops: Biosynthesis and chemical diversity. *Phytochemistry*, *172*, 112289.

Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., Hill, L., Qi, X., Bakht, S., Melton, R., & Osbourn, A. (2008). Sad3 and sad4 are required for saponin biosynthesis and root development in oat. *The Plant Cell*, *20*(1), 201–212.

Naithani, S., Dikeman, D., Garg, P., Al-Bader, N., & Jaiswal, P. (2021). Beyond gene

ontology (GO): using biocuration approach to improve the gene nomenclature and functional annotation of rice S-domain kinase subfamily. *PeerJ*, *9*, e11052.

Neal, A. L., Ahmad, S., Gordon-Weeks, R., & Ton, J. (2012). Benzoxazinoids in root exudates of maize attract Pseudomonas putida to the rhizosphere. *PloS One*, *7*(4), e35498.

Neal, A. L., & Ton, J. (2013). Systemic defense priming by Pseudomonas putida KT2440 in maize depends on benzoxazinoid exudation from the roots. *Plant Signaling & Behavior*, *8*(1), e22655.

Nelson, D., & Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *The Plant Journal: For Cell and Molecular Biology*, *66*(1), 194–211.

Nemoto, T., Cho, E.-M., Okada, A., Okada, K., Otomo, K., Kanno, Y., Toyomasu, T., Mitsuhashi, W., Sassa, T., Minami, E., Shibuya, N., Nishiyama, M., Nojiri, H., & Yamane, H. (2004). Stemar-13-ene synthase, a diterpene cyclase involved in the biosynthesis of the phytoalexin oryzalexin S in rice. *FEBS Letters*, *571*(1–3), 182–186.

Nie, S., Wang, B., Ding, H., Lin, H., Zhang, L., Li, Q., Wang, Y., Zhang, B., Liang, A., Zheng, Q., Wang, H., Lv, H., Zhu, K., Jia, M., Wang, X., Du, J., Zhao, R., Jiang, Z., Xia, C., … Zhang, Z. (2021). Genome assembly of the Chinese maize elite inbred line RP125 and its EMS mutant collection provide new resources for maize genetics research and crop improvement. *The Plant Journal: For Cell and Molecular Biology*, *108*(1), 40–54.

Niehaus, T. D., Kinison, S., Okada, S., Yeo, Y.-S., Bell, S. A., Cui, P., Devarenne, T. P., & Chappell, J. (2012). Functional identification of triterpene methyltransferases from Botryococcus braunii race B. *The Journal of Biological Chemistry*, *287*(11), 8163–8173.

Nomura, T., Ishihara, A., Imaishi, H., Endo, T., Ohkawa, H., & Iwamura, H. (2002). Molecular characterization and chromosomal localization of cytochrome P450 genes involved in the biosynthesis of cyclic hydroxamic acids in hexaploid wheat. *Molecular Genetics and Genomics: MGG*, *267*(2), 210–217.

Nomura, Taiji, Ishihara, A., Imaishi, H., Ohkawa, H., Endo, T. R., & Iwamura, H. (2003). Rearrangement of the genes for the biosynthesis of benzoxazinones in the evolution of Triticeae species. *Planta*, *217*(5), 776–782.

Ødum, M. T., Teufel, F., Thumuluri, V., Almagro Armenteros, J. J., Johansen, A. R., Winther, O., & Nielsen, H. (2024). DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. *Nucleic Acids Research*, *52*(W1), W215–W220.

Ogo, Y., Kakei, Y., Itai, R. N., Kobayashi, T., Nakanishi, H., Takahashi, H., Nakazono, M., & Nishizawa, N. K. (2014). Spatial transcriptomes of iron-deficient and cadmium-stressed rice. *The New Phytologist*, *201*(3), 781–794.

Okada, A., Okada, K., Miyamoto, K., Koga, J., Shibuya, N., Nojiri, H., & Yamane, H. (2009). OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *The Journal of Biological Chemistry*, *284*(39), 26510–26518.

Okada, K., Kawaide, H., Miyamoto, K., Miyazaki, S., Kainuma, R., Kimura, H., Fujiwara, K., Natsume, M., Nojiri, H., Nakajima, M., Yamane, H., Hatano, Y., Nozaki, H., & Hayashi, K.-I. (2016). HpDTC1, a Stress-Inducible Bifunctional Diterpene Cyclase Involved in Momilactone Biosynthesis, Functions in Chemical Defence in the Moss Hypnum plumaeforme. *Scientific Reports*, *6*, 25316.

Olofsdotter, M., Navarez, D., & Rebulanan, M. (1999). 1983501. Weed-suppressing rice

cultivars--does allelopathy play a role? *Weed*. https://pubag.nal.usda.gov/?page=19836&per_page=100&search_field=all_fields

Olofsdotter, Maria, Navarez, D., & Moody, K. (1995). Allelopathic potential in rice (Oryza sativa L.) germplasm. *The Annals of Applied Biology*, *127*(3), 543–560.

Olofsdotter, Maria, Rebulanan, M., Madrid, A., Dali, W., Navarez, D., & Olk, D. C. (2002). Why phenolic acids are unlikely primary allelochemicals in rice. *Journal of Chemical Ecology*, *28*(1), 229–242.

