Investigating chlorsulfuron toxicity and resistance in the liverwort *Marchantia polymorpha*



Alexandra Casey

St Cross College

Department of Plant Sciences

University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity 2021

Investigating chlorsulfuron toxicity and resistance in the liverwort *Marchantia polymorpha*

Alexandra Casey
St Cross College
Department of Plant Sciences
University of Oxford
Thesis submitted for the degree of Doctor of Philosophy
Trinity 2021

Abstract

Herbicide metabolic resistance is defined as the enhanced ability to metabolically detoxify herbicides. Enzymes that metabolise herbicides belong to the 71 and 72 clans of the cytochrome P450 (CYP) enzyme family and the Tau and Phi classes of the glutathione s-transferase (GST) enzyme family. I show that the 71 and 72 clans and Tau and Phi classes arose before the evolution of land plants and rapidly expanded in seed plants.

Using RNA-Seq and over-expression constructs I show that a 71 clan CYP up-regulated by the herbicide chlorsulfuron in the liverwort *Marchantia polymorpha* conferred a low level of resistance to chlorsulfuron when over-expressed. Chlorsulfuron treatment of sensitive plants caused the differential expression of hundreds of genes, while target site resistant (TSR) plants that possess a point mutation in the chlorsulfuron molecular target Mp*ALS* did not exhibit a transcriptional response to chlorsulfuron. This indicates that changes in steady state mRNA levels after chlorsulfuron treatment are caused by the chlorsulfuron-inhibition of ALS and not by the detection of the herbicide chemical in the plant. This suggests that the transcriptome response to chlorsulfuron represents a stress response to the disruption of the ALS enzyme, and is not a xenobiotic response. To test this, four of the most transcriptionally induced genes in response to chlorsulfuron were over-expressed in *M. polymorpha*. Only one gene conferred weak chlorsulfuron tolerance when over-expressed. This supports the hypothesis that plants do not possess a xenobiotic detoxifying suite of genes.

To determine how Mp*ALS* inhibition leads to plant death I generated loss of function (LOF) mutations in the Mp*ALS* gene. I show that the deletion of Mp*ALS* is lethal but the growth of Mp*als*^{lof} mutants can be partially restored on medium supplemented with branched chain amino acids. Using targeted and untargeted metabolomics I show that chlorsulfuron-treatment reduces the abundance of some branched chain amino acids and increases pools of other amino acids and metabolites. I propose the hypothesis that chlorsulfuron lethality is due to a lack of branched chain amino acids which leads to large changes in gene expression.

Table of contents

Acknowledgements 4

Abbreviations 5

Chapter 1 Introduction 7

Chapter 2 Evolution of cytochrome P450 enzymes in the Archaeplastida 19

2.1. Abstract 20

2.2. Introduction 21

2.3. Materials and Methods 23

2.4. Results 25

2.5. Discussion 38

2.6. Appendix 41

Chapter 3 Evolution of glutathione s-transferase enzymes in the Archaeplastida 48

3.1. Abstract 49

3.2. Introduction 50

3.3. Materials and Methods 52

3.4. Results 54

3.5. Discussion 67

3.6. Appendix 70

Chapter 4 Mp*CYP822A1* up-regulated by chlorsulfuron confers low-level chlorsulfuron resistance when over-expressed 77

4.1. Abstract 78

4.2. Introduction 79

4.3. Materials and Methods 81

4.4. Results 88

4.5. Discussion 108

4.6. Appendix 111

Chapter 5 Chlorsulfuron lethality is not due to the accumulation of quinate or 2-ketobutyrate 119

5.1. Abstract 120

5.2. Introduction 121

5.3. Materials and Methods 123

5.4. Results 127

5.5. Discussion 138

Chapter 6 General Discussion 142

Acknowledgements

First I would like to extend my thanks to my supervisor Prof. Liam Dolan for taking me in as a lost DPhil. student and entrusting me with this exciting and ambitious project. Thank you for your support, encouragement and advice throughout the years. Thanks to Prof. Nick Kruger, Prof. Phil Poole and Prof. Gail Preston for their valued feedback and support. A huge thank you to Roni for working extremely hard to allow researchers back in the lab as quickly as possible during the national lockdown. Thank you to the BBSRC and the Gregor Mendel Institute for funding my DPhil.

I am grateful to the whole Dolan group, past and present, for making my time in the Department so enjoyable. Thank you Dr. Clement Champion for your advice and helping me set up my first experiments. To Dr. Reka Toth, Dr. Anna Thamm, and Dr. Susanna Streubel, thank you for your input during lab presentations, meetings, journal clubs and more. They challenged me and greatly enriched my project. Dr. Sandy Heatherington, thanks for always having the time to help with my gene trees. Special thanks to Hugh Muvley for his help with cloning. To Dr Radka Slovak, thank you for your advice and guidance. To Chloe Casey, thank you for your eternally happy presence, for making the conference in Hawaii so fun, and for saving my experiments by supplying spores. Thanks to Sam Caygill, who did his undergraduate project with me, for extracting DNA and making dose response curves. Thanks to Pippa Sinclair who did her rotation project with me, for all your cloning and plating countless gemmae. Thanks to Dylan Cleveland who did a summer project with me, for propagating and genotyping the Mp*als*^{lof} lines. To Helen Prescott and Lida Chen, thank you for helping me in the lab, and making the enormous quantities of media that I regularly consumed. Thank you to Dr. Thomas Köcher, for doing the measurements and analysis of the metabolomics experiments. Thank you to Sam and Chloe for proofreading and providing comments for this thesis.

I would also like to thank my close friends and family. To Chloe, Sam, Pippa, Julius, Flo, Ellie, Brandon, Bea, Hayley and Tim, thank you for all the fun times. Finally, to my parents and Milly for being my number 1 fans, and to Nick, for your calming influence and making sure I'm fed, thank you.

Abbreviations

aLRT	approximate likelihood ratio test
Cr1	*Chlamydomonas reinhardtii* GST 1
Cr2	*Chlamydomonas reinhardtii* GST 2
CDNB	1-chloro-2,4-dinitrobenzene
CYP	Cytochrome P450
DHAR	dehydroascorbate reductase
EF1Bγ	eukaryotic translation elongation factor 1B-γ subunit
FOP	aryloxyphenoxypropionate
GHR	glutathionyl hydroquinone reductase
GR_{50}	herbicide dose at which the growth of the treated plant is reduced by 50% relative to the untreated plant
GSH	glutathione
GST	glutathione s-transferase
GSTB	Beta GST
GSTC8	C-8 GST
GSTF	Phi GST
GSTH	Hemerythrin GST
GSTI	Iota GST
GSTL	Lambda GST
GSTM	Metaxin GST
GSTMu	Mu GST
GSTP	Pi GST
GSTSL	Sigma-like GST

GSTT	Theta GST
GSTU	Tau GST
GSTZ	Zeta GST
G-site	GSH-binding site
GT	Glycosyl transferase
GTT-1	glutathione-transferase-1
GTT-2	glutathione-transferase-2
HPLC	High performance liquid chromatography
H-site	substrate binding site
Kn1	*Klebsormidium nitens* GST 1
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LOF	Loss of function
MAPEG	membrane associated proteins in eicosanoid and glutathione metabolism
mPGES2	microsomal prostaglandin E-synthase type 2
NTSR	Non-target site resistance
RNA-Seq	RNA sequencing
RT-qPCR	real time quantitative polymerase chain reaction
Tak-1	Takaragaike-1
Tak-2	Takaragaike-2
TCHQD	tetrachlorohydroquinone dehalogenase
TSR	Target site resistance

Chapter 1

Introduction

1.1. Summary

In this literature review I begin with introducing the problems presented by herbicide resistance in weeds. I describe how the strong selection pressure imposed by herbicides on weeds has led to an exponential increase in weed species evolving resistance, and combined with a lack of new herbicides on the market this presents a threat to food production. I follow with a section describing the general features of herbicides and their modes of action, with a particular focus on acetolactate synthase (ALS) inhibiting herbicides such as chlorsulfuron. I show that the extensive use of ALS inhibiting herbicides has caused a dramatic increase in resistant weeds, to the point that there are more species resistant to ALS inhibitors than to any other herbicide mode of action.

In the 3rd section I describe the two mechanisms of herbicide resistance: target site resistance and non-target site resistance, and show that while target site resistance is more common, non-target site resistance presents a bigger threat to food production due to the ability to confer resistance to herbicides not encountered by the plant. I follow with a review of metabolic resistance and the main enzyme families involved, namely cytochrome P450s, glutathione s-transferases, and glucosyl-transferases. I review the evidence for their involvement in metabolic resistance, and show that although there is substantial proof of the involvement of cytochrome P450s and glutathione s-transferases in metabolic resistance to various herbicides, few genes conferring metabolic resistance have been isolated. Therefore the genetic basis of metabolic resistance is not well known.

In the final section I explain why *Marchantia* has an advantage over weeds for studying complex genetic traits such as herbicide metabolic resistance, and I conclude with a summary of how I used *M. polymorpha* as a system to identify genes that can confer herbicide resistance, and to characterise the lethal effect of chlorsulfuron on sensitive plants.

1.2. The herbicide resistance problem

Weeds threaten arable agriculture by competing with the crop plant for light, water and nutrients. They also contaminate harvests, resulting in significant food waste. Weeds are estimated to cause the same crop losses as those from pests and diseases combined (Oerke, 2006). For decades farmers have relied on herbicides to control weed contaminations of their crops. This strong selection pressure has led to the rapid evolution of herbicide resistance in virtually all areas where herbicides are used (Heap, 2021).

A weed can be defined as any plant that is growing in the 'wrong' place. However weeds usually comprise a small number of species that thrive in agricultural environments. Examples of weeds include blackgrass (*Alopecurus myosuroides*), shortawn foxtail (*Alopecurus aequalis*) and ryegrass (*Lolium rigidum*). The UK's most widely grown crop is wheat, thus in the UK and in many other European countries, the major problem weed is blackgrass, which thrives in wheat fields. Populations of blackgrass show resistance to a wide range of modes of action, with both target-site resistance and metabolic resistance now widespread (Heap, 2021; Marshall et al., 2013). Resistant blackgrass has been found in virtually all areas in the UK where herbicides are applied for its control (Heap, 2021).

Herbicide resistance can be defined as the natural ability of some biotypes within a population to survive and reproduce in the presence of the herbicide. The first cases of herbicide resistance were reported in 1957, which represented cases of resistance to the synthetic auxin 2,4-D, 17 years after it was first introduced to the market (Heap, 2021). Since then, there have been numerous reports of resistant weeds emerging worldwide. Today, resistance has evolved to 23 of the 26 known herbicide sites of action, in 522 unique cases across 254 species (148 dicots and 106 monocots) (Heap, 2021). Many weed populations are resistant to multiple herbicide classes (Walsh et al., 2004; Yu et al., 2009) and different resistance mechanisms have been discovered in the same plant (Yang et al., 2016).

The over-reliance on herbicides, lack of diversity of management practices and the limited herbicide discoveries (Duke, 2012) has led to the widespread evolution of resistant weed populations and enhanced herbicide metabolism. The weed resistance problem is likened to the rapid evolution of multiple drug resistance in human pathogens, and involves potentially similar metabolic pathways. Climate change could accelerate the evolution of weed metabolic resistance, since detoxification enzymes have been shown to have higher activities at higher ambient temperatures (Matzrafi et al., 2016). Understanding the metabolic pathways underpinning herbicide metabolism would help to elucidate the evolution of NTSR in plants. This would inform the future development of weed management practices, the development of new herbicides and of detection methods for NTSR.

1.3. Herbicides: general properties and modes of action

Herbicides are organic molecules designed to be absorbed into plant tissues where they exert a toxic effect, and are used to control unwanted vegetation. Once in the cell, herbicides work in a diversity of ways to impair the target plant. Their mode of action usually involves the inhibition of an essential process needed for the plant's survival such as growth, photosynthesis, cell division, amino acid synthesis or cell membrane biogenesis. A few of the main classes of herbicides and their modes of action are summarised in Table 1.1.

Essential cellular process	Target site for inhibition	Example active ingredient	Toxic effect	Reference
Photosynthesis	PSII	atrazine linuron	accumulation of electrons in the chloroplast	Pfister & Arntzen, 1979
Amino acid synthesis	EPSPS enzyme	glyphosate	Inhibit synthesis of tryptophan, phenylalanine and tyrosine	Amrhein, et al., 1980
	ALS enzyme	chlorsulfuron tribenuron	Inhibit synthesis of branch-chain aas	Ray, 1982, 1984
Lipid biosynthesis	ACCase enzyme	fenoxaprop sethoxydim	Inhibit lipid synthesis	Burton et al., 1987
Growth regulation (synthetic auxins)	Auxin transport	diflufenzopyr	distorted cell division & expansion, deregulation of growth	Grossmann, 2010

Table 1.1. Herbicide modes of action.

aa; amino acid, ACCase; acetyl CoA carboxylase, ALS; acetolactate synthase, EPSPS; enolpyruvylshikimate 3-phosphate synthase, PSII; photosystem II.

1.3.1. ALS inhibitors e.g. chlorsulfuron

ALS-inhibiting herbicides are some of the most widely used weed controlling chemicals since their introduction in the early 1980s (Ray, 1982). These form five classes: sulfonylureas, imidazolinones, triazolopyrimidines, pyrimidinylthio (or oxy)-benzoates and sulfonylamino-carbonyltrazolinones (Zhou et al., 2007). Chlorsulfuron (2-chloro-N-[(4-methyl-1,3,5-triazin-2-yl)aminocarbonyl]benzenesulfonamide) is a compound of the sulfonylurea class of ALS inhibitors used to control both grass and broad-leaved weeds in small grain crops such as wheat, barley and oat crops. It is a selective compound and translocated to the whole plant after being absorbed by the leaves and roots.

Chlorsulfuron kills the target plant by inhibiting the activity of the acetolactate synthase (ALS) enzyme involved in the synthesis of the branch-chain amino acids (Ray, 1982). ALS catalyses two reactions: the condensation of two pyruvate molecules to produce CO_2 and 2-acetolactate, a precursor of valine and leucine, and the condensation of pyruvate and 2-ketobutyrate to form CO_2 and 2-acetohydroxybutyrate, a precursor of isoleucine (Figure 1.1.). Once absorbed by the plant, chlorsulfuron blocks the activity of ALS by binding to it, preventing the synthesis of these amino acids (Ray, 1984). This leads to the inhibition of DNA synthesis leading to impaired growth and cell division (Ray, 1982, 1984; Zhou et al., 2007)

ALS inhibiting herbicides such as chlorsulfuron (marketed as 'Gleen' by DuPont) emerged in the early 1980s as new herbicides for cereals. The unique feature of this new class of herbicides was its very high herbicidal activity at extremely low application rates -grams per hectare as opposed to the kilograms per hectare rates typical of previous herbicides, and its low mammalian toxicity, as ALS is not present in animals (Bellinder et al., 1994; Levitt et al., 1981).

Due to their popularity, ALS inhibitors have been used extensively to control weeds in wheat, barley and oat fields. Today, more than 50 commercial ALS-inhibiting compounds are used for selective weed control in an immense variety of grass and broadleaf crops worldwide. (Scarabel et al., 2015). As a result, there's been a dramatic increase in resistant biotypes to ALS-inhibitors (Figure 1.2). Today, more than 170 species have evolved resistance to ALS-inhibiting herbicides, far more than towards any other herbicide mode of action (Heap, 2021).

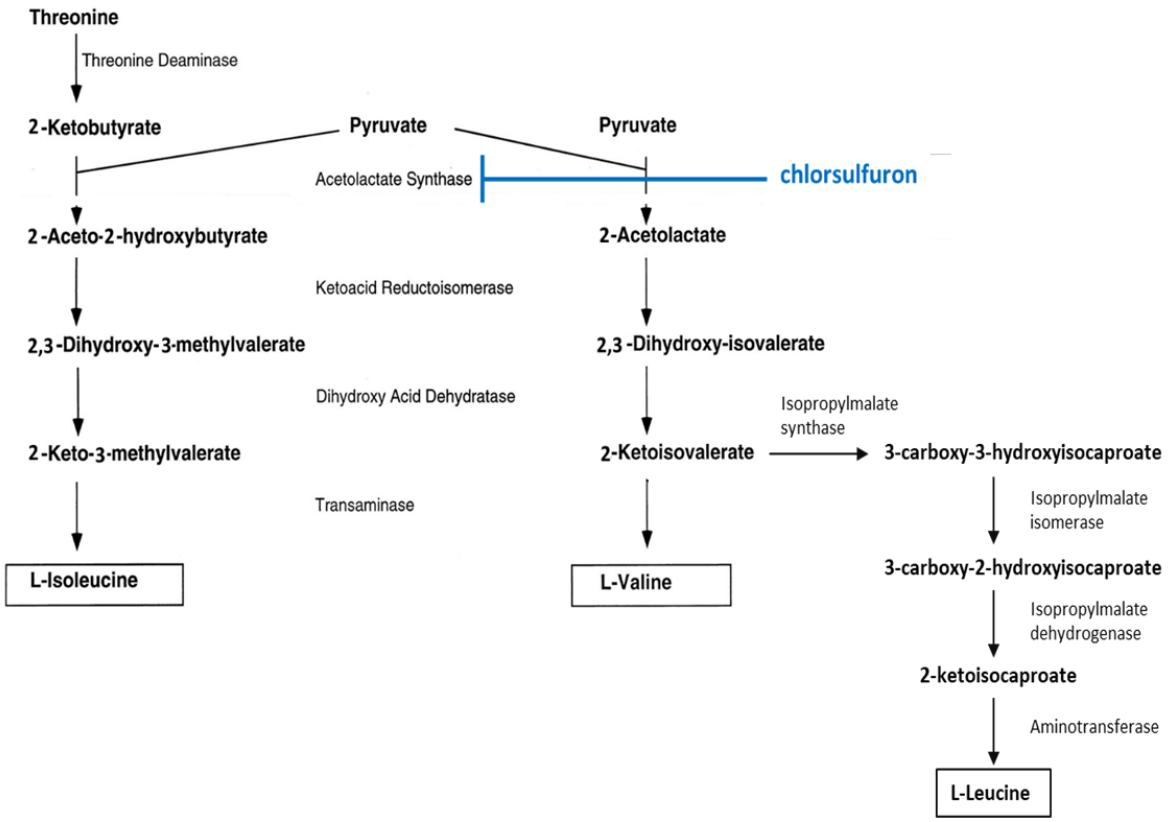


Figure 1.1. Biochemical pathway of branched chain amino acids leucine, valine and isoleucine.
The enzyme targeted by chlorsulfuron is acetolactate synthase (ALS), the first common enzyme in the biochemical pathway of branched chain amino acids, and is indicated by the blue line.

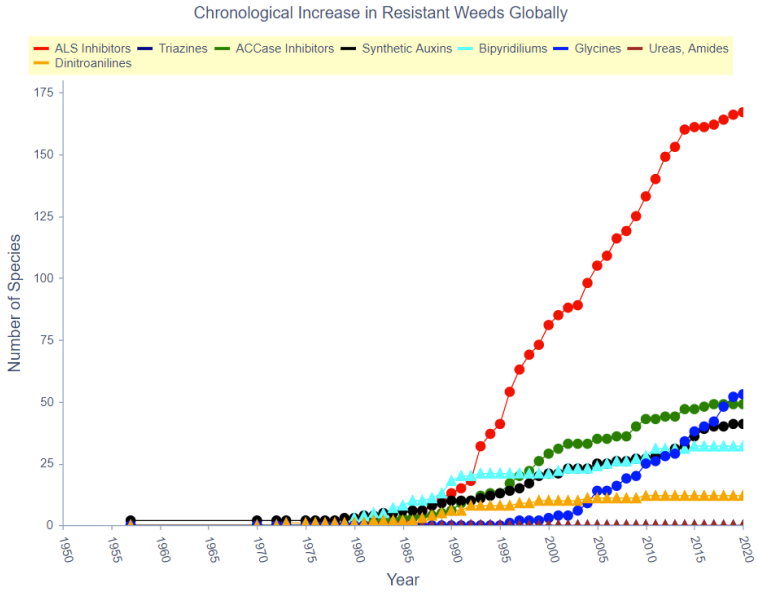


Figure 1.2. Global increase in resistant weed species to different herbicide classes.
From Heap, (2021).

1.4. Target site resistance and non-target site resistance

Resistance to herbicides exists in two forms; target-site resistance (TSR) and non-target-site resistance (NTSR). Target-site resistance occurs by a mutation in the gene coding for the herbicide's target-site enzyme/protein complex (preventing herbicide binding), or by increasing the production of the target protein. Non-target-site resistance is attributed to a reduction in herbicide molecules reaching the target site by reducing herbicide uptake or translocation, increasing herbicide sequestration or increased herbicide metabolism. An enhanced rate of herbicide metabolism is the dominant mechanism of NTSR (Yu & Powles, 2014c).

Traditional ways to manage TSR weeds is to rotate the use of herbicides with different target sites. This is a temporary measure as it imposes the same evolutionary pressures as the previous herbicide, and the weed will eventually evolve resistance. NTSR involving enhanced herbicide metabolism has been found to be effective against herbicides with different chemistries and modes of action, including herbicides not yet marketed (Burnet et al., 1994; Burns et al., 2018; Délye et al., 2011). The numbers of weed species reported to be resistant to multiple herbicides with different modes of action is increasing (Figure 1.3), suggesting that NTSR is becoming more prevalent. Therefore, although NTSR is less widespread than TSR, it may pose a greater threat to food production. Since no new herbicides have been brought to the market in 20 years, it is unwise to rely on the development of new herbicides to solve the problem of herbicide resistance.

It is comparatively easy to study the molecular basis of target-site resistance, as it tends to involve a point-mutation in the target gene. NTSR involves multiple genes and the mechanisms are varied due to the genomic plasticity among weeds (Letouzé & Gasquez, 2003; Preston, 2004; Preston et al., 1996). These complex biochemical processes are still poorly understood (Preston, 2004; Preston et al., 1996).

1.5. Herbicide metabolism: current knowledge

The detoxification of xenobiotics is grouped in four main phases as defined by Kriton K. Hatzios (1997). These include phase I conversion (or functionalisation), phase II conjugation of the chemical, and phase III compartmentation. Conversion of the xenobiotic usually occurs by oxidation or hydrolysis. Oxidation can take place using cytochrome P450s or other mixed function oxidases/peroxidases. These reactions increase the polarity of the molecule or provide linkage sites to make it suitable for conjugation to a hydrophilic molecule, namely glutathione, glucose or an amino acid (Cummins et al., 2013; Reade et al., 2004; Yang et al., 2016). Resulting conjugates are no longer active towards the target site of the chemical, and are more hydrophilic and less mobile than the parent herbicide. This allows them to be transported into the vacuole or deposited outside of the cell (phase III) by ABC transporters (Klein et al., 2006). The sequestered compounds may undergo further detoxification in the vacuole or extra cellular space (phase IV). (Hatzios, 1997).

Plants show a myriad of detoxification pathways that can metabolise a wide diversity of herbicides, making their elucidation and characterisation a complex task (Preston, 2004). The enzymes most frequently found to be involved in the detoxification of ALS inhibitors such as chlorsulfuron are cytochrome P450 monooxygenases, glutathione-S-transferases and glycosyltransferases (Cummins et al., 2011; Hatzios, 1997; Preston, 2004; Siminszky, 2006; Yuan et al., 2007).

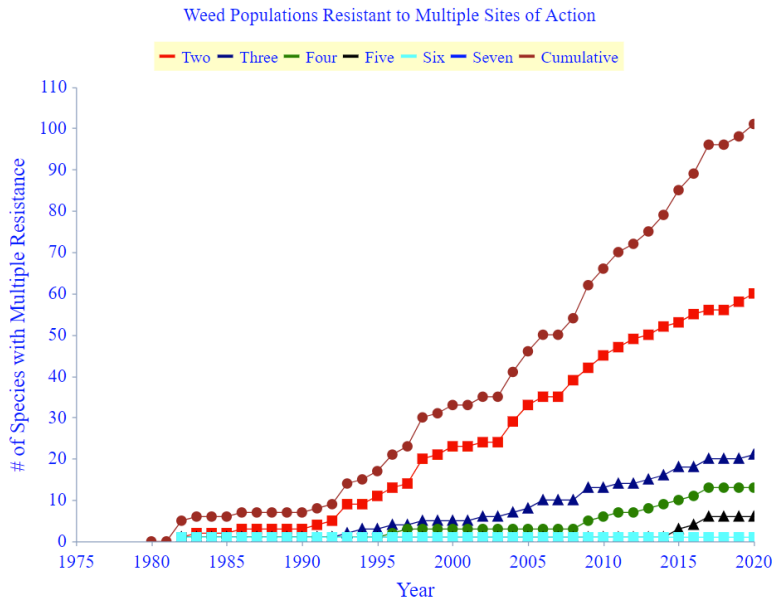


Figure 1.3. Number of weed populations exhibiting resistance to herbicides with multiple modes of action.
From (Heap, 2021).

1.5.1. Cytochrome P450s

Cytochrome P450s (CYPs) are the largest family of enzymes in plants. CYPs are heme-protein dependent oxidases with a wide diversity of functions (Xu et al., 2015). The most common reactions catalysed by cytochrome P450s are mono-oxygenase reactions that use electrons from NADPH to cleave atmospheric dioxygen using NADPH-CYP reductase; the first oxygen atom hydroxylates with the (inert, hydrophobic) substrate (RH) and the second oxygen atom forms water. This reaction produces a more reactive and soluble product molecule:

$$\mathbf{R\text{-}H} + O_2 + NADPH + H+ \rightarrow NADP+ + H_2O + \mathbf{R\text{-}OH}$$

Plant CYPs are far more numerous and diverse than in animals, which is likely a vital adaptation to a sedentary lifestyle. In eukaryotes, CYPs are membrane-bound and primarily located in the endoplasmic reticulum. They are already well known in human drug metabolism and insecticide resistance in insects (Danielson, 2002; Li et al., 2007).

CYPs play a primary role in phase I herbicide metabolism. The first evidence for CYP involvement in herbicide metabolism was published by Frear et al. (1969) on the metabolism of monuron (PSII inhibitor) using a microsomal fraction from cotton seedlings. By the 1990s an extensive body of evidence had built up supporting the role of CYPs in the metabolism of a range of herbicides in weed and crop species (Burnet et al., 1993; Powles et al., 1991; Cole & Owen, 1987; Fonné-Pfister et al., 1990; Kreuz et al., 1991; Mougin et al., 1991; Ohkawa et al., 1999; Siminszky, 2006). CYPs have been shown to play a primary role in the metabolism of various herbicide classes, particularly in the aryl-hydroxylation of members of the sulfonylurea group of the ALS-inhibitors (of which chlorsulfuron is a member) (illustrated in Fig.3) (Fonné-Pfister et al., 1990; Frear et al., 1991; Xiang et al., 2005), and the N-demethylation of many phenylurea herbicides (photosystem II inhibitors) (Didierjean et al., 2002; Siminszky et al., 1999; Siminszky, 2006).

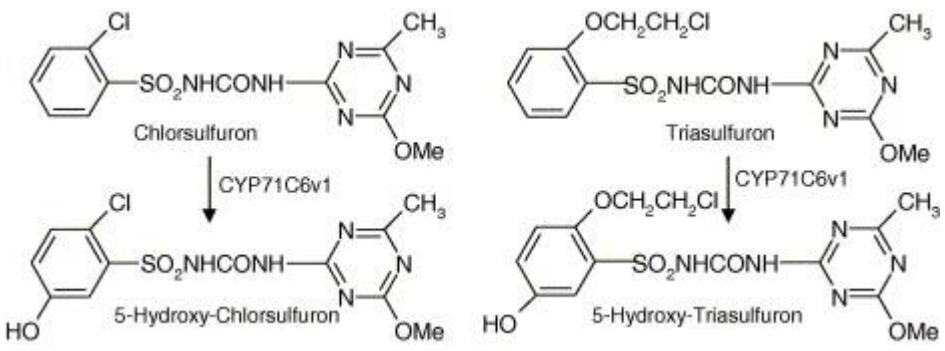


Figure 1.4. Diagram of a wheat cytochrome P450 catalysing the phenyl-ring hydroxylation of two sulfonylureas.
The two herbicides are chlorsulfuron and triasulfuron. Reaction determined from NMR and mass-spectrometry data of the herbicide metabolites (Xiang et al., 2005).

A few CYPs that metabolise sulfonylurea herbicides have been isolated. A full list of CYPs that metabolise or confer resistance to herbicides can be found in Chapter 2, Table 2.1. CYP72A31 from rice (*Oryza sativa*) confers tolerance to acetolactate synthase-inhibiting herbicides in rice and *A. thaliana* (Saika et al., 2014). CYP81A12 and CYP81A21 in *Echinochloa phyllopogon* are associated with resistance to two acetolactate synthase inhibitors (Iwakami et al., 2014). CYP94A1, CYP94A2, CYP71A4, and CYP734A6 were all linked to resistance to mesosulfuron-methyl in shortawn foxtail (*Alopecurus aequalis*) (Zhao et al., 2017). Two maize (*Zea mays*) CYPs, CYP92A1 and CYP72A5 were induced in response to triasulfuron (Persans et al., 2001). CYP81A9 in maize confers tolerance to nicosulfuron (Li et al., 2013). A wheat (*Triticum aestivum*) CYP, CYP71C6v1 cloned and expressed in yeast catalysed the metabolism of some sulfonylurea herbicides (Xiang et al., 2005). CYP81A6 in rice conferred resistance to benzaton and three sulfonylurea herbicides (Pan et al., 2006; Zhang et al., 2006). The ectopic expression of all CYP81A genes in *Echinochloa phyllopogon* demonstrated that 6/9 CYP81A genes conferred resistance to sulfonylurea herbicides (Dimaano et al., 2020). Finally CYP749A16 in cotton (*Gossypium hirsutum*) confers tolerance to trifloxysulfuron sodium (Thyssen et al., 2018).

1.5.2. Glutathione-S-transferases

The second most well-documented herbicide metabolising gene family is the glutathione-S-transferase (GST) group. GSTs play a role in phase II herbicide metabolism, catalysing the conjugation of the hydrophilic molecule glutathione to a hydrophobic compound via a substitution or addition reaction. Conjugates are no longer active towards the chemical's target site. Like CYPs, the GST gene family is large and diverse, and appears to have undergone dynamic evolution with multiple gene duplications. GSTs show a diverse substrate specificity, allowing them to detoxify a wide range of chemicals as well as synthesise various secondary metabolites (Cummins et al., 2011; Yuan et al., 2007). A full list of GSTs that metabolise or confer resistance to herbicides can be found in Chapter 3, Table 3.4. GSTs act on several herbicide classes including triazines (PSII inhibitors), chloroacetanilides (gibberellin pathway inhibitors) and aryloxyphenoxy-propionates (ACCase inhibitors) (Cummins et al., 2013, 2011; Karavangeli et al., 2005; Li et al., 2016; Shah et al., 1986; Thom et al., 2002). Although there is some metabolic evidence that glutathione s-transferase mediated GSH conjugation of a sulfonylurea herbicide rendered it inactive to ALS (Wittenbach et al., 1994), only 5 GSTs have been isolated that metabolise sulfonylurea. GSTU4/19 from maize and GSU5, GSTU9 and GSTU10 from soybean (*Glycine max*) showed GSH conjugating activity towards chlorimuron ethyl (McGonigle et al., 2000), and over-expression of GSTL2 in rice conferred some tolerance towards chlorsulfuron (Hu et al., 2009).

1.5.3. Glycosyltransferases

Glycosyltransferases (GTs) are the second group of phase II conjugating enzymes. These enzymes catalyse the conjugation of glucose to a hydroxylated compound, completing the detoxification process for that compound. It was suggested from HPLC studies using C^{14}-labelled chlorsulfuron-treated tolerant wheat leaves that after phenyl ring-hydroxylation chlorsulfuron is rapidly conjugated to glucose (Sweetser et al., 1982). There are very few studies on the role of GTs in herbicide metabolism. Studies of GT activity in blackgrass have been found to be involved in the metabolism of several herbicides (Brazier et al., 2002), and a GT was consistently highly expressed in NTSR rigid ryegrass (*Lolium rigidum)* populations to diclofop (ACCase inhibitor) (Gaines et al., 2014). GT has also been found to metabolise diclofop in wheat in coordination with CYTP450 aryl hydroxylation (Helvig et al., 1996; Shimabukuro et al., 1987). Two GTs (GT83A1 and GT75D1) were recently isolated in NTSR shortawn foxtail (*Alopecurus aequalis)* resistant to mesosulfuron-methyl (a sulfonylurea) and (Zhao et al., 2017).

1.6. *Marchantia polymorpha* as a system to study herbicide resistance

The identification of genes conferring metabolic herbicide resistance has so far only been done in weed and crop species. However, transcriptome analysis of these species requires de novo assembly and extensive validation in different tolerant and susceptible populations due to their high genetic diversity. Instead I used the liverwort species *M. polymorpha* to determine the genetic basis of herbicide metabolism.

M. polymorpha is part of the liverwort lineage which represents one of the earliest diverging lineages in land plants, with hornworts and mosses, having first evolved around 450 million years ago (Morris et al., 2018). *M. polymorpha* is dioecious, producing male and female plants (Shimamura, 2016). It also has a haploid dominant lifestyle, spending the majority of its lifecycle as a haploid plant called a gametophyte. Male gametophytes produce antheridiophores and female gametophytes produce archegoniophores. Antheridiophores contain antheridia which produce sperm, and archegoniophores contain archegonia which produce eggs. A fertilised egg develops into a diploid sporophyte in which spore mother cells divide by mitosis to produce haploid spores. These spores grow into new male and female plants. *M. polymorpha* can also reproduce asexually. Each gametophyte produces 100s of vegetative clonal propagules called gemmae within specialised structures called gemmae cups. Gemmae provide a simple way to measure changes in plant growth, because they grow flat on solid media and can be imaged from above.

The liverwort plant has several characteristics that provide advantages over weeds for functional genomic studies of multi-gene processes such as herbicide metabolism. *M. polymorpha* has a short lifecycle, taking 3 months to be able to reproduce via sexual reproduction using spores, or 2-3 weeks to reproduce via asexual reproduction using gemmae. *M. polymorpha* plants can be propagated with ease in liquid or solid medium, and each cross produces 1000s of spores. The dominant haploid lifecycle is an advantage for generating loss of function mutants, because the defective phenotype caused by the loss of function mutation is visible within the first generation of transformed plants, therefore there is no need for the long process of establishing homozygous lines that is required in diploid recessive plants such as *A. thaliana*. Other advantages for genetic manipulation include a high frequency of transformation, low genetic redundancy and a small genome size (approximately 280 Mb) (Bowman et al., 2017).

The development of transformation technologies (Ishizaki et al., 2008; Ishizaki et al., 2016; Tsuboyama-Tanaka & Kodama, 2015), targeted genome modification (Sugano et al., 2014b) and the recent completion of the whole genome sequencing project for *M. polymorpha* (Bowman et al., 2017) make *M. polymorpha* a powerful genetic model. These characteristics make it a good system to study complex phenomena such as herbicide metabolism in vivo.

1.7. CYP, GST and GT gene families in *Marchantia polymorpha*

The CYP, GST and GT gene families are large in *M. polymorpha*, and contain similar numbers or slightly fewer genes than in *A. thaliana* (for example, there are 149 CYP450s in *M. polymorpha*, 244 in *A. thaliana*, 328 in *O. sativa*). The recent sequencing of the *M. polymorpha* genome revealed that the majority of CYPs belong to *Marchantia*-specific subfamilies, which suggests that CYPs in *Marchantia* have lineage specific functions (Bowman et al., 2017). The functions of the liverwort CYPs, GSTs and GTs are largely uncharacterised.

Only one study has looked at GSTs in liverworts in response to herbicide treatment, by treating liverwort cultured cells with photobleaching herbicides (Iwata et al., 1992). Young treated cells had 4-fold higher levels of glutathione than untreated cells, which peaked after 8h, however they concluded that GSH protected the plant from reactive oxygen species produced as a stress response to the herbicide, not by metabolising the herbicide. No studies have characterised herbicide metabolism in the liverwort.

In this thesis I use *M. polymorpha* to investigate the genetic basis of metabolic resistance towards the herbicide chlorsulfuron, and to investigate chlorsulfuron toxicity in plants. I conducted phylogenetic analyses of the GST and CYP enzyme families in the Archaeplastida, where I show that GSTs and CYPs that metabolise herbicides belong to the two expanded GST classes and CYP clans in vascular plants. I also show that these classes and clans predate the evolution of land plants, suggesting that the ability to metabolise herbicides may be conserved across land plants. I then investigated

the transcriptional response of *M. polymorpha* to chlorsulfuron treatment. I identified a CYP up-regulated by chlorsulfuron that confers a low level of resistance to chlorsulfuron when over-expressed. This result lead me to hypothesise that non-target site herbicide resistance (NTSR) evolves from genes that form part of the transcriptional response to herbicide treatment. Finally I investigated how chlorsulfuron causes plant death using growth assays, metabolomic analyses and reverse genetics. These results provide insights into how the transcriptional and metabolic changes caused by ALS inhibition cause lethality or lead to resistance in plants. This is important for the development of tools to overcome NTSR and to develop better herbicides.

Chapter 2

Evolution of cytochrome P450 enzymes in the Archaeplastida

2.1. Abstract

Cytochrome P450s enzymes (CYPs) are a large protein family and have essential roles in the biosynthesis of hormones, primary and secondary metabolites, and the detoxification of xenobiotics. To characterise the evolution of CYP proteins in Archaeplastida I identified CYP sequences from 9 species representing the rhodophyte, chlorophyte, streptophyte algae and land plant lineages and constructed gene trees.

The CYP protein family is subdivided into clans. The topology of the gene tree indicates that there were two CYP clans in the common ancestor of the Archaeplastida. One of these clans, CYP51, is conserved across eukaryotes and is likely to be the source of the ancestral, eukaryotic CYP. The number of CYP clans increased during the course of Archaeplastida evolution and there are now 9 clans in the chlorophyte *Chlamydomonas reinhardtii* and 12 clans in the streptophyte alga *Klebsormidium nitens*. All 12 streptophyte alga CYP clans are present in the genomes of land plants and no new clans evolved in after plants colonised land. Three clans were lost during vascular plant evolution (CYP727, CYP746 and CYP747) and three were lost among the early diverging land plants (CYP711, CYP746 and CYP747). Despite these clan losses, the total number of CYP genes in seed plants is almost ten times greater than in streptophyte alga genomes; there are 242 CYP genes in *Arabidopsis thaliana* and 29 in *Klebsormidium nitens*. The large increase in CYP gene number that occurred after the colonisation of the land results from gene duplications in four clans; CYP71, CYP72, CYP85 and CYP86. CYPs that confer resistance to herbicides belong to two of these CYP clans, CYP71 and CYP72.

Together these data demonstrate that the major diversification of CYPs into distinct clans occurred in the Proterozoic before the Palaeozoic radiation of land plants. Subsequent gene duplication followed by neofunctionalization in four of the twelve clans, early in embryophyte evolution, resulted in the high CYP diversity among land plants. It is among the CYP clans that diversified on land that genes conferring herbicide resistance evolved in the last fifty years.

2.2. Introduction

Cytochrome p450 enzymes (CYPs) are a superfamily of enzymes that are present in plants, fungi, bacteria and animals. CYPs are heme-thiolate proteins that use molecular oxygen and NADPH to modify substrates by diverse chemical reactions such as oxidations, hydroxylations, dealkylations and reductions (Mizutani & Sato, 2011) and are implicated in a wide array of biochemical pathways. CYPs participate in the synthesis and modification of primary metabolites such as sterols and fatty acids, secondary metabolites such as phenylpropanoids, glucosinolates and carotenoids, and the synthesis and catabolism of hormones such as gibberellins, jasmonic acid, abscisic acid, brassinosteroids and strigolactones (Mizutani & Ohta, 2010; Mizutani & Sato, 2011; Wakabayashi et al., 2019).

CYP proteins are usually bound to the cytoplasmic side of the endoplasmic reticulum membrane, with some proteins predicted to be targeted to the chloroplast and mitochondria (Schuler et al., 2006; Werck-Reichhart & Feyereisen, 2000). Their protein sequences are characterised by a conserved heme-binding domain, an oxygen binding domain, two conserved regions (X-E-X-X-R and P-E-R-F) that form what is known as the ERR triad, which is involved in positioning and stabilising the heme pocket, and several highly variable substrate positioning and recognition sites (Rupasinghe & Schuler, 2006). Proteins belonging to the CYP74 family are unique in their amino acid sequence, having a nine amino acid insertion in the heme-binding domain, and lacking an oxygen binding motif. Despite CYP proteins sharing as little as 16% of their amino acid sequences, their structural fold is highly conserved (Hasemann et al., 1995).

Plants have many more CYPs than animals. There are 55 CYPs in humans, while 245 have been identified in *Arabidopsis thaliana* and 459 in grape (Nelson et al., 2008; Nelson, 2009; Schuler et al., 2006). CYPs represent 1% of total genes in many plant species. Plant CYPs are polyphyletic, initially classified into two groups, the monophyletic A-type group, and the polyphyletic non-A type group. With the addition of new CYPs as more genomes are sequenced they have since been reclassified into 11 clans and over 250 families (Bak et al., 2011; D. R. Nelson, 2018) based on their phylogeny and sequence similarity (Nebert et al., 1987; Nelson et al., 1996). The A-type group was renamed CYP71 clan and contains enzymes mostly involved in the biosynthesis of natural products. CYPs have been found to have important roles in stress responses and herbicide detoxification (Dimaano & Iwakami, 2021; Pandian et al., 2020). As the function of the majority of CYPs is still unknown, understanding the phylogeny of herbicide-detoxifying enzymes and CYPs in general is crucial for identifying potential orthologs in other species or closely related genes in the same species, that may share the same function.

Previous studies on plant CYP evolution have classified CYPs in different species (Du et al., 2016; Nelson, 2006, 2018; Nelson et al., 2008; Nelson & Werck-Reichhart, 2011) but an analysis of the CYP superfamily across the

Archaeplastida was lacking. In this chapter I characterise the evolution of the CYP superfamily in plants. I find that plant CYPs are polyphyletic and form multiple clans. Land plants group into 12 clans, all of which are present in the genome of the streptophyte algae *Klebsormidium nitens*, indicating that these clans evolved before the colonisation of plants on land. CYP747 is a new clan present in the chlorophyte alga and streptophyte alga and retained in the hornwort genome. The CYP71, CYP72, CYP85, CYP86 clans expanded in land plants, and contain all identified plant CYPs that are responsible for herbicide metabolic resistance. These results provide a basis for the functional investigation of plant CYPs, particularly in relation to herbicide resistance.

2.3. Materials and Methods

2.3.1. Data resources

Protein sequences from *A. thaliana* were retrieved from TAIR10 (Huala et al., 2001) (https://www.arabidopsis.org/) Protein sequences from *Oryza sativa* were retrieved from the rice genome annotation project (Kawahara et al., 2013) (http://rice.plantbiology.msu.edu/).Protein sequences from the liverwort *M. polymorpha* were obtained from MarpolBase (http://marchantia.info/). Protein sequences from the hornwort *Anthoceros agrestis* were obtained from (Li et al., 2020) (https://www.hornworts.uzh.ch/en/download.html). Protein sequences from the streptophyte alga *Klebsormidium nitens* were obtained from the *K. nitens* genome webpage (Hori et al., 2014) (http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/). Protein sequences from the moss *Physcomitrella patens* and the chlorophyte alga *Chlamydomonas reinhardtii* were retrieved from Phytozome 12 (Goodstein et al., 2012) (https://phytozome.jgi.doe.gov/pz/portal.html). Protein sequences from the red alga *Cyanidioschyzon merolae* (Matsuzaki et al., 2004) were retrieved from the *C. merolae* genome webpage (https://www.genome.jp/kegg-bin/show_organism?org=cme).

A classification of CYP genes from *A. thaliana, S. moellendorffii*, *P. patens*, *M. polymorpha*, *C. reinhardtii* and *C. merolae* is available on The Cytochrome P450 Homepage (Nelson, 2009) (http://drnelson.utmem.edu/cytochromeP450.html). Two other Arabidopsis CYP databases can be found on the Arabidopsis Cytochrome P450 List (Paquette et al., 2000) (http://www.p450.kvl.dk/At_cyps/table.shtml) and CyPEDIA (Ehlting et al., 2008) (http://www-ibmp.u-strasbg.fr/~CYPedia/). The classification of *O. sativa* CYPs is available on the University of California, Davis Rice CYP Database (https://ricephylogenomics.ucdavis.edu/p450/).

2.3.2. Sequence collection

CYP proteins from *A. thaliana* and *O. sativa* are already well characterised (Liu et al., 2018; Nelson et al., 2004) and were used for the retrieval of CYP proteins from the other species. Sequences from these species were used to perform BlastP searches using an e value cut-off of $1e^{-10}$ against the proteomes of *S. moellendorffii*, *M. polymorpha*, *A. agrestis*, *P. patens*, *K. nitens*, *C. reinhardtii*, and *C. merolae*. This initial list of sequences was self-blasted against the respective genomes in each species to retrieve any additional sequences belonging to species-specific clans. The results from these searches were analysed for the presence of the cytochrome p450 domain (PF00067) using PFAM v26.0 (Punta et al., 2012).

2.3.3. Sequence alignment

Sequences were aligned in MAFFT (Katoh & Standley, 2013) using the FFT-NS-2 algorithm and visualised in Bioedit (Hall, 1999). Sequences lacking conserved regions of the cytochrome P450 domain were removed and labelled as pseudogenes.

To trim ambiguous regions and large gaps, four approaches to alignment cleaning were undertaken. A manual approach using knowledge of the location of the conserved P450 domains was undertaken., as well as an automated approach using the trimming software trimAl v.1.2. (Capella-Gutiérrez et al., 2009) using the three automated modes (-gappyout, -strict and -strictplus).

2.3.4. Phylogenetic analysis

The final alignments were subjected to a maximum-likelihood analysis conducted by PHyML 3.0 (Guindon et al., 2010) using an estimated gamma distribution parameter, the LG+G+F model of amino acid substitution and a Chi^2-based approximate likelihood ratio test (aLRT). The resulting unrooted trees were visualised in Figtree (http://tree.bio.ed.ac.uk/software/ figtree/), and annotated in Inkscape (https://inkscape.org/release/inkscape-1.0.2/).

2.4. Results

Protein sequences from the genomes of 9 species representing major plant lineages were collected by BlastP searches of known plant CYPs, checked for the presence of the cytochrome p450 heme-binding domain using PFAM (PF00067) and manually filtered for pseudogenes, resulting in a final list of 1297 cytochrome P450s and 316 pseudogenes. These species include the red alga *Cyanidioschyzon merolae* (5 genes), the chlorophyte alga *Chlamydomonas* reinhardtii (40 genes), the streptophyte alga *Klebsormidium nitens* (29 genes), the liverwort *Marchantia polymorpha* (115 genes), the moss *Physcomytrella patens* (69 genes), the hornwort *Anthoceros agrestis* (144 genes), the lycophyte *Selaginella moellendorfii* (202 genes), and the angiosperms *Oryza sativa* (292 genes) and *Arabidopsis thaliana* (241 genes). The sequences were named following the CYP nomenclature developed by Nelson *et al.* (1996). The *A. agrestis* and *K. nitens* CYPs had not yet been identified or classified.

The sequences were aligned in MAFFT using the FFT-NS-2 algorithm and were trimmed manually to remove poorly conserved regions and retain the conserved cytochrome P450 domains and variable substrate recognition sites (Figure 2.5B). A more stringent trimming approach was also tested where the alignment was cleaned by the automatic trimming software trimAl v.1.2. (Capella-Gutiérrez et al., 2009) using the three automated modes (-gappyout, -strict and -strictplus) (Figure 2.1). The maximum-likelihood (ML) trees resulting from the trimAl -strict and -strictplus alignments had very low approximate likelihood ratio test (Chi^2-based aLRT) support values for the deepest clades and CYP clans were not monophyletic. The manually trimmed and trimAl -gappyout alignments possessed the expected monophyletic clans. Of these two trees, the tree based on the manually trimmed alignment had the overall highest aLRT values (0.97 and above) for each clan and was selected as the representative tree.

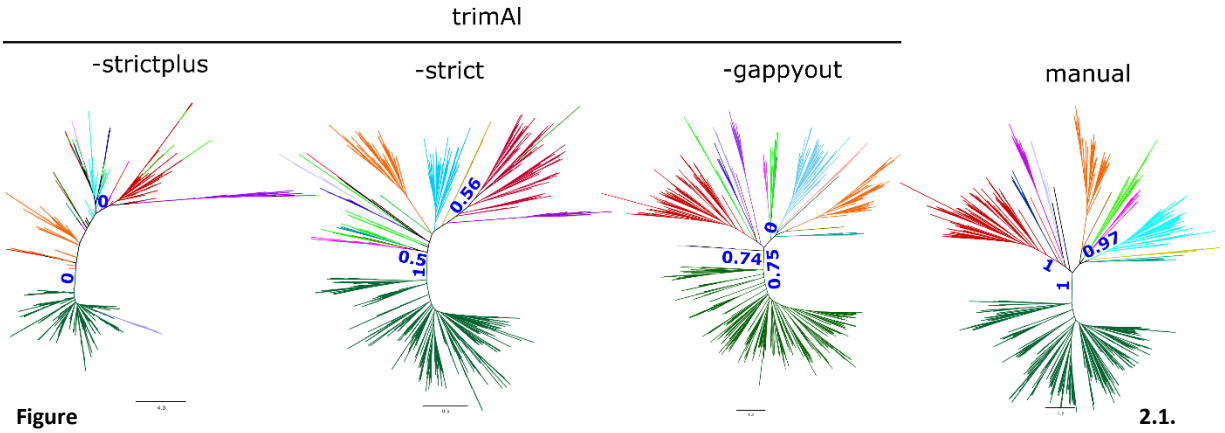


Figure 2.1. Plant CYP phylogenetic analysis using automatic and manual trimming approaches. Unrooted cladogram of a maximum likelihood (ML) analysis conducted by PHyML 3.0 (Guindon et al., 2010) using an estimated gamma distribution parameter, the LG+G+F model of amino acid substitution and a Chi^2-based approximate likelihood ratio (aLRT) test. Sequences were aligned in MAFFT and trimmed with the automatic trimming software trimAl using the automatic modes -strictplus, -strict, -gappyout or by manual trimming. Branches are coloured to show the different CYP clans. aLRT Support values for some of the clades are shown for comparison.

2.4.1. All land plant CYP clans are present in streptophyte algae

CYPs are heme-containing enzymes that form one of the largest enzyme families in plants. They are involved in many aspects of plant metabolism including hormone, sterol and fatty acid biosynthesis. CYPs are classified into clans, families and subfamilies according to a nomenclature committee (Nelson et al., 1993) based on their amino acid sequence identity and evolutionary relationship. Proteins need to share 40% amino acid identity to be classed in the same family, and 55% identity to be grouped in the same subfamily. Clans are monophyletic clades containing one or more families and are named after their lowest numbered family members. To elucidate the evolution of CYP clans in plants, a phylogenetic tree was constructed using a maximum likelihood approach (Figure 2.2). This analysis identified that CYP proteins from 9 species of Archaeplastida grouped into 18 monophyletic clans (Figure 2.2). The phylogeny identified 17 previously defined monophyletic clans and are consistent with previous studies on plant CYP phylogeny (Nelson, 2006, 2018). The inclusion of the *A. agrestis* and *K. nitens* CYP protein sequences changed the tree topology resulting in a monophyletic clade that corresponds to a clan that has not been identified previously. CYP747 is a monophyletic clade supported by a by Chi^2-aLRT value of 1, and comprises single genes from the genomes of *C. reinhardtii* , *K. nitens* and *A. agrestis*. There are no CYP747 representatives in any other land plant genomes. This suggests that CYP747 evolved among the common ancestor of the green algae and was lost early in land plant evolution.

The CYPs encoded by the genomes of land plant species *A. agrestis*, *M. polymorpha, P. patens*, S*. moellendorffii*, *O. sativa* and *A. thaliana* grouped in 12 - CYP51, CYP71, CYP72, CYP74, CYP85, CYP86, CYP97, CYP710, CYP711, CYP727, CYP746 and CYP747 - of the 18 Archaeplastida clans. All land plant clans also contained sequences of the streptophyte alga *K. nitens,* indicating that they were present before the divergence of charophyte algae and land plants. Members of 6/12 land plant clans – CYP71, CYP72, CYP74, CYP85, CYP86, CYP727 – were found in the streptophyte algal genome but not in the chlorophyte alga *C. reinhardtii*. This suggests that they originated in the streptophyte lineage after the divergence of chlorophytes and streptophytes and before the divergence of streptophyte algae and land plants (Figure 2.5A). *C. reinhardtii* CYP sequences represent 6/12 of the land plant clans – CYP51, CYP97, CYP710, CYP711, CYP746 and CYP747 – indicating that they were present before the divergence of streptophytes and chlorophytic algae for their last common ancestor. Two of the clans that are found in land plants were also present in red algae. These clans contain in addition one (CYP51) and two (CYP710) members from red alga. This places their origin before the divergence of Rhodophyta and Viridiplantae (Figure 2.5A). CYP51 and CYP710 were present in the last common ancestor of Archaeplastida and therefore constitute the most ancient CYP clans.

Three of the 18 Archaeplastida clans – CYP55, CYP737 and CYP741 –contained only chlorophyte sequences. There is a single CYP55 clan member in *C. reinhardtii*, CrCYP55B1, which is thought to have been inherited from bacteria through lateral transfer (Nelson, 2018). CrCYP55B1 was sister to the CYP51 clan. Thirty CrCYP sequences formed a single

monophyletic clade named CYP737 and was sister to the clade containing CYP85, CYP97, CYP741 and CYP747 clans. Two CrCYP protein sequences – CrCYP741A1 and CrCYP768A1 – formed a monophyletic clade named CYP741, that was sister to the clade containing clans CYP85, CYP97, CYP737, CYP741 and CYP747. Therefore, CYPs in the chlorophyte alga *Chlamydomonas reinhardtii* are not similar to other plant CYPs with many in chlorophyte-specific clans. Two clans - Cm1 and Cm2 - comprised single red algae proteins. Cm1 was sister to the clade containing clans CYP72, CYP85, CYP97, CYP711, CYP727, CYP737, CYP746 and CYP747. Cm1 may therefore be ancestral to these clans. Cm2 was sister to the CYP710 clan but shares very low amino acid identity (20%) with members of CYP710. Cm2 is possibly ancestral to CYP710 or represents a red-algae specific clan.

In summary, the phylogenetic analysis shows that all of the land plant clans are also found in streptophyte algae which indicates that the large diversity in CYP genes in plants evolved in an aquatic environment before plants colonised land at least 470 million years ago (Morris et al., 2018). No new clans appeared after plants colonised land. Five land plant CYP clans were encoded in the genome of the chlorophyte algae *C. reinhardtii*, which places their origin before the separation of the chlorophyte and streptophyte lineages. Two land plant clans, CYP51 and CYP710, contain red algae sequences. This suggests that these clans are the most ancient land plant clans and evolved before the divergence of Rhodophyta and Viridiplantae.

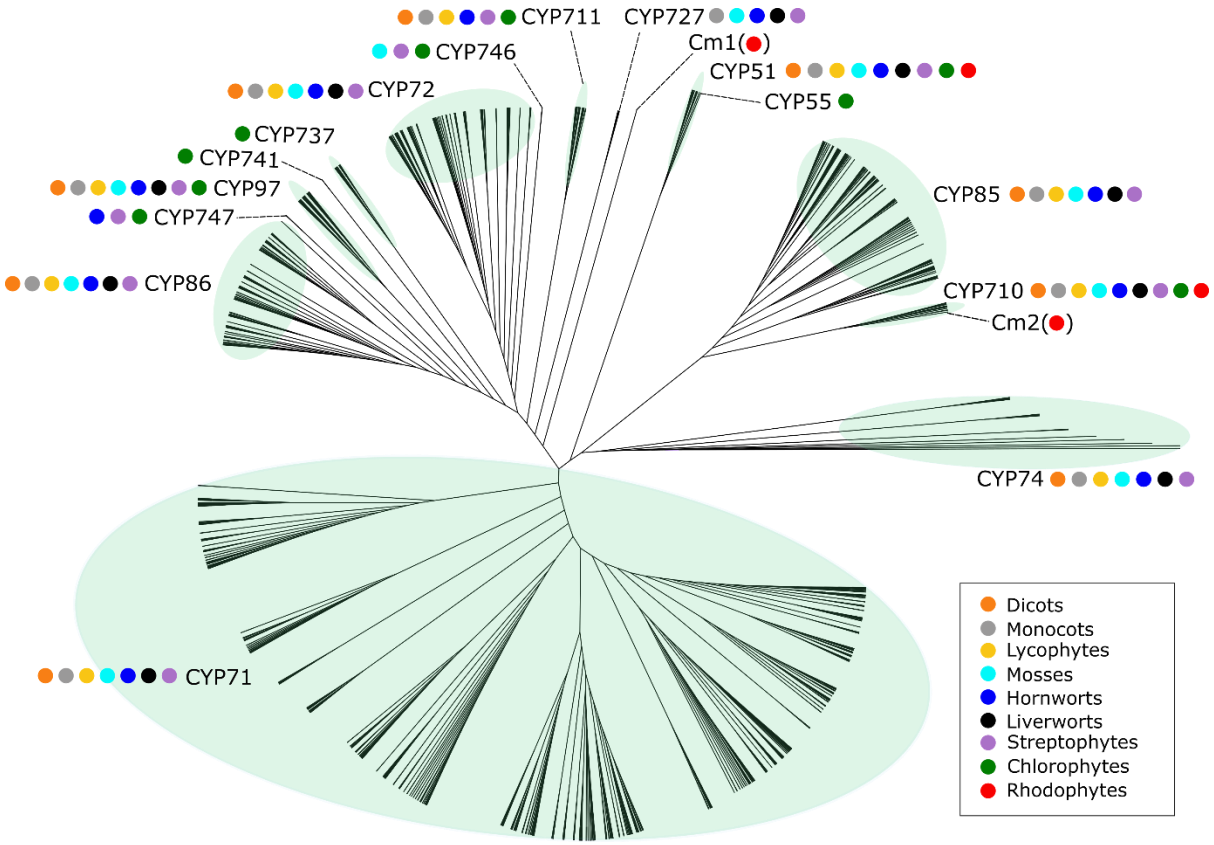


Figure 2.2. Cytochrome P450s have diverged into 18 clans during the evolutionary history of the Archaeplastida. Unrooted cladogram of a maximum likelihood (ML) analysis conducted by PHyML 3.0 (Guindon et al., 2010) using an estimated gamma distribution parameter, the LG+G+F model of amino acid substitution and a Chi^2-based approximate likelihood ratio (aLRT) test. Protein sequences were aligned using MAFFT with the L-INS-i algorithm. Clans are indicated by green circles and numbers. Coloured dots indicate the presence of sequences from different species in each clan. *A. thaliana* (orange); *O. sativa ssp. japonica* (grey); *Selaginella moellendorffii* (yellow); *Physcomytrella patens* (cyan); *Anthoceros agrestis* (blue); *Marchantia polymorpha* (black); *Klebsormidium nitens* (purple); *Chlamydomonas reinhardtii* (green); *Cyanidioschyzon merolae* (red).

2.4.2. *Klebsormidium nitens* encodes 29 cytochrome P450s

Klebsormidium nitens is a streptophyte alga and represents a key evolutionary step in the transition of plants from an aquatic to a terrestrial environment. To determine the number of CYP sequences in the genome of *K. nitens*, a BlastP search was performed on the *K. nitens* genome using known members from all Archaeplastida CYP clans. All sequences above a $1e^{-10}$ E-value threshold were retained. This initial list of sequences was self-blasted to identify any species-specific CYPs. The sequences were further checked for the presence of the cytochrome p450 heme-binding domain using PFAM (PF00067). The sequences were aligned in MAFFT v.7 (Katoh & Standley, 2013) and any sequences missing key conserved domains (Figure 2.5B) were labelled as pseudogenes and removed.

29 full length CYP proteins and 4 pseudogenes were identified in the genome of *Klebsormidium nitens* (Figure 2.3, Supplementary Table S2.1). Fewer CYP proteins were identified in the genome of *K. nitens* than the chlorophyte species *C. reinhardtii,* which encodes 40 CYP proteins in its genome. This suggests that since the divergence of the chlorophyte and streptophyte algae lineages there were either losses in the CYP protein family in the streptophyte algae lineage, or expansions in the chlorophyte lineage. *K. nitens* encodes a much smaller number of CYP proteins than the land plant species in this study. Three times more CYP proteins (70) were found in the genome of the moss *P. patens*, and nearly five times more (144) in the genome of the hornwort *A. agrestis*. This suggests that the number of CYP proteins rapidly increased after the divergence of the streptophyte algae and land plants from their common ancestor.

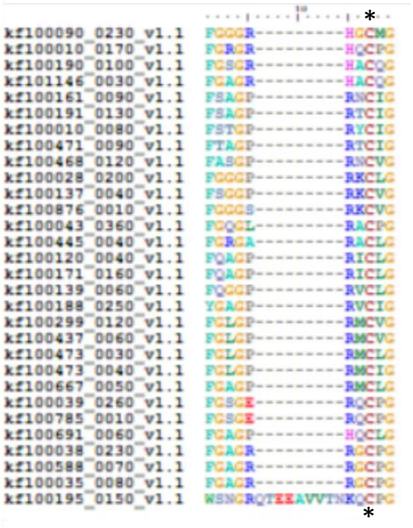


Figure 2.3. Amino acid alignment of the heme-binding domain of the 29 *Klebsormidium nitens* CYPs. The absolutely conserved cysteine which acts as ligand to the heme iron is marked with an asterisk.

2.4.3. *Anthoceros agrestis* encodes 144 cytochrome P450s

The hornwort. *A. agrestis* is a member of the earliest diverging lineage of extant land plants, the Bryophyte clade, which also contains mosses and liverworts. The hornwort lineage has been recently resolved as sister to the setaphytes (mosses and liverworts) (Morris et al., 2018). The inclusion of a hornwort species in the phylogeny of Archaeplastida CYP protein family provides a more complete picture of the evolution and distribution of CYP proteins in the land plants. *Anthoceros agrestis* has been shown to have a unique genome composition compared to mosses and liverworts. Despite a small genome (117 to 133 Mb), the proportion of transposable and repetitive elements (36-38%) is comparable to other plant species (Li et al., 2020). To determine the number of CYP sequences in the genome of *Anthoceros agrestis*, a BlastP search was performed on the *A. agrestis* genome using known members from all Archaeplastida CYP clans. All sequences above a $1e^{-10}$ E-value threshold were retained. This initial list of sequences was self-blasted to identify any species-specific CYP sequences. The sequences were further checked for presence of the cytochrome p450 heme-binding domain using PFAM (PF00067). The sequences were aligned in MAFFT v.7 (Katoh & Standley, 2013) and any sequences missing key conserved domains (Figure 2.5B) were labelled as pseudogenes and removed.

A total of 144 CYP proteins and 109 pseudogenes were identified in the genome of *Anthoceros agrestis* (Figure 2.4, Supplementary Table S2.1), the highest number of the three bryophyte species included. There are 115 and 70 CYP proteins in the genomes of *M. polymorpha* and *P. patens*, respectively. This indicates that the number of CYP proteins increased during the evolution of the hornwort lineage after the divergence of hornworts and setaphytes (mosses and liverworts). The genomes of *A. agrestis*, *M. polymorpha* and *P. patens* encode fewer CYP proteins than the vascular plants. There are 202, 295 and 242 CYPs in the genomes of *S. moellendorffii*, *O. sativa* and *A. thaliana*, respectively. This suggests that the number of CYP proteins increased during land plant evolution in association with the radiation of land plants.

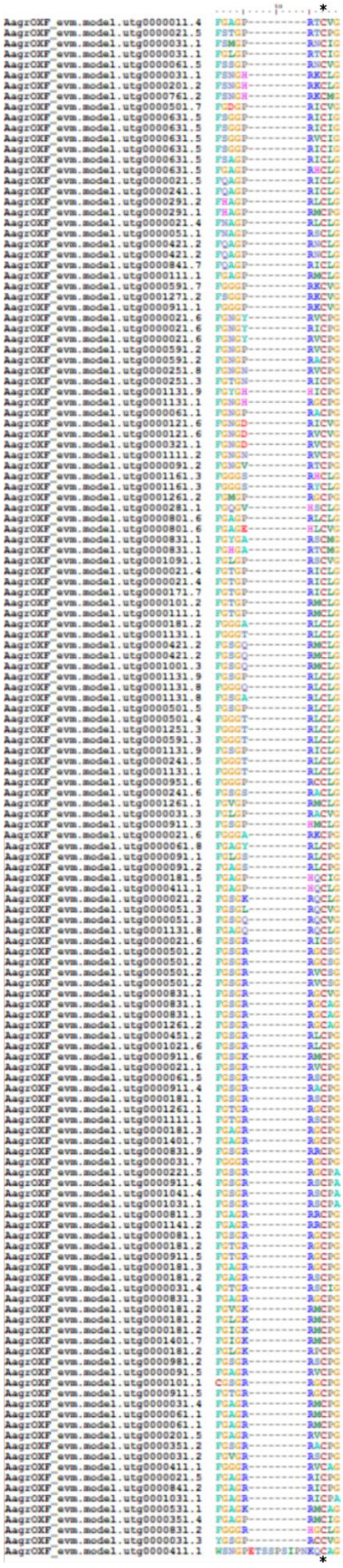


Figure 2.4. Amino acid alignment of the heme-binding domain of the 144 *Anthoceros agrestis* CYP proteins. The absolutely conserved cysteine which acts as ligand to the heme iron is marked with an asterisk.

2.4.4. CYP71, CYP72, CYP85 and CYP86 clans expanded in land plants

The divergence of land plants from streptophyte algae is correlated with an increase in the percent of total CYP coding genes in lineages on land than in the charophyte lineage. CYP coding proteins represent 0.17% of total genes in the streptophyte alga genome, 0.2-0.6% of total genes in bryophytes, 0.9% of total genes in lycophytes and 0.9% of total genes in angiosperm genomes (Supplementary Table S2). The number of genes in the CYP71, CYP72, CYP85 and CYP86 clans increased considerably after plants colonised land, and then again after the evolution of vascular plants (Figure 2.5B). The number of CYPs in the CYP71 clan increased more than any other clan in land plants after land colonisation and evolution of vascular plants. There are 3 CYP71 clan members in the genome of the streptophyte algae *K. nitens.* Among the early diverging land plants there are 59 CYP71 members in the hornwort *A. agrestis*, 68 in the liverwort *M. polymorpha* and 38 in the moss *P. patens*. Among the vascular plants there are 101 in the lycophyte *S. moellendorffii*, 150 in *A. thaliana* and 164 in *O. sativa* (Supplementary Table S2.1). CYP71 clan proteins represent 10% of all CYPs in *K. nitens* and 40-60% of all CYPs in the land plants.

The number of CYP72 clan genes in the genomes of land plants also increased in two phases, first after the colonisation of land and second with the evolution of vascular plants. There are 5 CYP72 clan members in the genome of *K. nitens.* Among the early diverging land plants there are 13 in *A. agrestis*, 8 in *M. polymorpha* and 6 in *P. patens*. Among the vascular plants there are 37 in the lycophyte *S. moellendorffii,* 34 in *O. sativa* and 19 in *A. thaliana* (Figure 2.5B, Supplementary Table S2.1). CYP85 clan also expanded among land plants, particularly in hornworts. There are two CYP85 clan members in the genome of *K. nitens*, 19 in the liverwort *M. polymorpha*, 4 in the moss *P. patens*, but 53 in the hornwort *A. agrestis*. Among the vascular plants there are 32 in the lycophyte *S. moellendorffii*, 28 in *A. thaliana* and 30 in rice. There are 53 CYP85 clan members and 59 CYP71 clan members in *A. agrestis*, thus *A. agrestis* is the only species in which the two largest clans have a similar number of genes. Finally, the number of genes in the CYP86 clan of land plant species is also higher than in algae. There are 3 CYP86 clan genes in *K. nitens*, 9 in *A. agrestis*, 11 in *M. polymorpha* and 11 in *P. patens*. Among the vascular plants there are 14 CYP86 clan genes in *S. moellendorffii*, 33 in *A. thaliana* and 41 in rice (Figure 2.5B, Supplementary Table S2.1). The increase in the number of genes in the CYP86 clan occurred after the divergence of *S. moellendorffii* and *A. thaliana* from a last common ancestor, unlike the CYP85 clan where the there was an increase in clan size per genome preceding the divergence of the *S. moellendorffii* and *A. thaliana.*

CYP members belonging to 8 clans - CYP51, CYP74, CYP97, CYP710, CYP711, CYP727, CYP746 and CYP74 - were present in low numbers in the genomes of all streptophyte species. Generally, there were fewer than 10 CYP members from any one of these clans in any one streptophyte species (Supplementary Table S2.1). Clans that were present in the genome of the streptophyte alga *K. nitens* and land plants did not have representatives in some land plant species, suggesting that these clans were lost in these species or lineages (Figure 2.5A). CYP727 was missing in both the lycophyte *S.*

moellendorffii and the dicot *A. thaliana* but present in bryophytes, and rice. CYP711 was present in streptophyte algae but absent from liverworts and mosses, suggesting that it was lost in the common ancestor of liverworts and mosses but conserved in the hornwort and vascular plants. CYP746 was present in streptophyte algae and the moss *P. patens* but absent from hornworts, liverworts and vascular plants. This suggests that this clan was lost independently in both hornwort and liverwort lineages, and also in the common ancestor of lycophytes and angiosperms. CYP747 was present in the hornwort genome but was absent from the lycophyte and angiosperm genomes. This suggests the CYP747 clan was retained in the hornwort lineage but lost in the common ancestor of liverwort and mosses and also in the common ancestor of angiosperms and lycophytes (Figure 2.3A).

In summary, the phylogenetic analysis shows that the more than four-fold increase in the percent of total CYP coding genes in land plants (from less than 0.2% in the streptophyte *K. nitens* to 0.9% lycophyte and angiosperm genomes) is caused by an increase in the number of genes in the CYP71, CYP72, CYP85 and CYP86 clans in land plant genomes. The expansion of these four clans suggests that these clans are under positive selection to diversify and evolve new functions. The lack of expansion of gene numbers in the 8 other single-family clans suggests that these clans are under purifying selection in all species, and are involved in essential functions in all plants. Some clans were lost during land plant evolution in certain lineages. These clans may contain genes that specialised into functions that are no longer required in these lineages.

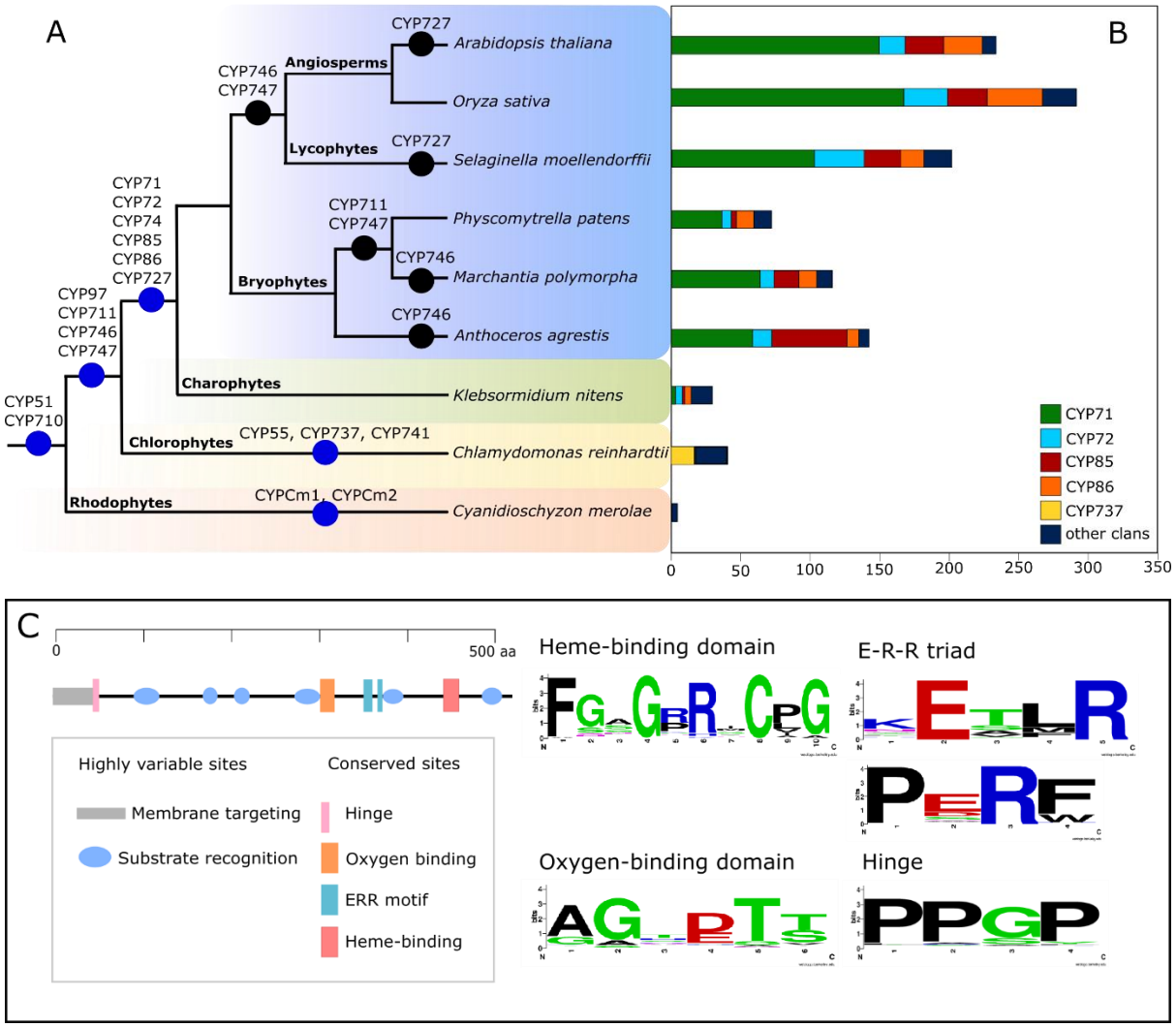


Figure 2.5. Overview of cytochrome P450 clan evolution and protein features in plants. A. Cladogram of Archaeplastida phylogeny showing CYP clan origins and losses in plants. Blue circles represent first appearance of a clan, black circles represent the absence of a clan in a particular lineage. **B.** Number of CYP proteins in each species showing expansions in four clans in the land plants (not including pseudogenes). **C.** Diagram of a typical CYP protein showing recognisable amino acid sites. Adapted from Werck-Reichhart & Feyereisen, 2000. Weblogos of the four conserved CYP amino acid motifs in plants are shown.

2.4.5. Two of the four diversified land plant CYP clans contain proteins with functions in herbicide metabolism or resistance

CYP enzymes metabolically modify a broad range of xenobiotic chemicals. The extensive use of herbicides has resulted in the natural selection of weeds that are resistant to herbicides (Balazs Siminszky, 2006; Danièle Werck-Reichhart et al., 2000; Yu & Powles, 2014a). CYP-mediated herbicide resistance results from the increased expression of CYP genes or increased CYP activity in resistant weed populations (Batard et al., 1998; Gaines et al., 2014). To identify the clans containing herbicide metabolising CYPs, a literature search was performed to identify CYPs that have been experimentally shown to metabolise or confer resistance to one or multiple herbicides in plants (Table 2.1) These studies used either sensitivity assays using transgenic plants or in vitro metabolism assays to demonstrate the involvement of the enzyme(s) in herbicide metabolism or resistance. A large number of plant CYPs were up-regulated in NTSR weeds but weren't experimentally tested for their involvement in herbicide metabolism or resistance. These genes were called candidate NTSR genes and are listed in Supplementary Table S2.3.

A total of 24 plant CYP proteins have so far been experimentally shown to metabolise or confer resistance to one or more herbicides using sensitivity or metabolism assays (Table 2.1). The CYP proteins were identified in the model species thale cress (*A. thaliana*), the grass weeds barnyard grass (*Echinochloa phyllopogon*) and annual ryegrass (*Lolium rigidum*) the gymnosperm western red cedar (*Thuja plicata*), and the crops barley (*Hordeum vulgare*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), maize (*Zea mays*), cotton (*Gossypium hirsutum*), soybean (*Glycine max*), ginseng (*Panax ginseng*), jerusalem artichoke (*Helianthus tuberosus*) and tobacco (*Nicotiana tabacum*). The 24 proteins metabolised or conferred resistance to diverse herbicide chemotypes including aryloxyphenoxypropionates (FOP), benzothiadiazinones, clomazone and pelargonic acid (unclassified), cyclohexanediones (DIM), isoxazolidinones, phenylpyrazolins (DEN), phenylurea, pyrazoles, pyridazinones, pyrimidinyl thiobenzoates, sulfonylaminocarbonyl-triazolinones, sulfonylurea, thiadiazines, triazolopyrimidine and triketones, either individually or in mixtures. The majority (19 of 24) of CYPs metabolised phenylureas or sulfonylureas. All herbicide-metabolising enzymes belong to two of the multiple-family CYP clans in land plants: CYP71 and CYP72. 21 CYP71 clan members from 12 different species conferred resistance to phenylureas, sulfonylureas, isoxazolidinones, thiadiazines, FOP and DIM herbicide chemicals. There are many CYP71 clan members encoded in the genomes of all land species. There are 150 members of CYP71 clan in *A. thaliana* and 164 members of CYP71 clan in *O. sativa*. These data indicate that the majority of genes conferring resistance to herbicides belong to the CYP71 clan and metabolise herbicides from mainly the phenylurea and sulfonylurea classes. Three CYP72 clan members from three species conferred resistance to thiobenzoates, pelargonic acid or sulfonylureas. There are fewer CYP72 members than CYP71 in the genomes of land plants. There are 19 members of CYP72 in *A. thaliana* and 34 members of CYP72 in *O. sativa*. These data indicate that the CYP proteins that have been found to be implicated in herbicide

metabolism so far belong to the CYP71 and CYP72 classes, which represent two of the four expanded CYP clans in land plants.

There are twice as many genes in CYP71 and CYP72 clans combined than in all other clans in all land plant genomes in this study except *A. agrestis* (Figure 2.5A). For example, in *A. thaliana*, there are 169 genes in CYP71 and CYP72 clans and 73 genes in all other clans. In *O. sativa*, there are 198 genes in CYP71 and CYP72 clans and 97 CYPs in all other clans. In other land plants this proportion is even larger. Li & Wei (2020) identified 233 genes in CYP71 and CYP72 clans and 30 in other clans in maize, and 933 genes in CYP71 and CYP72 clans and 352 in other clans in wheat. This suggests that herbicide metabolising CYPs are more likely to belong to the CYP71 or CYP72 clans than to the other clans, simply because there are more genes in these clans. Therefore we cannot reject the hypothesis that CYPs in all clans can metabolise herbicide chemicals.