Olofsdotter, Olofsdotter, Navarez, Rebulanan, & Streibig. (1999). Weed-suppressing rice cultivars - does allelopathy play a role? In *Weed Research* (Vol. 39, Issue 6, pp. 441–454). https://doi.org/10.1046/j.1365-3180.1999.00159.x

Orme, A., Louveau, T., Stephenson, M. J., Appelhagen, I., Melton, R., Cheema, J., Li, Y., Zhao, Q., Zhang, L., Fan, D., Tian, Q., Vickerstaff, R. J., Langdon, T., Han, B., & Osbourn, A. (2019). A noncanonical vacuolar sugar transferase required for biosynthesis of antimicrobial defense compounds in oat. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(52), 27105–27114.

Ortiz, D., de Guillen, K., Cesari, S., Chalvon, V., Gracy, J., Padilla, A., & Kroj, T. (2017). Recognition of the Magnaporthe oryzae effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5. *The Plant Cell*, *29*(1), 156–168.

Otomo, K., Kanno, Y., Motegi, A., Kenmoku, H., Yamane, H., Mitsuhashi, W., Oikawa, H., Toshima, H., Itoh, H., Matsuoka, M., Sassa, T., & Toyomasu, T. (2004). Diterpene cyclases responsible for the biosynthesis of phytoalexins, momilactones A, B, and oryzalexins A-F in rice. *Bioscience, Biotechnology, and Biochemistry*, *68*(9), 2001–2006.

Owatworakit, A., Townsend, B., Louveau, T., Jenner, H., Rejzek, M., Hughes, R. K., Saalbach, G., Qi, X., Bakht, S., Roy, A. D., Mugford, S. T., Goss, R. J. M., Field, R. A., & Osbourn, A. (2013). Glycosyltransferases from oat (Avena) implicated in the acylation of avenacins. *The Journal of Biological Chemistry*, *288*(6), 3696–3704.

Page, D. R., & Grossniklaus, U. (2002). The art and design of genetic screens: Arabidopsis thaliana. *Nature Reviews. Genetics*, *3*(2), 124–136.

Pan, L., Yu, Q., Wang, J., Han, H., Mao, L., Nyporko, A., Maguza, A., Fan, L., Bai, L., & Powles, S. (2021). An ABCC-type transporter endowing glyphosate resistance in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(16). https://doi.org/10.1073/pnas.2100136118

Pan, T., Jin, H., Zhou, C., & Yan, M. (2024). Rice Serine Hydroxymethyltransferases: Evolution, Subcellular Localization, Function and Perspectives. *Plants*, *13*(8). https://doi.org/10.3390/plants13081116

Pang, K., Li, Y., Liu, M., Meng, Z., & Yu, Y. (2013). Inventory and general analysis of the ATP-binding cassette (ABC) gene superfamily in maize (Zea mays L.). *Gene*, *526*(2), 411–428.

Passricha, N., Saifi, S., Ansari, M. W., & Tuteja, N. (2017). Prediction and validation of cis-regulatory elements in 5' upstream regulatory regions of lectin receptor-like kinase gene family in rice. *Protoplasma*, *254*(2), 669–684.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire,

C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418–D427.

Pelot, K. A., Chen, R., Hagelthorn, D. M., Young, C. A., Addison, J. B., Muchlinski, A., Tholl, D., & Zerbe, P. (2018). Functional Diversity of Diterpene Synthases in the Biofuel Crop Switchgrass. *Plant Physiology*, *178*(1), 54–71.

Perochon, A., Benbow, H. R., Ślęczka-Brady, K., Malla, K. B., & Doohan, F. M. (2021). Analysis of the chromosomal clustering of Fusarium-responsive wheat genes uncovers new players in the defence against head blight disease. *Scientific Reports*, *11*(1), 7446.

Peters, R. J. (2006). Uncovering the complex metabolic network underlying diterpenoid phytoalexin biosynthesis in rice and other cereal crop plants. *Phytochemistry*, *67*(21), 2307–2317.

Phillips, D. R., Rasbery, J. M., Bartel, B., & Matsuda, S. P. (2006). Biosynthetic diversity in plant triterpene cyclization. *Current Opinion in Plant Biology*, *9*(3), 305–314.

Pierman, B., Toussaint, F., Bertin, A., Lévy, D., Smargiasso, N., De Pauw, E., & Boutry, M. (2017). Activity of the purified plant ABC transporter NtPDR1 is stimulated by diterpenes and sesquiterpenes involved in constitutive and induced defenses. *The Journal of Biological Chemistry*, *292*(47), 19491–19502.

Polturak, G., Dippe, M., Stephenson, M. J., Chandra Misra, R., Owen, C., Ramirez-Gonzalez, R. H., Haidoulis, J. F., Schoonbeek, H.-J., Chartrain, L., Borrill, P., Nelson, D. R., Brown, J. K. M., Nicholson, P., Uauy, C., & Osbourn, A. (2022). Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(16), e2123299119.

Powles, S. B., & Yu, Q. (2010). Evolution in action: plants resistant to herbicides. *Annual Review of Plant Biology*, *61*, 317–347.

Qi, J., Li, J., Han, X., Li, R., Wu, J., Yu, H., Hu, L., Xiao, Y., Lu, J., & Lou, Y. (2016). Jasmonic acid carboxyl methyltransferase regulates development and herbivory-induced defense response in rice. *Journal of Integrative Plant Biology*, *58*(6), 564–576.

Qin, Y.-J., Wu, W.-H., & Wang, Y. (2019). ZmHAK5 and ZmHAK1 function in K+ uptake and distribution in maize under low K+ conditions. *Journal of Integrative Plant Biology*, *61*(6), 691–705.