Eight members of the in the CYP81A subfamily of CYP71 clan confer herbicide resistance in five species: rice, maize, barley, barnyard grass and annual ryegrass (Guo et al., 2019; Han et al., 2021; Iwakami et al., 2014, 2019; Zhang et al., 2006). This is more than any other subfamily in this or any other clan (Table 2.1). CYP81A enzymes have been shown to catalyse hydroxylations and N-/O- demethylations of herbicide substrates (Dimaano et al., 2020) but the *in vivo* function of CYP81A enzymes is not known. The CYP81A enzymes metabolise herbicides from five chemical classes, more than any other CYP subfamily. This suggests that the increase in catalytic diversity that was acquired during the evolution of the CYP81A subfamily pre-adapted plants to evolve herbicide resistance when subjected to extreme selection.

Table 2.1. Plant CYPs that metabolise or confer resistance to herbicides are found within the CYP71 and CYP72 clans. Table adapted from Dimaano & Iwakami (2021).

Clan	Sub-family	Gene name and species	Evidence	Herbicide class	Genes per clan	References
71	71A	CYP71A11 (tobacco)	Metabolism in yeast	Phenylurea	26	Yamada et al., 2000
	71A	CYP71A12 (thale cress)	Metabolism in yeast	Pyrazole		Hayashi et al., 2007
	71A	CYP71A10 (soybean)	Transformation in tobacco	Phenylurea		Siminszky et al., 1999
	71C	CYP71C6v1 (wheat)	Metabolism in yeast	Sulfonylurea		Xiang et al., 2005
	73A	CYP73A1 (Jerusalem artichoke)	Metabolism in yeast	Phenylurea		Pierrel et al., 1994
	76A	CYP76AA20 (western redcedar)	Metabolism in yeast	Phenylurea		Gesell et al., 2015
	76A	CYP76AA21 (western redcedar)	Metabolism in yeast	Phenylurea		Gesell et al., 2015
	76A	CYP76AA22 (western redcedar)	Metabolism in yeast	Phenylurea		Gesell et al., 2015
	76A	CYP76AA25 (western redcedar)	Metabolism in yeast	Phenylurea		Gesell et al., 2015
	76B	CYP76B1 (Jerusalem artichoke)	Metabolism in yeast	Phenylurea		Robineau et al., 1998
	76C	CYP76C1 (thale cress)	Metabolism in yeast	Phenylurea		Höfer et al., 2014
	76C	CYP76C2 (thale cress)	Metabolism in yeast	Phenylurea		Höfer et al., 2014
	76C	CYP76C4 (thale cress)	Metabolism in yeast	Phenylurea		Höfer et al., 2014
	81A	**CYP81A6 (rice)**	Knock-out in rice	Thiadiazine, Sulfonylurea		Zhang et al., 2006
	81A	**CYP81A9 (maize)**	Gene-silencing in maize	Sulfonylurea		Li et al., 2013
	81A	**CYP81A12 (barnyard grass)**	Transformation in *A. thaliana, E. coli* and yeast	Benzothiadiazinone, DEN, DIM, Pyrimidinyl(thio)benzoate, Sulfonylaminocarbonyl-triazolinone Sulfonylurea, Triazolopyrimidine, Triketone		Dimaano et al., 2020, Iwakami et al., 2014
	81A	**CYP81A14 (barnyard grass)**	Transformation in *A. thaliana, E. coli* and yeast	Pyrimidinyl(thio)benzoate, Sulfonylaminocarbonyl-triazolinone, Sulfonylurea		Dimaano et al., 2020
	81A	**CYP81A15 (barnyard grass)**	Transformation in *A. thaliana, E. coli,* yeast, and rice	Benzothiadiazinone, Clomazone, DEN, DIM, Pyrimidinyl(thio)benzoate, Sulfonylurea		Dimaano et al., 2020, Guo et al., 2019
	81A	**CYP81A18 (barnyard grass)**	Transformation in *A. thaliana E. coli* and yeast	Pyrimidinyl(thio)benzoate, Sulfonylaminocarbonyl-triazolinone, Sulfonylurea		Dimaano et al., 2020
	81A	**CYP81A21 (barnyard grass)**	Transformation in *A. thaliana E. coli* and yeast	Benzothiadiazinone, Clomazone, Pyrimidinyl(thio)benzoate Sulfonylaminocarbonyl-triazolinone, Sulfonylurea, Triazolopyrimidine		Dimaano et al., 2020, Iwakami et al., 2014
	81A	**CYP81A24 (barnyard grass)**	Transformation in *A. thaliana, E. coli,* yeast, and rice	Benzothiadiazinone, Clomazone, Pyrazole Pyridazinone, Pyrimidinyl(thio)benzoate, Sulfonylurea, Triazolopyrimidine, Triketone		Dimaano et al., 2020, Guo et al., 2019
	81A	**CYP81A10v7 (annual ryegrass)**	Transformation in rice	DIM, FOP, Sulfonylurea		Han et al., 2021
	81A	**CYP81A63 (barley)**	Transformation in rice	DEN, DIM, FOP		Iwakami et al., 2019
	81B	CYP81B2 (tobacco)	Metabolism in tobacco cells	Phenylurea		Yamada et al., 2000
	81E	CYP81E22 (soybean)	Transformation in soybean	Benzothiadiazinone		Kato et al., 2020
	736A	CYP736A12 (ginseng)	Over-expression in *A. thaliana*	Phenylurea		Khanom et al., 2019
72	72A	CYP72A31 (rice)	Over-expression in *A. thaliana*	Pyrimidinyl(thio)benzoate	3	Saika et al., 2014
	72A	CYP72A18 (rice)	Metabolism in yeast microsomes	Pelargonic acid		Imaishi & Matumoto, 2007
	749A	CYP749A16 (cotton)	Gene-silencing in cotton	Sulfonylurea		Thyssen et al., 2018

2.5. Discussion

The phylogenetic analysis of the cytochrome P450 superfamily in the Archaeplastida revealed all land plant CYP clans originated in the charophyte lineage, before the evolution of land plants. No new clans evolved after plants colonised land. The evolution of CYPs in land plants was instead associated with an increase in the number of CYPs within some clans, particularly the CYP71 clan. A small number of clans were lost in some plant lineages. Members of the diversified land plant clans have gained the ability to metabolise herbicides from different chemical classes.

The transition from the aquatic environment to the terrestrial environment was accompanied by the radiation of plant phenotypic diversity in the Phanerozoic (Bateman et al., 1998; Morris et al., 2018; Wickett et al., 2014). This morphological radiation did not coincide with an increase in the number of CYP clans. All land plant clans evolved before the divergence of the streptophyte algae and land plants from their last common ancestor. Instead, the evolution of land plants was associated with an increase in the number of CYPs in the CYP71, CYP72, CYP85 and CYP86 clans. This is consistent with the observation during the sequencing of the *K. nitens* genome that the numbers of genes in gene families with functions in hormone signalling and environmental stress responses are greater in land plants than in *K. nitens* (Hori et al., 2014). Most land plant clans (CYP71, CYP72, CYP74, CYP85, CYP86 and CYP727) evolved in the streptophyte algal ancestor of land plants, placing their origin over 629 million years ago (Morris et al., 2018), and four others (CYP97, CYP711, CYP746 and CYP747) evolved in the common ancestor of Viridiplantae, placing their origin over 670 million years ago (Morris et al., 2018).

CYP51 and CYP710 clans were the only land plant clans encoded in the red algae genome, making them the most ancient land plant clans. Phylogenetic analyses of CYPs across phyla have noted that CYP51 is the only clan that is conserved in plants, fungi, bacteria and animals and is thought to be the original CYP from which all other CYPs evolved (Nelson, 1998). CYP51 and CYP710 are single family clans that function in the synthesis of sterols which form the cell membranes of eukaryotes (Kim et al., 2005; Morikawa et al., 2006). These enzyme families are therefore thought to have appeared in early eukaryote evolution (Nelson, 2018).

The inclusion of *A. agrestis* and *K. nitens* sequences in the analysis enabled the identification of a previously unrecognised CYP clan, CYP747, which contained a single member from *A. agrestis*, *K. nitens* and *C. reinhardtii*. CYP747 therefore appears to have originated in the common ancestor of chlorophyte algae and streptophytes, and was retained in the hornwort lineage but was lost in the common ancestor of liverworts and mosses and also in the common ancestor of angiosperms and lycophytes. The function of these proteins is not known. CYPs in the same branch or family usually have similar or related substrates (Nelson & Werck-Reichhart, 2011) therefore it is likely that CYP747 genes share the same function in the three species. CYP747 enzymes presumably function in biochemical pathways unique to algae and

hornworts but not found in other land plants. One unique feature of *A. agrestis* is the presence of a pyrenoid based carbon concentrating mechanism which also exists in algae (Li et al., 2020). A possible function of CYP747 enzymes may be related to this mechanism.

In land plants, there are many more genes in the CYP71, CYP72, CYP85 and CYP86 clans than in other clans. There are multiple families in each of these clans which perform functions in both general and specialised metabolism (Bak et al., 2011). The CYP71 clan contains half of all plant CYPs. Over 40% of CYPs in the hornwort genome and over 50% of CYPs in the moss genome are members of the CYP71 clan, indicating that this clan was already dominant in early land plants. The eight other streptophyte CYP clans - CYP51, CYP74, CYP97, CYP710, CYP711, CYP727, CYP746 and CYP747 - are much smaller in land plants. These clans are represented in all plant species in this study in low numbers, suggesting that they play critical functions and are under purifying selection. This is consistent with the observation that the ratios of non-synonymous to synonymous substitutions in rice single family clans average at (dN/dS) 0.35 which indicates strong purifying selection (Wei & Chen, 2018). By contrast the dN/dS ratios for several genes within multi-family clans was >1, indicating positive selection. Enzymes in these clans are involved in the synthesis of sterols (CYP51, CYP710), jasmonic acid (CYP74), strigolactones (CYP711) and carotenoids (CYP97). The functions associated with these clans are conserved in all land plants (Mizutani & Ohta, 2010).

CYPs play an important role in the metabolic detoxification of herbicides, and a literature search found that all named plant CYPs that confer herbicide resistance are members of 3 of the 4 expanded land plant clans. Enzymes belonging to 5 families in the CYP71 clan and 2 families in the CYP72 clan are implicated in herbicide metabolism. This demonstrates that the ability to metabolise herbicides is not restricted to a single clan or family. The CYP81A subfamily which is part of the CYP71 clan contains the largest number of herbicide resistance conferring genes identified in the literature. CYP81A enzymes modify herbicides from numerous chemical classes (Guo et al., 2019; Han et al., 2021; Iwakami et al., 2014; Zhang et al., 2006) which demonstrates a large catalytic diversity in this subfamily. Of the species in this analysis CYP81A sequences are only found in rice. The endogenous function of CYP81A enzymes in plants is not yet known. These enzymes represent priority targets for the identification of candidate herbicide metabolising enzymes in other species.

The attributes needed for a CYP to be able to metabolise a herbicide are still not known. It is not clear whether the larger clans contain specialised enzymes that can chemically modify xenobiotics or whether these clans simply contain more genes, therefore the likelihood of identifying a herbicide metabolising CYP belonging to these clans is greater. Herbicide-metabolising enzymes may exist in the smaller clans but haven't yet been identified. Another possibility is these enzymes are involved in pathways that are more likely to become co-opted for herbicide resistance, such as stress

response pathways. Finally, enzymes from the diversified clans may have evolved more variable substrate recognition sites which allow access of a wider range of substrates (Supplementary Figure S2.1).

As more plant and algal genomes are sequenced, we will be able to view the evolution of large protein families at a higher resolution. New clans, families and subfamilies are likely to emerge from continued in depth analyses of plant CYPs. One area that is still lacking information is in the green algae CYP phylogeny. CYPs in the green alga *Chlamydomonas reinhardtii* are highly divergent in their sequence and share low relatedness to other plant CYPs. The sequencing of more green algal genomes and identification of algal CYP sequences will help to classify these proteins into new clans. The early evolution of the CYP51 and CYP70 clans is still unclear and would benefit from an in-depth phylogenetic analysis including species from plants, animals, and fungi, and was not the scope of this chapter.

In conclusion Archaeplastida CYP clans are more ancient than previously thought (Nelson & Werck-Reichhart, 2011), having evolved in algae before plants emerged onto land. Genes that have the potential to confer NTSR may have arisen by two ways. One scenario is expansions of certain clans and neo-functionalisation of new genes in land plants conferred the ability of members of these clans to detoxify herbicide chemicals. The other scenario is that there is an equal possibility that a CYP from any clan can metabolise herbicides. In the next chapter I investigate the phylogeny of glutathione s-transferases in Archaeplastida in the context of herbicide metabolic resistance.

2.6. Appendix

Table S2.1. Cytochrome P450 clans and gene numbers in green plants and red algae

Species	CYP51	CYP55	CYP71	CYP72	CYP74	CYP85	CYP86	CYP97	CYP710	CYP711	CYP727	CYP737	CYP741	CYP746	CYP747	Cm1	Cm2	Total
At	1	0	150	19	2	28	33	3	4	1	0	0	0	0	0	0	0	242
Os	8	0	164	34	5	30	41	3	4	5	1	0	0	0	0	0	0	295
Sm	1	0	101	37	12	32	14	3	1	1	0	0	0	0	0	0	0	202
Aa	1	0	59	13	1	53	9	3	2	1	1	0	0	0	1	0	0	144
Pp	1	0	39	6	3	4	11	3	2	0	1	0	0	1	0	0	0	70
Mp	1	0	68	8	2	19	11	3	2	0	1	0	0	0	0	0	0	115
Kn	3	0	3	5	1	2	3	3	1	4	3	0	0	1	1	0	0	29
Cr	1	1	0	0	0	0	0	4	1	13	0	16	2	1	1	0	0	40
Cm	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1	5

Numbers of CYP proteins in each clan, excluding pseudogenes. *At, Arabidopsis thaliana; Os, Oryza sativa; Sm, Selaginella moellendorffii; Aa, Anthoceros agrestis; Pp, Physcomitrella patens; Mp, Marchantia polymorpha; Kn, Klebsormidium nitens; Cr, Chlamydomonas reinhardtii; Cm, Cyanidioschyzon merolae*.

Table S2.2. CYP genes as a proportion of total genes in different plant lineages.

Species	Total CYPs	Protein-coding genes in genome	CYPs as a % of total genes	References
Arabidopsis thaliana	242 (18)	25,498	0.94	Kaul et al., 2000
Oryza sativa	300 (62)	35,681	0.84	Kawahara et al., 2013
Selaginella moellendorffii	202 (76)	22,285	0.91	Banks et al., 2011
Anthoceros agrestis	144 (109)	24,700	0.58	Li et al., 2020
Physcomitrella patens	71 (20)	35,938	0.19	Rensing et al., 2008
Marchantia polymorpha	115 (31)	19,138	0.60	Bowman et al., 2018
Klebsormidium nitens	29 (4)	16,215	0.17	Hori et al., 2014
Chlamydomonas reinhardtii	39 (2)	15,143	0.26	Merchant et al. 2007
Cyanidioschyzon merolae	5	5,331	0.1	Matsuzaki et al., 2004

Note: pseudogenes are in parentheses and not included in calculations.

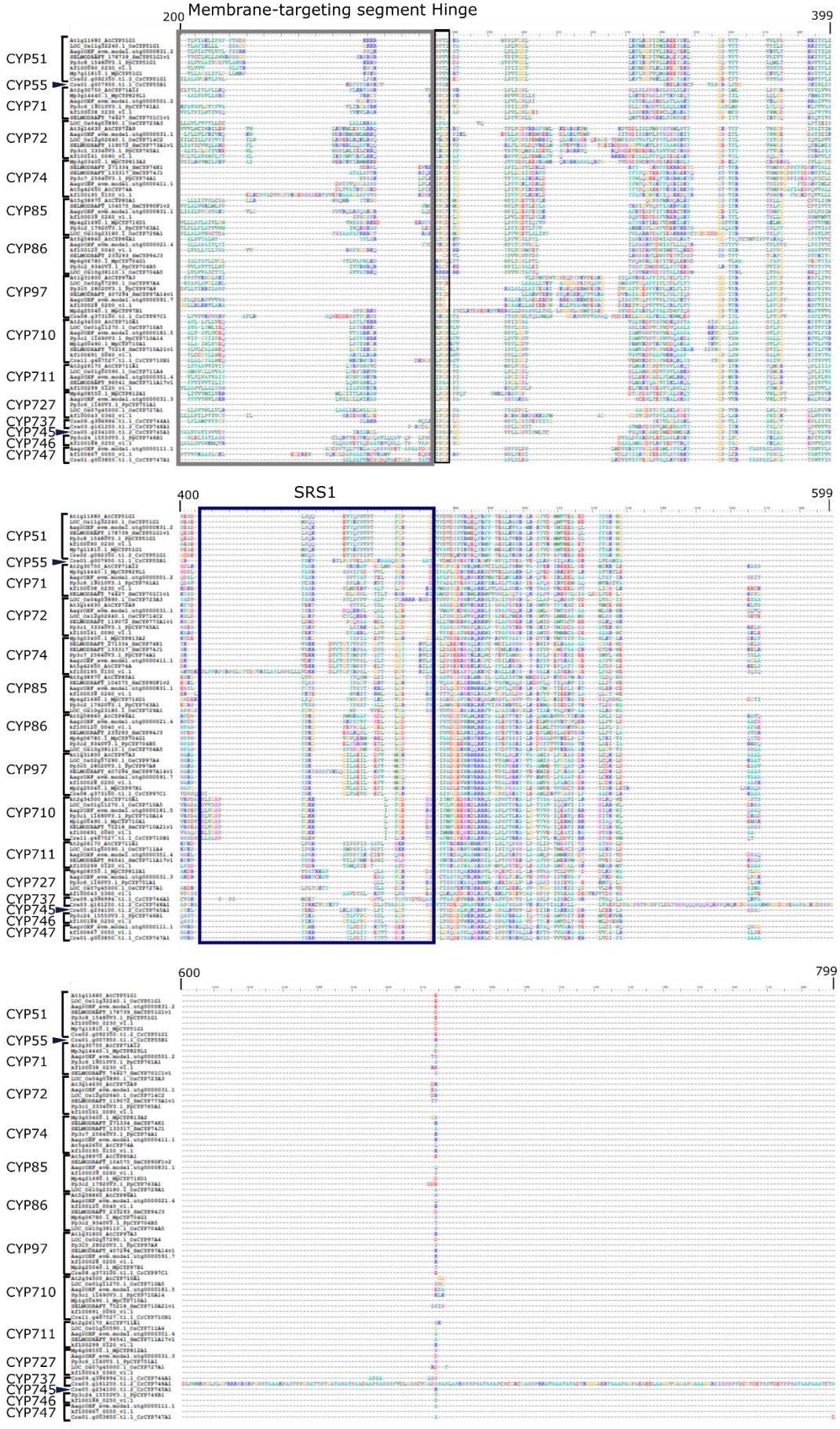
Table S2.3. Candidate NTSR CYPs are found within clans 71, 72 and 86 and represent candidate herbicide metabolising enzymes.

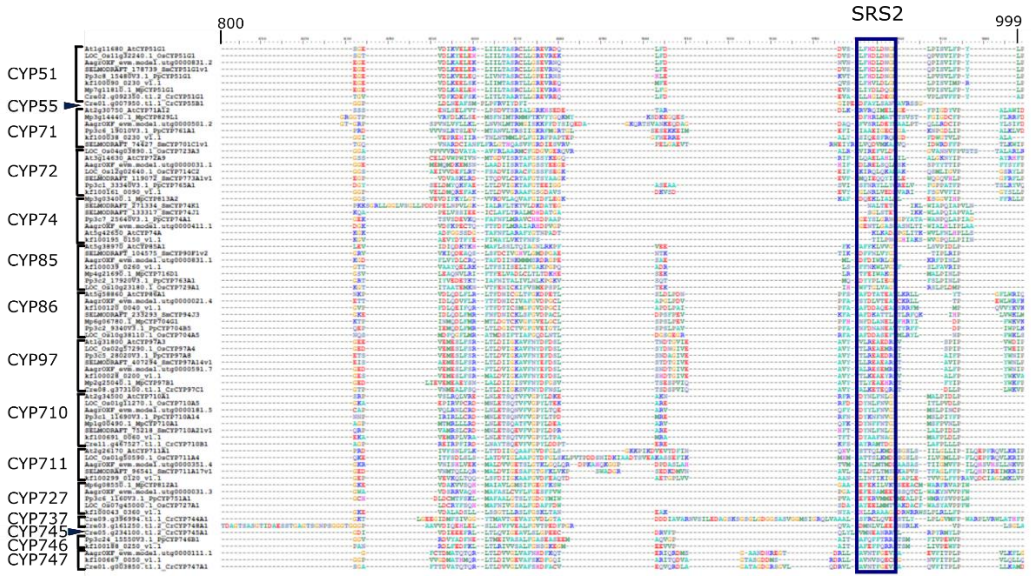
Clan	Sub-family	Gene name	Evidence	Herbicide chemical class	genes per clan	References
71	71A	AmCYP71A (blackgrass)	up-regulated in NTSR plants compared to sensitive plants	DIM, dinitroalanines, FOP, sulfonylurea	27	(Gardin et al., 2015)
	71A	CbCYP71A26 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	71B	AmCYP71B3 (blackgrass)	up-regulated in NTSR plants compared to sensitive plants	DIM, dinitroalanines, FOP, sulfonylurea		(Gardin et al., 2015)
	71B	CbCYP71B7 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	71B	CbCYP72B7 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	71C	ZmCYP71C2 (maize)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Liu et al., 2018)
	71C	ZmCYP71C3v2 (maize)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Liu et al., 2018)
	71C	ZmCYP71C4 (maize)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Liu et al., 2018)
	71D	BsCYP71D7 (American slough grass)	up-regulated in NTSR plants compared to sensitive plants	FOP		(Bai et al., 2020)
	71D	BsCYP71D10 (American slough grass)	up-regulated in NTSR plants compared to sensitive plants	FOP		(Bai et al., 2020)
	71K	EpCYP71K2 (barnyard grass)	up-regulated in NTSR plants compared to sensitive plants	Pyrimidinylthiobenzoic acid		(Iwakami et al., 2014)
	76C	Ma76C1 (water chickweed)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Liu et al., 2018)
	76C	CbCYP76C2 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	78A	CbCYP78A-1 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	78A	CbCYP78A-2 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	81A	AmCYP81A2 (blackgrass)	up-regulated in NTSR plants compared to sensitive plants	DIM, dinitroalanines, FOP, sulfonylurea		(Franco-Ortega et al., 2021)
	81A	AmCYP81A3 (blackgrass)	up-regulated in NTSR plants compared to sensitive plants	DIM, dinitroalanines, FOP, sulfonylurea		(Franco-Ortega et al., 2021)
	81A	AmCYP81A4 (blackgrass)	up-regulated in NTSR plants compared to sensitive plants	DIM, dinitroalanines, FOP, sulfonylurea		(Franco-Ortega et al., 2021)
	81A	ZmCYP81A9 (maize)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Liu et al., 2018)
	81B	LrCYP81B1 (annual ryegrass)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Duhoux et al., 2017)
	81D	AmCYP81D (blackgrass)	up-regulated in NTSR plants compared to sensitive plants	DIM, dinitroalanines, FOP, sulfonylurea		(Gardin et al., 2015)
	82A	CbCYP82A3 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	82A	82A3-like (purple morning glory)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Leslie & Baucom, 2014)
	82A	CbCYP82A4 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	92A	ZmCYP92A1 (maize)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Liu et al., 2018)
	93A	CbCYP93A3 (Hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	99A	BsCYP99A2 (American slough grass)	up-regulated in NTSR plants compared to sensitive plants	FOP		(Bai et al., 2020)
72	72A	LrCYP72A1 (annual ryegrass)	up-regulated in NTSR plants compared to sensitive plants	two sulfonylurea	10	(Duhoux et al., 2017)
	72A	LrCYP72A2 (annual ryegrass)	up-regulated in NTSR plants compared to sensitive plants	two sulfonylurea		(Duhoux et al., 2017)
	72A	LrCYP72A1 (annual ryegrass)	up-regulated in NTSR plants compared to sensitive plants	FOP		(Gaines et al., 2014)
	72A	LrCYP72A2 (annual ryegrass)	up-regulated in NTSR plants compared to sensitive plants	FOP		(Gaines et al., 2014)
	72A	ZmCYP72A5 (maize)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Liu et al., 2018)
	72A	ApCYP72A219 (palmers amaranth)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Salas-Perez et al., 2018)
	72A	BhCYP72A14 (hybridous Mediterranean false brome)	up-regulated in NTSR plants compared to sensitive plants	DEN		(Matzrafi et al., 2017)
	72A	BhCYP72A15 (hybridous Mediterranean false brome)	up-regulated in NTSR plants compared to sensitive plants	DEN		(Matzrafi et al., 2017)
	72A	EpCYP72A254 (Barnyard grass)	up-regulated in NTSR plants compared to sensitive plants	Pyrimidinylthiobenzoic acid		Iwakami et al (2014)
	749A	GhCYP749A16 (cotton)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Thyssen et al., 2018)

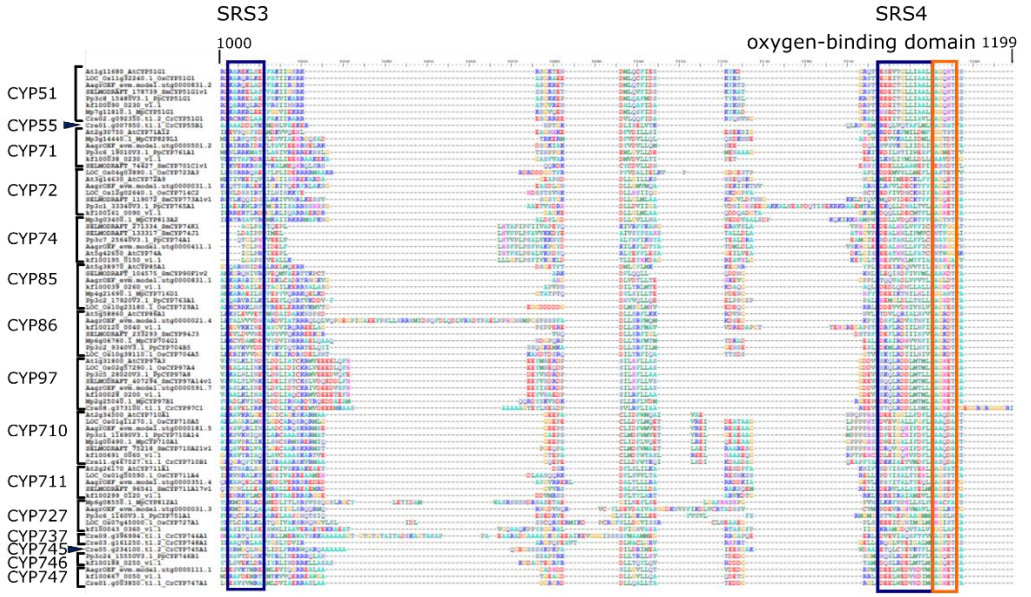
Table S2.3. Candidate NTSR CYPs are found within clans 71, 72 and 86 and represent candidate herbicide metabolising enzymes (continued).

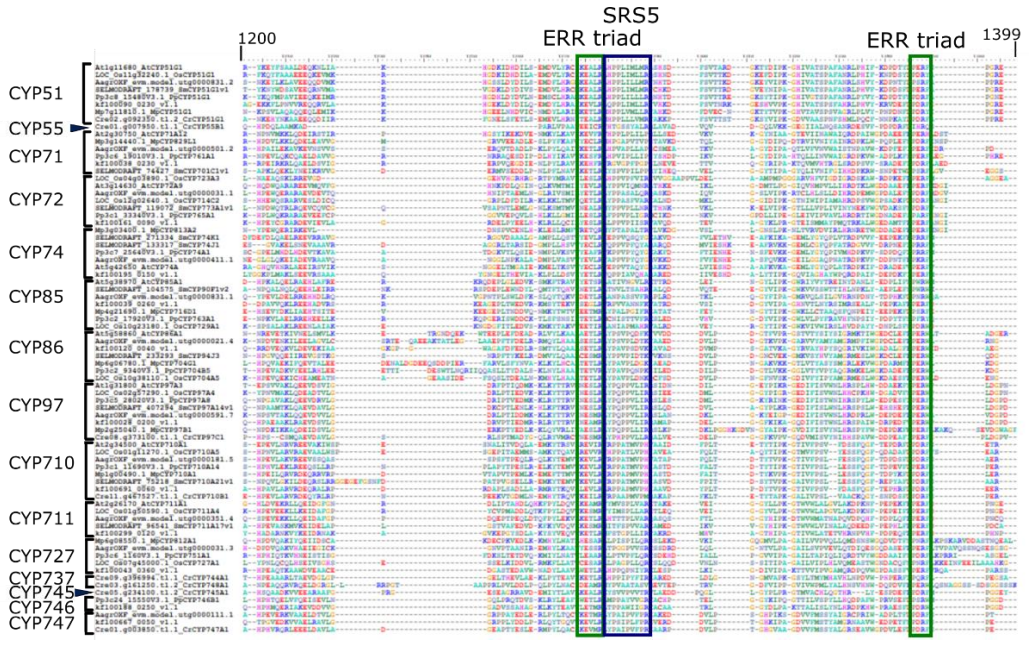
Clan	Sub-family	Gene name	Evidence	Herbicide chemical class	genes per clan	References
85	87A	BsCYP87A3 (American slough grass)	up-regulated in NTSR plants compared to sensitive plants	FOP	1	(Pan et al., 2016)
86	86B	MaCYP86B1 (water chickweed)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea	5	(W. Liu et al., 2018)
	86B	BcCYP86B1 (American slough grass)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Wang et al., 2021)
	94A	AaCYP94A2 (shortawn foxtail)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Zhao et al., 2017)
	94A	CbCYP94A1 (hairy fleabane)	up-regulated in NTSR plants compared to sensitive plants	Organophosphorus		(Piasecki et al., 2019)
	96A	DsCYP96A13 (flixweed)	up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		(Yang et al., 2016)

Note: The organophosphorus class refers to the herbicide glyphosate.


Membrane-targeting segment Hinge
200
399
SRS1
400
599
600
799
CYP51
CYP55
CYP71
CYP72
CYP74
CYP85
CYP86
CYP97
CYP710
CYP711
CYP727
CYP737
CYP745
CYP746
CYP747


SRS2
800
999
CYP51
CYP55
CYP71
CYP72
CYP74
CYP85
CYP86
CYP97
CYP710
CYP711
CYP727
CYP737
CYP745
CYP746
CYP747


SRS3
SRS4
1000
oxygen-binding domain
1199
CYP51
CYP55
CYP71
CYP72
CYP74
CYP85
CYP86
CYP97
CYP710
CYP711
CYP727
CYP737
CYP745
CYP746
CYP747


SRS5
ERR triad
ERR triad
1200
1399
CYP51
CYP55
CYP71
CYP72
CYP74
CYP85
CYP86
CYP97
CYP710
CYP711
CYP727
CYP737
CYP745
CYP746
CYP747

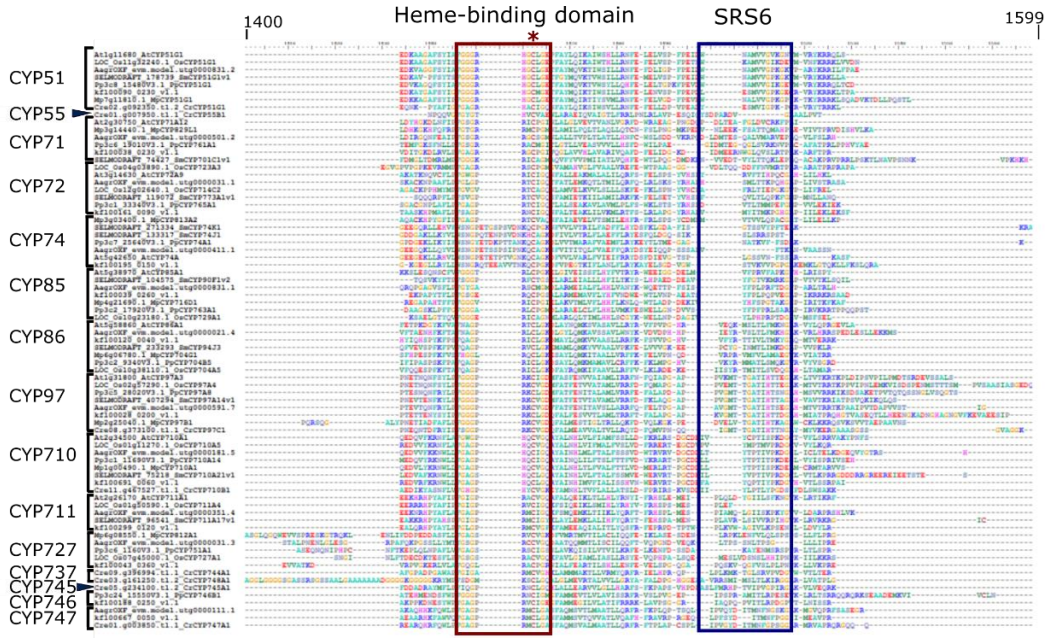


Figure S2.1. Untrimmed amino acid alignment of representative CYP proteins from each clan showing the location of conserved CYP domains. Representative sequences from each plant species in this study are included for each clan. The locations of the substrate recognition sites are based on those identified in Arabidopsis CYPs in Rupasinghe & Schuler, 2006. The absolutely conserved cysteine that binds the heme within the heme-binding domain is marked with an asterisk.

Chapter 3

Evolution of glutathione s-transferase enzymes in the Archaeplastida

3.1. Abstract

Glutathione s-transferases (GSTs) are an ancient multifunctional protein family with roles in abiotic and biotic stress response, hormone signalling and secondary metabolism. To characterise the evolution of GSTs in Archaeplastida a phylogenetic analysis was undertaken of GSTs collected from 9 species representing the rhodophyte, chlorophyte, streptophyte algae and land plant lineages and gene trees were constructed.

The GST protein family is divided into classes. The Theta class has been proposed as the ancestral cytosolic class which evolved in the ancestor of eukaryotes and prokaryotes. Many GST classes are widespread among eukaryotes and are thought to have originated early in eukaryote evolution. 5 classes are considered plant-specific: DHAR, Iota, Tau, Lambda and Hemerythrin. The topology of the tree indicates that the Iota class was present the common ancestor of Archaeplastida. The DHAR class was present in the common ancestor of Viridiplantae. The Tau and Lambda classes were present in the common ancestor of streptophytes. The Hemerythrin class was present in the common ancestor of land plants, indicating a stepwise evolution of plant-specific GST classes. There are fewer GST classes in land plants than in the streptophyte alga *Klebsormidium nitens*. There are 15 GST classes in *Klebsormidium nitens* and 11 in *Oryza sativa* and *Arabidopsis thaliana*. Despite the smaller number of classes the total number of GSTs is 3 times higher in seed plants. There are 86 GSTs in *O. sativa* and 26 GSTs in *K. nitens*. The larger number of GSTs in land plants is a result of gene duplications in the Tau class in early land plants and in the Phi class in vascular plants. GSTs that metabolise herbicides belong to these two classes.

Together these data demonstrate that the 5 GST classes to originate in Archaeplastida did so in a stepwise manner with the most recent class arising before the diversification of land plants. Gene duplication and neofunctionalization in the Tau and Phi classes in land plants led to the large numbers of GSTs present in land plant genomes. It is among the GST classes that diversified on land that enzymes that metabolise herbicides have evolved.

3.2. Introduction

Glutathione-s-transferases (GSTs) are an ancient superfamily of enzymes found in virtually all organisms. GSTs are most well-known for catalysing the conjugation of glutathione (GSH) to both endogenous and exogenous electrophilic, hydrophobic substrates to create more polar, hydrophilic compounds. GSTs also catalyse GSH-dependent peroxidase, isomerase and deglutathionylation reactions. GSTs are still largely uncharacterised but are associated with functions in abiotic and biotic detoxification pathways (Kumar et al., 2013; Riechers et al., 2005), ascorbic acid metabolism (Dixon et al., 2002), hormone signalling such as auxin and cytokinin homeostasis (Bilang & Sturm, 1995; Gonneau et al., 1998; Zettl et al., 1994), secondary metabolism of anthocyanins and flavonoids (Dixon & Edwards, 2010; Liu et al., 2019), tyrosine catabolism (Fernández-Cañón et al., 2002) and in preventing apoptosis (Kampranis et al., 2000).

Plant GSTs are divided into three broad groups based on their localisation; cytosolic, mitochondrial and microsomal GSTs. Within these groups GSTs are separated into classes based on amino acid identity, gene structure and active site residues. GSTs were originally classified into four types: type I, II, III and IV GSTs (Droog, 1997). This classification was later updated to Greek letter designations which matched the system used for animals GSTs (Edwards et al., 2000). Plant GSTs are divided into 12 cytosolic classes: Tau (GSTU), Phi (GSTF), Theta (GSTT), Lambda (GSTL), Zeta (GSTZ), Iota (GSTI), Hemerythrin (GSTH), tetrachlorohydroquinone dehalogenase (TCHQD), eukaryotic translation elongation factor 1B-γ subunit (Ef1Bγ), URE2p, glutathionyl hydroquinone reductase (GHR) and dehydroascorbate reductases (DHAR). There is one microsomal GST class in plants, microsomal prostaglandin E-synthase type 2 (mPGES2) and one mitochondrial GST class in plants named Metaxin (GSTM). Phi, Tau, DHAR, Hemerythrin, Iota and Lambda classes are considered specific to plants (Liu et al 2013). In 2016 members of the Phi class were found in fungi, bacteria and protists (Munyampundu et al., 2016). However plant Phi proteins formed a distinct clade from Phi GSTs from other kingdoms, indicating possible polyphyletic origins of Phi class.

In plants, Tau and Phi GSTs are the most abundant classes, with members catalysing the detoxification of a wide range of xenobiotics and with peroxidase activities (Cummins et al., 2011; Kumar & Trivedi, 2018). DHAR and Lambda GSTs have no detectable activity towards xenobiotics (Dixon et al., 2002). DHARs function as GSH-dependent oxidoreductases and regulate ascorbic acid while Lambda GSTs have been shown to catalyse the deglutathionylation of flavonoids (Dixon et al., 2002). Hemerythrin GSTs may function in the detoxification of heavy metals in plants (Liu et al., 2013). Zeta and Theta classes are found in bacteria, plants, fungi and animals (Board et al., 1997) and have limited activity towards xenobiotics. Zeta GSTs catalyse GSH-dependent isomerase reactions and function in tyrosine catabolism (Fernández-Cañón et al., 2002). Theta GSTs function as glutathione peroxidases during oxidative stress (Edwards & Dixon, 2005). GHR class is found in bacteria, fungi, archaea and plants and functions in quinone catabolism (Lallement et al., 2015). Metaxin class is found in animals and plants and forms part of the mitochondrial translocation system (Lister et al., 2007). The plant functions of

mPGES2, TCQHD, EF1By and URE2p GSTs have not yet been characterised. EF1By GSTs are found in plants, fungi and animals (Koonin et al., 1994), TCHQD proteins are found in plants and bacteria (Liu et al., 2013). URE2p class is found in plants, fungi and bacteria and is thought to have been acquired in plants through lateral transfer from bacteria (Liu et al., 2013). mPGES2 class is found in animals, plants and bacteria (Tanikawa et al., 2002; Yamada et al., 2005). Ef1By class is found in plants, animals and fungi (Liu et al., 2013).