Quan, N. T., & Xuan, T. D. (2018). Foliar application of vanillic and p-hydroxybenzoic acids enhanced drought tolerance and formation of phytoalexin momilactones in rice. *Archives of Agronomy and Soil Science*, *64*(13), 1831–1846.

Rakoczy-Trojanowska, M., Orczyk, W., Krajewski, P., Bocianowski, J., Stochmal, A., & Kowalczyk, M. (2017). ScBx gene based association analysis of hydroxamate content in rye (Secale cereale L.). *Journal of Applied Genetics*, *58*(1), 1–9.

Rasmann, S., Köllner, T. G., Degenhardt, J., Hiltpold, I., Toepfer, S., Kuhlmann, U., Gershenzon, J., & Turlings, T. C. J. (2005). Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature*, *434*(7034), 732–737.

Ren, F., Mao, H., Liang, J., Liu, J., Shu, K., & Wang, Q. (2016). Functional characterization of ZmTPS7 reveals a maize τ-cadinol synthase involved in stress response. *Planta*, *244*(5), 1065–1074.

Righini, S., Rodriguez, E. J., Berosich, C., Grotewold, E., Casati, P., & Falcone Ferreyra, M. L. (2019). Apigenin produced by maize flavone synthase I and II protects plants against UV-B-induced damage. *Plant, Cell & Environment*, *42*(2), 495–508.

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-response analysis using R. *PloS One*, *10*(12), e0146021.

Rivera, S. B., Swedlund, B. D., King, G. J., Bell, R. N., Hussey, C. E., Jr, Shattuck-Eidens, D. M., Wrobel, W. M., Peiser, G. D., & Poulter, C. D. (2001). Chrysanthemyl diphosphate synthase: isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from Chrysanthemum cinerariaefolium. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4373–4378.

Röbbelen, G. (1957). Untersuchungen An Strahleninduzierten Blattfarbmutanten Von Arabidopsis Thaliana (L) Heynh. *Zeitschrift Fur Induktive Abstammungs- Und Vererbungslehre*, *88*(2), 189–252.

Rudolf, J. D., & Chang, C.-Y. (2020). Terpene synthases in disguise: enzymology, structure, and opportunities of non-canonical terpene synthases. *Natural Product Reports*, *37*(3), 425–463.

Saldivar, E. V., Ding, Y., Poretsky, E., Bird, S., Block, A. K., Huffaker, A., & Schmelz, E. A. (2023). Maize Terpene Synthase 8 (ZmTPS8) Contributes to a Complex Blend of Fungal-Elicited Antibiotics. *Plants*, *12*(5). https://doi.org/10.3390/plants12051111

Schluttenhofer, C., & Yuan, L. (2015). Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiology*, *167*(2), 295–306.

Schmelz, E. A., Kaplan, F., Huffaker, A., Dafoe, N. J., Vaughan, M. M., Ni, X., Rocca, J. R., Alborn, H. T., & Teal, P. E. (2011). Identity, regulation, and activity of inducible diterpenoid phytoalexins in maize. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(13), 5455–5460.

Schnee, C., Köllner, T. G., Gershenzon, J., & Degenhardt, J. (2002). The maize gene terpene synthase 1 encodes a sesquiterpene synthase catalyzing the formation of (E)-beta-farnesene, (E)-nerolidol, and (E,E)-farnesol after herbivore damage. *Plant Physiology*, *130*(4), 2049–2060.

Schneeberger, K. (2014). Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature Reviews. Genetics*, *15*(10), 662–676.

Schreiber, K. J., Austin, R. S., Gong, Y., Zhang, J., Fung, P., Wang, P. W., Guttman, D. S., & Desveaux, D. (2012). Forward chemical genetic screens in Arabidopsis identify genes that influence sensitivity to the phytotoxic compound sulfamethoxazole. *BMC Plant Biology*, *12*, 226.

Sekido, H., & Akatsuka, T. (1987). Mode of Action of Oryzalexin D against Pyricularia oryzae. *Agricultural and Biological Chemistry*, *51*(7), 1967–1971.

Serra Serra, N., Shanmuganathan, R., & Becker, C. (2021). Allelopathy in rice: a story of momilactones, kin recognition, and weed management. *Journal of Experimental Botany*, *72*(11), 4022–4037.

Shen, B., Zheng, Z., & Dooner, H. K. (2000). A maize sesquiterpene cyclase gene induced by insect herbivory and volicitin: characterization of wild-type and mutant alleles. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14807–14812.

Shibuya, M., Adachi, S., & Ebizuka, Y. (2004). Cucurbitadienol synthase, the first committed enzyme for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol biosynthesis. *Tetrahedron*, *60*(33), 6995–7003.

Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.-W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N., Koga, J., Nojiri, H., & Yamane, H. (2007). Identification of a biosynthetic gene cluster in rice for momilactones. *The Journal of*

Biological Chemistry, *282*(47), 34013–34018.

Shitan, N., Nishitani, S., Inagaki, A., Nakahara, Y., Yamada, Y., & Koeduka, T. (2023). Gene expression analysis of the ATP-binding cassette transporter ABCD1 in petunia (Petunia hybrida) and tobacco (Nicotiana spp.). *Plant Gene*, *33*, 100391.

Singh, S. K., Shrivastava, S., Mishra, A. K., Kumar, D., Pandey, V. K., Srivastava, P., Pradhan, B., Behera, B. C., Bahuguna, A., & Baek, K.-H. (2023). Friedelin: Structure, Biosynthesis, Extraction, and Its Potential Health Impact. *Molecules* , *28*(23). https://doi.org/10.3390/molecules28237760

Soh, D., Dong, D., Guo, Y., & Wong, L. (2011). Finding consistent disease subnetworks across microarray datasets. *BMC Bioinformatics*, *12 Suppl 13*(Suppl 13), S15.

St Johnston, D. (2002). The art and design of genetic screens: Drosophila melanogaster. *Nature Reviews. Genetics*, *3*(3), 176–188.

Stein, J. C., Yu, Y., Copetti, D., Zwickl, D. J., Zhang, L., Zhang, C., Chougule, K., Gao, D., Iwata, A., Goicoechea, J. L., Wei, S., Wang, J., Liao, Y., Wang, M., Jacquemin, J., Becker, C., Kudrna, D., Zhang, J., Londono, C. E. M., … Wing, R. A. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. *Nature Genetics*, *50*(2), 285–296.

Stephens, C., Hammond-Kosack, K. E., & Kanyuka, K. (2022). WAKsing plant immunity, waning diseases. *Journal of Experimental Botany*, *73*(1), 22–37.

Sun, X., Zhang, L., Tang, Z., Shi, X., Ma, J., & Cui, R. (2019). Transcriptome analysis of roots from resistant and susceptible rice varieties infected with Hirschmanniella mucronata. *FEBS Open Bio*, *9*(11), 1968–1982.

Sun, Y., Huang, X., Ning, Y., Jing, W., Bruce, T. J. A., Qi, F., Xu, Q., Wu, K., Zhang, Y., & Guo, Y. (2017). TPS46, a Rice Terpene Synthase Conferring Natural Resistance to Bird Cherry-Oat Aphid, Rhopalosiphum padi (Linnaeus). *Frontiers in Plant Science*, *8*, 110.

Surhone, L. M., Tennoe, M. T., & Henssonow, S. F. (Eds.). (2010). *Uniprot*. Betascript Publishing.

Svyantek, A. W., Aldahir, P., Chen, S., Flessner, M. L., McCullough, P. E., Sidhu, S. S., & Scott McElroy, J. (2016). Target and Nontarget Resistance Mechanisms Induce Annual Bluegrass (Poa annua) Resistance to Atrazine, Amicarbazone, and Diuron. *Weed Technology: A Journal of the Weed Science Society of America*, *30*(3), 773–782.

Swaminathan, S., Morrone, D., Wang, Q., Fulton, D. B., & Peters, R. J. (2009). CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *The Plant Cell*, *21*(10), 3315–3325.

Takeuchi, K., Gyohda, A., Tominaga, M., Kawakatsu, M., Hatakeyama, A., Ishii, N., Shimaya, K., Nishimura, T., Riemann, M., Nick, P., Hashimoto, M., Komano, T., Endo, A., Okamoto, T., Jikumaru, Y., Kamiya, Y., Terakawa, T., & Koshiba, T. (2011). RSOsPR10 expression in response to environmental stresses is regulated antagonistically by jasmonate/ethylene and salicylic acid signaling pathways in rice roots. *Plant & Cell Physiology*, *52*(9), 1686–1696.

Tamizhselvan, P., Madhavan, S., Constan-Aguilar, C., Elrefaay, E. R., Liu, J., Pěnčík, A., Novák, O., Cairó, A., Hrtyan, M., Geisler, M., & Tognetti, V. B. (2023). Chloroplast Auxin Efflux Mediated by ABCB28 and ABCB29 Fine-Tunes Salt and Drought Stress Responses in Arabidopsis. *Plants*, *13*(1). https://doi.org/10.3390/plants13010007

Tang, D., Wang, G., & Zhou, J.-M. (2017). Receptor kinases in plant-pathogen interactions: More than pattern recognition. *The Plant Cell*, *29*(4), 618–637.

Tang, H., Krishnakumar, V., Zeng, X., Xu, Z., Taranto, A., Lomas, J. S., Zhang, Y., Huang, Y., Wang, Y., Yim, W. C., Zhang, J., & Zhang, X. (2024). JCVI: A versatile toolkit for comparative genomics analysis. *IMeta*, *3*(4), e211.

Tang, Y., Guo, J., Zhang, T., Bai, S., He, K., & Wang, Z. (2021). Genome-Wide Analysis of WRKY Gene Family and the Dynamic Responses of Key WRKY Genes Involved in Ostrinia furnacalis Attack in Zea mays. *International Journal of Molecular Sciences*, *22*(23). https://doi.org/10.3390/ijms222313045

Taniguchi, S., Hosokawa-Shinonaga, Y., Tamaoki, D., Yamada, S., Akimitsu, K., & Gomi, K. (2014). Jasmonate induction of the monoterpene linalool confers resistance to rice bacterial blight and its biosynthesis is regulated by JAZ protein in rice. *Plant, Cell & Environment*, *37*(2), 451–461.

Taniguchi, S., Miyoshi, S., Tamaoki, D., Yamada, S., Tanaka, K., Uji, Y., Tanaka, S., Akimitsu, K., & Gomi, K. (2014). Isolation of jasmonate-induced sesquiterpene synthase of rice: product of which has an antifungal activity against Magnaporthe oryzae. *Journal of Plant Physiology*, *171*(8), 625–632.

Taylor, M. R., Reinders, A., & Ward, J. M. (2015). Transport Function of Rice Amino Acid Permeases (AAPs). *Plant & Cell Physiology*, *56*(7), 1355–1363.