GSTs that share greater than 40% amino acid identity are grouped within the same class, and GSTs in different classes can share less than 20% amino acid identity. Despite the low sequence identities between GSTs, their three-dimensional structure is highly conserved (Neuefeind et al., 1997a; Neuefeind et al., 1997b; Reinemer et al., 1996; Thom et al., 2002). Phi, Tau, Zeta and Theta GSTs typically function as dimers and Lambda and DHAR GSTs function as monomers. Each monomer is characterised by a conserved N-terminal domain which contains an active site and several GSH binding site residues (G-sites), and a less conserved C-terminal domain which is mainly composed of alpha helixes and contains class specific substrate binding sites (H-sites) (Edwards et al., 2000).

GST classification in plants is occurring very rapidly. In the last year alone GSTs have been classified in the genomes of melon (Wang et al., 2020), apple (Fang et al., 2020), chickpea (Ghangal et al., 2020) and radish (Gao et al., 2020). Recently, an analysis of GST phylogeny in green plants was undertaken (Monticolo et al., 2017). This study aims to determine the evolution of the GST superfamily in Archaeplastida, with a focus on the origins and expansions of plant-specific GST classes. GSTs have been linked to herbicide resistance in weeds and herbicide selectivity in crops, therefore elucidating the evolution of GSTs in plants will help to understand the evolution of GST mediated weed resistance.

3.3. Materials and Methods

3.3.1. Data resources

Protein sequences from *A. thaliana* were retrieved from TAIR10 (Huala et al., 2001) (https://www.arabidopsis.org/). In addition a classification of GSTs in *A. thaliana* can be found on the TAIR10 website (https://www.arabidopsis.org/browse/genefamily/gst.jsp). Protein sequences from *Oryza sativa* were retrieved from the rice genome annotation project RGAP7 (Kawahara et al., 2013) (http://rice.plantbiology.msu.edu/). Protein sequences from the liverwort *M. polymorpha* were obtained from MarpolBase (http://marchantia.info/). Protein sequences from the hornwort *Anthoceros agrestis* were obtained from (Li et al., 2020) (https://www.hornworts.uzh.ch/en/download.html). Protein sequences from the streptophyte alga *Klebsormidium nitens* were obtained from the *K. nitens* genome webpage (Hori et al., 2014) (http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/). Protein sequences from the lycophyte *Selaginella moellendorffii*, the moss *Physcomitrella patens* and the chlorophyte alga *Chlamydomonas reinhardtii* were retrieved from Phytozome 12 (Goodstein et al., 2012) (https://phytozome.jgi.doe.gov/pz/portal.html). Protein sequences from the red alga *Cyanidioschyzon merolae* (Matsuzaki et al., 2004) were retrieved from the *C. merolae* genome webpage (https://www.genome.jp/kegg-bin/show_organism?org=cme). A summary of data resources and genome sizes for each species can be found in Table 3.1.

Species	Classification	Genome (Mb)	Protein-coding genes	GSTs	References
Arabidopsis thaliana	Angiosperm dicot	135	25,498	61	Kaul et al., 2000
Oryza sativa	Angiosperm monocot	321	35,681	85	Kawahara et al., 2013
Selaginella moellendorffii	Lycophyte	212.6	22,285	57	Banks et al., 2011
Anthoceros agrestis	Hornwort	133	24,700	26	Li et al., 2020
Physcomitrella patens	Moss	480	35,938	42	Rensing et al., 2008
Marchantia polymorpha	Liverwort	225.8	19,138	35	Bowman et al., 2018
Klebsormidium nitens	Streptophyte alga	117.1	16,215	24	Hori et al., 2014
Chlamydomonas reinhardtii	Chlorophyte alga	120	15,143	19	Merchant et al., 2007
Cyanidioschyzon merolae	Rhodophyte alga	16.5	5,331	9	Matsuzaki et al., 2004

Table 3.1. List of species used in the analysis. Including their phylum (classification), genome size (Genome), the total number of protein-coding genes in the genome (Protein-coding genes) and the bibliographical reference.

3.3.2. Sequence collection and nomenclature

GST protein sequences were retrieved by BlastP searches using representative GSTs from each class in *A. thaliana* (Dixon et al., 2002; Dixon & Edwards, 2010a; Wagner et al., 2002), *O. sativa* (Jain et al., 2010; Soranzo et al., 2004) and *P. patens* (Liu et al., 2013). Sequences from these species were used to perform BlastP searches using an e value cut-off of $1e^{-10}$ against

the predicted proteomes of *S. moellendorffii*, *M. polymorpha*, *A. agrestis*, *K. nitens*, *C. reinhardtii*, and *C. merolae*. This initial list of sequences was self-blasted against the genome databases of each species to retrieve any additional sequences belonging to species-specific clans. In addition, GST keyword searches were performed on all genome databases to retrieve potential additional GSTs from species-specific classes. The results from these searches were analysed for the presence of a glutathione s-transferase N-terminal domain (IPR004045, IPR019564, PF13409, PF17172, PF13417 and PF02798) and C-terminal domain (IPR010987, PF13410, PF00043, PF14497 and PF17171) using InterProScan 84.0 (Blum et al., 2021). Kappa and MAPEG GSTs lack GST N-terminal thioredoxin-like and C-terminal domains site found in all other GSTs. Therefore protein sequences belonging to these classes weren't included in the phylogenetic analysis. An additional group of sequences identified by BlastP searches were found to have two GST N-terminal domains but lack a C-terminal domain. These were named 2N GSTs and were also not included in the phylogenetic analysis. GSTs were named following the system described by (Dixon et al., 2002) each GST gene was named starting with two letters denoting the Latin binomial of the source organism, followed by the GST class, and a progressive number for each gene.

3.3.3. Sequence alignment

Sequences were aligned in MAFFT (Katoh & Standley, 2013) using the FFT-NS-1 algorithm and visualised in Bioedit (Hall, 1999). To trim ambiguous regions and large gaps, four approaches to alignment cleaning were undertaken. Sequences were trimmed manually or with the automatic trimming software trimAI v.1.2. (Capella-Gutiérrez et al., 2009) using the three automated modes (-gappyout, -strict and -strictplus).

3.3.4. Phylogenetic analysis

The final alignments were subjected to a maximum-likelihood analysis conducted by PHyML 3.0 (Guindon et al., 2010) using an estimated gamma distribution parameter, the LG+G+F model of amino acid substitution and a Chi^2-based approximate likelihood ratio test (aLRT). The resulting unrooted trees were visualised in Figtree (http://tree.bio.ed.ac.uk/software/ figtree/) , and annotated in Inkscape (https://inkscape.org/release/inkscape-1.0.2/) (Yuan et al., 2016).

3.3.5. Analyses of phylogenetic relationships

The online MEME tool (http:// memesuite.org/) (Bailey et al., 2009) was used to identify conserved protein domains in a list of sequences to aid in assigning class membership. In addition classification of GSTs was confirmed by BlastP searches against the UniProt protein database (http://www.uniprot.org) (Apweiler et al., 2004). To further assign class membership, the list of GSTs identified by (Liu et al., 2013) was used as a BlastP database to aid in the classification of sequences that did not phylogenetically group with any known classes in Archaeplastida.

3.4. Results

3.4.1. 392 GST sequences were identified in the genomes of 9 species of Archaeplastida

Protein sequences from the genomes of 9 species representing major plant lineages were collected by BlastP searches of GSTs from *A. thaliana*, checked for the presence of the GST N-terminal and GST C-terminal domains, and manually filtered for pseudogenes, resulting in a final list of 392 glutathione-s-transferases. These species include the red alga *Cyanidioschyzon merolae* (9 genes), the chlorophyte alga *Chlamydomonas reinhardtii* (19 genes), the streptophyte alga *Klebsormidium nitens* (24 genes), the liverwort *Marchantia polymorpha* (35 genes), the moss *Physcomytrella patens* (42 genes), the hornwort *Anthoceros agrestis* (26 genes), the lycophyte *Selaginella moellendorffii* (57 genes), and the angiosperms *Oryza sativa* (85 genes) and *Arabidopsis thaliana* (61 genes). Two GST classes, mitochondrial Kappa and microsomal MAPEG, don't possess the typical GST N-terminal thioredoxin-like domain or C-terminal domain and lack a conserved N-terminal active site found in all other GSTs. Therefore protein sequences belonging to these classes weren't included in the phylogenetic analysis. An additional group of sequences identified by BlastP searches were found to have two GST N-terminal domains but lack a C-terminal domain. These were named 2N GSTs and were also not included in the phylogenetic analysis.

The sequences were aligned in MAFFT using the FFT-NS-2 algorithm and trimmed manually. A more stringent trimming approach was also tested where the alignment was cleaned by the automatic trimming software trimAl v.1.2. (Capella-Gutiérrez et al., 2009) using the three automated modes (-gappyout, -strict and -strictplus) (Figure 3.1). The maximum-likelihood (ML) tree resulting from the trimAl -strictplus alignment had low approximate likelihood ratio test (Chi^2-based aLRT) support values (as low as 0) for some clades and several GST classes were not monophyletic. The ML trees resulting from the trimAl -strict and -gappyout alignments had monophyletic classes, but very low aLRT support values (0.05-0.23) for the main clades. The ML tree generated from the manually trimmed alignment had highest support values for the classes (>0.8) of the four trees, therefore this tree was selected for downstream analysis.

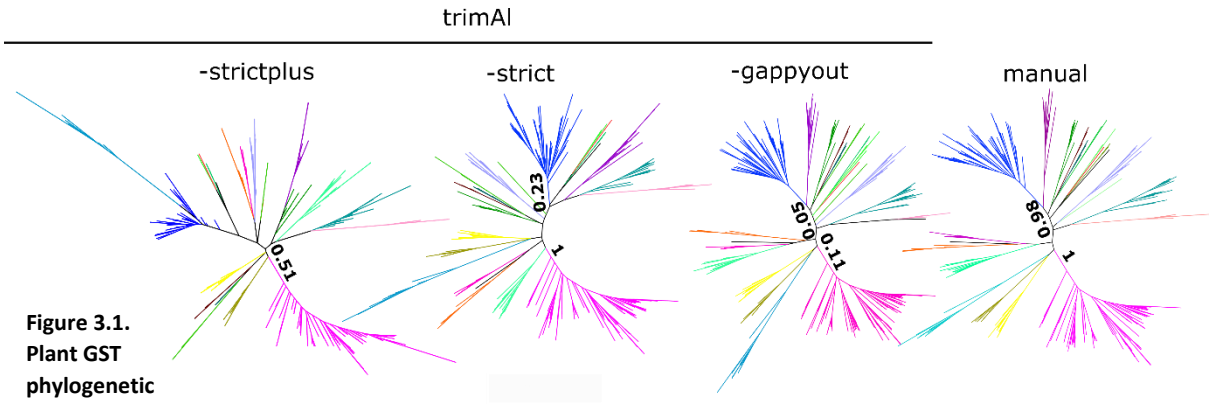


Figure 3.1. Plant GST phylogenetic analysis using automatic and manual trimming approaches.

Unrooted cladogram of a maximum likelihood (ML) analysis conducted by PHyML 3.0 (Guindon et al., 2010) using an estimated gamma distribution parameter, the LG+G+F model of amino acid substitution and a Chi^2-based aLRT test. Sequences were aligned in MAFFT and trimmed with the automatic trimming software trimAl using the automatic modes -strictplus (), -strict (), -gappyout () or by manual trimming. Branches are coloured to show the different GST classes. aLRT Support values for some of the clades are shown for comparison.

3.4.2. Five GST classes arose in a stepwise manner during Archaeplastida evolution

GSTs are multifunctional thioredoxin-like enzymes that are present in all groups of life, which catalyse the conjugation of glutathione (GSH) to a wide variety of endogenous and exogenous substrates. GSTs are classified based on their amino acid identity. As a general rule, proteins in the same class share at least 40% amino acid identity, but this number can be lower with sequences from distantly related species. Plant GSTs have been previously classified into 13 classes (Monticolo et al., 2017). To elucidate the evolutionary history of GST classes in Archaeplastida, a phylogenetic tree was constructed using a maximum likelihood approach (Figure 3.2). This analysis identified that GST proteins from 9 species of Archaeplastida grouped into 22 monophyletic classes: C8-like, Cr1, Cr2, DHAR, EF1B-γ, GHR, GTT1L, Hemerythrin, Iota, Kn1, Lambda, Metaxin, mPGES2, Mu, Phi, Pi, Sigma-like, Tau, TCHQD, Theta, URE2p and Zeta.

The GSTs encoded by the genomes of land plant species *A. agrestis*, *M. polymorpha, P. patens*, S. *moellendorffii*, *O. sativa* and *A. thaliana* grouped in 14 – DHAR, EF1B-γ, GHR, Hemerythrin, Iota, Lambda, Metaxin, mPGES2, Phi, Tau, TCHQD, Theta, URE2p and Zeta - of the 22 Archaeplastida classes (Figure 3.2). 5 GST classes present in land plants- DHAR, Hemerythrin, Iota, Lambda and Tau – have not been identified in non-plant genomes and are therefore considered to be plant-specific. URE2p GSTs are also found in bacteria and fungi. Plant URE2p proteins form a monophyletic clade with some bacterial Ure2p proteins, which suggests that they were inherited in plants by lateral transfer from bacteria, rather than having evolved in the common ancestor of bacteria and plants (Liu et al., 2013). Phi GSTs have been found in fungi, bacteria and protists. Plant Phi GSTs form a distinct monophyletic clade from Phi proteins from other kingdoms, suggesting that Phi GSTs originated separately in plant and non-plant lineages. (Munyampundu et al., 2016).

Members of the Hemerythrin class were found in the genomes of the bryophytes *P. patens, M. polymorpha* and *A. agrestis* and the lycophyte *S. moellendorffii*, but not in the angiosperm genomes and not in *K. nitens, C. reinhardtii* or *C. merolae*. This suggests that the Hemerythrin class originated in the common ancestor of bryophytes and vascular plants. Members of the Lambda, Phi and Tau classes were present in the genome of the streptophyte alga *K. nitens* but not in the chlorophyte alga *C. reinhardtii* or the rhodophyte alga *C. merolae*. This suggests that these classes originated in the streptophyte lineage after the divergence of chlorophytes and streptophytes and before the divergence of streptophyte algae and land plants. URE2p proteins were also present in land plants and *K. nitens* but not in *C. reinhardtii or C. merolae*. This suggests that URE2p was acquired from bacteria after the divergence of chlorophytes and streptophytes, and before the divergence of streptophytes and land plants. The two other plant-specific classes - Iota and DHAR - each contain members belonging to the streptophyte alga *K. nitens* and the green alga *C. reinhardtii*. This suggests that Iota and DHAR GSTs were present in the last common ancestor of chlorophytes and streptophytes. The Iota class contains in addition a GST sequence belonging to the red algae *C. merolae*. This indicates that Iota class enzymes originated before the divergence of rhodophytes and chlorophytes, making Iota the most ancient Archaeplastida-specific GST class.

Several algal sequences did not group with the land plant classes (Table 3.2). To determine their classification, sequences were analysed by BlastP searches against the UniProt protein database (Supplementary Table S3.4). Sequences were also analysed by BlastP searches against the list of 152 sequences representing 36 GST classes from plants, animals fungi and bacteria identified by Liu et al. (2013) (Supplementary Table S3.4). The catalytic motifs of the algal sequences were compared to those of known GST classes which represented the closest BlastP hits (Table 3.2). A clade, sister to the URE2p GST class, contained sequences from *C. reinhardtii* and *C. merolae*. BlastP analyses of these sequences against the UniProt and Liu et al. (2013) databases found that their best hits belong to the bacterial C-8 class (Supplementary Table S3.4), and amino acid alignments found that the catalytic site of these sequences and other C-8 proteins was mostly conserved (Table 3.2). This clade was thus named GSTC8-like. A clade sister to the C8-like and URE2p classes contained one member each from *K. nitens*, *C. reinhardtii* and *C. merolae*. BlastP analyses of these sequences against the UniProt and Liu et al. (2013) databases found that their best hits were GTT1-like sequences (Supplementary Table S3.4), a class previously reported in fungi, bacteria and yeast (Choi et al., 1998; Liu et al., 2013). The catalytic domain is very variable in GTT1-like proteins, therefore GTT1-like proteins can't be confidently classified by their catalytic motif (Table 3.2). These sequences were tentatively classified as GTT1-like proteins. The presence of Sigma-like and Pi GSTs in *C. reinhardtii*, first identified by Chatzikonstantinou et al. (2017), was confirmed in this analysis by BlastP searches and comparisons of catalytic motifs (Table 3.2, Supplementary Table S3.4). A red algal GST (CMT103C) was sister to the Pi GSTs, but a BlastP search against the UniProt and Liu et al. (2013) databases found that the best hits were Mu GSTs (Supplementary Table S3.4). The catalytic

domain of Mu GSTs was also conserved in CMT103C (Table 3.2), supporting the classification of this GST within the Mu class.

Three sequences could not be assigned to a GST class. One streptophyte algal GST named Kn1 (Kfl00304_0120_v1) was sister to the Lambda class and has the same catalytic motif as Lambda proteins (Table 3.2) however it shares low amino acid identity (20%) with Lambda GSTs and a domain search using the online MEME tool (http:// memesuite.org/) (Bailey et al., 2009) found that it lacks the C-terminal protein motifs present in all other Lambda class GSTs (Supplementary Figure S3.1). Due to its similarity with Lambda GSTs, Kn1 may be the ancestral Lambda GST. A chlorophyte algal GST named Cr1 (Cre02.g142200.t1) was sister to the clade containing the Sigma-like, Mu, and Pi GSTs. The top hits resulting from a BlastP search of the Cr1 protein sequence against the GST database from Liu et al (2013) are Pi GSTs, however the active site motif of this protein is different to that of Pi GSTs (Table 3.2), and Cr1 shares low amino acid identity (20%) with the other Pi GSTs. Another chlorophyte algal GST named Cr2 (Cre12.g508850.t1) was sister to the TCHQD class. However, a meme search found that Cr2 lacks TCHQD motifs (Supplementary Figure S3.2). The closest BlastP hits for Cr2 against the UniProt database and GST list from Lui et al. (2013) were to Phi class GSTs (Supplementary Table S3.4). Additionally, Cr2 doesn't have the same active site motif as Phi or TCHQD GSTs (Table 3.2). Therefore Kn1, Cr1 and Cr2 were not assigned to a GST class.

In summary, this analysis shows that Archaeplastida GSTs are divided into 22 classes. 7 of the classes found in Archaeplastida genomes originated since the divergence of Archaeplastida from other eukaryotes. The earliest GST class to arise in Archaeplastida was the Iota class which originated before the divergence of rhodophytes and chlorophytes. The DHAR class originated after the divergence of rhodophytes and chlorophytes and before the divergence of chlorophytes and streptophytes. Lambda, Tau, plant Phi and plant URE2p GSTs originated in the last common ancestor of streptophyte algae and land plants. The most recently diverging plant GST class, the Hemerythrin class, originated in the last common ancestor of land plants. Some algal sequences did not group within the land plant GST classes and their closest BlastP hits included animal Sigma-like, Mu and Pi GSTs and bacterial C8-like and GTT1-like GSTs.

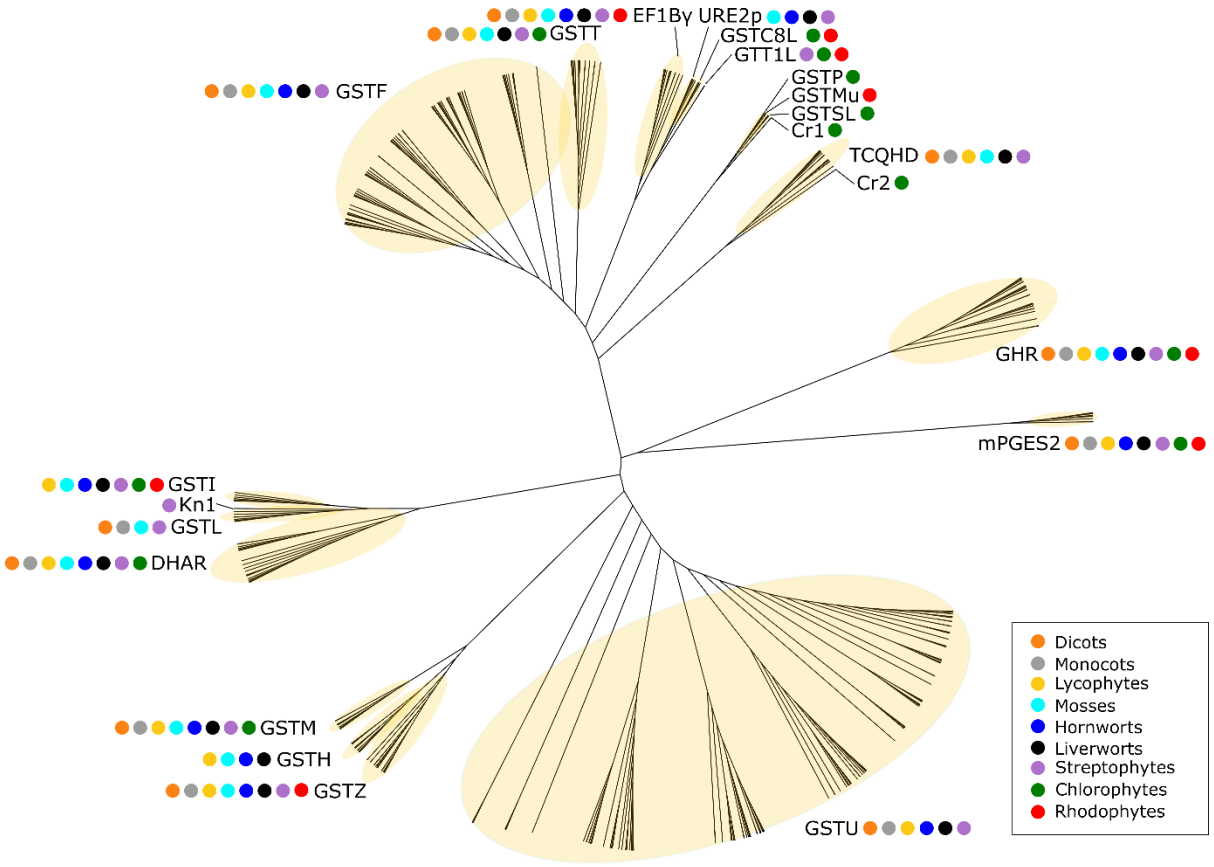


Figure 3.2. GSTs have diverged into 22 distinct classes during the evolutionary history of Archaeplastida.

Unrooted cladogram of a maximum likelihood (ML) analysis conducted by PHyML 3.0 (Guindon et al., 2010) using an estimated gamma distribution parameter, the LG+G+F model of amino acid substitution and a Chi^2-based aLRT test. Protein sequences were aligned using MAFFT with the L-INS-i algorithm. Classes are indicated by orange circles and numbers. Coloured dots indicate the presence of sequences from different species in each clan. *A. thaliana* (orange); *O. sativa ssp. japonica* (grey); *Selaginella moellendorffii* (yellow); *Physcomitrella patens* (cyan); *Anthoceros agrestis* (blue); *Marchantia polymorpha* (black); *Klebsormidium nitens* (purple); *Chlamydomonas reinhardtii* (green); *Cyanidioschyzon merolae* (red).

Accession number	Species	Catalytic motif of genes	Catalytic motif of closest class by BlastP searches	Assigned class	Reference
CMR110C	*C. merolae*	N-P-X-S	S/N-P-R/G-S (C_8)	C8-like	Liu et al., 2013
Cre15.g636800.t1	*C. reinhardtii*				
Cre15.g636750.t1	*C. reinhardtii*				
Cre13.g584850.t1	*C. reinhardtii*				
CMO318C	*C. merolae*	P/N-P/Q-T/S-R	X-X-S-R/Q (GTT1-like)	GTT1-like	Liu et al., 2013
kfl00075_0100_v1	*K. nitens*				
CMT103C	*C. merolae*	Y-W-D-I	Y-W-X-I/L (Mu)	Mu	Kalita et al., 2017
Cre17.g721350.t1	*C. reinhardtii*	Y-F/Y-P-A/I	Y-F-P/A-X	Pi	Feng & Singh, 2009
Cre17.g742300.t1	*C. reinhardtii*				
Cre17.g742450.t1	*C. reinhardtii*				
Cre16.g670973.t1	*C. reinhardtii*	Y-F-P/A-X	Y-F-N-X (Sigma)	Sigma-like	Liu et al., 2020
Cre16.g682725.t1	*C. reinhardtii*				
Cre16.g688550.t1	*C. reinhardtii*				
AagrOXF_evm.model.utg000005l.356	*A. agrestis*	S-T-F-G	S-X-F-X (Tau)	Tau	This analysis
Unclassified GST proteins:					
Cre02.g142200.t1	*C. reinhardtii*	W-P-G-M	Y-F-P/A-X (Pi)	Cr1	Liu et al., 2013
Cre12.g508850.t1	*C. reinhardtii*	S-N-N-G	A/S-X-X-S/P (TCHQD)	Cr2	This analysis
			S/Y-X-A/C-T/Q (Phi)		
Kfl00304_0120_v1	*K. nitens*	C-P-Y-A	C-P-F/Y-A (Lambda)	Kn1	This analysis

Table 3.2. Analysis of catalytic motifs of algal and hornwort GST protein sequences to assign class membership.

Comparisons of the catalytic motif of unclassified GSTs to the catalytic motif of the closest class identified by BlastP searches of the UniProt and the Liu et al. (2013) databases. The catalytic motifs of the known classes were identified from amino acid alignments in the literature (Reference). Species full names are the hornwort *Anthoceros agrestis*, and algae *Cyanidioschyzon merolae*, *Chlamydomonas reinhardtii* and *Klebsormidium nitens.*

3.4.3. *Anthoceros agrestis* encodes 26 GSTs including one Tau GST

The hornwort *A. agrestis* is a member of the earliest diverging lineage of extant land plants, the bryophyte clade, which also contains mosses and liverworts. The hornwort lineage has been recently resolved as sister to the setaphytes (mosses and liverworts) (Morris et al., 2018). The inclusion of a hornwort species in the phylogenetic analysis of the Archaeplastida GST protein family provides a more complete picture of the evolution and distribution of GST proteins in land plants. To determine the number of GST sequences in the genome of *A. agrestis*, a BlastP search was performed on the *A. agrestis* genome (Li et al., 2020) using known members from all Archaeplastida GST classes. All sequences above a $1e^{-10}$ E-value threshold were retained. This initial list of sequences was self-blasted to identify any species-specific GST sequences. The sequences were further checked for presence of the GST N-terminal and C-terminal domains using InterProScan 84.0 (Blum et al., 2021). The sequences were aligned in MAFFT v.7 (Katoh & Standley, 2013) and any sequences missing conserved domains (Supplementary Figure S3.2) were removed.

The genome of *A. agrestis* was found to encode a total of 26 GST proteins belonging to 12 classes (Figure 3.3, Supplementary Tables S3.1). Additionally, two MAPEG sequences and one 2N GST sequence were identified (Supplementary Table S3.2). There are fewer GSTs in *A. agrestis* than in the other bryophytes. 42 and 35 GSTs were identified in *P. patens* and *M. polymorpha*, respectively. This suggests that there has been either an increase in GST number in setaphytes or gene losses in hornworts after the divergence of setaphytes and hornworts. The genomes of *A. agrestis*, *P. patens* and *M. polymorpha* all encode fewer GSTs than vascular plants. 55, 85 and 61 GST sequences were collected from the genomes of *S. moellendorffii*, *O. sativa* and *A. thaliana*, respectively. This indicates that the number of GSTs increased during land plant evolution.

Monticolo et al. (2017) reported Tau GSTs are absent from the mosses *P. patens* and *Sphagnum fallax*, but present in *M. polymorpha,* lycophytes and vascular plants. This observation led to the suggestion that liverworts represent a separate clade from bryophytes (which include mosses and hornworts) and is more closely related to vascular plants. However my analysis identified a GST in *A. agrestis* (AagrOXF_evm.model.utg000005l.356.1) that groups phylogenetically within the Tau GST clade. It also possesses the classical Tau class catalytic motif (Table 3.2) and BlastP searches of this protein sequence against the UniProt database and the list of GSTs obtained from Liu et al. (2013) results in closest hits to Tau GSTs (Table 3.3). Therefore this is strong evidence that this is a Tau GST, and supports the model of a monophyletic origin of bryophytes in which the Tau GST class was present in the last common ancestor of bryophytes and subsequently lost in the moss lineage (Figure 3.4B).



Figure 3.3. Amino acid alignment of part of the N-terminal domain of the 26 *Anthoceros agrestis* GST proteins.

G-site domains involved in GSH binding are indicated with pink shading, and the catalytic residue within the active site is marked with an asterisk.

3.4.4. GST Phi class radiated in land plants and GST Tau class radiated in vascular plants

The number of GSTs collected differed between the 9 Archaeplastida genomes. The algal genomes encode fewer GSTs than the bryophyte genomes. 9, 19 and 24 GST sequences were collected from the genomes of the red alga *C. merolae*, the chlorophyte alga *C. reinhardtii*, and the streptophyte alga *K. nitens*, respectively, while 35, 42 and 26 GST sequences were collected from the genomes of *A. agrestis*, *P. patens* and *M. polymorpha*, respectively (Figure 3.4B, Supplementary Table S3.1). The bryophytes encode fewer GSTs than the vascular plants. 55, 85 and 61 GST sequences were collected from the genomes of *S. moellendorffii*, *O. sativa* and *A. thaliana*, respectively (Figure 3.4B, Supplementary Table S3.1). This indicates that Archaeplastida evolution is associated with an increase in the number of GST proteins in plant genomes. The total number of GSTs in the genome of *O. sativa* is 3 times higher than in the streptophyte alga *K. nitens*. Despite the larger number of GSTs there are fewer GST classes in land plants than in *K. nitens*. There are 15 GST classes in *K. nitens* and 11 in *O. sativa* and *A. thaliana*.

The number of GSTs in the Phi and Tau classes differed considerably between species. There are 3 Phi class members in the genome of the streptophyte alga *K. nitens*. Among the early diverging land plants there are 18 Phi class genes in the genome of *M. polymorpha*, 10 in *P. patens* and 11 in *A. agrestis*. Only 1 Phi class member was identified in the genome of the lycophyte *S. moellendorffii*. Among the angiosperms, there are 19 Phi class genes in *O. sativa* and 13 in *A. thaliana*. This suggests that the Phi class expanded during Archaeplastida evolution after the divergence of streptophyte algae and land plants and before the divergence of bryophytes and vascular plants. There are more Tau class GSTs in vascular plant genomes than in the bryophyte and algal genomes (Figure 3.3B). There are 3 Tau class genes in the genome of *K. nitens.* Among the early diverging land plants there are 2 Tau class members in *M. polymorpha*, 1 in *A. agrestis* and none in *P. patens*. Among the vascular plants there are 34 in *S. moellendorffii*, 49 in *O. sativa* and 28 in *A. thaliana*. This suggests that the Tau class expanded in vascular plants after the divergence of bryophytes and vascular plants.

In the other 20 classes in Archaeplastida – C8-Like, Cr1, Cr2, DHAR, EF1B-γ, GHR, GTT1L, Hemerythrin, Iota, Kn1, Lambda, Metaxin, mPGES2, Mu, Pi, Sigma-like, TCHQD, Theta, URE2p and Zeta – gene numbers remain below 10 in each species, and usually number 1-3 per species (Supplementary Table S3.1). Several classes did not have representatives in some land plant species, suggesting that they were lost in these species or lineages (Figure 3.4A). Members of the Lambda GST class were present in the genomes of streptophyte alga *K. nitens* and the moss *P. patens* but absent from the genomes of *A. agrestis* and *M. polymorpha*, suggesting that it was lost in the hornwort and liverwort lineages. The Tau class was present in *K. nitens* and in all land plants in this study except *P. patens* which suggests that it was lost in the moss lineage. Iota GSTs were found in all Archaeplastida genomes except *A. thaliana* and *O. sativa* which suggests that Iota was lost in the common ancestor of angiosperms. URE2p class was present in the genomes of all land plants except *O. sativa* and *A. thaliana* suggesting it was lost in the last common ancestor of angiosperms. The Hemerythrin class was absent from *A.*

thaliana and *O. sativa*, but present in the lycophyte *S. moellendorffii* and the three bryophyte genomes, suggesting that it was lost in the last common ancestor of angiosperms.

In summary, the total number of GSTs collected in each species differed significantly. Algal genomes encode fewer GSTs than bryophyte genomes, which encode fewer GSTs than vascular plants. This suggests that the number of GSTs increased during Archaeplastida evolution. Two classes are responsible for the large numbers of GST in land plants. GST Tau class expanded in vascular plants after their divergence from bryophytes, while GST Phi class expanded in early land plant evolution, after the divergence of streptophyte algae and land plants. The 20 other GST classes in Archaeplastida have low gene numbers in each species. The lack of expansions in the 20 other GST classes in Archaeplastida suggests that these classes may perform fundamental functions in all plants. Several different classes did not have members in some lineages, suggesting that they were lost in these species or lineages.

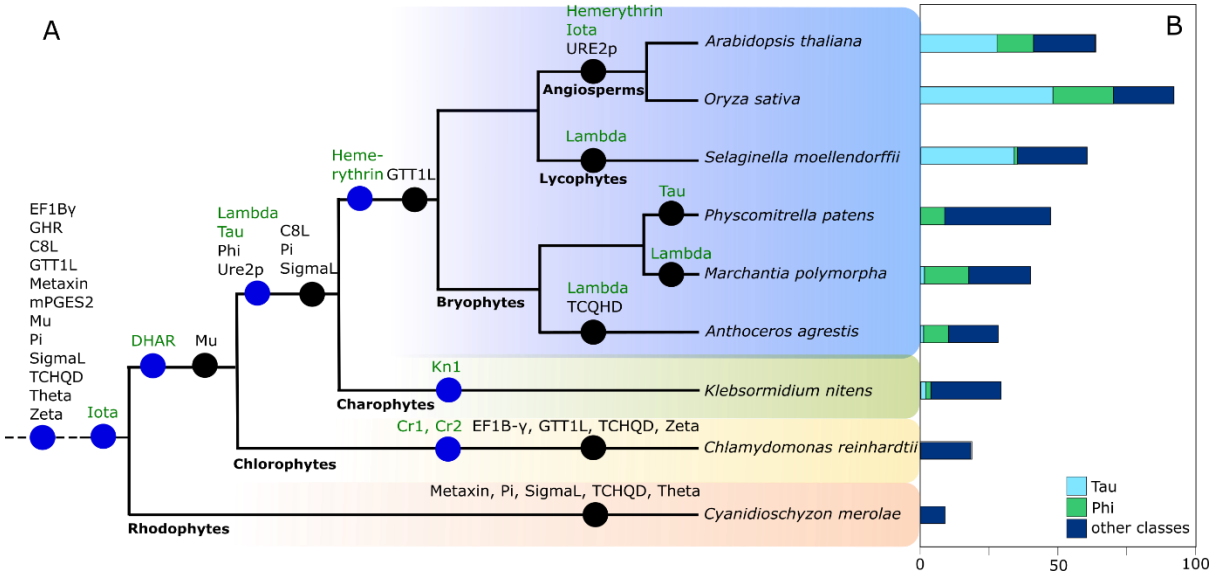


Figure 4. Tau and Phi GST classes expanded during land plant evolution.

A. Cladogram of Archaeplastida phylogeny showing putative origins and losses of GST classes in plants. Blue circles represent first appearance of a class, black circles represent the absence of a class in a particular lineage. Green classes represent classes that are only found in Archaeplastida genomes. Black classes represent classes that are found in non-Archaeplastida genomes. The Phi class is thought to have multiple origins in bacteria, fungi, protists and plants (Munyampundu et al., 2016), and the URE2p class was likely acquired by lateral transfer from bacteria (Liu et al., 2013) **B.** Number of GST proteins in each species showing expansions of the GST Phi and Tau classes in the land plants.

3.4.5. Herbicide resistance-conferring GSTs are members of the two expanded Tau and Phi classes

From the time of their discovery 50 years ago, GST enzymes have been associated with the chemical transformation of herbicides (Frear & Swanson, 1970). The repeated use of herbicides has resulted in the evolution of resistant weeds globally (Heap, 2021). GSTs can confer herbicide resistance via the increased expression or activity of GSTs which catalyse GSH-dependent conjugations of the herbicide chemical in the plant (Cummins et al., 2011). GSTs are also thought to confer herbicide resistance by regulating the plant to produce protective flavonoids (Cummins et al., 2013), or by enhanced GSH peroxidase activity to reduce harmful hydroperoxides produced in response to herbicide treatment (Benekos et al., 2010; Cummins et al., 1999, 2003). To understand which classes of GSTs are most likely to contain enzymes that protect against herbicides, a literature review was undertaken to identify GSTs that have been experimentally shown to metabolise herbicides or to confer herbicide resistance (Table 3.4). Several plant GSTs are transcriptionally up-regulated or have enhanced protein levels in non-target site resistant plants, but haven't been tested for their role in herbicide resistance. These genes weren't included in this analysis and can be found in Supplementary Table S3.3.

A total of 35 plant GST proteins were identified in the literature to have conjugating activity or confer resistance towards one or multiple herbicides (Table 3.2). The GSTs were identified in the model species thale cress (*Arabidopsis thaliana*), moss (*P. patens)*, the crops maize (*Zea mays*), rice (*Oryza sativa*), sorghum (*Sorghum bicolor*), wheat (*Triticum aestivum*) and soybean (*Glycine max*), the weed species annual rye-grass (*Lolium rigidum*) and blackgrass (*Alopecurus myosuroides*). The 35 proteins conferred resistance to or metabolised a wide diversity of herbicide substrates from chemical classes such as aryloxyphenoxypropionates (FOP), bipyridyliums, chloroacetanilides, diphenyl ethers, organophosphorus chemicals, phenylurea, thiocarbamates, sulfonylurea and thiadiazines, either individually or in mixtures. The majority (28 of 35 genes) metabolised chloroacetanilide herbicides, such as alachlor and metolachlor. All GSTs with conjugating or peroxidase activity towards herbicides characterised so far belonged to the two largest classes in plants, Phi and Tau. 22 Tau GSTs from 6 different species metabolised chloroacetanilide, diphenyl ether, FOP, sulfonylurea and triazine herbicide chemicals. There are many Tau GSTs encoded in the genomes of vascular plants. There are 28 Tau members in *A. thaliana* and 49 in *O. sativa*. 13 Phi class GSTs from 7 different species metabolised chloroacetanilide, diphenyl ether, thiocarbamate and triazine herbicide chemicals, and conferred resistance by peroxidase activity towards bipyridylium, DIM, organophosphorus, phenylurea, sulphonylurea and FOP herbicide chemicals. Fewer Phi GSTs are encoded in the genomes of some land plants. There are 18 Phi members in *M. polymorpha* and 19 in *O. sativa*. These numbers indicate that the GSTs that metabolise or protect against herbicide chemicals belong to the Phi and Tau classes, which represent the two expanded GST classes in land plants. One single Lambda GST from rice was found to exhibit moderate tolerance towards the sulphonylurea herbicide chlorsulfuron and the organophosphorus herbicide glyphosate,

by catalysing the reduction of superoxides (Hu et al., 2009). Lambda is a plant-specific class with 1-3 members present in plant genomes. Lambda GSTs were lost in the genomes of *S. moellendorffii, A. agrestis* and *M. polymorpha*.