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., & Osbourn, A. (2014). Triterpene biosynthesis in plants. *Annual Review of Plant Biology*, *65*, 225–257.

Thomson, S. M., Pulido, P., & Jarvis, R. P. (2020). Protein import into chloroplasts and its regulation by the ubiquitin-proteasome system. *Biochemical Society Transactions*, *48*(1), 71–82.

Toyomasu, T., Kagahara, T., Okada, K., Koga, J., Hasegawa, M., Mitsuhashi, W., Sassa, T., & Yamane, H. (2008). Diterpene phytoalexins are biosynthesized in and exuded from the roots of rice seedlings. *Bioscience, Biotechnology, and Biochemistry*, *72*(2), 562–567.

Toyomasu, T., Usui, M., Sugawara, C., Otomo, K., Hirose, Y., Miyao, A., Hirochika, H., Okada, K., Shimizu, T., Koga, J., Hasegawa, M., Chuba, M., Kawana, Y., Kuroda, M., Minami, E., Mitsuhashi, W., & Yamane, H. (2014). Reverse-genetic approach to verify physiological roles of rice phytoalexins: characterization of a knockdown mutant of OsCPS4 phytoalexin biosynthetic gene in rice. *Physiologia Plantarum*, *150*(1), 55–62.

Tsunakawa, M., Ohba, A., Sasaki, N., Kabuto, C., Kato, T., Kitahara, Y., & Takahashi, N. (2006). MOMILACTONE-C, A MINOR CONSTITUENT OF GROWTH INHIBITORS IN RICE HUSK. *Chemistry Letters*, *5*(11), 1157–1158.

Turner, E. M. (1953). The Nature of the Resistance of Oats to the Take-all Fungus. *Journal of Experimental Botany*, *4*(3), 264–271.

Umemura, K., Ogawa, N., Shimura, M., Koga, J., Usami, H., & Kono, T. (2003). Possible role of phytocassane, rice phytoalexin, in disease resistance of rice against the blast fungus Magnaporthe grisea. *Bioscience, Biotechnology, and Biochemistry*, *67*(4), 899–902.

Upadhyay, N., Kar, D., Deepak Mahajan, B., Nanda, S., Rahiman, R., Panchakshari, N., Bhagavatula, L., & Datta, S. (2019). The multitasking abilities of MATE transporters in plants. *Journal of Experimental Botany*, *70*(18), 4643–4656.

van der Biezen, E. A., & Jones, J. D. G. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current Biology: CB*, *8*(7), R226–R228.

Van Quan, N., Xuan, T. D., Tran, H.-D., & Dieu Thuy, N. T. (2019). Inhibitory Activities of

Momilactones A, B, E, and 7-Ketostigmasterol Isolated from Rice Husk on Paddy and Invasive Weeds. *Plants*, *8*(6). https://doi.org/10.3390/plants8060159

Vaughan, M. M., Christensen, S., Schmelz, E. A., Huffaker, A., McAuslane, H. J., Alborn, H. T., Romero, M., Allen, L. H., & Teal, P. E. A. (2015). Accumulation of terpenoid phytoalexins in maize roots is associated with drought tolerance. *Plant, Cell & Environment*, *38*(11), 2195–2207.

Veenstra, A., Moola, N., Wighard, S., Korsman, J., Christensen, S. A., Rafudeen, M. S., & Murray, S. L. (2019). Kauralexins and zealexins accumulate in sub-tropical maize lines and play a role in seedling resistance to Fusarium verticillioides. *European Journal of Plant Pathology / European Foundation for Plant Pathology*, *153*(1), 223–237.

Venado, R. E., Owens, B. F., Ortiz, D., Lawson, T., Mateos-Hernandez, M., Ferruzzi, M. G., & Rocheford, T. R. (2017). Genetic analysis of provitamin A carotenoid β-cryptoxanthin concentration and relationship with other carotenoids in maize grain (Zea mays L.). *Molecular Breeding: New Strategies in Plant Improvement*, *37*(10), 127.

Venturelli, S., Belz, R. G., Kämper, A., Berger, A., von Horn, K., Wegner, A., Böcker, A., Zabulon, G., Langenecker, T., Kohlbacher, O., Barneche, F., Weigel, D., Lauer, U. M., Bitzer, M., & Becker, C. (2015). Plants Release Precursors of Histone Deacetylase Inhibitors to Suppress Growth of Competitors. *The Plant Cell*, *27*(11), 3175–3189.

Viana, V. E., Carlos da Maia, L., Busanello, C., Pegoraro, C., & Costa de Oliveira, A. (2021). When rice gets the chills: comparative transcriptome profiling at germination shows WRKY transcription factor responses. *Plant Biology* , *23 Suppl 1*, 100–112.

Wan, W., Wu, Y., Hu, D., Ye, F., Wu, X., Qi, X., Liang, H., Zhou, H., Xue, J., Xu, S., & Zhang, X. (2023). Genome-wide association analysis of kernel nutritional quality in two natural maize populations. *Molecular Breeding: New Strategies in Plant Improvement*, *43*(3), 18.

Wang, C.-T., Ru, J.-N., Liu, Y.-W., Li, M., Zhao, D., Yang, J.-F., Fu, J.-D., & Xu, Z.-S. (2018). Maize WRKY Transcription Factor ZmWRKY106 Confers Drought and Heat Tolerance in Transgenic Plants. *International Journal of Molecular Sciences*, *19*(10). https://doi.org/10.3390/ijms19103046

Wang, H., Li, Y., Guo, Z., Zhou, X., Zhao, Y., Han, Y., & Lin, X. (2024). Genome-wide identification of AAAP gene family and expression analysis in response to saline-alkali stress in foxtail millet (Setaria italica L.). *Scientific Reports*, *14*(1), 3106.

Wang, Jia, Song, L., Gong, X., Xu, J., & Li, M. (2020). Functions of Jasmonic Acid in Plant Regulation and Response to Abiotic Stress. *International Journal of Molecular Sciences*, *21*(4). https://doi.org/10.3390/ijms21041446

Wang, Jian, Hu, H., Jiang, X., Zhang, S., Yang, W., Dong, J., Yang, T., Ma, Y., Zhou, L., Chen, J., Nie, S., Liu, C., Ning, Y., Zhu, X., Liu, B., Yang, J., & Zhao, J. (2024). Pangenome-Wide Association Study and Transcriptome Analysis Reveal a Novel QTL and Candidate Genes Controlling both Panicle and Leaf Blast Resistance in Rice. *Rice* , *17*(1), 27.

Wang, Junzhu, Han, M., & Liu, Y. (2021). Diversity, structure and function of the coiled-coil domains of plant NLR immune receptors. *Journal of Integrative Plant Biology*, *63*(2), 283–296.

Wang, L., Fu, J., Shen, Q., & Wang, Q. (2023). OsWRKY10 extensively activates multiple rice diterpenoid phytoalexin biosynthesis genes to enhance rice blast resistance. *The Plant Journal: For Cell and Molecular Biology*, *115*(3), 758–771.

Wang, Q., Hillwig, M. L., Okada, K., Yamazaki, K., Wu, Y., Swaminathan, S., Yamane, H., & Peters, R. J. (2012). Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *The Journal of Biological Chemistry*, *287*(9), 6159–6168.

Wang, Q., Hillwig, M. L., & Peters, R. J. (2011). CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *The Plant Journal: For Cell and Molecular Biology*, *65*(1), 87–95.

Wang, Q., Hillwig, M. L., Wu, Y., & Peters, R. J. (2012-3). CYP701A8: A Rice ent-Kaurene Oxidase Paralog Diverted to More Specialized Diterpenoid Metabolism1[W][OA]. *Plant Physiology*, *158*(3), 1418–1425.

Wang, X.-Q., Kim, K.-W., Chu, S.-H., Phitaktansakul, R., Park, S.-W., Chung, I.-M., Lee, Y.-S., & Park, Y.-J. (2019). Genome-Wide Association Study for Squalene Contents and Functional Haplotype Analysis in Rice. *ACS Omega*, *4*(21), 19358–19365.

Wang, Yapeng, Li, T., Sun, Z., Huang, X., Yu, N., Tai, H., & Yang, Q. (2022). Comparative transcriptome meta-analysis reveals a set of genes involved in the responses to multiple pathogens in maize. *Frontiers in Plant Science*, *13*, 971371.

Wang, Yiqin, Teng, Z., Li, H., Wang, W., Xu, F., Sun, K., Chu, J., Qian, Y., Loake, G. J., Chu, C., & Tang, J. (2023). An activated form of NB-ARC protein RLS1 functions with cysteine-rich receptor-like protein RMC to trigger cell death in rice. *Plant Communications*, *4*(2), 100459.

Wang, Yiting, Zou, J., Li, J., Kong, F., Xu, L., Xu, D., Li, J., Yang, H., Zhang, L., Li, T., & Fan, H. (2023). Identification and functional analysis of ZmDLS associated with the response to biotic stress in maize. *Frontiers in Plant Science*, *14*, 1162826.

Wang, Yongqin, Xu, H., Wei, X., Chai, C., Xiao, Y., Zhang, Y., Chen, B., Xiao, G., Ouwerkerk, P. B. F., Wang, M., & Zhu, Z. (2007). Molecular cloning and expression analysis of a monosaccharide transporter gene OsMST4 from rice (Oryza sativa L.). *Plant Molecular Biology*, *65*(4), 439–451.

Wang, Yupeng, Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49.

Watanabe, Manabu, Kono, Y., Watanabe, M., Uzawa, J., Teraoka, T., Hosokawa, D., Suzuki, Y., Sakurai, A., & Teraguchi, M. (1992). Structures of Oryzalic Acid B and Three Related Compounds, a Group of Novel Antibacterial Diterpenes, Isolated from Leaves of a Bacterial Leaf Blight-Resistant Cultivar of Rice. *Bioscience, Biotechnology, and Biochemistry*, *56*(1), 113–117.

Watanabe, Minoru, Sakai, Y., Teraoka, T., Abe, H., Kono, Y., Uzawa, J., Kobayashi, K., Suzuki, Y., & Sakurai, A. (1990). Novel C19-Kaurane Type of Diterpene (Oryzalide A), a New Antimicrobial Compound Isolated from Healthy Leaves of a Bacterial Leaf Blight-resistant Cultivar of Rice Plant. *Agricultural and Biological Chemistry*, *54*(4), 1103–1105.

Wei, K., Wang, Y., Zhong, X., & Pan, S. (2014). Protein kinase structure, expression and regulation in maize drought signaling. *Molecular Breeding: New Strategies in Plant Improvement*, *34*(2), 583–602.