In vascular plants, one half to two thirds of GSTs belong to Tau and Phi classes. In *A. thaliana*, there are 41 GSTs in Tau and Phi and 20 GSTs in other classes. In *O. sativa*, there are 68 GSTs in Phi and Tau and 17 GSTs in other classes. This suggests that it is possible that the majority of GSTs that confer resistance to herbicides belong to Phi and Tau because there are more genes in these classes. Therefore the hypothesis that GSTs from any class can confer herbicide resistance cannot be rejected. The finding that a Lambda class GST when over-expressed was able to confer resistance to two herbicides supports this. In addition, candidate herbicide resistance conferring GSTs that were up-regulated in crops and non-target site resistant weeds belong to a larger number of classes, including DHAR, MAPEG and Lambda (Supplementary Table S3.3).

In summary, GSTs were found to confer herbicide resistance via multiple mechanisms. All GST enzymes demonstrated to have GSH-conjugating activities towards herbicides belong to the Phi and Tau classes, which represent the two expanded GST classes in land plants. These enzymes metabolised chloroacetanilide, diphenyl ether, FOP, thiocarbamate and triazine herbicide chemicals. Some GSTs belonging to the Phi and Tau classes and a Lambda class GST showed GSH-peroxidase activities that conferred indirect herbicide resistance towards bipyridylium, chloroacetanilide, FOP, organophosphorus, phenylurea, sulphonylurea and triazine herbicides. One GST from blackgrass, AmGSTF1, was found to confer herbicide resistance by regulating the plant to produce protective flavonoids (Cummins et al., 2013). Therefore multiple mechanisms of resistance can be conferred by GSTs across classes and within the same class.

Class	Gene name	Evidence	Herbicide chemical class	Genes per class	References
Lambda	OsGSTL1 (rice)	Over-expression in rice	Organophosphorus, Sulphonylurea	1	Hu et al., 2009
Phi	ZmGSTF1 (maize)	Conjugating activity in vitro, transformed in tobacco	Chloroacetanilide	13	Karavangeli et al., 2005, Mozer et al., 1983, Shah et al., 1986
	ZmGSTF2 (maize)	Conjugating activity in vitro, transformed in tobacco & wheat	Chloroacetanilide, Diphenyl ether, Thiocarbamate		Dixon et al., 1997, Fuerst et al., 1993, Milligan et al., 2001
	ZmGSTF3 (maize)	Conjugating activity in vitro & in *E. coli*	Chloroacetanilide, Diphenyl ether		Fuerst et al., 1993, Sommer & Böger, 1999
	ZmGSTF4 (maize)	Conjugating activity in vitro, Transformed in tobacco	Chloroacetanilide, Thiocarbamate		Milligan et al., 2001, Irzyk & Fuerst, 1993
	TaGSTF2-2 (wheat)	Conjugating activity in *E. coli*	Chloroacetanilide, Diphenyl ether		Cummins et al., 2003
	TaGSTF3-3 (wheat)	Conjugating activity & peroxidase activity in *E. coli*	Chloroacetanilide, Diphenyl ether		Cummins et al., 2003
	TaGST2-3 (wheat)	Conjugating activity in vitro	Diphenyl ether		Pascal & Scalla, 1999
	OsGSTF3-3 (rice)	Conjugating activity in *E. coli*	Chloroacetanilide		Cho & Kong, 2005
	SbGSTB1-B2 (sorghum)	Conjugating activity in vitro	Chloroacetanilide		Gronwald & Plaisance, 1998
	PpGSTF7 (moss)	Conjugating activity in *E. coli*	Diphenyl ether		Liu et al., 2013
	LrGSTF1 (annual rye-grass)	Glutathione peroxidase activity in *E. coli*	Organophosphorus, Sulphonylurea, FOP		Cummins et al., 2013
	AmGSTF1 (blackgrass)	Conjugating and peroxidase activity in *E. coli,* flavonoid metabolism, transformation in *A. thaliana*	Chloroacetanilide, Diphenyl ether, Triazine, Phenylurea, Bipyridylium		Cummins et al., 1999, Cummins et al., 2013
	AmGSTF2 (blackgrass)	Glutathione peroxidase activity in *E. coli*	Bipyridylium, Phenylurea, Diphenyl ether		Cummins et al., 1999
Tau	AmGSTU1-1 (blackgrass)	Metabolism in *E. coli*	Diphenyl ether, FOP	22	Cummins et al., 1999
	AtGSTU19 (thale cress)	Metabolism in *E. coli*	Chloroacetanilide		Deridder et al., 2002
	GmGSTU2 (soybean)	Conjugating activity in *E. coli*	Triazine		McGonigle et al., 2000
	GmGSTU4 (soybean)	Conjugating and peroxidase activity in *E. coli,* Overexpressed in tobacco	Chloroacetanilide, Diphenyl ether		Benekos et al., 2010, McGonigle et al., 2000
	GmGSTU5 (soybean)	Conjugating activity in *E. coli*	Chloroacetanilide, Sulphonylurea		McGonigle et al., 2000
	GmGSTU7 (soybean)	Conjugating activity in *E. coli*	Triazine		McGonigle et al., 2000
	GmGSTU8 (soybean)	Conjugating activity in *E. coli*	Chloroacetanilide		McGonigle et al., 2000
	GmGSTU9 (soybean)	Conjugating activity in *E. coli*	Chloroacetanilide, Sulphonylurea		McGonigle et al., 2000
	GmGSTU10 (soybean)	Conjugating activity in *E. coli*	Chloroacetanilide, Sulphonylurea		McGonigle et al., 2000
	GmGSTU21 (soybean)	Conjugating activity in *E. coli,* transformed in tobacco	Diphenyl ether		Skipsey et al., 2005
	OsGSTU3 (rice)	Conjugating activity in *E. coli*	Chloroacetanilide		Lee et al., 2011
	OsGSTU4 (rice)	Conjugating activity in *E. coli*	Chloroacetanilide		Jo et al., 2011
	ZmGSTU1 (maize)	Conjugating activity in vitro & in *E. coli*	Chloroacetanilide, Diphenyl ether		Dixon et al., 1998, McGonigle et al., 2000
	ZmGSTU2 (maize)	Conjugating activity in *E. coli*	Chloroacetanilide, Diphenyl ether		Dixon et al., 1999
	ZmGSTU3 (maize)	Conjugating activity in *E. coli*	Chloroacetanilide		Dixon et al., 1999
	ZmGSTU4/19 (maize)	Conjugating activity in *E. coli*	Chloroacetanilide, Sulphonylurea		McGonigle et al., 2000
	TaGSTU1-1 (wheat)	Conjugating and peroxidase activity in vitro	Chloroacetanilide, Diphenyl ether		Cummins et al., 1997
	TaGSTU1-2 (wheat)	Conjugating and peroxidase activity in vitro	Chloroacetanilide, Diphenyl ether, FOP, Triazine		Cummins et al., 1997
	TaGSTU1-3 (wheat)	Conjugating activity in vitro	Chloroacetanilide, Diphenyl ether, FOP, Triazine		Cummins et al., 1997
	TaGSTU1-4 (wheat)	Conjugating and peroxidase activity in vitro	Chloroacetanilide, Diphenyl ether, FOP		Cummins et al., 1997
	TaGSTU4-4 (wheat)	Conjugating activity in vitro & in *E. coli*	Chloroacetanilide and FOP		Thom et al., 2002

Table 3.4. Plant GSTs involved in herbicide metabolism or resistance are members of the Phi and Tau and Lambda classes.
Maize and *A. thaliana* genes were renamed according to nomenclature proposed by Edwards et al. (2000).

3.5. Discussion

The phylogenetic analysis of the glutathione s-transferase superfamily in Archaeplastida revealed that GST classes that evolved in Archaeplastida arose in a stepwise manner with the most recently diverging plant GST class originating in the common ancestor of land plants. The evolution of land plants was associated with an increase in the number of GSTs in the Tau and Phi classes, with the Phi class expanding in the ancestor of land plants, and the Tau class expanding in the vascular plants. Several classes were lost in certain lineages during Archaeplastida GST evolution. Members of the diversified Tau and Phi classes and one Lambda class member are able to confer resistance to herbicides from several herbicide classes. Algal GSTs are also highly divergent from land plant GSTs and were found to possess members from animal and bacteria classes.

The transition from the aquatic environment to the terrestrial environment was accompanied by the radiation of plant phenotypic diversity in the Phanerozoic (Bateman et al., 1998; Morris et al., 2018; Wickett et al., 2014). This morphological radiation did not coincide with an increase in the number of GST classes. Only one class originated after the emergence of land plants, the Hemerythrin class. Instead the colonisation of terrestrial habitats by plants coincided with an increase in the number of genes in the Phi class, and the evolution of vascular plants was associated with an increase in gene numbers in the Tau class. GST classes emerged in a stepwise manner, with Lambda, Tau and Phi classes first originating in the streptophyte ancestor of land plants over 629 million years ago (Morris et al., 2018). The DHAR class evolved in the common ancestor of Viridiplantae, placing its origin over 670 million years ago (Morris et al., 2018). The Iota class was present before the divergence of rhodophytes and chlorophytes, placing its origin over 1.4 billion years ago (Betts et al., 2018).

GSTs are widely distributed in aerobic organisms and are thought to have evolved in bacteria in response to oxygen toxicity (Fahey & Sundquist, 1991). Due to the structure of their N-terminal domains GSTs are suggested to have originated from a thioredoxin ancestor to which a C-terminal helical domain was added (Frova, 2006; Martin, 1995). Different models have been presented for the evolution of GST classes. Sheehan et al. (2001) hypothesised that the ancestral GST diverged first into the mitochondrial, microsomal and cytosolic forms, followed by the diversification of cytosolic GSTs. This is consistent with the cytosolic Theta, Zeta and GHR classes, the mitochondrial Kappa class and microsomal MAPEG and mPGES2 classes being distributed among eukaryotes and prokaryotes, suggesting that these classes arose before the prokaryote-eukaryote split. A model based on GST catalytic residues and protein structures (Frova 2006) proposed that all soluble GSTs are derived from the *E. coli* glutaredoxin 2 (GRX2) protein, and the first GST cytosolic classes possessed the GRX2 monomeric form and cysteine active site residue. According to this model DHAR and Lambda classes are among the most ancient. A subsequent dimerization event and the replacement of the cysteine active site residue with a serine led to the Theta, Zeta, EF1Bγ, Phi and Tau classes. Finally a change from serine to tyrosine led to the

classes Sigma, Mu, Alpha and Pi. However since Lambda and DHAR GSTs are only found in plants, this scenario seems unlikely. Based on their distribution in red and brown algae, oomycetes, vertebrates and invertebrates (Hervé et al., 2008), Sigma GSTs are thought to be ancestral to Mu, Pi and Alpha GSTs, emerging in the ancestor of eukaryotes. The discovery of a Mu GST in the red alga *C. merolae* and Pi GST in the chlorophyte alga *C. reinhardtii* in this analysis suggests the Mu and Pi GST classes also originated in early eukaryote evolution.

A number of classes conserved in Archaeplastida were missing from some land plant species which suggests that they were lost in these species or lineages. One explanation is these classes contain genes that specialised into functions that are not required in these lineages. URE2p regulates nitrogen catabolism in yeast (Coschigano & Magasanik, 1991). URE2p GSTs were absent from the angiosperm genomes, suggesting this function was not needed in angiosperms. TCHQD GSTs were absent in *C. merolae*, *C. reinhardtii* and *A. agrestis*. TCHQD from the soil bacterium *Sphingobium chlorophenolicum* catalyses the degradation of the pesticide pentachlorophenol (Warner & Copley, 2007). This function may not have been required in *C. merolae, C. reinhardtii* and *A. agrestis*. Some GSTs may have been lost due to functional overlap of different classes leading to redundancy of certain classes of enzymes. The Tau class was present in all Archaeplastida genomes except the moss *P. patens.* Tau GSTs are associated with abiotic stress responses. The Hemerythrin class is expanded in moss (there are 8 GSTH members in *P. patens* and 1 in other species). Hemerythrin GSTs have been suggested to be involved in heavy metal binding and detoxification (Liu et al., 2013), therefore the lack of Tau GST in *P. patens* may be due to functional overlap of Hemerythrin and Tau enzymes. Hemerythrin GSTs were absent from the angiosperms *A. thaliana* and *O. sativa*. Hemerythrin GSTs may have become redundant as Phi and Tau GST numbers increased in plants. Lambda GSTs regulate flavonoid metabolism and have peroxidase activity. A Phi GST has been found to also regulate flavonoid metabolism in blackgrass (Cummins et al., 2013), therefore there may be functional overlap between these classes which caused Lambda GSTs to be lost in *A. agrestis, S. moellendorffii*, and *M. polymorpha*.

In seed plants, more than 50% of GSTs belong to the Tau and Phi classes. GST Phi class expanded in early land plant evolution, after the divergence of streptophyte algae and land plants, while GST Tau class expanded in vascular plants after their divergence from bryophytes. This suggests that they are under positive selection in these lineages. The expansion of the Phi class in land plants suggests that Phi class diversification was important for the adaption to terrestrial habitats. Phi GSTs are strongly induced in response to abiotic and biotic stresses (Sylvestre-Gonon et al., 2019). Phi GSTs have functions in protection against salinity, drought and oxidative stress, anthocyanin transport and accumulation, and the synthesis of plant defence compounds (Chen et al., 2012; Kitamura et al., 2004; Sappl et al., 2009). The functions of Phi GSTs are consistent with the adaption to new stresses associated with life on land. There is only one Phi GST in the lycophyte *S. moellendorffii*, suggesting that the expansion of Phi GSTs in vascular plants is restricted to seed plants. One Tau GST was found in the hornwort *A. agrestis*, two in the liverwort *M. polymorpha*, and 34 in the lycophyte *S.*

moellendorffii, which indicates a rapid expansion of the Tau class after the divergence of bryophytes and vascular plants. The diversification of Tau GSTs in vascular plants suggests that they provided novel functions needed for vascular plant lifestyles. Tau GSTs are induced in response to environmental stresses such as salinity, temperature, drought, UV radiation, oxidative stress and pathogens (Dixon & Edwards, 2010a; Sylvestre-Gonon et al., 2019). Tau GSTs have functions in secondary metabolism such as synthesis of defence compounds, and binding of porphyrin derivatives, anthocyanins, and fatty acids (Dixon et al., 2008; Dixon & Edwards, 2009; Piślewska-Bednarek et al., 2018). The functions of Tau GSTs are consistent with an increase in exposure to environmental stresses associated with a vascular lifestyle.

Tau and Phi enzymes catalyse the detoxification of diverse substrates by GSH conjugation. GSTs from other classes show no GSH-conjugating activity towards herbicides (Dixon et al., 2009; He et al., 2016; Lan et al., 2010; Liu et al., 2013). GSTs with a cysteine residue in the start of the active site (Lambda, DHAR, GHR, Hemerythrin, Iota) perform deglutathionylation, thiol-transferase and DHAR reactions and cannot conjugate GSH (Lallement et al., 2014), while GSTs with a serine active site residue (Tau, Phi, TCHQD, Theta and Zeta) catalyse GSH conjugation, peroxidase or isomerase reactions (Sylvestre-Gonon et al., 2019). No Theta, Zeta or TCHQD GSTs were found to metabolise herbicides, therefore herbicide metabolism may be a Phi and Tau class-specific trait. However, not all Phi and Tau GSTs can conjugate herbicides (Dixon et al., 2009; Liu et al., 2013). A number of Tau and Phi GSTs from soybean and maize had no activity towards the herbicides chlorimuron ethyl, alachlor and atrazine (McGonigle et al., 2000). One study showed that while all poplar GSTs are under purifying selection, some residues in the C-terminal of poplar Tau GSTs are under positive selection. C-terminals of GSTs contain the substrate binding sites, therefore positive selection in these sites could lead to more diverse substrate specificities in Tau GSTs (Lan et al., 2010). This suggests that rather than any GST being able to metabolise herbicides, the larger diversity of enzymes in the Tau and Phi classes combined with the GSH conjugating activity of Phi and Tau GSTs allows for some GST from these classes to be able to use some herbicides as substrates. Finally, some Phi, Tau and Lambda GSTs confer herbicide tolerance by catalysing GSH peroxidase reactions which reduce toxic ROS which accumulates after herbicide treatment. Therefore there are be multiple mechanisms by which GSTs can confer non-target site herbicide resistance, and this is consistent with GSTs from several other classes being upregulated in NTSR plants (Supplementary Table S3.3).

In conclusion, the evolution of GSTs in Archaeplastida involves a stepwise emergence of new plant-specific classes and many lineage specific losses of GST classes. The classes containing enzymes that detoxify herbicides belong to the Tau and Phi classes with evolved before the divergence of streptophyte algae and land plants. Tau and Phi are the largest GST classes in land plants, having undergone duplications and diversifications at key evolutionary transitions, namely the colonisation of land and evolution of vasculature, and abiotic and biotic stress responses. GST-mediated herbicide resistance is conferred by different mechanisms. The ability to conjugate herbicide chemicals seems restricted to

certain GSTs in the Phi and Tau classes, and is possibly due to the large diversity of enzymes in these classes in some land plants.

3.6. Appendix

Species	C8-like	DHAR	EF1B-γ	GHR	GTT1-like	Heme-rythrin	Iota	Lambda	Metaxin	mPGES2	Mu
Arabidopsis thaliana	0	3	2	4	0	0	0	3	1	1	0
Oryza sativa	0	2	2	2	0	0	0	3	1	1	0
Selaginella moellendorffii	0	4	1	5	0	1	1	0	1	1	0
Anthoceros agrestis	0	2	1	2	0	1	1	0	1	1	0
Physcomitrella patens	0	3	4	3	0	8	1	1	2	0	0
Marchantia polymorpha	0	2	1	2	0	1	1	0	1	1	0
Klebsormidium nitens	0	2	1	3	1	0	1	1	1	1	0
Chlamydomonas reinhardtii	3	1	0	3	0	0	1	0	1	1	0
Cyanidioschyzon merolae	1	0	1	1	1	0	1	0	0	2	1

Species	Phi	Pi	Sigma-like	Tau	TCHQD	Theta	URE2p	Zeta	Kn1	Cr1	Cr2	Total
Arabidopsis thaliana	13	0	0	28	1	3	0	2	0	0	0	61
Oryza sativa	19	0	0	49	1	1	0	4	0	0	0	85
Selaginella moellendorffii	1	0	0	34	1	3	2	2	0	0	0	57
Anthoceros agrestis	11	0	0	1	0	1	2	2	0	0	0	26
Physcomitrella patens	10	0	0	0	5	3	1	1	0	0	0	42
Marchantia polymorpha	18	0	0	2	2	2	1	1	0	0	0	35
Klebsormidium nitens	3	0	0	3	3	1	1	1	1	0	0	24
Chlamydomonas reinhardtii	0	3	3	0	0	1	0	0	0	1	1	19
Cyanidioschyzon merolae	0	0	0	0	0	0	0	1	0	0	0	9

Table S3.1. Glutathione-S-transferase classes and gene numbers in green plants and red algae.

Species	2N	Kappa	MAPEG
Arabidopsis thaliana	2	0	1
Oryza sativa	1	0	1
Selaginella moellendorffii	2	0	2
Anthoceros agrestis	1	0	2
Physcomitrella patens	1	1	4
Marchantia polymorpha	2	1	2
Klebsormidium nitens	3	1	1
Chlamydomonas reinhardtii	1	0	1
Cyanidioschyzon merolae	1	0	1

Table S3.2. Number of GSTs identified from the 2N, Kappa, and MAPEG classes in green plants and red algae. Sequences from these classes were not included in the phylogenetic analysis due to lack of classical N-terminal and C-terminal GST domains. 2N GST sequences have two N-terminal domains and lack a C-terminal domain. Kappa GSTs lack the two classical GST domains and instead have one single thioredoxin-like kappa GST domain (InterPro domain IPR014440). MAPEG GSTs lack both C and N-terminal GST domains and have instead a single 'MAPEG' domain.

Class	Gene name	Evidence	Herbicide chemical class	Genes per clan	References
DHAR	MaDHAR2 (parrot's feather)	Up-regulated in NTSR plants compared to sensitive plants	Sulphonylurea	1	Bai et al., 2018
Lambda	TaGSTL (wheat)	Up-regulated in safener-treated plants	FOP, Sulphonylurea	1	Taylor et al., 2013
MAPEG	MaGSTM1 (parrot's feather)	Up-regulated in NTSR plants compared to sensitive plants	Sulphonylurea	1	Bai et al., 2018
Phi	BsGSTF1 (American sloughgrass)	Up-regulated in NTSR plants compared to sensitive plants	FOP	12	Bai et al., 2020
	MaGSTF11 (parrot's feather)	Up-regulated in NTSR plants compared to sensitive plants	Sulphonylurea		Bai et al., 2018
	ZmGSTI (maize)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetamide		Li et al., 2016
	ZmGSTII (maize)	Induced by safener treatment, up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide		Fuerst et al., 1993, Li et al., 2016
	ZmGSTIII (maize)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide		Fuerst et al., 1993, Li et al., 2016
	AmGSTF1 (blackgrass)	Increased glutathione peroxidase activity and higher protein levels in NTSR plants	Chloroacetanilide, Chloro-s-triazine, DIM, Dinitroaniline , FOP, Phenylurea, Paraquat		Franco-Ortega et al., 2021, Tétard-Jones et al., 2018
	AmGSTF2 (blackgrass)	Enhanced protein levels in NTSR plants compared to sensitive plants	Chloroacetanilide, Chloro-s-triazine, DIM, Dinitroaniline, FOP, Phenylurea, Paraquat		Tétard-Jones et al., 2018
	OsGSTII (rice)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide		Wu et al., 2002
	TaGSTF I-II (wheat)	Up-regulated in safener-treated plants	FOP, Sulphonylurea		Taylor et al., 2013
	LrGSTF1 (annual rye-grass)	Enhanced protein levels in NTSR plants compared to sensitive plants	Organophosphorus, Sulphonylurea, FOP		Cummins et al., 2013
	LrGSTFA (annual rye-grass)	Up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		Duhoux et al., 2017
	Contig16302 (annual rye-grass)	Up-regulated in NTSR plants compared to sensitive plants	FOP		Gaines et al., 2014
Tau	ZmGST5 (maize)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide	14	Li et al., 2017
	ZmGST6 (maize)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide		Li et al., 2017
	ZmGST7 (maize)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide		Li et al., 2017
	EcGST1 (barnyard grass)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide, Carboxylic acid		Li et al., 2013
	OsGSTI (rice)	Up-regulated in NTSR plants compared to sensitive plants	Chloroacetanilide		Wu et al., 2002
	TaGSTUI-I (wheat)	Up-regulated in safener-treated plants	FOP, Sulphonylurea		Taylor et al., 2013
	AmGSTU2 (blackgrass)	Enhanced protein levels in NTSR plants compared to sensitive plants	Chloroacetanilide, Chloro-s-triazine, DIM, Dinitroalanines, FOP, Phenylurea, Paraquat		Tétard-Jones et al., 2018
	LrGSTUI (rye-grass)	Up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		Duhoux et al., 2017
	LrGSTUJ (rye-grass)	Up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		Duhoux et al., 2017
	LrGSTUK (rye-grass)	Up-regulated in NTSR plants compared to sensitive plants	Sulfonylurea		Duhoux et al., 2017
	Contig5390 (rye-grass)	Up-regulated in NTSR plants compared to sensitive plants	FOP		Gaines et al., 2014
	Contig13326 (rye-grass)	Up-regulated in NTSR plants compared to sensitive plants	FOP		Gaines et al., 2014
	BsGSTU1 (American sloughgrass)	Up-regulated in NTSR plants compared to sensitive plants	FOP		Pan et al., 2016
	BsGSTU3 (American sloughgrass)	Up-regulated in NTSR plants compared to sensitive plants	FOP		Pan et al., 2016
Theta	BsGSTT3 (American sloughgrass)	Up-regulated in NTSR plants compared to sensitive plants	FOP	1	Pan et al., 2016

Table S3.3. Enzymes associated with NTSR plants or crops belong to several GST classes.

The organophosphorus class refers to the herbicide glyphosate.

Best Lui 2013 BlastP hits	Species	Group	GST class	Score	E value	Best UniProt BlastP hits	Species	Group	GST class	Score	E value
Query: Kfl00304_0120_v1 (Kn1) – *Klebsormidium nitens*											
AT5G02790	*Arabidopsis thaliana*	Dicot	Lambda	110	6E-27	A0A2P6TTZ1	*Chlorella sorokiniana*	Chlorophyte alga	N.A.	493	9E-58
LOC_Os03g17460	*Oryza sativa*	Monocot	Lambda	109	1E-26	A0A2P6V651	*Micractinium conductrix*	Chlorophyte alga	N.A.	461	5E-53
KC119459	*Physcomitrella patens*	Moss	Lambda	108	1E-26	E1Z4I7	*Chlorella variabilis*	Chlorophyte alga	N.A.	459	9E-49
Query: Cre02.g142200.t1 (Cr1) – *Chlamydomonas reinhardtii*											
NP_001082252	*Xenopus laevis*	Amphibian	Pi	70	5E-15	A0A150FZW3	*Gonium pectorale*	Chlorophyte alga	N.A.	915	5E-123
NP_038569	*Mus musculus*	Mammal	Pi	63	1E-12	A0A383VC28	*Tetradesmus obliquus*	Chlorophyte alga	N.A.	834	1E-110
BAA76974	*Oncorhynchus nerka*	Fish	Pi	57	4E-11	A0A2J7ZHK9	*Tetrabaena socialis*	Chlorophyte alga	Pi	542	2E-67
Query: Cre12.g508850.t1.1 (Cr2) – *Chlamydomonas reinhardtii*											
pdb|1AW9	*Zea mays*	Monocot	Phi	72	3E-15	A0A150H1P1	*Gonium pectorale*	Chlorophyte alga	Phi	1,035	2E-140
KC119467	*Physcomitrella patens*	Moss	Phi	58	3E-11	D8TP51	*Volvox carteri*	Chlorophyte alga	N.A.	790	5E-101
YP_159901	*Aromatoleum aromaticum*	Proteo-bacterium	Beta	56	1E-10	C1FGB4	*Micromonas commoda*	Chlorophyte alga	N.A.	777	6E-101
Query: Cre17.g721350.t1 (CrGSTP1) – *Chlamydomonas reinhardtii*											
NP_001082252	*Xenopus laevis*	Amphibian	Pi	82	2E-18	A0A150GPV5	*Gonium pectorale*	Chlorophyte alga	N.A.	801	2E-104
BAA76974	*Oncorhynchus nerka*	Fish	Pi	78	3E-17	D8U504	*Volvox carteri*	Chlorophyte alga	N.A.	753	2E-97
NP_038569	*Mus musculus*	Mammal	Pi	77	4E-17	D8UF81	*Volvox carteri*	Chlorophyte alga	Pi	291	2E-28
Query : Cre16.g670973.t1 (CrGSTSL1) - *Chlamydomonas reinhardtii*											
AAP79878	*Solenopsis invicta*	Arthropod	Sigma-Like	89	1E-20	D8U137	*Volvox carteri*	Chlorophyte alga	N.A.	777	1E-102
NP_508625	*Caenorhabditis elegans*	Nematode	Sigma-Like	84	3E-19	A0A150FTV9	*Gonium pectorale*	Chlorophyte alga	N.A.	740	4E-97
CAA71620	*Ascaris suum*	Nematode	Sigma-Like	82	1E-18	A0A2J7ZU59	*Tetrabaena socialis*	Chlorophyte alga	N.A.	637	2E-81
Query: CMT103C (CmGSTMu1) - *Cyanidioschyzon merolae*											
NP_000839	*Homo sapiens*	Primate	Mu	99	2E-23	A9UU35	*Monosiga brevicollis*	Choano-flagellate	N.A.	299	4E-29
AAO15607	*Sarcoptes scabiei*	Arthropod	Mu	96	1E-22	P15626	*Trichechus manatus*	Mammal	N.A.	285	3E-27
AAM64045	*Taenia solium*	Flatworm	Mu	93	1E-21	A0A2Y9E6B3	*Mus musculus*	Mammal	Mu	269	6E-25
Query: Cre15.g636750.t1.2 (CrGSTC8L1) - *Chlamydomonas reinhardtii*											
NP_767413	*Bradyrhizobium japonicum*	Rhizobium	C-8	81	3E-18	A0A150FZN5	*Gonium pectorale*	Chlorophyte alga	N.A.	793	1E-105
YP_353980	*Rhodobacter sphaeroides*	Proteo-bacterium	C-8	80	5E-18	D8UG08	*Volvox carteri*	Chlorophyte alga	N.A.	786	4E-104
YP_367374	*Burkholderia*	Proteo-bacterium	N2_C2	78	2E-17	A0A383WFY7	*Tetradesmus obliquus*	Chlorophyte alga	N.A.	709	2E-92
Query: CMO318C (CmGTT1L1) - *Cyanidioschyzon merolae*											
AAU38224	*Mannheimia succiniciproducens*	Proteo-bacterium	GTT1L	196	4E-53	Q13JR6	*Paraburkholderia xenovorans*	Proteo-bacterium	N.A.	743	3E-97
AAZ19390	*Psychrobacter arcticus*	Proteo-bacterium	GTT1L	174	2E-46	A0A2Z7C5U5	*Dorcoceras hygrometricum*	Streptophyte	N.A.	723	1E-93
XP_002171454	*Schizosaccharomyces japonicus*	Fungus	GTT1L	160	4E-42	A0A239SHF3	*Pandoraea sputorum*	Proteo-bacterium	N.A.	704	2.E-91
Query: AagrOXF_evm.model.utg000005l.356.1 (AaGSTU1) – *Anthoceros agrestis*											
AAF64450	*Euphorbia esula*	Dicot	Tau	155	8E-41	A0A199VMR2	*Ananas comosus*	Monocot	Tau	442	3E-51
AT1G27130	*Arabidopsis thaliana*	Dicot	Tau	140	3E-36	A0A2R6WAI6	*Marchantia polymorpha*	Liverwort	Tau	437	2E-50
Medtr5g076900	*Medicago truncatula*	Dicot	Tau	137	3E-35	A0A5B6V3U1	*Gossypium australe*	Dicot	Tau	430	2E-49

Table S3.4. Results of BlastP searches of unclassified GST protein sequences against the list of 152 GST protein sequences representing 36 GST classes from Liu et al. (2013) and the UniProt database.

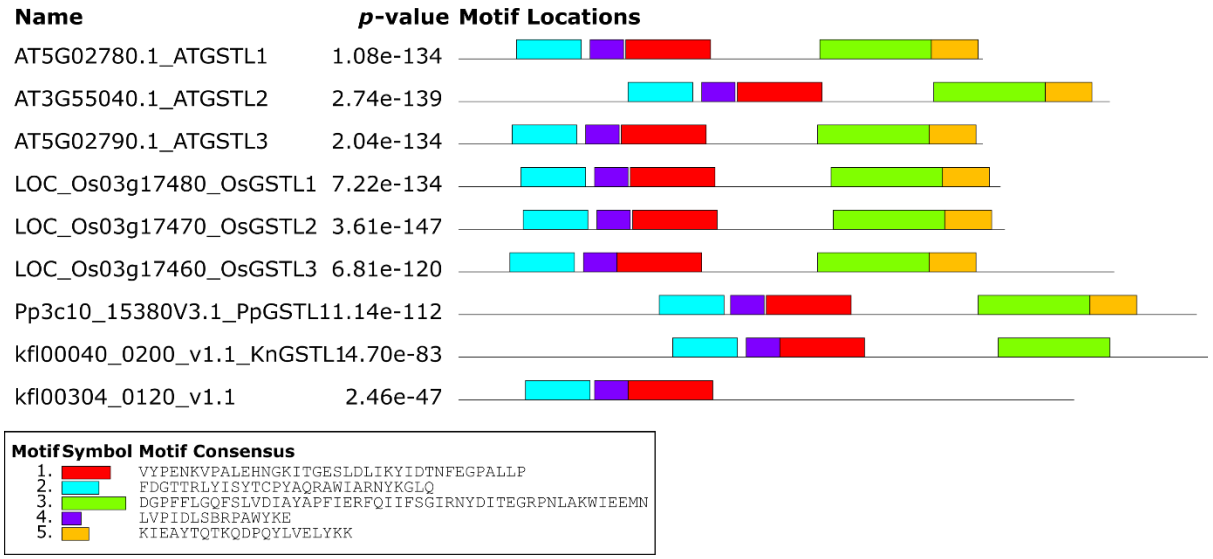


Figure S1. Motif search of Kn1 and all Lambda GST sequences confirms Kn1 is not a Lambda GST.

The MEME protein domain search (http:// memesuite.org/) (Bailey et al., 2009) confirms that the Kn1 protein (kfl00304_0120_v1.1) possesses the Lambda N terminal motifs but does not possess the C terminal motifs found in lambda class GST protein sequences.

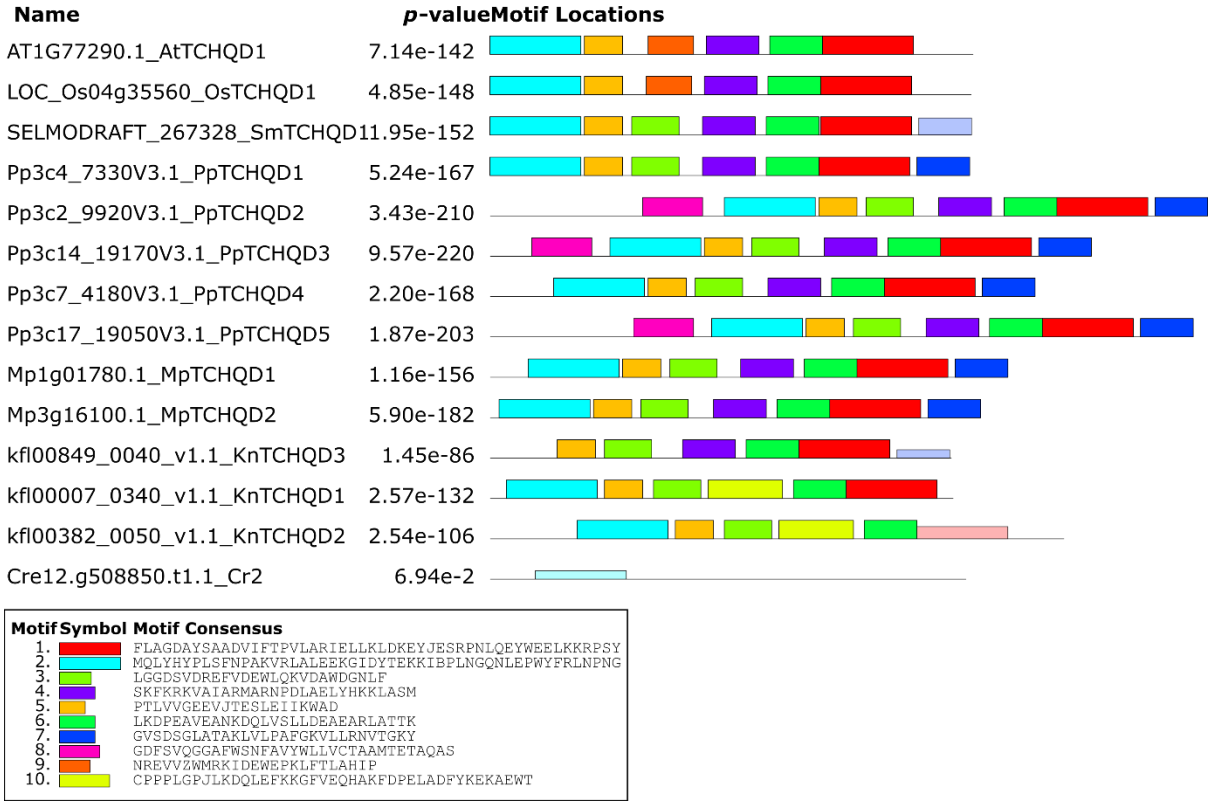
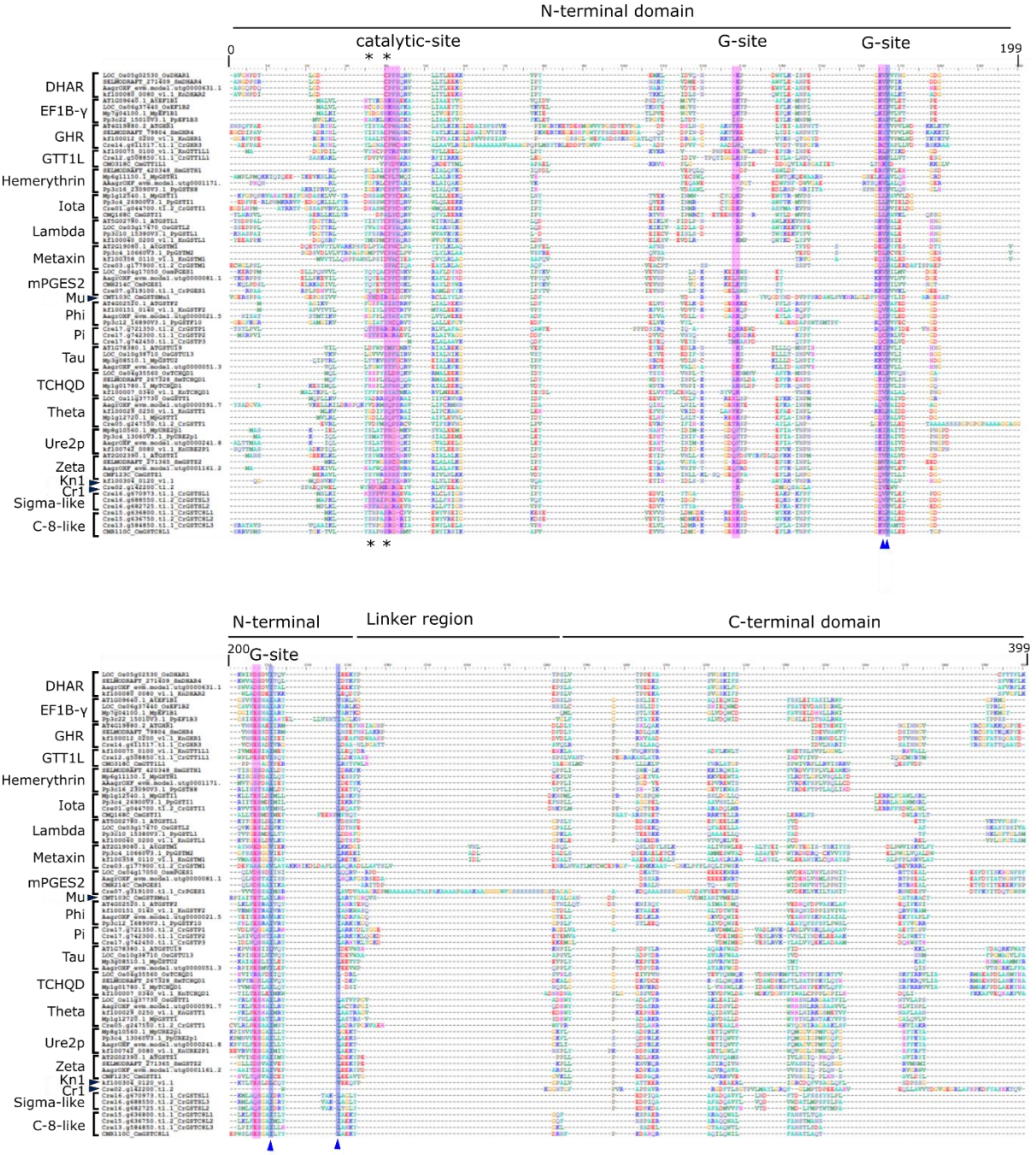


Figure S2. Motif search of Cr2 and all TCHQD GST sequences confirms Cr2 is not a TCQHD class GST.