Wu, G., Zhang, Y., Wang, B., Li, K., Lou, Y., Zhao, Y., & Liu, F. (2021). Proteomic and Transcriptomic Analyses Provide Novel Insights into the Crucial Roles of Host-Induced Carbohydrate Metabolism Enzymes in Xanthomonas oryzae pv. oryzae Virulence and Rice-Xoo Interaction. *Rice* , *14*(1), 57.

Wu, Y., Hillwig, M. L., Wang, Q., & Peters, R. J. (2011). Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS Letters*, *585*(21), 3446–3451.

Wu, Y., Wang, Q., Hillwig, M. L., & Peters, R. J. (2013). Picking sides: distinct roles for CYP76M6 and CYP76M8 in rice oryzalexin biosynthesis. *Biochemical Journal*, *454*(2), 209–216.

Wyatt, G., Zerbe, P., & Tiedge, K. (2024). Characterization of switchgrass (Panicum virgatum L.) PvKSL1 as a levopimaradiene/abietadiene-type diterpene synthase. *Plant Biology* . https://doi.org/10.1111/plb.13708

Xiao, Y., Wang, Q., Erb, M., Turlings, T. C. J., Ge, L., Hu, L., Li, J., Han, X., Zhang, T., Lu, J., Zhang, G., & Lou, Y. (2012). Specific herbivore-induced volatiles defend plants and determine insect community composition in the field. *Ecology Letters*, *15*(10), 1130–1139.

Xie, Z., Zhang, Z.-L., Zou, X., Huang, J., Ruas, P., Thompson, D., & Shen, Q. J. (2005). Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiology*, *137*(1), 176–189.

Xu, M., Galhano, R., Wiemann, P., Bueno, E., Tiernan, M., Wu, W., Chung, I.-M., Gershenzon, J., Tudzynski, B., Sesma, A., & Peters, R. J. (2012). Genetic evidence for natural product-mediated plant-plant allelopathy in rice (Oryza sativa). *The New Phytologist*, *193*(3), 570–575.

Xuan, T. D., Minh, T. N., Anh, L. H., & Khanh, T. D. (2016). Allelopathic momilactones A and B are implied in rice drought and salinity tolerance, not weed resistance. *Agronomy for Sustainable Development*, *36*(3), 52.

Xue, Z., Duan, L., Liu, D., Guo, J., Ge, S., Dicks, J., ÓMáille, P., Osbourn, A., & Qi, X. (2012). Divergent evolution of oxidosqualene cyclases in plants. *The New Phytologist*, *193*(4), 1022–1038.

Xue, Z., Tan, Z., Huang, A., Zhou, Y., Sun, J., Wang, X., Thimmappa, R. B., Stephenson, M. J., Osbourn, A., & Qi, X. (2018). Identification of key amino acid residues determining product specificity of 2,3-oxidosqualene cyclase in Oryza species. *The New Phytologist*, *218*(3), 1076–1088.

Yactayo-Chang, J. P., Broadhead, G. T., Housler, R. J., Resende, M. F. R., Jr, Verma, K., Louis, J., Basset, G. J., Beck, J. J., & Block, A. K. (2024). Maize terpene synthase 1 impacts insect behavior via the production of monoterpene volatiles β-myrcene and linalool. *Phytochemistry*, *218*, 113957.

Yamada, K., Saijo, Y., Nakagami, H., & Takano, Y. (2016). Regulation of sugar transporter activity for antibacterial defense in *Arabidopsis*. *Science*, *354*(6318), 1427–1430.

Yamamura, C., Mizutani, E., Okada, K., Nakagawa, H., Fukushima, S., Tanaka, A., Maeda, S., Kamakura, T., Yamane, H., Takatsuji, H., & Mori, M. (2015). Diterpenoid phytoalexin factor, a bHLH transcription factor, plays a central role in the biosynthesis of diterpenoid phytoalexins in rice. *The Plant Journal: For Cell and Molecular Biology*, *84*(6), 1100–1113.

Yan, J., Su, P., Meng, X., & Liu, P. (2023). Phylogeny of the plant receptor-like kinase (RLK) gene family and expression analysis of wheat RLK genes in response to biotic and abiotic stresses. *BMC Genomics*, *24*(1), 224.

Yang, P., Fu, J., Liang, J., Shen, Q., Liu, L., Wang, L., & Wang, Q. (2020). Promoter Variation Results in Differential Phytoalexin Accumulation in Two Maize Inbred Lines. *Plant Molecular Biology Reporter / ISPMB*, *38*(1), 165–174.

Yang, T., Zhang, S., Hu, Y., Wu, F., Hu, Q., Chen, G., Cai, J., Wu, T., Moran, N., Yu, L., & Xu, G. (2014). The role of a potassium transporter OsHAK5 in potassium acquisition and transport from roots to shoots in rice at low potassium supply levels. *Plant Physiology*, *166*(2), 945–959.

Yimer, H. Z., Nahar, K., Kyndt, T., Haeck, A., Van Meulebroek, L., Vanhaecke, L., Demeestere, K., Höfte, M., & Gheysen, G. (2018). Gibberellin antagonizes jasmonate-induced defense against Meloidogyne graminicola in rice. *The New Phytologist*, *218*(2), 646–660.

Yoon, J., Baek, G., Pasriga, R., Tun, W., Min, C. W., Kim, S.-T., Cho, L.-H., & An, G. (2023). Homeobox transcription factors OsZHD1 and OsZHD2 induce inflorescence meristem activity at floral transition in rice. *Plant, Cell & Environment*, *46*(4), 1327–1339.