The MEME protein domain search (http:// memesuite.org/) (Bailey et al., 2009) confirms that the Cr2 protein (Cre12.g508850.t1.1_Cr2) lacks the motifs found in all TCHQD class GST protein sequences.


N-terminal domain
catalytic-site
G-site
G-site
0
199
DHAR
EF1B-γ
GHR
GTT1L
Hemerythrin
Iota
Lambda
Metaxin
mPGES2
Mu
Phi
Pi
Tau
TCHQD
Theta
Ure2p
Zeta
Kn1
Cr1
Sigma-like
C-8-like
N-terminal
Linker region
C-terminal domain
200
G-site
399

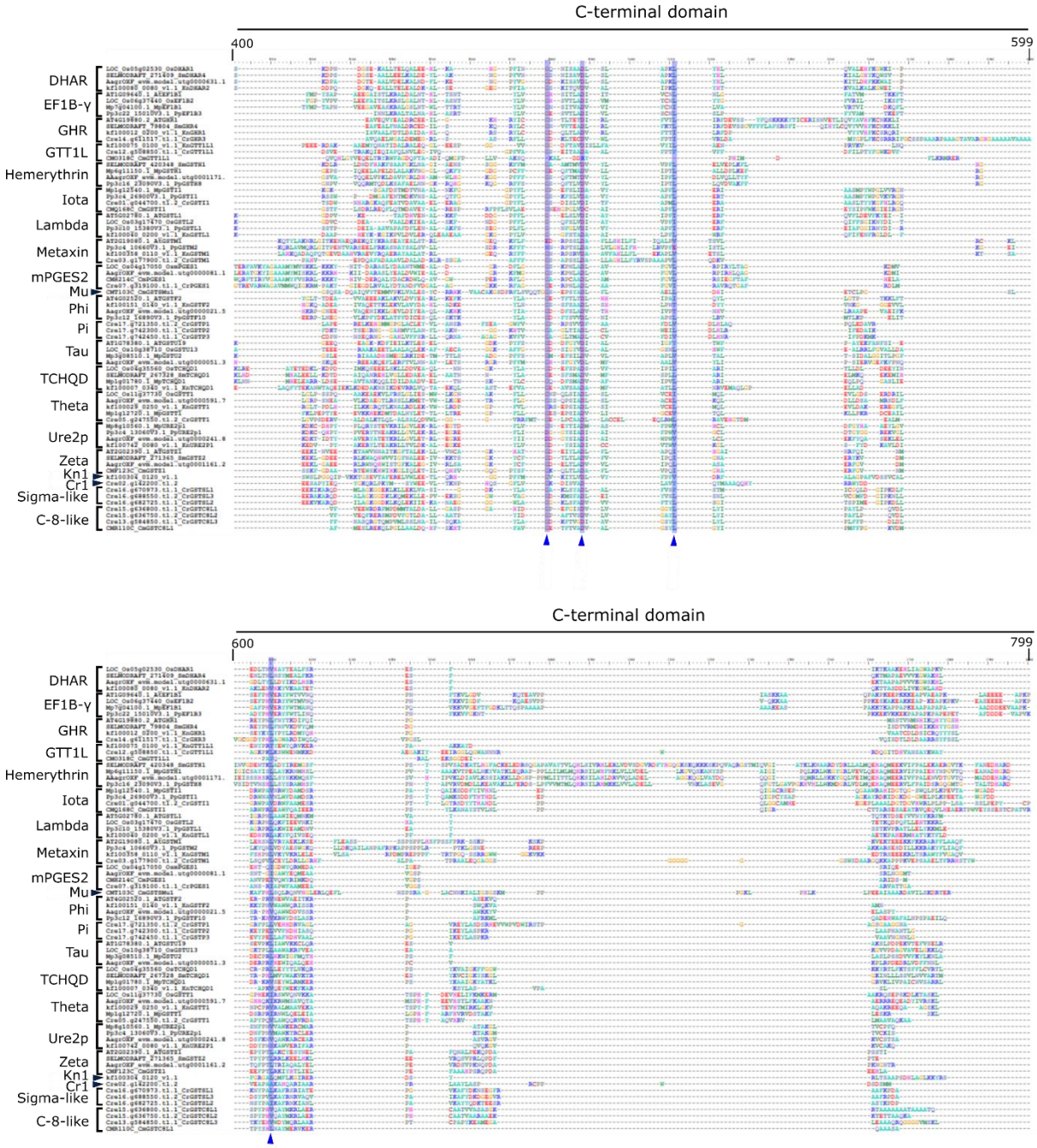


Figure S3. Amino acid alignment of representative plant GST proteins showing the location of conserved GST domains. GSTs are formed of a N-terminal domain and a less conserved C-terminal domain. Four representative sequences from different species are shown for each GST class. The locations of the catalytic residues in different classes are indicated with an asterisk. Sites that bind GSH (G-sites) are marked in pink. Residues conserved in at least 80% of samples are indicated by blue arrows. GSTHs and GSTIs have large domains that extend past the C-terminal domain end which haven't been included in the figure. Large gaps caused by single sequences were removed for clarity. The catalytic motifs are consistent with those identified by (P. A. Lallement et al., 2014) G-site residues are based on the crystal structure of TaGSTU4 (Thom et al., 2002).

Chapter 4

Mp*CYP822A1* up-regulated by chlorsulfuron confers low-level chlorsulfuron resistance when over-expressed in *Marchantia polymorpha*

4.1. Abstract

Herbicide resistance can result from mutations in the gene encoding the herbicide target (target site resistance, TSR), or by reducing the amount of herbicide reaching the target (non-target site resistance, NTSR). NTSR presents a considerable threat to agriculture because it can confer resistance to many herbicides, and is poorly characterised, likely involving several genes and biochemical pathways. This research aims to identify genes encoding NTSR proteins in the haploid liverwort *Marchantia polymorpha* that are induced by the herbicide chlorsulfuron which inhibits acetolactate synthase (ALS), the first common enzyme in the biosynthetic pathway of branched-chain amino acids. Transcriptomes were generated to identify transcriptomic differences between treated and untreated susceptible plants and between treated and untreated chlorsulfuron target-site resistant (TSR) plants.

Differential gene expression analysis identified 2982 genes that were differentially expressed in chlorsulfuron-treated compared to untreated susceptible plants. GO enrichment analysis identified branched-chain amino acid biosynthetic and metabolic processes as the most significantly enriched GO terms in up-regulated genes, while photosynthesis and DNA replication were the most significantly enriched GO terms in down-regulated genes. Several glutathione-s transferase and cytochrome P450 genes were expressed at higher levels in herbicide-treated susceptible *M. polymorpha* than in untreated plants. Four of these genes – Mp*CYP813A5*, Mp*CYP822A1*, Mp*GSTF10* and Mp*GSTF15* – were over-expressed in *M. polymorpha* to test their role in chlorsulfuron sensitivity. One gene, Mp*CYP822A1*, belonging to the 71 clan, conferred resistance to a non-lethal chlorsulfuron dose when over-expressed in *M. polymorpha*. This is the first laboratory experiment identifying a NTSR gene from herbicide-treated sensitive plants, and the first NTSR gene identified in the CYP822 family. These findings suggest that the ability of CYPs to confer herbicide resistance is conserved in land plants, and the low level of resistance acquired by the over-expression of one gene suggests that NTSR to a lethal chlorsulfuron dose requires the increased expression of several genes.

Fewer than 20 genes were differentially expressed in chlorsulfuron treated TSR plants than in untreated plants. Of these, none were from families with known roles in NTSR. This result suggests that an inducible suite of herbicide detoxifying enzymes does not exist in plants, and NTSR evolves from the genes that form part of the herbicide transcriptional stress response.

4.2. Introduction

The over-reliance on herbicides over several decades has led to the global evolution of resistant weeds (Heap, 2021). There are two main mechanisms of herbicide resistance: resistance resulting from mutations in the gene encoding the herbicide target (target site resistance, TSR), or resistance conferred by the reduction of the amount of herbicide reaching the herbicide target (non-target site resistance, NTSR). The most studied form of NTSR is metabolic resistance, defined as the increased ability to chemically modify herbicides into non-toxic forms. Gene families that participate in herbicide metabolic resistance are glutathione s-transferases (GSTs), cytochrome P450s (CYPs), glycosyl transferases (GTs) and ATP-binding cassette (ABC) transporters (Yu & Powles, 2014a; Yuan et al., 2007). Peroxidases, esterases, and hydrolases have also been implicated in NTSR mechanisms (Cummins et al., 1999; Duhoux et al., 2017). NTSR presents a considerable threat to agriculture because it can confer resistance to herbicides with different chemical structures including novel herbicides, making weed control more challenging. Furthermore, the complex biochemical pathways that enable NTSR are likely to involve many genes and are poorly understood (Délye, 2013).

Acetolactate synthase (ALS) is the first common enzyme in the biosynthetic pathway of branched-chain amino acids valine, leucine and isoleucine. Inhibition of ALS stops synthesis of branched chain amino acids which leads to lack of protein synthesis, arrested growth and eventually plant death (Liu et al., 2017). ALS-inhibiting herbicides such as chlorsulfuron are widely used in agriculture. There have been several reported cases of NTSR to ALS inhibitors in monocot and dicot weeds (Tranel & Wright, 2002; Yu & Powles, 2014b). RNA-Seq has allowed the identification of candidate NTSR genes to ALS inhibiting herbicides in weed species such as *Echinochloa phyllopogon* (Iwakami et al., 2014, 2019) water chickweed (*Myosoton aquaticum*) (Liu et al., 2018), blackgrass (*Alopecurus myosuroides*) (Gardin et al., 2015) and rye-grass (*Lolium* spp.) (Duhoux et al., 2015) and in the crops maize (*Zea mays*) (Liu et al., 2018) and cotton (Thyssen et al., 2018). However the high genetic diversity and lack of lab techniques in weedy species makes the functional characterisation of candidate NTSR genes challenging.

It is generally acknowledged that NTSR mechanisms arise from pathways involved in herbicide response (Christophe Délye, 2013; J. S. Yuan et al., 2007), however the timing of the activation of the response pathways before or after the herbicide binds to the target is not clear. I predict that a subset of the genes up-regulated by the ALS-inhibitor chlorsulfuron in *M. polymorpha* will function in chlorsulfuron metabolism, and that chlorsulfuron treatment of Mp*ALS* target-site resistant (TSR) plants will induce specific herbicide detoxifying enzymes and won't alter the expression of genes related to the stress caused by Mp*ALS* inhibition. A recent study comparing transcriptional responses towards ALS-inhibitors of *A. thaliana* and the monocot weeds blackgrass (*Alopecurus aequalis*) and rye-grass (*Lolium* spp.) found that 71% of ALS-inhibitor response marker genes in grasses matched their expression in *A. thaliana* (Délye et al., 2018), suggesting that pathways affected by ALS inhibition are highly conserved across evolutionarily distant species. As I

demonstrated in chapters 2 and 3, *M. polymorpha* has smaller CYP and GSTs gene families than angiosperms, while retaining the same CYP clans and GST classes that contain NTSR genes in angiosperm genomes. Thus I predict that there is less functional redundancy between genes in the CYP and GSTs gene families in *M. polymorpha* than in weeds, therefore the genetic manipulation of an Mp*CYP* or Mp*GST* may more likely yield a phenotype. *M. polymorpha* therefore represents a good system for studying a complex genetic trait such as herbicide NTSR.

In this chapter I aim to identify genes with functions in chlorsulfuron metabolism in *M. polymorpha*. I describe the transcriptomic analysis of sensitive (Tak-2) and TSR *M. polymorpha* plants treated with a non-lethal dose of chlorsulfuron during a 0-48 hour time-course. A large number of genes were differentially expressed in Tak-2 herbicide treated plants compared to untreated plants, while only a few genes were differentially expressed in the two TSR mutants. Several Mp*CYP* and Mp*GST* genes were up-regulated in chlorsulfuron treated Tak-2 plants but not in chlorsulfuron-treated TSR plants, and their expression was confirmed using quantitative real time PCR. Five of the most strongly induced genes in Tak-2 – Mp*CYP813A5*, Mp*CYP822A1*, MpGSTF7, Mp*GSTF10* and Mp*GSTF15* – were cloned into over-expression vectors and transformed in *M. polymorpha* spores. Lines over-expressing Mp*CYP813A5*, Mp*CYP822A1*, Mp*GSTF10* and Mp*GSTF15* were tested in a chlorsulfuron sensitivity growth assay. One gene, Mp*CYP822A1*, conferred a low level of resistance to chlorsulfuron when over-expressed, compared to control plants. These findings show that genes conferring NTSR are part of the herbicide transcriptional stress response, and the low level of resistance conferred by the over-expression of one gene suggests that for NTSR to be observed in the field, NTSR to chlorsulfuron is likely to require the increased expression of multiple genes in *M. polymorpha*.

4.3. Materials and Methods

4.3.1. Plant materials and growth conditions

The two laboratory *Marchantia polymorpha* lines Takaragaike-1 (Tak-1) (male) and Takaragaike-2 (Tak-2) (female) were used as wild-type plants (Ishizaki et al., 2008). Several target-site resistant (Mp*ALS*^{tsr}) plants to chlorsulfuron were previously generated in the lab via UV-B mutagenesis. Plants were propagated asexually from gemmae and grown in light cabinets under 30 $\mu molm^{-2}s^{-1}$ continuous white light at 23°C.

Plants used in the transcriptome analysis, quantitative real time PCR verification of transcriptome data and chlorsulfuron dose response assay with cellulose discs were grown on 100x21mm round petri dishes on top of autoclaved cellulose discs (A.A. Packaging Ltd, Preston, UK) placed on the surface of 30 ml Johnson's medium with 0.8% agar (see Appendix 4.6.2). Plants were transferred from the herbicide-free medium to herbicide-containing medium by lifting the discs on which plants were growing from one plate to another. Plants used in the other experiments were grown on 120x120mm square petri dishes with 60mL ½ Gamborg medium with 1% sucrose and 0.8% agar (see Appendix 4.6.2).

To generate spores for transformations, two-week old Tak-1 and Tak-2 plants were grown on a 1:3 mixture of vermiculite and John Innes No.2 compost and incubated in a growth chamber under a 16h light:8h dark photoperiod at 23°C. Gametophore production was induced under far red light irradiation as described in (Chiyoda et al., 2008).

4.3.2. Sequencing Mp*ALS* in Tak-2 and Mp*ALS*^{tsr} mutants

Several chlorsulfuron Mp*ALS*^{tsr} mutant *M. polymorpha* plants were previously generated in the lab via UV-B mutagenesis. To determine the mutations in their ALS gene, whole plant DNA was extracted from two mutants (Mp*ALS*^{tsr} _7 and Mp*ALS*^{tsr} _10) and wild type (Tak-2) using the CTAB method adapted from Porebski et al., (1997). A 1x1 cm piece of plant tissue was harvested in a 1.5mL Eppendorf tube and frozen in liquid N_2, then ground in a tissuelyser (Qiagen). 500µL of CTAB buffer was added to the tube and incubated at room temperature for 15 min. 250µL of chloroform: isoamylalcohol (24:1) was added, mixed well and centrifuged for 1 min at 13000rpm. The upper phase was placed in a new tube to which another 250µL of chloroform: isoamylalcohol (24:1) was added, mixed and centrifuged. The upper phase was moved to a new tube with 500µL isopropanol and centrifuged for 10 min at 13000rpm. The supernatant was discarded and 500µL 70% ethanol was added, followed by centrifugation 5 min at full speed. The ethanol was removed to dry the pellet, and the DNA was dissolved in 100µL H_20. DNA concentration was measured on a Nanodrop ND-1000 spectrophotometer.

A 140 bp region of the *M. polymorpha* ALS gene that contains the most commonly found mutations conferring resistance to ALS-inhibitors was amplified by polymerase chain reaction (PCR) using the PCRBIO HiFi polymerase and buffer (PCR Biosystems, Cat. No. PB10.41-02). The amplification was carried out using a BIO-RAD Dyad Peltier thermal cycler. Each

reaction was conducted in a 50 µl mixture, consisting of 10 µl reaction buffer, 2 µl of forward primer, 2 µl of reverse primer, 5 µl of DNA, 0.5 µl of HiFi Polymerase, and 30.5 µl of DNase free milli-Q H2O, with three replicates per sample. The PCR program consisted of 1 min initial denaturation phase at 95°C, followed by 32 cycles, with each cycle consisting of a 15s incubation time at 95°, a 15s incubation at 56°C annealing temperature, and a 30s extension time at 72°C. Amplified DNA was extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The quantity of DNA was measured on a Nanodrop ND-1000 spectrophotometer. Samples were sequenced by Sanger sequencing (Source Bioscience, Oxford, UK). The primers used for PCR amplification and sequencing can be found in section 4.3.13.

4.3.3. Chlorsulfuron effect on *M. polymorpha* growth

A whole plant dose-response experiment was carried out to determine the sensitivity of wild-type and target-site resistant plants to chlorsulfuron. Chlorsulfuron, 1-(2-chlorophenylsulfonyl)-3-(4-methyoxy-6-methyl-1,3,5-triazin-2-yl)urea, was supplied by Sigma Aldrich and diluted in milli-Q H_2O, then filter sterilized with a 0.2uM sterile syringe filter. Chlorsulfuron was pipetted into melted medium and mixed thoroughly before pouring into plates. The female *M. polymorpha* accession (Tak-2) was chosen as the susceptible biotype to chlorsulfuron, because it tends to grow flatter on the medium than Tak-1 making it easier to image. Tak-2, Mp*ALS*^{tsr}_7 and Mp*ALS*^{tsr}_10 gemmae were grown for 7 days then transferred onto herbicide treated plates and replaced in the growth cabinet for 7 days before imaging. Tak-2 plants were treated with ten herbicide doses: 0nM, 0.1nM, 0.2nM, 0.5nM, 1nM, 1.5nM, 2.5nM, 5nM, 10nM, and 30nM chlorsulfuron. Mp*ALS*^{tsr} plants were treated with ten herbicide doses: 1nM, 2nM, 5nM, 10nM, 20nM, 50nM, 100nM and 300nM chlorsulfuron. There were two replicate plates per herbicide dose and five plants grown per plate. The plants were imaged using a Panasonic Lumix camera (model no. DMC-FS7) on day 7 after transfer to herbicide plates. Lateral growth was quantified using Image J thresholding. Data were expressed as % of the average area of untreated plants and were plotted with a non-linear fit, sigmoidal, four-parameter regression curve using GraphPad Prism 9.1.0 (GraphPad Software, Inc., San Diego, CA, USA). The herbicide rate or dose resulting in 50% growth reduction (GR_{50}) was calculated as follows:

$$y = C + \frac{D - C}{1 + (x/ED_{50})^b}$$

In which C is the lower limit, D is the upper limit, ED_{50} is the effective dose which reduced growth by 50%, and b is the slope at ED_{50}. Resistance index was then calculated as follows:

$$RI = \frac{GR_{50}(S)}{GR_{50}(R)}$$

A second chlorsulfuron dose response analysis was conducted to determine the effect of chlorsulfuron on growth and survival in wild type plants grown from day 0 on chlorsulfuron. Tak-1 and Tak-2 gemmae were grown on 120x120mm square petri dishes and left to grow for 14 days before imaging. Plants were treated with 10 herbicide doses: 0nM, 0.1nM,

0.3nM, 1nM, 3nM, 10nM, 30nM, 100nM, 300nM, and 1µM chlorsulfuron. There were two replicate plates per herbicide dose and nine plants grown per plate. Chlorophyll autofluorescence was imaged using a fluorescence imaging system (NightOWL II LB983, Berthold Technologies, Bad Wildbad, Germany) and was used to measure plant area (mm^2) as a proxy for growth.

4.3.4. RNA-Seq experimental design and RNA extraction and sequencing

Total RNA was extracted from Tak-2 and target-site resistant mutants Mp*ALS*^{tsr} _7 and Mp*ALS*^{tsr} _10 across eight timepoints: 0h, 2h, 4h, 6h, 12h, 24h, 36h and 48h after transferring plants to herbicide-treated (5nM) or untreated plates. The criteria for the selection of the herbicide dose was a dose causing a toxic effect without killing the plant, and was chosen based on the results from the dose response curve. There were two replicate samples per timepoint, and six plants per sample. Total RNA was extracted from 90 samples.

Gemmae were grown for 8 days on cellulose discs on untreated medium then moved onto chlorsulfuron-treated medium or untreated control plates. Plants were collected at the different timepoints after transfer onto chlorsulfuron and immediately frozen in liquid N_2 and kept at -80°C. Samples were pulverised with a tissuelyser (Qiagen) and total RNA was extracted using RNeasy plant mini kits (Qiagen, Hilden, Germany). To remove DNA, RNA samples were DNase treated using the DNA-free kit by Ambion (Life Technologies) following manufacturer instructions. The quantity of RNA was measured on a Nanodrop ND-1000 spectrophotometer and RNA integrity was measured on an Agilent 2100 BioAnalyser (Agilent Technologies). The RNA integrity number (RIN) provided by the BioAnalyzer software ranges from 1 (a very degraded sample) to 10 (a mostly intact sample). Samples with RIN numbers ≥8 were suitable for cDNA library preparation following the Illumina RNA prep protocols (Welcome Trust Oxford Genomics Centre, Oxford). Illumina sequencing of 75 bp paired-end reads was conducted on the 90 cDNA libraries by an Illumina HiSeq4000. Samples were run over five lanes, and coverage was approximately 12 million reads per sample.

4.3.5. Read pre-processing and mapping to reference transcriptome

Raw reads were returned as fastq.gz files. These were unzipped and their quality was viewed using FastQC v.0.11.7. Illumina adaptors and low quality tails were quality trimmed using Trimmomatic-0.38 (Bolger et al., 2014). Paired-end reads were interleaved and ribosomal RNA was removed using Sortmerna-2.1 (Kopylova et al., 2012) and error corrected using BayesHammer (SPAdes-3.10.0) (Nikolenko et al., 2013) with setting–only-error-correction. Clean reads were mapped to a reference transcriptome (JGI 3.1, obtained from https://marchantia.info/download/v31/) using Salmon (Patro et al., 2017).

4.3.6. Differential expression analysis

The differences in expression of each gene per transcript between sample pairs were calculated using the DESeq 2 v.1.21.3 package (Love et al., 2014) in the statistical software R. P-values were adjusted following the Benjamini-Hochberg procedure for controlling the false discovery rate (Benjamini & Hochberg, 1995). The criteria for differential gene expression was an adjusted p-value < 0.05 and a $|\log_2(\text{fold-change})| \geq 1$ (a fold-change of expression ≥ 2) between compared groups. Volcano plots were produced for each comparison. Expression differences were compared between untreated and treated Tak-2 2h-48h after chlorsulfuron treatment.

4.3.7. Gene Ontology term enrichment analysis

The VISEAGO R package (Brionne et al., 2019) was used to identify GO terms significantly enriched in differentially expressed genes between treated and untreated plants, using the classic algorithm and Fisher's exact test. The functional annotations obtained from Marpolbase (https://marchantia.info/download/MpTak_v6.1/) for the MpTakv6.1 transcriptome were used to create a custom GO database, because the transcriptome used for the RNA-Seq analysis (JGI 3.1) is poorly annotated with a large number of obsolete GO annotations. Takv6.1 annotations were associated with JGI 3.1 gene names using the gene correspondence table provided by Marpolbase. Multidimensional plots were generated showing clusters of GO terms into functional groups based on Wang's semantic similarity distance and ward.D2 aggregation criterion.

4.3.8. Real time quantitative PCR validation of RNA-Seq results

The RNA-Seq experimental setup was repeated with fewer timepoints for real time quantitative PCR (RT-qPCR) validation. Plants were transferred to control or herbicide treated plates eight days after growth and collected 6, 12, 24, and 36 hours after treatment. Samples were collected and RNA was extracted and DNase treated as described in section 4.3.4. 1µg RNA was converted into cDNA using ProtoScript II RT (NEB), oligo(dT)s and Murine RNAse inhibitor (NEB) in a 20µL reaction volume according to manufacturer instructions. Specific primers for the Mp*CYP* and Mp*GST* genes identified in the transcriptome analysis were designed using the Primer3Plus tool (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi). Mp*APT* and Mp*ACT* were selected as reference genes (Saint-Marcoux et al., 2015). The list of primers used for RT-qPCR can be found in 4.3.13. Primer efficiencies were calculated for each primer pair by building a standard curve with a 1:5 dilution series of pooled cDNA (50, 10, 2, 0.4, 0.08ng) and ranged between 1.8-2.2.

RT-qPCR experiments were performed with three biological replicates using a QuantStudio 7 cycler (Applied Biosystems). The 10µL reaction mixture contained 5µL 2X SYBR Green/ROX master mix (Applied Biosystems), 2µL of 1:20 diluted cDNA, and 0.6µL each of the forward and reverse primer, and 2.4µL of RNAse free water. Each biological replicate was run in two technical replicates. The QuantStudio7 cycling method was as follows: 50°C for 2 min, 95°C for 10 min, then

40 cycles of 95°C for 15s and 60°C for 1 min. A melt curve analysis was conducted after PCR amplification to check primer specificity, with temperatures ranging from 60°C to 95°C increasing in 0.05°C s^{-1} increments. Relative expression was calculated using an adapted $2^{-(\Delta\Delta Ct)}$ method that accounts for differences in primer efficiencies between the gene of interest and the reference gene(s), and incorporates multiple control genes described in (Hellemans et al., 2008). Technical replicate Ct values were averaged for each biological replicate. Each biological replicate Ct was then subtracted from the average Ct of the 3 untreated biological replicates to give ΔCt values. The calculated primer efficiency of each primer pair to the power of ΔCt gave the relative quantity (RQ). Expression of the gene of interest in each sample was calculated by dividing the RQ of that gene by the geometric mean of the RQ of the two reference genes in the same sample to give the normalised relative quantity (NRQ) (Vandesompele et al., 2002). Significant differences between treated and untreated groups for each gene and timepoint were calculated by students t-tests.

4.3.9. Cloning of overexpression constructs

DNA was extracted from the Tak-2 wild type line using a CTAB extraction method adapted from (Porebski et al., 1997) and described in section 4.3.4. Each of the up-regulated genes (Mp*GSTF7*, Mp*GSTF10*, Mp*GSTF15*, Mp*CYP813A5*, Mp*CYP822A1*) was PCR amplified from genomic Tak-2 DNA using CloneAmp HiFi Premix (Takara Bio). The primers used to amplify gene sequences can be found in 4.3.10. PCR products were run on 1% agarose gel and fragments of the correct size were extracted and purified using the GeneJET Gel Extraction Kit (ThermoFisher) following manufacturer instructions. The purified products were individually cloned into a directional entry vector pENTR D-TOPO and transformed into *E. coli*. The plasmid was extracted and purified using a Minprep Kit (Qiagen), and then cloned by LR reaction into destination vector pMpGWB403 carrying the constitutive endogenous elongation factor alpha (Mp*EF1α)* promoter (Ishizaki et al., 2015). The final vectors were transformed into *Agrobacterium tumefaciens* by electroporation.

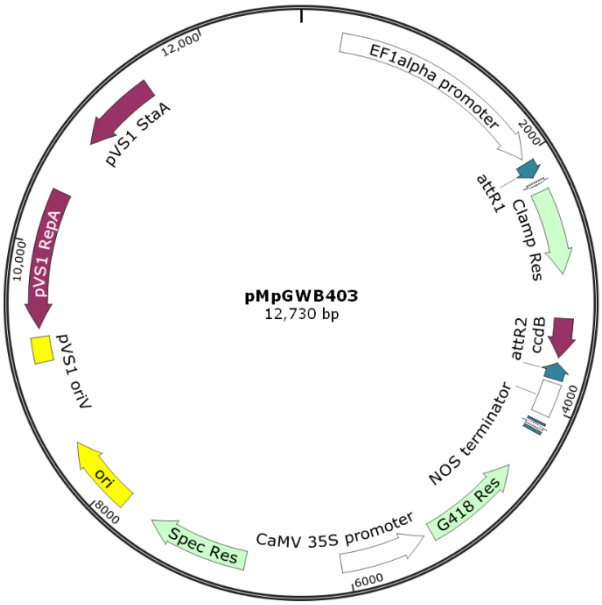


Figure 4.1. Vector map of $_{pro}$Mp*EF1α* overexpression construct. The full length DNA sequences of the genes of interest (MpGSTF7, MpGSTF10, MpGSTF15, MpCYP813A5, MpCYP822A1) were inserted into a pENTR D-TOPO entry vector which was then cloned into pMpGWB403 by Gateway cloning.

4.3.10. Agrobacterium-mediated transformation of *M. polymorpha* sporelings

M. polymorpha spores were transformed following the method described by (Ishizaki et al., 2008). In summary, fresh sporangia from a Tak-1 x Tak-2 cross were sterilised with 1% sodium dichloroisocyanurate (NADCC, Sigma). 1-2 sterilised sporangia per transformation were burst and spores were grown in 25ml M51C medium in 125ml vented tissue culture flasks for 7 days under continuous white light (30 $\mu molm^{-2}s^{-1}$) at 23°C. One colony of Agrobacterium strain GV3101 containing the desired vector was inoculated in 5 mL LB medium and 50$ugmL^{-1}$ of gentamycin, rifampicin and spectinomycin, and was cultured shaking at 28°C overnight. The 5mL culture was pelleted by centrifugation at 4000 rpm for 15 minutes, and resuspended in 10 ml M51C liquid medium containing 100µM acetosyringone (Sigma Aldrich). The culture was grown for 6 hours shaking at 28°C. 1mL of the induced agrobacterium culture and 100uM acetosyringone were added to the 7 day old spore liquid culture. Spores and bacteria were left to co-cultivate for two days shaking at 120 rpm at 18°C under continuous white light (56 $\mu molm^{-2}s^{-1}$). Sporelings were subsequently collected on a sterile 40µM cell strainer (Fisher Scientific) and washed with 150 mL sterile H_20. The sporelings were spread on Johnson's medium with 150$ugmL^{-1}$ cefotaxime to kill the bacteria and 20$ugmL^{-1}$ geneticin to select for transformants. After 2 weeks transformants were moved onto fresh selection plates.

4.3.11. Real time quantitative PCR of Mp*GST* and Mp*CYP* over-expression transformants

RNA was extracted from 14-day old gemmalings from five independent lines transformed with an overexpression construct containing one of five chlorsulfuron-induced genes (Mp*GSTF7*, Mp*GSTF10*, Mp*GSTF15*, Mp*CYP813A5*, Mp*CYP822A1*). RNA was extracted using the RNeasy plant mini kit (Qiagen, Hilden, Germany). RNA samples were DNase treated using the DNA-free kit by Ambion (Life Technologies) according to manufacturer instructions. RNA was converted to cDNA as described in section 4.3.8. The list of primers used for RT-qPCR can be found in 4.3.13. RT-qPCR experiments were performed with three biological replicates using a QuantStudio7 cycler (Applied Biosystems). Each biological replicate sample was run in two technical replicates. The reaction mixture, cycling method and relative expression calculations were conducted as described in 4.3.6.

4.3.12. Herbicide sensitivity assay of over-expression lines

To characterise the sensitivity of Mp*GST* and Mp*CYP* over-expressing plants to chlorsulfuron, gemmae over-expressing one of the four candidate genes (Mp*GSTF7*, Mp*GSTF10*, Mp*CYP813A5* and Mp*CYP822A1*) and from an empty vector control were grown on medium supplemented with 30nM chlorsulfuron (non-lethal dose), 200nM chlorsulfuron (lethal dose) or no

herbicide. Significant differences in growth between each line and the empty vector control within each treatment were calculated by students t-tests.

4.3.13. Primers

Gene	Experiment	Forward primer	Reverse primer	Product size (bp)
MpALS	PCR amplification	TCATTGGTGTTATAATCGTTGAAGTGG	GCTAGTGGTTCGTGTTAATGGA	140
	Sequencing	TTGGTCCGACTCATCATTTG		
MpACT	qPCR	AGGCATCTGGTATCCACGAG	ACATGGTCGTTCCTCCAGAC	108
MpAPT	qPCR	CGAAAGCCCAAGAAGCTACC	GTACCCCCGGTTGCAATAAG	146
MpCYP73A109	qPCR	GCAAACAACCCCGAACACTG	ACAGCGGAATGGCGATGATG	154
MpCYP73A110	qPCR	GAAAGACCAGGGGCAATCGA	GAATCACCATCAGCGTTGCC	145
MpCYP94L1	qPCR	AGCGATGGTGCTGACATTCA	GTACCAGTTTGCTCTCGGCT	122
MpCYP704AH1	qPCR	ACCTGCGACGGAAACATCTT	CAGCTGCCGTTGATTCACAC	125
MpCYP704N1	qPCR	CCGGATGGCACCCGTATTAA	CTGGTTGGAAGACTCCGTCC	139
MpCYP710A16	qPCR	TCCAATGACCATGCGGCTAC	GCTTCGTGGTATGCAAAGCC	177
MpCYP813A5	qPCR	CGATGGTCACGGTGGTCTTG	AATCAGAAGTGCTGACGCGA	129
	Overexpression	CACCTCGGGTAAGACTGGTC	TAGCAGTCAGGCCAGGATAT	2892
MpCYP822A1	qPCR	CCACACGTGAACATGGAGGA	GCTTGAGCACAAAAGGGTCG	127
MpCYP822B2	qPCR	GTTGTCGGTGTAGGGAAGCA	GTGAGAACGGGGATAGCTCG	110
	Overexpression	CACC**ATG**GAGAATGCTACGTTT	CGGCTGTGCCAATGAATT **CTA**	2113
MpCYP823A2	qPCR	ACAAGGACGCGATGGACTTC	CCTGCCAAAACCATGTCGTG	109
MpCYP829E1	qPCR	TGAGTGTGTGGCAGCCTTAG	TTTCGTCGTGGGCTCAATGA	94
MpGSTF7	qPCR	ATACAACGGCCAAGGCAAGA	TCTGTCACGCCCTAGATGGA	92
	Overexpression	CACC**ATG**GCGATCAAGATTCAT	AGCTTCAATTCGACAGTGATT **TCA**	1189
MpGSTF8	qPCR	TCACTATCGAATGGTAGTTGCA	ATGCAGGAAAGCCCTCAAGG	142
	Overexpression	CACC**ATG**GCGATCAAGATTCAT	CACAAAGAGTCTTAGCGGCCG	1246
MpGSTF10	qPCR	GGGTCAGGGTACTCCTCTGT	TTCCTGAGTTCCTCCCCGAT	123
	Overexpression	CACC**ATG**GCGAACACGATTTTT	CGATCGATG **TCA**ATAATTGTTGGTGTA	1093
MpGSTF15	qPCR	ATCGGGATATTTCCAGGGCATG	GTTCAAGAGGCGTGCTGAAAG	105
	Overexpression	CACC**ATG**GCCATTCAGATCTAT	**TCA**TGGGATATGTCCTCCCAG	1544
MpGSTH1	qPCR	CGCTCAGACTGTATGGCGAT	TCTGCGTCCACTCTTTGTCC	114

The overhang for Gateway cloning is underlined. Start and stop codons are in bold.

4.4. Results

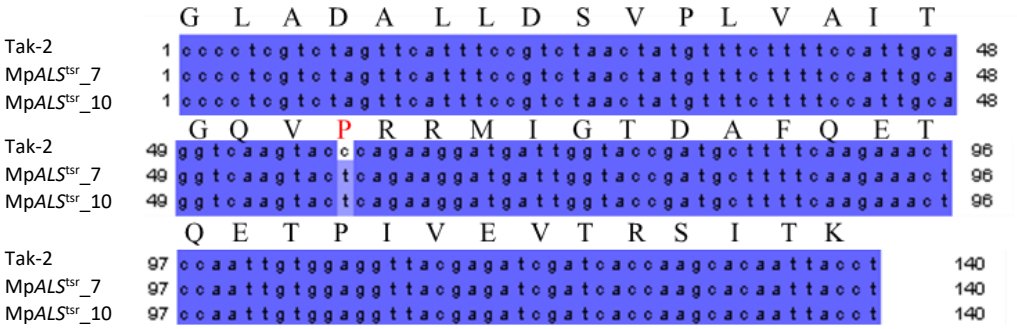
4.4.1. Sensitive and target site resistant *M. polymorpha* plants show differential growth response to chlorsulfuron

To determine the response of young wild type (Tak-2) *Marchantia polymorpha* plants to chlorsulfuron treatment, a whole plant dose-response experiment was carried out using 9 doses of chlorsulfuron (Figure 4.2B). Plants were frown for 7 days on control plates then moved onto chlorsulfuron-treated plates for another 7 days before imaging. Growth reductions were observed by 0.2nM chlorsulfuron treatment under which plants were 20% smaller than untreated plants, while 0.34nM chlorsulfuron reduced plant growth by 50%. There was no additional growth inhibition at concentrations above 10nM. By 30nM chlorsulfuron plants were going brown after one week on the herbicide treated plates. This shows that very low chlorsulfuron doses are sufficient to stop growth in *M. polymorpha* and is consistent with the potency of ALS inhibiting herbicides at very low application rates in agriculture.

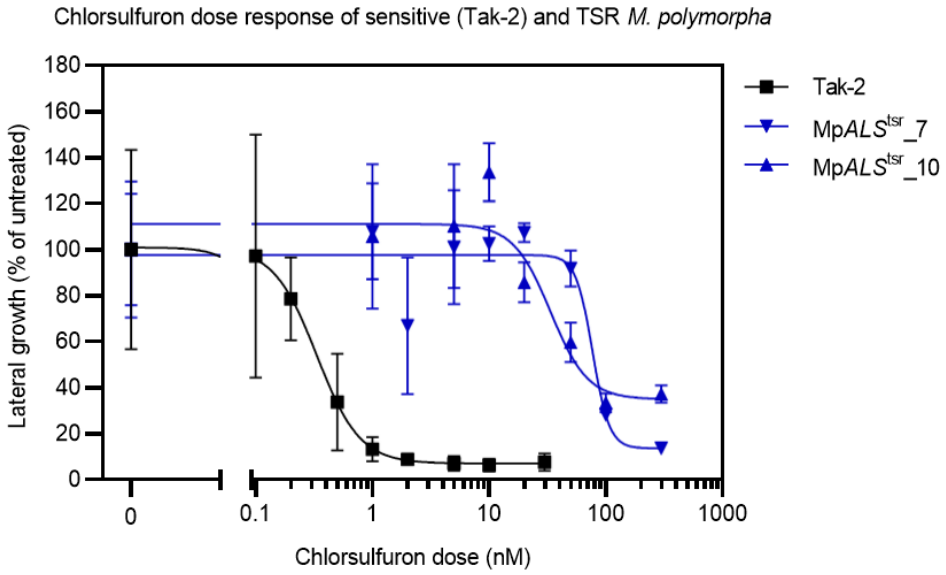
M. polymorpha plants that are resistant to chlorsulfuron because of mutations in the gene encoding the herbicide target Mp*ALS* were also tested for their sensitivity to chlorsulfuron. Both Mp*ALS* target-site resistant (TSR) plants, Mp*ALS*^{tsr}_7 and Mp*ALS*^{tsr}_10, were generated by UV-B mutagenesis, and their ALS gene was sequenced to identify the TSR mutation. A 140 bp region of the *M. polymorpha* ALS gene was amplified and sequenced in two mutants, with Tak-2 used as a control. Both mutants share the same mutation in their target site, a single base change of a cytosine to a thymine leading to the translation of a leucine instead of a proline, corresponding to Pro197 in the *A. thaliana* ALS protein sequence (Figure 4.2A). A mutation in Pro197 is the most commonly found TSR mutation in ALS in weeds (Tranel & Wright, 2002; Yu & Powles, 2014a). Both Mp*ALS*^{tsr}_7 and Mp*ALS*^{tsr}_10 showed much higher levels of resistance to chlorsulfuron compared to Tak-2 plants (Figure 4.2B). 76.45nM and 31.45nM chlorsulfuron were required to reduce growth by 50% compared to untreated plants in Mp*ALS*^{tsr}_7 and Mp*ALS*^{tsr}_10, respectively.

In summary, chlorsulfuron was toxic at very low concentrations in wild type sensitive *M polymorpha* plants. Plants with a point mutation in their Mp*ALS* gene sequence causing an amino acid change are 101 (SE = 56.98) and 234 (SE = 69.42) fold more resistant to chlorsulfuron than sensitive plants (Figure 4.2C). Based on the results from this assay, 5nM was selected as the herbicide dose for the transcriptomic analysis, because it was considered toxic enough in Tak-2 plants to cause a strong (90%) growth inhibition without killing the plants, and did not cause any detectable growth reduction in Mp*ALS*^{tsr} plants.

A.



B.



C.

M. polymorpha biotype	GR_{50} ± SE (nM)	RI ± SE
Tak-2	0.33 ± 0.06	1
Mp*ALS*^{tsr}_7	77.19 ± 8.92	233.91 ± 69.42
Mp*ALS*^{tsr}_10	33.65 ± 12.94	100.59 ± 56.98

Figure 4.2. Effect of chlorsulfuron on growth of sensitive (Tak-2), and target-site resistant (MpALS^{tsr}_7 and MpALS^{tsr}_10) *M. polymorpha*.

A. Alignment of the 140bp sequenced region of the *Marchantia* ALS gene containing the point mutation in mutant plants that are resistant to chlorsulfuron. Tak-2 plants are susceptible and represent the wild type sequence. Colouring is based on % identity. The corresponding wild-type amino acid sequence is above the alignment. The substituted proline is in red.

B. Dose response curve of the growth of two-week-old wild-type (Tak-2) *M. polymorpha* gemmae after seven days on chlorsulfuron-treated medium. Data plotted with a non-linear fit, sigmoidal, four-parameter regression curve. n =5. Error bars are ±SE.

C. GR_{50} and resistance index (RI) values for susceptible (Tak-2) and resistant (Mp*ALS*^{tsr}_7 and Mp*ALS*^{tsr}_10) *M. polymorpha* plants calculated from B in GraphPad Prism. GR_{50} is the herbicide dose that causes a 50% reduction in growth. RI = GR_{50}(TSR)/GR_{50}(Tak-2).

4.4.2. A large induction of differentially expressed genes occurs 6-48h after chlorsulfuron treatment in sensitive plants

An RNA-Seq experiment was performed using susceptible (Tak-2) and Mp*ALS* target-site resistant (TSR) *M. polymorpha* plants to identify genes that are differentially expressed following treatment with the herbicide chlorsulfuron. Our hypothesis was that the herbicide detoxification response would be induced in chlorsulfuron-treated Mp*ALS*^{tsr} plants, but genes involved in the general stress response that occurs in treated sensitive plants would not be induced. We reasoned that comparing the responses of treated Mp*ALS*^{tsr} plants with treated susceptible plants would help us to select potential chlorsulfuron metabolising enzymes. Because the Mp*ALS*^{tsr} plants were generated by UV-B mutagenesis, two mutants with the same mutation in their target-site were chosen to control for differences caused by mutations in the background generated by the UV-B rays. Samples were collected before transfer (timepoint 0) and 2, 4, 6, 12, 24, 36 and 48 hours after transferring plants to chlorsulfuron-treated or control plates. 90 cDNA libraries were generated and the 75 bp paired-end reads were sequenced by Illumina sequencing. Samples were run over five lanes, and coverage was approximately 12 million reads per sample. Reads were error-corrected and ribosomal RNA was removed. The RNA-Seq analysis had a coverage of approximately 12 million reads per sample. Comparisons of gene expression were made between treated and untreated plants at each timepoint after treatment.

The numbers of differentially expressed genes between treated and untreated plants in the three biotypes can be found in Figure 4.3. 2982 genes were significantly differentially expressed (significance < 0.05 and a Log_2(fold change)≥1 or ≤-1) between chlorsulfuron-treated and untreated Tak-2 plants in at least one timepoint, which represents 1/3 of all expressed genes (11953) in the samples (Figure 4.3A & D). Only a handful of genes – 20 in Mp*ALS*^{tsr}_7 (Supplementary Figure S4.2) and 2 in Mp*ALS*^{tsr}_10 (Supplementary Figure S4.3) - were differentially expressed in treated compared to untreated plants in the Mp*ALS*^{tsr} mutants (Figure 4.3A & B). None of the differentially expressed genes were shared between the three biotypes (Figure 4.3B).

Of the 2982 differentially expressed genes between treated and untreated Tak-2 plants, 1612 and 1370 were significantly up-regulated or down-regulated in treated plants in at least one timepoint, respectively. Induced differential expression was clear from 6 hours after treatment, with 675 and 241 genes transcriptionally up-regulated and down-regulated, respectively, in treated compared to untreated plants (Figure 4.3C). This transcriptional response reduced slightly 12 (612 DEGs) and 24 hours (221 DEGs) after treatment and increased in the two later timepoints after chlorsulfuron treatment, with 891 genes transcriptionally up-regulated in chlorsulfuron-treated plants and 1284 genes down-regulated, relative to untreated plants 48 hours after treatment.

Volcano plots for each Tak-2 comparison can be found in Figure 4.4. Volcano plots are scatter plots with Log_2(fold change) on the x axis and significance (adjusted p-value) on the y axis, and are used to visualise the most biologically significant genes i.e. genes with large fold changes that are also strongly significant. Most biologically significant up-

regulated genes include Mapoly0037s0014 (ABC transporter, 6 hours after treatment), Mapoly0045s0149 (Molybdenum cofactor biosynthesis pathway protein, 6 hours after treatment), Mapoly0098s0002 (abscisic stress-ripening protein, 12 hours after treatment), Mapoly0048s0068 (cyclin-B related protein, 24 hours after treatment), Mapoly0002s0043 (SAP motif containing protein, 24 hours after treatment), Mapoly0063s0075 (S-adenosylmethionine synthetase, 36 and 48 hours after treatment), Mapoly0193s0021 (Cupin, 36 hours after treatment), Mapoly0190s0002 (NAD dependent epimerase/dehydratase family protein, 48 hours after treatment), Mapoly0180s0030 (Cupin, 48 hours after treatment). Genes that are most significantly down-regulated include Mapoly0172s0018 (Histone H4, 12 hours after treatment), Mapoly190s0002 (NAD dependent epimerase/dehydratase family, 36 hours after treatment), Mapoly0009s0175 (Peroxidase, 48 hours after treatment) and Mapoly0044s0029 (Aquaporin, 48 hours after treatment).

The functions of the above listed genes are consistent with an abiotic stress response. ATP-binding cassette (ABC) transporters are associated with herbicide NTSR by sequestering in the vacuole detoxified products of herbicide chemicals and/or toxic molecules accumulating as a result of herbicide treatment (Kang et al., 2011). SAP is a DNA binding motif involved in DNA repair, apoptosis, and oxidative stress (Aravind & Koonin, 2000). Cupins are associated with seed germination and are linked to general stress responses (Gábrišová et al., 2016). S-adenosylmethionine (SAM) is a methyl donor for DNA methylation (Yu et al., 2012). Over-expression of SAM synthetase is associated with increased tolerance to abiotic stress (Ezaki et al., 2016).

In summary, chlorsulfuron treatment led to the differential expression of hundreds of genes in Tak-2 plants, whereas chlorsulfuron treatment resulted in the differential expression of only a handful of genes in Mp*ALS*^{tsr} plants. None of the differentially expressed genes in Mp*ALS*^{tsr} plants had obvious functions in herbicide detoxification. In addition, none of the differentially expressed genes in Tak-2, Mp*ALS*^{tsr}_7 and Mp*ALS*^{tsr}_10 were shared between the three biotypes. There was a large induction of differential gene expression from 6 hours after treatment in Tak-2 plants, which generally increased throughout the 48 hours. Of the 2982 genes that were differentially expressed in Tak-2 plants at any one timepoint, genes that had the largest and most significant fold changes in expression had functions in membrane transport, DNA replication and repair and abiotic stress response.

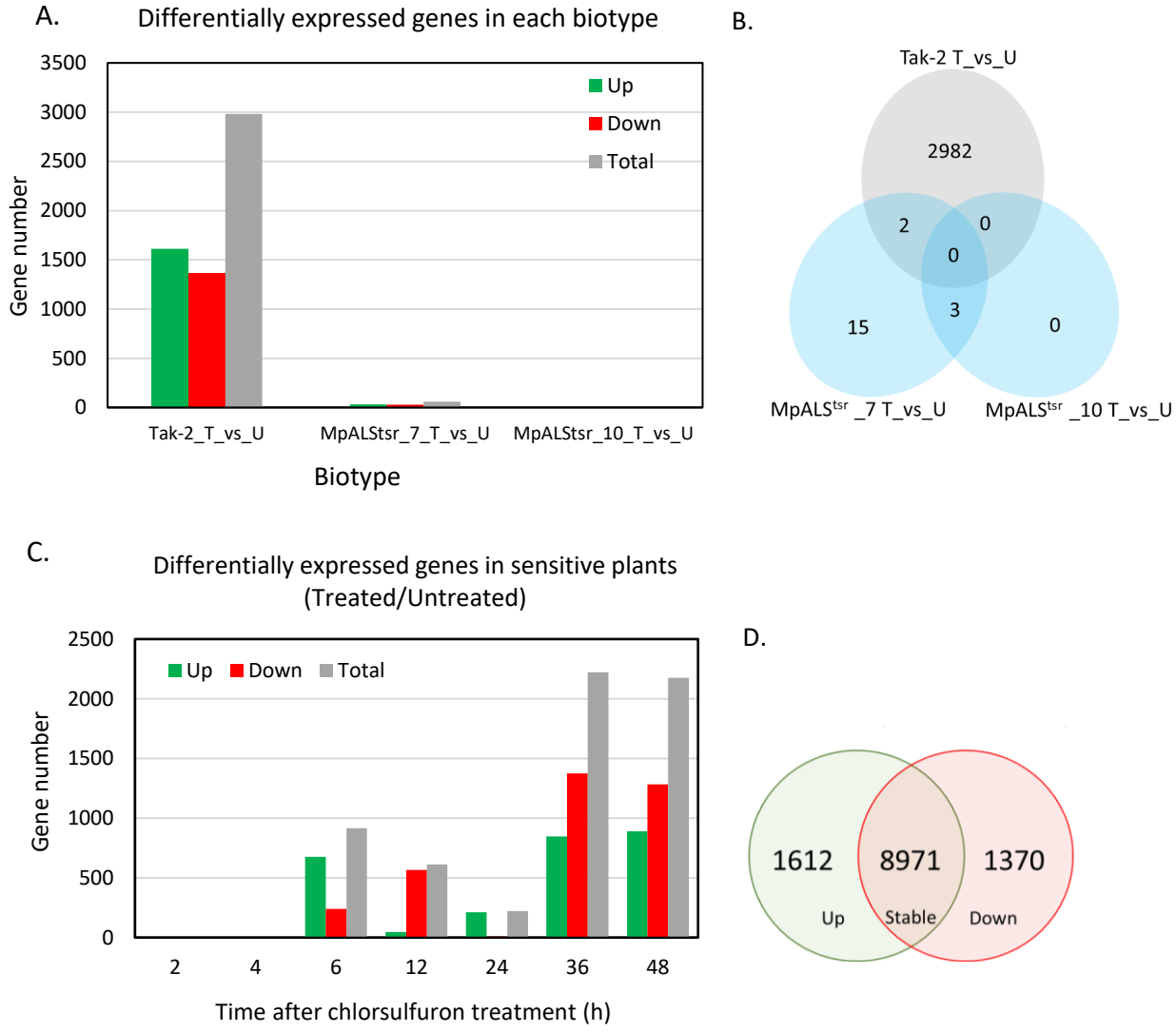


Figure 4.3. A large induction of differentially expressed genes occurs 6-48h after chlorsulfuron treatment in sensitive *M. polymorpha*

Total number of genes up-regulated (green), down-regulated (red), and differentially expressed (up- and down-regulated) (grey) in chlorsulfuron treated plants compared to untreated plants within each genotype (A) and in treated compared to untreated Tak-2 plants at each timepoint 2-48h after treatment (C). (B) Number of differentially expressed genes shared between Tak-2, Mp*ALS*^{tsr}_7 and Mp*ALS*^{tsr}_10 treated vs untreated plants. (D) Number of genes up-regulated, down-regulated or stably expressed between treated and untreated Tak-2 plants across the 48h time-course.

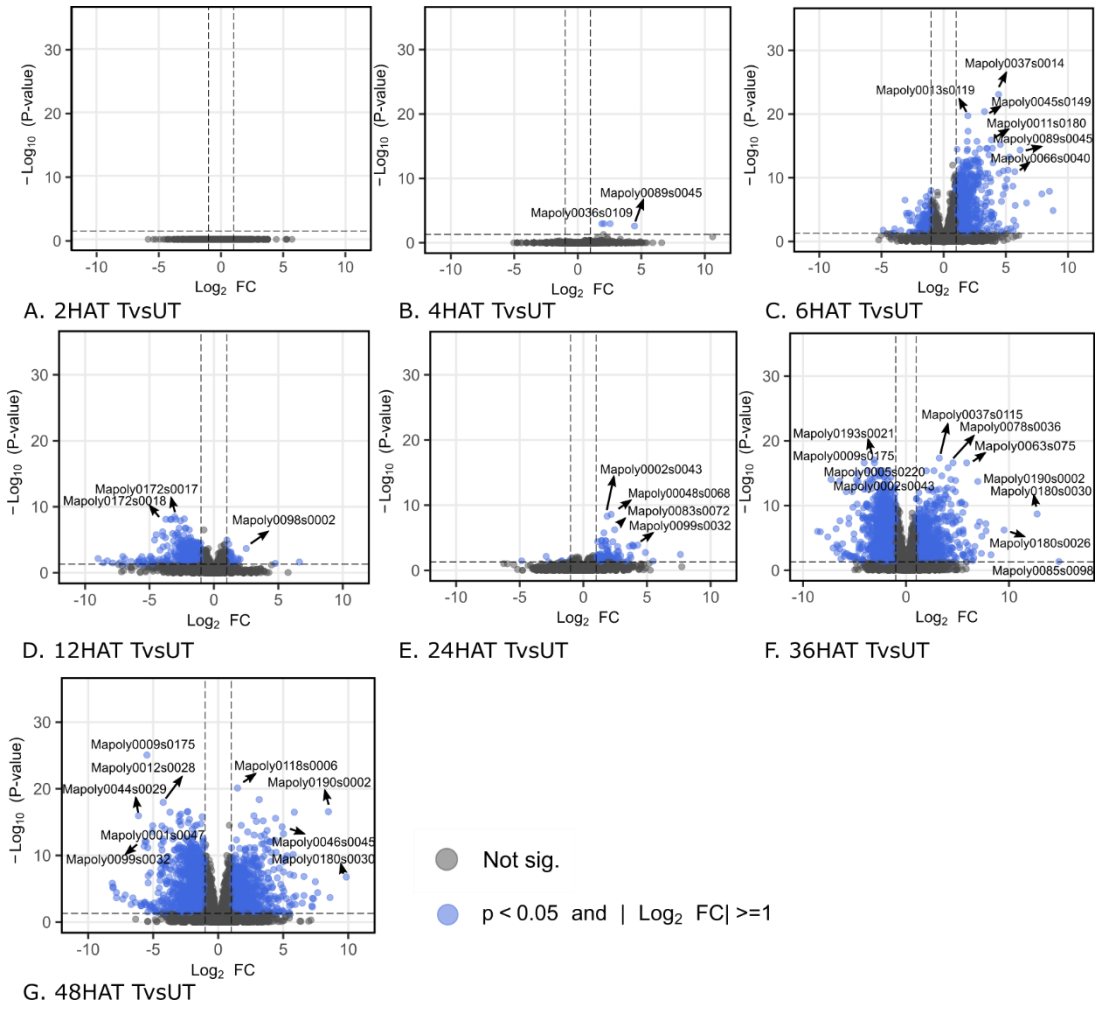


Figure 4.4. Volcano plots showing genes that are differentially expressed between chlorsulfuron-treated and untreated sensitive *M. polymorpha* 2-48 hours after treatment.

Blue dots represent genes with an adjusted p-value < 0.05 and showing a minimum 2-fold change in expression compared to untreated Tak-2 plants ($|Log_2FC| \geq 1$). Grey dots are genes that fall below the thresholds (broken lines) for the log_2FC and p-value. A negative Log_2(fold-change) indicates that the gene is down-regulated and a positive Log_2(fold-change) indicates that the gene is up-regulated.

4.4.3. GO terms associated with branched chain amino acid pathway, amino acids metabolism and defence response are significantly enriched in up-regulated genes in treated plants

GO term enrichment analysis allows the identification of functions or processes that are enriched in a list of genes. GO enrichment analysis using the VISEAGO package separated GO terms into three groups - biological processes, molecular functions and cellular components. 23 biological processes, 21 molecular functions and 9 cellular components were significantly enriched in up-regulated genes in chlorsulfuron treated plants compared to untreated plants (Figure 4.5). Most significantly enriched terms in up-regulated genes are branch chain amino acid metabolic processes (GO:0009081; *p*-value 4.5 10^{-07}, gene frequency 8/9) and biosynthetic processes (GO:0009082; *p*-value 3.2 10^{-06}, gene frequency 7/8). This is consistent with a transcriptional response to the perturbation of the branched chain amino acid pathway.

Defence response is the third most significantly enriched GO term (GO:006952; *p*-value 2.0 10^{-05}, gene frequency 12/26) in up-regulated genes in response to chlorsulfuron. The defence response GO term is defined as 'reactions triggered in response to the presence of a foreign body or the occurrence of an injury, which result in restriction of damage to the organism attacked or prevention/recovery from the infection caused by the attack.' Up-regulated genes annotated with the defence response GO id represent 8 genes encoding Bet v1-like proteins (Mapoly0151s0036, Mapoly0063s0018, Mapoly0063s0019, Mapoly0063s0020, Mapoly0063s0024, Mapoly0063s0025, Mapoly0064s0111, Mapoly0100s0057), two *barley wound induced* (Barwin) domain containing genes (Mapoly0040s0024, Mapoly0050s0033) and a *Mildew Locus O* (MLO) gene (Mapoly0175s0016). Bet v1 proteins are part of the pathogenesis-related (PR) protein superfamily and are involved in the biosynthesis of various compounds such as flavonoids and alkaloids (J. S. Morris et al., 2021). Barwin proteins are also part of the PR superfamily and are induced by fungal infection or wounding (Bravo et al., 2003). MLOs are membrane-bound proteins that are associated with stress-induced calcium signalling (M. C. Kim et al., 2002). They act as negative regulators of plant immunity in response to powdery mildew (Pessina et al., 2016). These genes are unlikely to function in herbicide detoxification but are possibly involved in signalling and general stress response pathways.

25 biological processes, 31 molecular functions and 22 cellular components were enriched in down-regulated genes (Figure 4.6). The most significantly enriched terms in down-regulated genes are photosynthesis (GO:0015979, *p*-value 1.6 10^{-10}, gene frequency 26/54), DNA replication (GO: 0006260, *p*-value 1.0 10^{-09}, gene frequency 25/54) and oxidoreductase activity (GO:0016491, *p*-value 8.5 10^{-16}, gene frequency 193/1031).

Enriched GO terms were clustered into functional groups by semantic similarity using the VISEAGO package based on Wang's semantic similarity distance (Figure 4.7), helping to visualise entire pathways that are disrupted. This analysis found that GO terms enriched in up-regulated genes group broadly into amino acid catabolic and biosynthetic processes (12 GO terms), macromolecule metabolic processes (5 GO terms), amino sugar metabolic processes (4 GO terms) and defence processes (2 GO terms). Molecular functions enriched in up-regulated genes group functionally into catalytic,

oxidoreductase and lyase activity (8 GO terms), binding activity (6 GO terms), transferase activity (4 GO terms) and acid ammonia ligase activity (3 GO terms). These results suggest that chlorsulfuron treatment is causing protein breakdown, combined with the induction of defence processes to mitigate damage caused to the plant. GO terms enriched in down-regulated genes group into DNA replication, mitosis and tetrapyrrole biosynthetic processes (10 GO terms), hydrogen peroxide, ROS and photosynthetic processes (5 GO terms), cell wall processes (4 GO terms), stress response and carbon utilisation (3 GO terms) and pigment metabolism (3 GO terms). This suggests that herbicide treatment is having an inhibitory effect on cell division and photosynthesis. Overall, the differentially expressed genes between chlorsulfuron treated and untreated plants have functions consistent with a response to abiotic stress.

In summary, several different GO terms were significantly enriched in differentially expressed genes between chlorsulfuron treated and untreated plants. The most significant GO terms in up-regulated genes are associated with the branched chain amino acid biosynthesis and metabolism, and defence response. The most significant GO terms represented in down-regulated genes are related to photosynthesis, DNA replication and oxidoreductase activity. Overall, the differentially expressed genes have functions that are consistent with plants experiencing abiotic stress and disruption of branched chain amino acid metabolism.

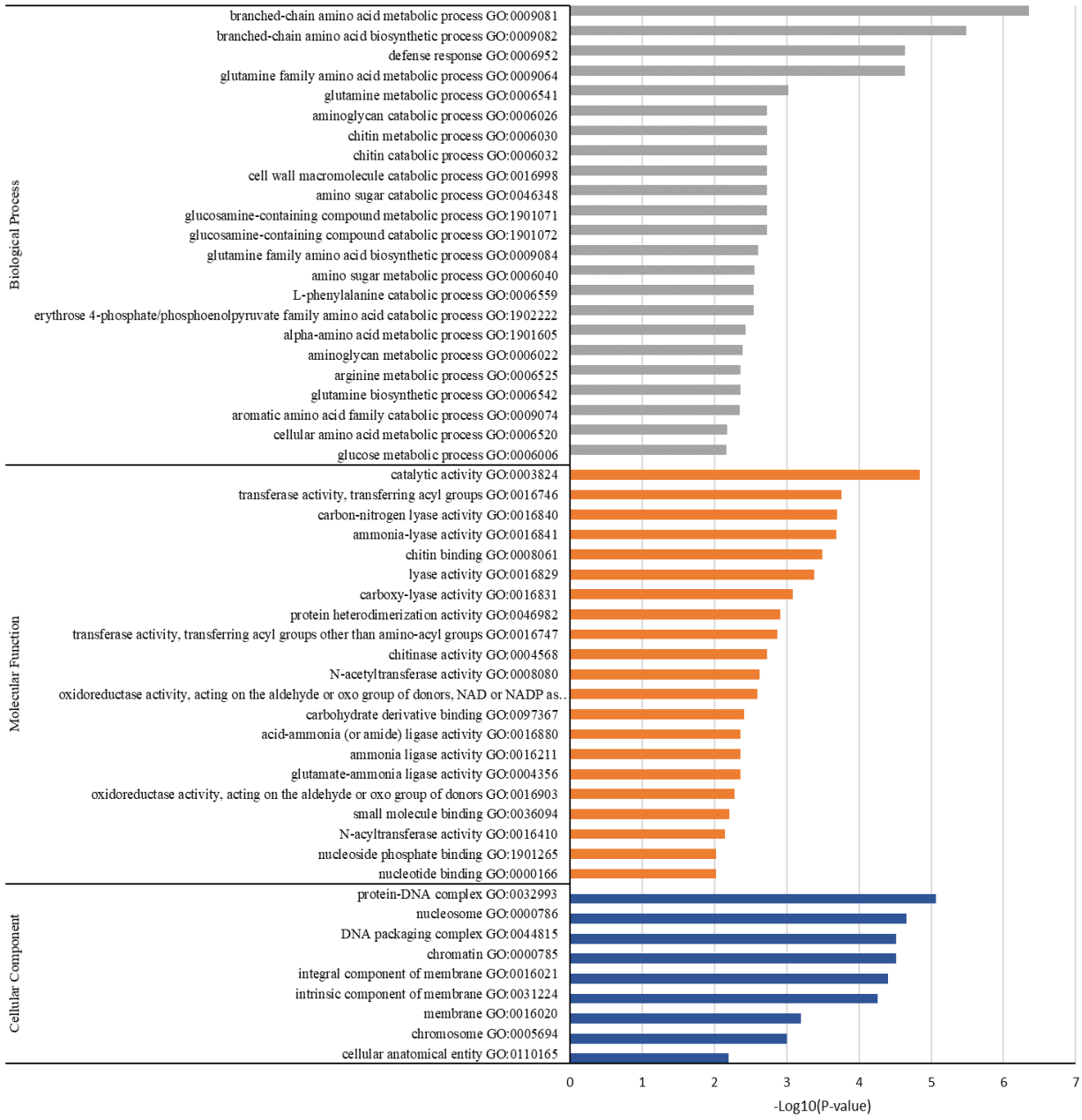


Figure 4.5. Enriched Gene Ontology (GO) terms of up-regulated genes in chlorsulfuron-treated sensitive *M. polymorpha* compared to untreated plants.

Up-regulated genes at every timepoint 2-48h after chlorsulfuron treatment were grouped into a single list and replicate gene names were removed to obtain a list of unique genes. The GO term enrichment analysis was conducted on this list using the VISEAGO R package (Brionne et al., 2019).

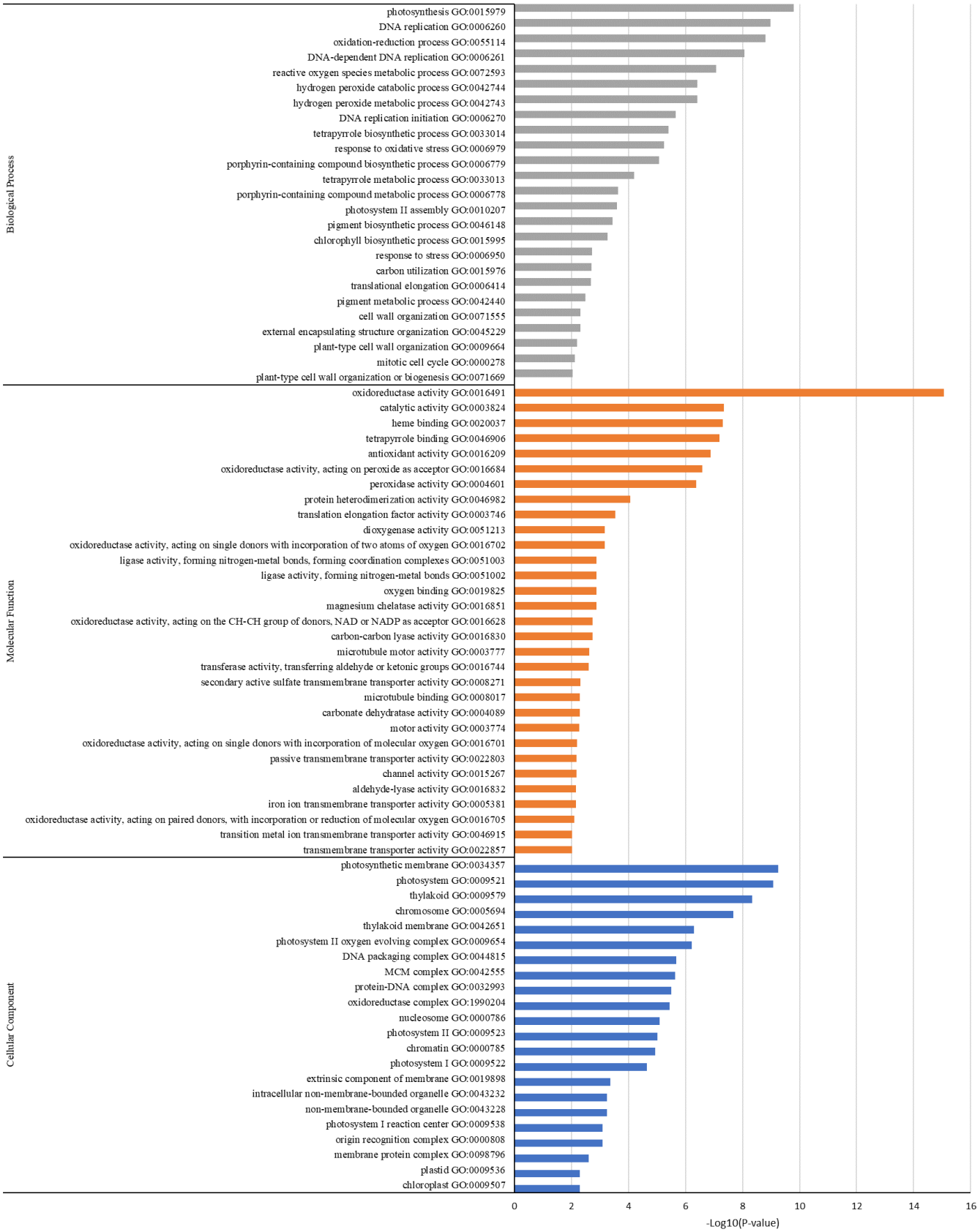


Figure 4.6. Enriched Gene Ontology (GO) terms of down-regulated genes in chlorsulfuron-treated sensitive *M. polymorpha* compared to untreated plants. Up-regulated genes at every timepoint 2-48h after chlorsulfuron treatment were grouped into a single list and replicate gene names were removed to obtain a list of unique genes. The GO term enrichment analysis was conducted on this list using the VISEAGO R package (Brionne et al., 2019).

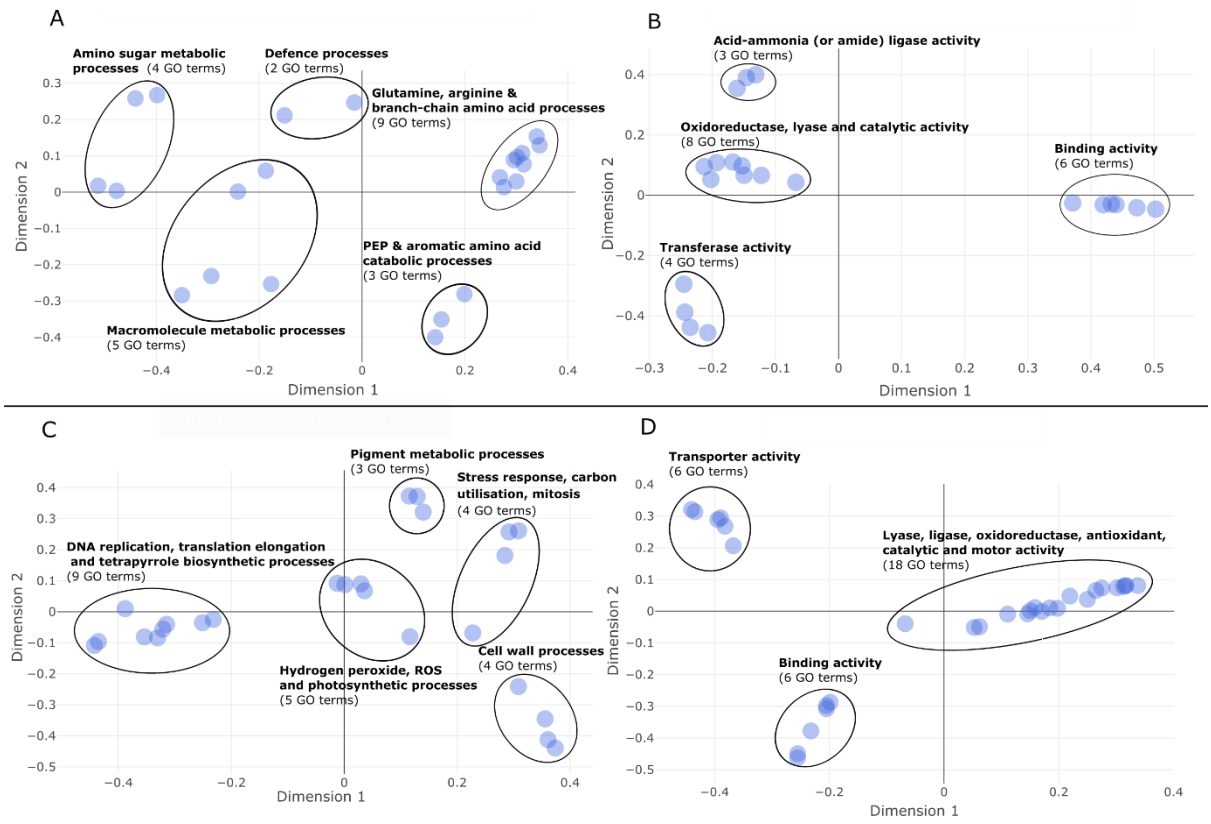


Figure 4.7. Semantic clusters of GO terms enriched in differentially expressed genes between chlorsulfuron-treated and untreated sensitive (Tak-2) *M. polymorpha*

Multidimensional scaling plots showing groups of biological processes (A & C) and molecular functions (B & D) - enriched in genes up-regulated (A & B) and down-regulated (C & D) in chlorsulfuron-treated sensitive *M. polymorpha* plants. Each blue circle represents an enriched GO term. GO terms were clustered into functional groups based on Wang's semantic similarity distance and ward.D2 aggregation criterion using the VISEAGO package (Brionne et al., 2019).

4.4.4. Several genes from herbicide detoxifying enzyme families are transcriptionally induced by chlorsulfuron treatment

The genes up-regulated in chlorsulfuron treated plants in all timepoints were investigated in *M. polymorpha,* to identify possible NTSR genes induced by chlorsulfuron. NTSR genes involved in herbicide metabolism belong to the cytochrome P450 (CYP), glutathione s-transferase (GST), and glycosyl transferase (GT) families (Délye, 2013).These enzymes chemically modify herbicides by adding hydroxyl or glutathione groups which label them for sequestration in the vacuole by ATP transporters. I hypothesise that plants possess enzymes that are dedicated to the detoxification of xenobiotics, therefore these will be induced by herbicide treatment in both sensitive and Mp*ALS*^{tsr} plants. This analysis aimed to identify enzymes that have the potential to confer NTSR in plants.

Comparison of the transcriptomes of chlorsulfuron-treated and untreated Mp*ALS*^{tsr} plants identified only a handful of genes that were differentially expressed (Supplementary Figures S4.2 and S4.3), and none of the differentially expressed genes were members of known herbicide metabolising enzyme families. This indicates that a herbicide-induced up-regulation of xenobiotic detoxifying genes did not occur in Mp*ALS*^{tsr} plants. Analysis of the genes up-regulated in chlorsulfuron-treated plants compared to untreated Tak-2 plants identified 20 MpCYPs and 15 MpGSTs that were up-regulated in at least one timepoint (Figure 4.8). No glucosyl transferases (GTs) were up-regulated in treated Tak-2 plants compared to untreated plants. The genes formed two categories based on timing of up-regulation; (i) genes up-regulated in the first 24 hours, and (ii) genes up-regulated 36-48h after chlorsulfuron treatment. No genes were up-regulated in every time-point. This may be due to the transcriptional response to chlorsulfuron occurring in two stages: an initial detoxification response followed by a general stress response. To check this, the normalised counts for each gene in treated and untreated samples were plotted (Figure 4.9). The normalised counts show that expression of most Mp*CYP*s and Mp*GST*s in untreated plants increased at the 12h and 24h time-points, then decreased at 36-48 hours after treatment back to levels seen 0-6 hours after treatment. This explains why fold changes in expression are not significant in the 12-24h timepoints in treated plants. Hence an explanation for the two stage increase in some GSTs and CYPs in treated plants is due to untreated plants up-regulating these genes 12 and 24h after treatment, possibly as a delayed stress response to the cellulose disc-mediated transfer to new plates.

The expression of 16 Mp*GST*s and Mp*CYP*s was verified by real time quantitative PCR in chlorsulfuron-treated Tak-2 and Mp*ALS*^{tsr} plants, over a reduced time-course (6-36 hours after treatment), and with only one of the TSR mutants (Mp*ALS*^{tsr} _7). Mp*APT* (adenine phosphoribosyl transferase 3) and Mp*ACT* (actin 7) were used as internal control genes. Generally fold change in expression was lower between treated and untreated Tak-2 plants in RT-qPCR samples, but most (11/16) genes were significantly up-regulated 24-36 hours after chlorsulfuron treatment (Figure 4.8). The other five genes showed no significant change in expression between treated plants and un-treated plants. None of the genes were up-regulated in Mp*ALS*^{tsr} _7 plants (Supplementary Figure S4.4), confirming the results of the RNA-Seq experiment. Mp*GST*

and Mp*CYP* genes that were most significantly up-regulated in RNA-Seq and quantitative real time PCR experiments in Tak-2 plants were Mp*GSTF7*, Mp*GSTF8*, Mp*GSTF10*, Mp*GSTF15*, Mp*CYP813A5*, Mp*CYP822A1*, and Mp*CYP94L1*. Five of these genes – Mp*CYP813A5*, Mp*CYP822A1*, Mp*GSTF7* and Mp*GSTF15* - were selected to investigate their ability to confer NTSR to chlorsulfuron.

Gene ID (Marpolbase)	Functional name	RNA-Seq up-regulation (Fold change Treated/Untreated)[a]					qRT-PCR verification (relative fold change Treated/Untreated)[b]			
		6 HAT	12 HAT	24 HAT	36 HAT	48 HAT	6 HAT	12 HAT	24 HAT	36 HAT
Mapoly0022s0192	**Mp*CYP813A5***	10.37***			10.41***	7.41***	1.09	1.13	1.62*	1.09
Mapoly0041s0103	Mp*CYP822B2*	2.70***					0.94	1.03	1.24	1.16
Mapoly0057s0084	Mp*CYP821B1*	2.40**					n.a.			
Mapoly0103s0038	Mp*CYP710A16*	2.89***			2.10**		0.96	1.01	1.26	1.08
Mapoly0163s0018	Mp*CYP73A109*	2.18**			2.46***	2.24**	1.16	0.97	1.10	1.72**
Mapoly0094s0036	Mp*CYP823A2*	2.95***	2.07*		3.50***	3.09***	0.93	0.93	1.08	1.69**
Mapoly0066s0002	Mp*CYP704AH1*			5.29*			0.70	1.11	0.87	1.07
Mapoly0010s0184	**Mp*CYP822A1***				14.20***	7.83***	1.05	1.18	2.43*	2.83**
Mapoly0067s0051	Mp*CYP829E1*				3.65***	3.50***	0.82	1.02	1.06	1.42*
Mapoly0095s0013	Mp*CYP822B1*				3.65***	3.91***	n.a.			
Mapoly0141s0021	Mp*CYP94L1*				3.51***	3.00***	0.99	1.64	1.64	2.84*
Mapoly0021s0144	Mp*CYP704N1*				2.68***	2.84***	0.86	1.70	1.47	1.82
Mapoly0081s0008	Mp*CYP73A110*				2.61***	2.75***	0.84	1.51	1.34	2.11*
Mapoly0004s0177	Mp*CYP829B1*				2.44***	2.33***	n.a.			
Mapoly0004s0182	Mp*CYP829F1*				2.13*	3.22***	n.a.			
Mapoly0002s0331	Mp*CYP817A1*				2.02*	2.26**	n.a.			
Mapoly0024s0025	Mp*CYP813B1*					6.27*	n.a.			
Mapoly0144s0022	Mp*CYP704L1*					4.90*	n.a.			
Mapoly0014s0215	Mp*CYP880A1*					2.14***	n.a.			
Mapoly0090s0052	Mp*CYP716X1*					2.02*	n.a.			
Mapoly0046s0045	**Mp*GSTF15***	11.74**			27.95***	30.15***	1.08	1.39	6.16*	7.56**
Mapoly0031s0032	Mp*GSTU1*	4.02*			16.28***	10.38***	n.a.			
Mapoly0818s0001	Mp*GSTF8*	6.39**			11.02***	6.41***	0.95	1.57	4.82	3.87*
Mapoly0177s0013	Mp*GSTF5*	5.33***	3.68*		8.46***	7.23***	n.a.			
Mapoly0193s0023	**Mp*GSTF10***	3.73**			11.46***	14.09***	0.88	1.56	3.00**	5.91*
Mapoly1812s0001	**Mp*GSTF7***	3.22**			3.77***	2.81**	0.87	1.33	1.58	2.63**
Mapoly0016s0155	Mp*GSTH1*			2.01**			0.95	1.26	0.74	1.12
Mapoly0009s0025	Mp*GSTF18*				5.94***	7.46***	n.a.			
Mapoly0185s0005	Mp*GSTF13*				3.63***	3.30***	n.a.			
Mapoly0008s0167	Mp*URE2p1*				3.46***	3.68***	n.a.			
Mapoly0020s0134	Mp*GSTT2*				2.86***	3.22***	n.a.			
Mapoly0002s0202	Mp*GSTF1*				2.66*		n.a.			
Mapoly0001s0057	Mp*GSTK2*				2.37***		n.a.			
Mapoly0141s0017	Mp*GSTF11*					2.43***	n.a.			