Yoshitomi, K., Taniguchi, S., Tanaka, K., Uji, Y., Akimitsu, K., & Gomi, K. (2016). Rice terpene synthase 24 (OsTPS24) encodes a jasmonate-responsive monoterpene synthase that produces an antibacterial γ-terpinene against rice pathogen. *Journal of Plant Physiology*, *191*, 120–126.

Yu, S., Ligang, C., Liping, Z., & Diqiu, Y. (2010). Overexpression of OsWRKY72 gene interferes in the abscisic acid signal and auxin transport pathway of Arabidopsis. *Journal of Biosciences*, *35*(3), 459–471.

Yuan, J. S., Köllner, T. G., Wiggins, G., Grant, J., Degenhardt, J., & Chen, F. (2008). Molecular and genomic basis of volatile-mediated indirect defense against insects in rice. *The Plant Journal: For Cell and Molecular Biology*, *55*(3), 491–503.

Zdrzałek, R., Kamoun, S., Terauchi, R., Saitoh, H., & Banfield, M. J. (2020). The rice NLR pair Pikp-1/Pikp-2 initiates cell death through receptor cooperation rather than negative regulation. *PloS One*, *15*(9), e0238616.

Zhan, C., Lei, L., Guo, H., Zhou, S., Xu, C., Liu, Z., Wu, Z., Deng, Y., Miao, Y., Han, Y., Zhang, M., Li, H., Huang, S., Yang, C., Zhang, F., Li, Y., Liu, L., Liu, X., Abbas, H. M. K., … Luo, J. (2023). Disease resistance conferred by components of essential chrysanthemum oil and the epigenetic regulation of OsTPS1. *Science China. Life Sciences*, *66*(5), 1108–1118.

Zhan, C., Lei, L., Liu, Z., Zhou, S., Yang, C., Zhu, X., Guo, H., Zhang, F., Peng, M., Zhang, M., Li, Y., Yang, Z., Sun, Y., Shi, Y., Li, K., Liu, L., Shen, S., Wang, X., Shao, J., … Luo, J. (2020). Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nature Plants*, *6*(12), 1447–1454.

Zhan, C., Lei, L., Liu, Z., Zhou, S., Yang, C., Zhu, X., Guo, H., Zhang, F., Peng, M., Zhang, M., Li, Y., Yang, Z., Sun, Y., Shi, Y., Li, K., Liu, L., Shen, S., Wang, X., Shao, J., … Luo, J. (2021). Author Correction: Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nature Plants*, *7*(1), 100.

Zhang, F., Zhang, J., Ma, Z., Xia, L., Wang, X., Zhang, L., Ding, Y., Qi, J., Mu, X., Zhao, F., Ji, T., & Tang, B. (2021). Bulk analysis by resequencing and RNA-seq identifies candidate genes for maintaining leaf water content under water deficit in maize. *Physiologia Plantarum*, *173*(4), 1935–1945.

Zhang, Ting, Tan, D., Zhang, L., Zhang, X., & Han, Z. (2017). Phylogenetic analysis and drought-responsive expression profiles of the WRKY transcription factor family in maize. *Agri Gene*, *3*, 99–108.

Zhang, Tingting, Lv, W., Zhang, H., Ma, L., Li, P., Ge, L., & Li, G. (2018). Genome-wide analysis of the basic Helix-Loop-Helix (bHLH) transcription factor family in maize. *BMC Plant Biology*, *18*(1), 235.

Zhao, M., Cheng, J., Guo, B., Duan, J., & Che, C.-T. (2018). Momilactone and Related Diterpenoids as Potential Agricultural Chemicals. *Journal of Agricultural and Food Chemistry*, *66*(30), 7859–7872.

Zhou, F., & Pichersky, E. (2020). More is better: the diversity of terpene metabolism in plants. *Current Opinion in Plant Biology*, *55*, 1–10.

Zhou, P., Li, Z., Magnusson, E., Gomez Cano, F., Crisp, P. A., Noshay, J. M., Grotewold, E., Hirsch, C. N., Briggs, S. P., & Springer, N. M. (2020). Meta Gene Regulatory Networks in Maize Highlight Functionally Relevant Regulatory Interactions. *The Plant Cell*, *32*(5), 1377–1396.

Zhou, T., Yue, C.-P., Huang, J.-Y., Cui, J.-Q., Liu, Y., Wang, W.-M., Tian, C., & Hua, Y.-P. (2020). Genome-wide identification of the amino acid permease genes and molecular characterization of their transcriptional responses to various nutrient stresses in allotetraploid rapeseed. *BMC Plant Biology*, *20*(1), 151.

Zhuang, X., Köllner, T. G., Zhao, N., Li, G., Jiang, Y., Zhu, L., Ma, J., Degenhardt, J., & Chen, F. (2012). Dynamic evolution of herbivore-induced sesquiterpene biosynthesis in sorghum and related grass crops. *The Plant Journal: For Cell and Molecular Biology*, *69*(1), 70–80.

Jichi Sekido, Toshiji Endo, Tatsuo Suga, Jikodama, Tonmi Akatsuka, Yoshiki Kono, & Setsuo Takeuchi. (1986). Oryzalexin D (3,7-dihydroxy-(+)-sandarakopimaradiene), blast disease morbidity A novel phytoalexin isolated from rice leaves. Journal of Pesticide, Japan Science), 11(3), 369–372.