Figure 4.8. Identification of up-regulated cytochrome P450s and glutathione s-transferases in chlorsulfuron-treated sensitive (Tak-2) *M. polymorpha* compared to untreated plants.

Colouring is based on level of fold change (FC) with 0≤FC<1.5 or non-significant FC: white, 2.5≤FC<4: salmon, 4≤FC<10: orange, 10≤FC : brick.

[a] Fold change is calculated using the DESeq2 package (Love et al., 2014). Significance is calculated with an adjusted *p*-value using the Benjamini-Hochberg method to control the false discovery rate.

[b]Relative fold change in RT-qPCR was calculated from internal control genes. Significant differences are calculated from students t-tests between treated and untreated groups for each gene and timepoint (n=3).

*=p< 0.05, ** = p< 0.01, *** = p< 0.001, n.s. = not significant.

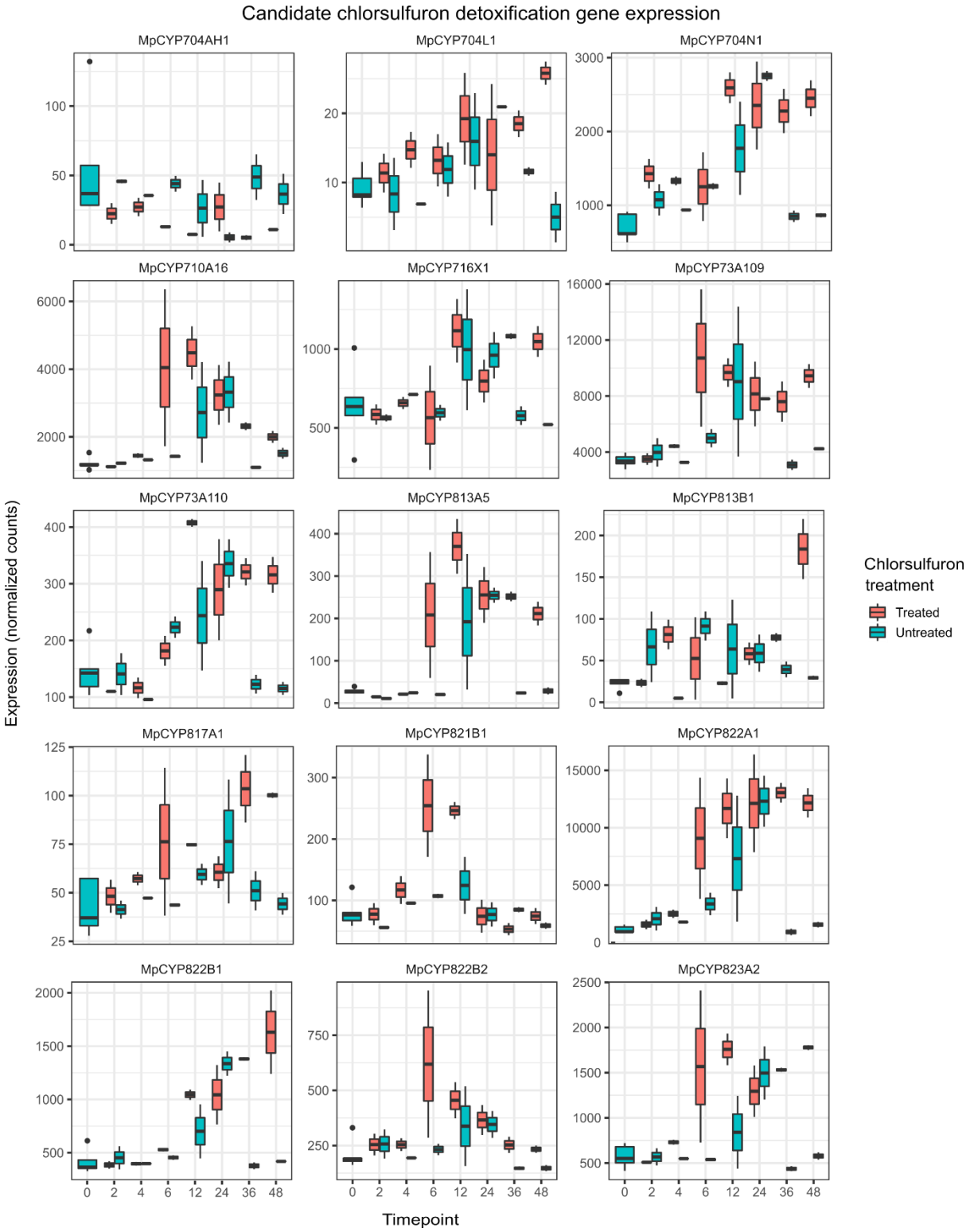


Figure 4.9. Expression of candidate chlorsulfuron detoxifying enzymes in treated and untreated *M. polymorpha* Tak-2 gemmalings. Y axis is count values normalised using DESeq2 mean of ratios method (Anders & Huber, 2010).

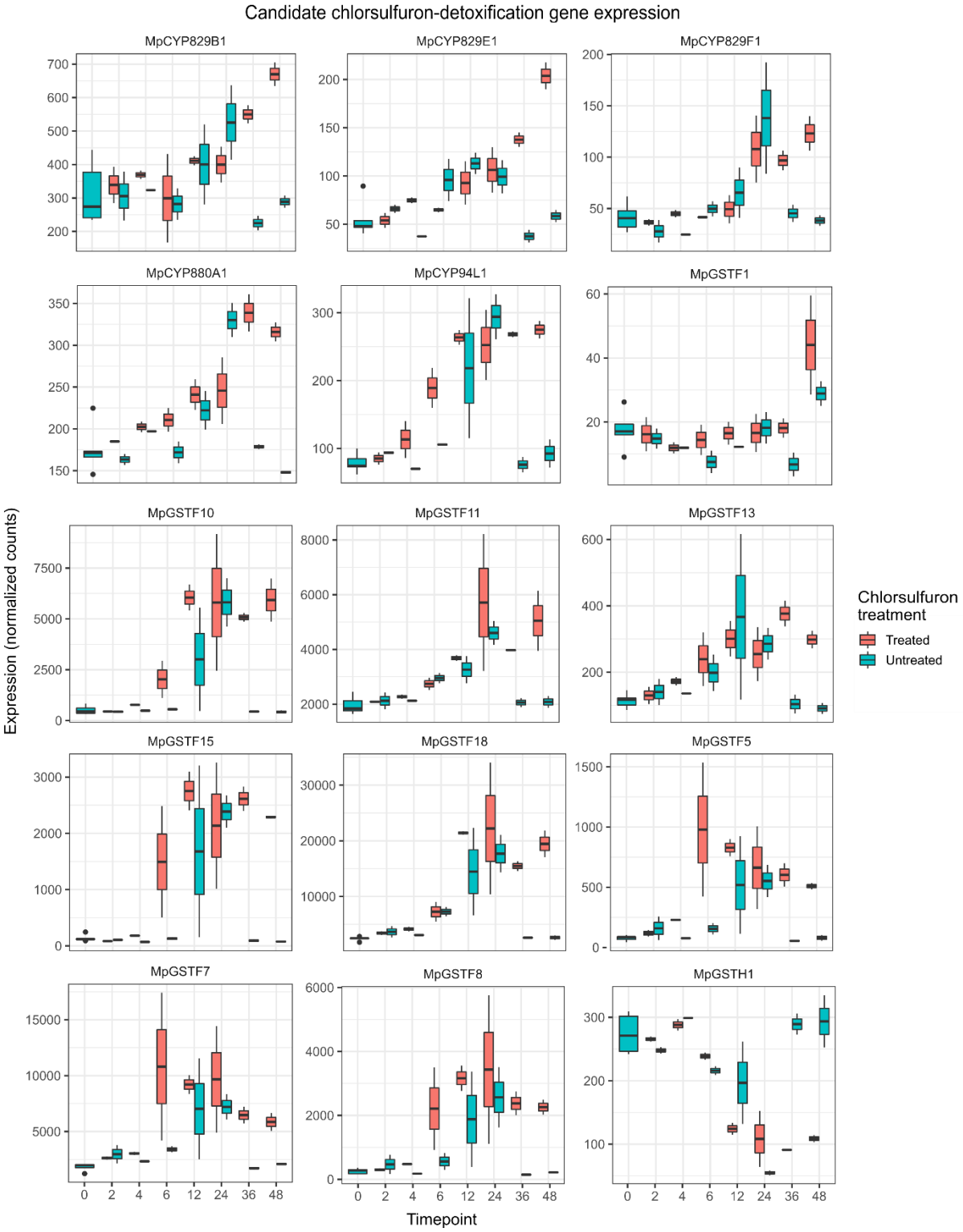


Figure 4.9. (continued). Expression of candidate chlorsulfuron detoxifying enzymes in treated and untreated *M. polymorpha* Tak-2 gemmalings. Y axis is count values normalised using DESeq2 mean of ratios method (Anders & Huber, 2010).

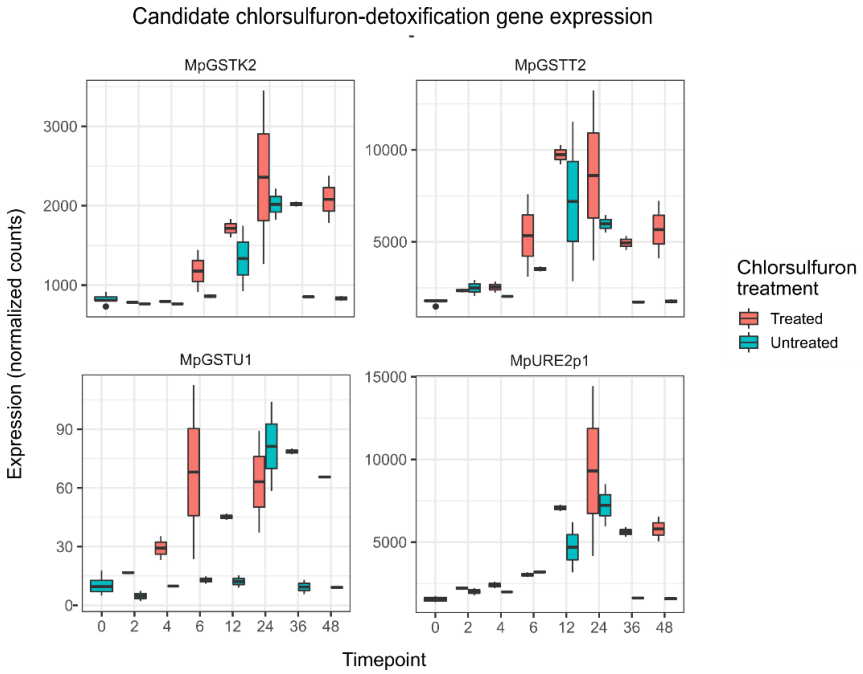


Figure 4.9. (continued). Expression of candidate chlorsulfuron detoxifying enzymes in treated and untreated *M. polymorpha* Tak-2 gemmalings. Y axis is count values normalised using DESeq2 mean of ratios method (Anders & Huber, 2010).

4.4.5. Over-expression of Mp*CYP822A1* confers a low level of resistance to chlorsulfuron at non-lethal dose

To test the hypothesis that genes up-regulated by chlorsulfuron treatment in sensitive *M. polymorpha* function in the detoxification of chlorsulfuron, five of the most up-regulated Mp*GST*s and Mp*CYP*s in chlorsulfuron treated sensitive plants were cloned into overexpressing vectors under the control of the endogenous elongation factor alpha (Mp*EF1α*) promoter, and tested for their sensitivity towards chlorsulfuron. Genes were PCR amplified from Tak-2 gDNA and cloned into pMpGWB403 over-expression vector using Gateway cloning, then transformed into *M. polymorpha* spores. Transformants were grown to adult size and gemmae from five independent lines were grown for 14 days. RNA was extracted from three biological replicates per line and converted to cDNA and the expression of the gene of interest was measured and compared to the two internal reference genes using quantitative real time PCR.

At least one line exhibited significantly increased levels of transcript in all overexpression constructs (Figure 4.10). The range in expression was very large, from 1.5 fold increase in $_{pro}$Mp*EF1α*:Mp*GSTF7* to 4000 fold increase in $_{pro}$Mp*EF1α*:Mp*CYP813A5*, compared to the empty vector controls. A 1.5 fold increase in expression was considered to be too low to confer a measurable phenotypic difference with wild type plants so none of the $_{pro}$Mp*EF1α*:Mp*GSTF7* lines were included in the subsequent growth analysis. The lines $_{pro}$Mp*EF1α*:Mp*GSTF10*_17, $_{pro}$Mp*EF1α*:Mp*GSTF15*_5, $_{pro}$Mp*EF1α*:Mp*CYP813A5*_7 and $_{pro}$Mp*EF1α*:Mp*CYP822A1*_3 exhibited 290, 2.4, 4000 and 58 fold higher expression than empty vector control plants, respectively, and were selected for characterisation of their sensitivity to chlorsulfuron compared to an empty vector control line.

Plants were grown from gemmae for 14 days either on untreated medium, medium supplemented with a non-lethal chlorsulfuron dose or a lethal chlorsulfuron dose then imaged in a NightOwl II LB983 fluorescence imaging system (Figure 4.11A-L). Chlorophyll autofluorescence was used to determine plant area (mm^2) which was treated as a proxy for growth of living tissue. Dead tissue is therefore not detected using this method. When grown on untreated medium, plants from three of the over-expressing lines were smaller than the empty vector control plants. Control plants were on average 182mm^2, while $_{pro}$Mp*EF1α*:Mp*CYP813A5*_7 plants were 25% smaller (135mm^2, *p*-value = 0.03), $_{pro}$Mp*EF1α*:Mp*CYP822A1*_3 plants were 30% smaller (126mm^2, *p*-value = 0.006) and $_{pro}$Mp*EF1α*:Mp*GSTF10*_17 plants were 25% smaller on average (135mm^2, *p*-value = 0.04). $_{pro}$Mp*EF1α*:Mp*GSTF15*_5 plants did not differ significantly in size from control plants (160mm^2, *p*-value =0.3). This suggests that Mp*CYP813A5*, Mp*CYP822A1* and Mp*GSTF10* over-expression confers a small fitness cost in control conditions.

Despite the smaller size of Mp*CYP822A1* over-expressing plants in control conditions, they were significantly larger (*p*-value =0.0002) than empty vector control plants on medium supplemented with a non-lethal chlorsulfuron dose (Figure 4.9Q). $_{pro}$Mp*EF1α*:Mp*CYP822A1*_3 plants were on average nearly double the size (4.3mm^2) of empty vector control plants (2.3mm^2) after 14 days on herbicide-treated medium, indicating that they had a higher tolerance to chlorsulfuron at

that dose. Plants over-expressing Mp*GSTF10*, Mp*GSTF15* and MpCYP*813A5* did not significantly differ in size to control plants on the non-lethal chlorsulfuron dose, although on average plants over-expressing Mp*GSTF10,* Mp*GSTF15* and Mp*CYP813A5* were 0.5-1.1mm^2 larger than control plants. No Mp*CYP* or Mp*GST* over-expressing plants or control plants survived when grown on the lethal chlorsulfuron dose (Figure 4.9P).

In summary, plants over-expressing one of four enzymes – Mp*CYP813A5*, MpC*YP822A1*, Mp*GSTF10* and Mp*GSTF15* - in *M. polymorpha* were tested for their sensitivity to chlorsulfuron. Plants over-expressing Mp*CYP822A1* were significantly larger than empty vector control plants, growing nearly twice as large as empty vector control plants after 14 days on a non-lethal chlorsulfuron dose, despite the smaller size of Mp*CYP822A1* over-expressing plants in control conditions compared to control plants. This indicates that Mp*CYP822A1* confers slight tolerance to chlorsulfuron, and suggests that Mp*CYP822A1* may be involved in chlorsulfuron detoxification.

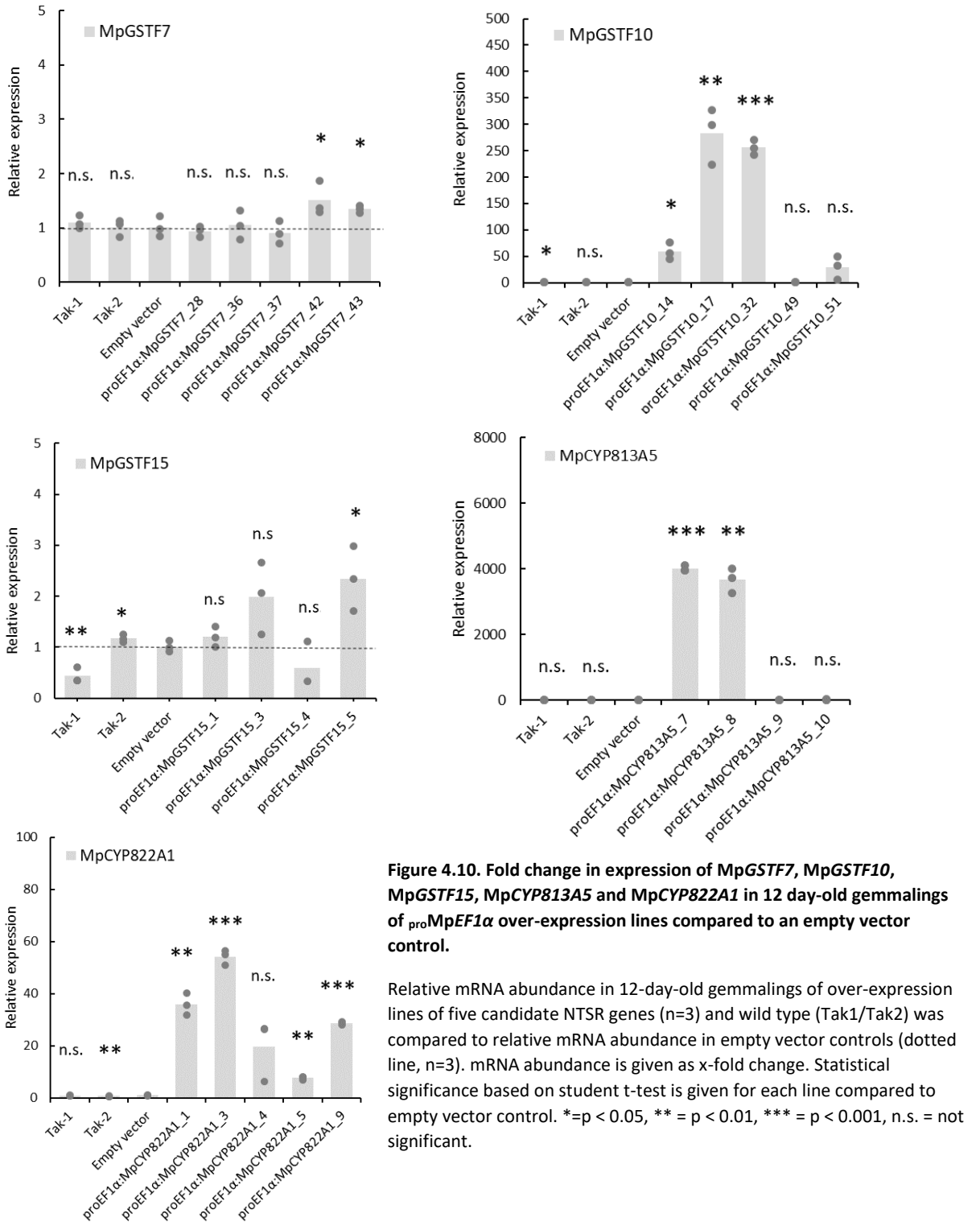


Figure 4.10. Fold change in expression of Mp*GSTF7*, Mp*GSTF10*, Mp*GSTF15*, Mp*CYP813A5* and Mp*CYP822A1* in 12 day-old gemmalings of $_{pro}$Mp*EF1α* over-expression lines compared to an empty vector control.

Relative mRNA abundance in 12-day-old gemmalings of over-expression lines of five candidate NTSR genes (n=3) and wild type (Tak1/Tak2) was compared to relative mRNA abundance in empty vector controls (dotted line, n=3). mRNA abundance is given as x-fold change. Statistical significance based on student t-test is given for each line compared to empty vector control. *=p < 0.05, ** = p < 0.01, *** = p < 0.001, n.s. = not significant.

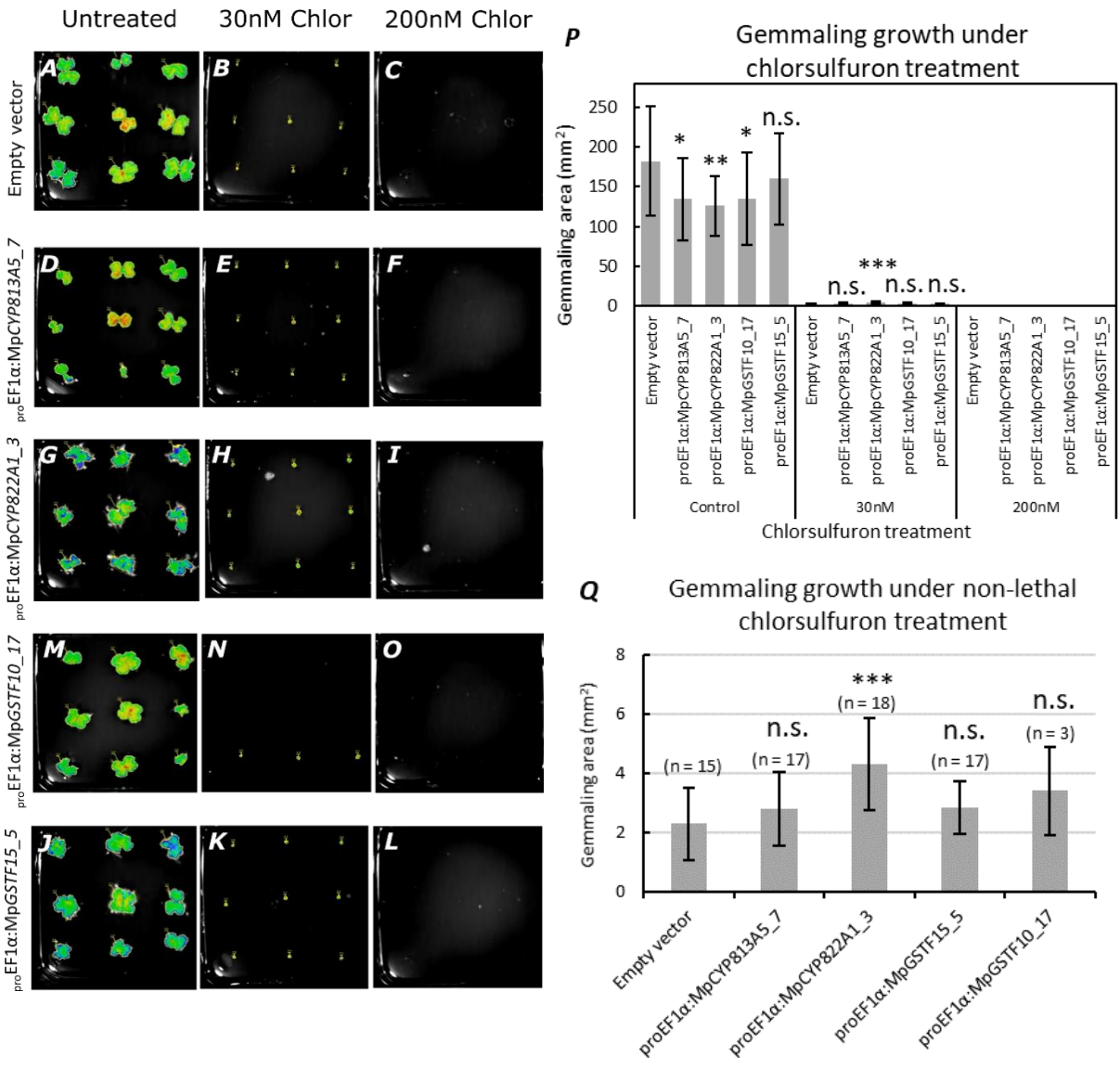


Figure 4.11. Chlorsulfuron sensitivity of $_{pro}$Mp*EF1α*:Mp*GST* and $_{pro}$Mp*EF1α*:Mp*CYP* over-expressing lines.

Growth of 14 day-old gemmalings of an empty vector line (A-C) and of Mp*CYP813A5*, Mp*CYP822A1*, Mp*GSTF10* and Mp*GSTF15* over-expressing lines (D-O) grown on untreated ½ Gamborg medium (Untreated), non-lethal dose (30nM) and a WT lethal dose (200nM) of chlorsulfuron (P). Chlorophyll autofluorescence is used to measure gemmae area (mm^2) which is treated as a proxy for growth. The same data for the 30nM chlorsulfuron treatment is shown on a smaller y axis scale in (Q). Statistical significance is measured by a student's t-test for each line compared to the empty vector control, and is shown above the bars. * = p<0.05, ** = p < 0.01, *** = p < 0.001, n.s. = not significant. Error bars are ±SD.

4.5. Discussion

Here I show that chlorsulfuron treatment caused the differential expression of hundreds of genes in sensitive *Marchantia polymorpha* plants, representing one third of all expressed genes detected or 15% of protein coding genes in the genome (Bowman et al., 2017). Up-regulated genes were associated with functions in amino acid metabolism and defence responses, while down-regulated genes had functions in photosynthesis, DNA replication and pigment biosynthesis. Chlorsulfuron treatment caused the up-regulation of several cytochrome P450s (CYPs) and glutathione s-transferases (GSTs) in sensitive Tak-2 *M. polymorpha*. One CYP, Mp*CYP822A1*, conferred a low level of resistance to chlorsulfuron compared to control plants when over-expressed. This is the first report of a non-target site gene conferring resistance to a herbicide in *Marchantia*. By contrast, chlorsulfuron treatment caused the differential expression of fewer than 20 genes in two Mp*ALS* target-site resistant (TSR) mutants compared to untreated plants, and none had possible functions in herbicide detoxification.

Chlorsulfuron blocks *M. polymorpha* growth at very low concentrations. A 10nM concentration was sufficient to inhibit growth by 95% in sensitive plants compared to untreated plants. A transcriptomic analysis of wild-type sensitive Tak-2 plants treated with a non-lethal chlorsulfuron dose identified 2982 genes that were differentially expressed in chlorsulfuron treated Tak-2 plants compared to untreated plants during a 48 hour time-course. This represents one third of all expressed genes in the samples. Large shifts in gene expression caused by ALS inhibition have been observed in other studies (Das et al., 2010; Gardin et al., 2015). For example, an investigation of the transcriptional response of blackgrass (*Alopecurus myosuroides*) to two ALS inhibitors found that 56% of the transcriptome was differentially expressed in treated compared to untreated sensitive plants over a 73 hour time-course (Gardin et al., 2015). Gene ontology terms enriched in up-regulated genes included amino acid metabolic processes, macromolecule metabolic processes, amino sugar metabolic processes and defence processes, while cell wall and photosynthesis processes were significantly enriched in down-regulated genes. This response bears a close similarity to global transcriptional responses to ALS inhibitors in other plant species (Das et al., 2010; Délye et al., 2018; Duhoux et al., 2015; Gardin et al., 2015; Liu et al., 2017). A comparison of transcriptional responses of *A. thaliana* treated with ALS-inhibitors and glyphosate, which disrupts the biosynthesis of aromatic amino acids, found that while glyphosate affected only 16 transcripts, ALS inhibitors caused changes in gene transcription to hundreds of genes related to secondary metabolism and growth (Das et al., 2010). Studies have also observed that genes and pathways up-regulated in ALS inhibited plants are also regulated by other stresses such as salicylic acid, salt stress, or bacterial infection (Glombitza et al., 2004; Manabe et al., 2007), but are distinct from herbicides with other modes of action, suggesting that ALS-inhibition triggers general stress response pathways in plants. This suggests that the toxic effect of ALS-inhibitors such as chlorsulfuron may go beyond a lack of specific amino acids.

Two target-site resistant lines Mp*ALS*^{tsr} _7 and Mp*ALS*^{tsr} _10 both harbour a point mutation in their Mp*ALS* gene sequence leading to the translation of a leucine instead of a proline at Pro197. Mp*ALS*^{tsr} _7 and Mp*ALS*^{tsr} _10 plants were 223 and 100-fold more resistant to chlorsulfuron than Tak-2 plants, respectively. We hypothesised that chlorsulfuron treatment of Mp*ALS*^{tsr} plants would induce specific herbicide detoxifying enzymes and wouldn't alter the expression of stress response genes. A transcriptome analysis of chlorsulfuron treated and untreated Mp*ALS*^{tsr} plants over a 48 hour time-course identified fewer than 20 genes that were differentially expressed in chlorsulfuron-treated Mp*ALS*^{tsr} plants in either mutant and none belonging to families associated with herbicide detoxification. No differentially expressed genes were common in Tak-2 and the two Mp*ALS*^{tsr} mutants, therefore it is unlikely that the genes with altered expression in Mp*ALS*^{tsr} plants are involved in chlorsulfuron response or signalling. Measurements of transcript abundance of CYPs and GSTs up-regulated in chlorsulfuron treated vs untreated Tak-2 plants using quantitative real time PCR confirmed that none of those genes showed differential expression in chlorsulfuron-treated vs untreated Mp*ALS*^{tsr} plants. Hence the hypothesis that a dedicated suite enzymes with specific functions in xenobiotic detoxification exists in plants is not supported by these results. This is also observed in imidazolinone-treated target-site resistant *A. thaliana* plants with a Ser653Asn mutation in the ALS gene, which didn't show any differential gene expression compared to untreated plants (Manabe et al., 2007).

Several Mp*CYP* and Mp*GST* genes were up-regulated in chlorsulfuron-treated sensitive Tak-2 plants compared to untreated plants and this was confirmed using quantitative real time PCR. One gene, Mp*CYP822A1*, conferred a low level of resistance to a non-lethal chlorsulfuron dose when over-expressed in *M. polymorpha*. This is the first lab experiment identifying a NTSR gene that was induced in herbicide-treated sensitive plants. Genes conferring NTSR are virtually always identified in NTSR weeds or crops (see chapter 2, table S2.3). The only other study that obtained NTSR resistant plants from sensitive plants is a study of *L. rigidum* populations treated with a reduced rate of diclofop-methyl which lead to the next generation having enhanced NTSR, although the genetic basis of the NTSR wasn't determined (Neve & Powles, 2005).

I show that a gene up-regulated in herbicide treated sensitive plants can confer tolerance to the herbicide when over-expressed. This indicates that enzymes that confer chlorsulfuron tolerance form part of the general herbicide induced stress response caused by the herbicide inhibiting its target, and hence NTSR evolves from random up-regulation of stress-response genes. The non-lethal dose used to test herbicide sensitivity is quite high, causing more than 95% reduction in growth in control plants, and a dose response curve would indicate whether the over-expression of Mp*CYP822A1* and of the other candidate genes confers higher tolerance to chlorsulfuron at lower doses. The low level of resistance conferred by the over-expression of Mp*CYP822A1* leads to the suggestion that NTSR to a lethal herbicide dose requires the increased expression of multiple genes in the same plant that each confer low level resistance. Therefore, I hypothesise that NTSR in the field is likely to be an additive trait conferred by multiple genes. This hypothesis is supported by a recent study on the evolution of herbicide resistance in *Amaranthus tuberculatus* which found that agricultural glyphosate-resistant

populations had larger CYP and GT gene copy numbers across the CYP and GT gene families than nearby glyphosate-sensitive populations (Kreiner et al., 2019). Mp*CYP822A1* has no direct homologs in vascular plants, but belongs to the CYP 71 clan which contains many previously reported enzymes that conferred resistance to ALS inhibitors (Dimaano et al., 2020; Han et al., 2021; J. Li et al., 2013; Xiang et al., 2005; Zhang et al., 2006). This is the first report of a gene in *M. polymorpha* conferring NTSR. Metabolomic experiments would determine whether Mp*CYP822A1* is directly detoxifying chlorsulfuron or other toxic molecules accumulating as a result of chlorsulfuron treatment. Whatever the mechanism, the hypothesis presented in chapter 2 that CYPs that metabolise herbicides or confer NTSR are found in the CYP 71 and 72 clans is supported by this result. Mp*CYP813A5* did not confer resistance to chlorsulfuron when over-expressed. Mp*CYP813A5* is part of the CYP 72 clan, and a couple of 72 clan enzymes have been shown to confer herbicide resistance in rice and cotton (Saika et al., 2014; Thyssen et al., 2018). Mp*CYP813A5* was one of the most strongly induced CYPs in chlorsulfuron treated Tak-2 plants, and had 4000 fold higher expression in Mp*CYP813A5* over-expressing plants, hence the degree of up-regulation of candidate NTSR genes in herbicide-treated plants doesn't appear to be correlated with their ability to confer resistance.

In conclusion, I observed that chlorsulfuron causes a large deregulation of gene expression which results in the up-regulation of genes with functions in amino acid metabolism and the down-regulation of genes involved in essential processes such as photosynthesis and DNA synthesis. Mp*CYP822A1* which conferred tolerance towards chlorsulfuron is part of the chlorsulfuron transcriptional stress response, and not part of a dedicated suite of xenobiotic detoxifying enzymes. Hence, I hypothesise that NTSR evolves by random up-regulation of stress response genes. The low level of resistance conferred by Mp*CYP822A1* suggests that NTSR is a polygenic trait, since mutations causing the increased expression of many genes would be required to confer high enough resistance to be visible in the field. These findings provide insight into the molecular basis and evolution of NTSR from a sensitive plant. This approach has allowed me to identify genes involved in low, incremental resistance which can contribute to NTSR, which is more challenging in less genetically tractable systems such as weeds. In the next chapter I study the effect of Mp*ALS* loss of function to investigate the mechanism of chlorsulfuron toxicity caused by Mp*ALS* inhibition in *M. polymorpha.*

4.6. Appendix

4.6.1. Supplementary figures

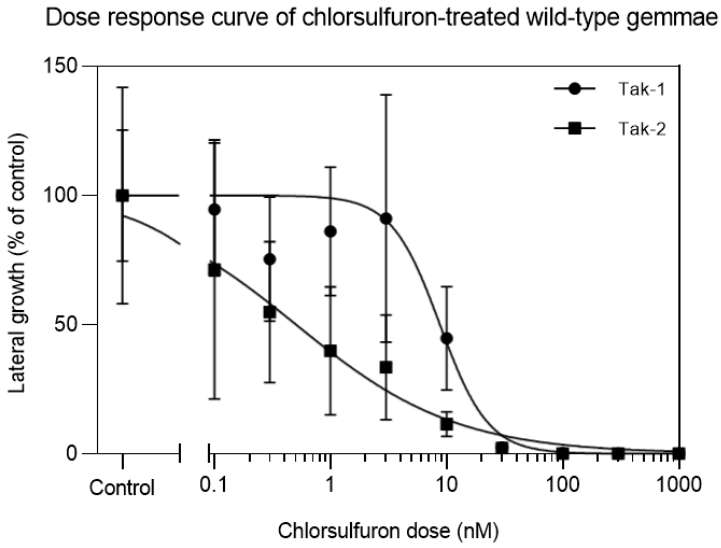


Figure S4.1. The lethal dose of chlorsulfuron on wild type *M. polymorpha* lines grown from gemmae is 100-300nM chlorsulfuron. Dose response curve of male (Tak-1) and female (Tak-2) gemmae grown on untreated and chlorsulfuron-treated ½ Gamborg medium for 14 days. Treatment means plotted with a non-linear fit, sigmoidal, four-parameter regression curve using GraphPad Prism. Error bars ± SD. n = 9.

Mp*ALS*^{tsr}_7		log_2FC	padj	Function
6 HAT	Mapoly0041s0139	29.24	$5.3\ 10^{-06}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0003s0267	14.87	$1.9\ 10^{-05}$	[PF00892] EamA-like transporter family
	Mapoly0041s0136	14.52	$1.3\ 10^{-02}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0147s0021	-17.08	$2.2\ 10^{-03}$	[KOG1577] Aldo/keto reductase family proteins
	Mapoly0085s0103	-18.95	$1.8\ 10^{-12}$	PF00153] Mitochondrial carrier protein
	Mapoly0091s0080	-21.16	$5.1\ 10^{-08}$	[PF00504] Chlorophyll A-B binding protein
12 HAT	Mapoly0041s0139	30.92	$3.3\ 10^{-08}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0018s0018	-15.09	$1.6\ 10^{-09}$	[PF14215] bHLH-MYC and R2R3-MYB transcription factors
	Mapoly0003s0267	-15.82	$7.8\ 10^{-11}$	[PTHR22911] Acyl-malonyl condensing enzyme-related
	Mapoly0018s0001	-16.30	$4.7\ 10^{-10}$	[KOG0627] Heat shock transcription factor
	Mapoly0017s0019	-16.36	$9.3\ 10^{-10}$	[PF07887] Calmodulin binding protein-like
	Mapoly0112s0056	-16.75	$1.2\ 10^{-11}$	[PF01814] Hemerythrin HHE cation binding domain
	Mapoly0017s0020	-17.21	$1.0\ 10^{-06}$	[K12850] pre-mRNA-splicing factor 38B
	Mapoly0018s0008	-17.68	$6.6\ 10^{-08}$	[KOG0087] GTPase Rab11/YPT3, small G protein superfamily
	Mapoly0018s0016	-20.08	$4.2\ 10^{-04}$	[KOG0648] Predicted NUDIX hydrolase FGF-2 related
24 HAT	Mapoly0041s0139	29.47	$2.8\ 10^{-06}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0050s0052	23.1	$2.6\ 10^{-02}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0017s0006	20.02	$5.8\ 10^{-12}$	[PTHR24351] RIBOSOMAL PROTEIN S6 KINASE
	Mapoly0091s0079	17.3	$4.0\ 10^{-05}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0017s0020	15.82	$9.5\ 10^{-08}$	[K12850] pre-mRNA-splicing factor 38B
	Mapoly0091s0081	15.38	$1.6\ 10^{-03}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0018s0008	14.11	$8.3\ 10^{-04}$	[KOG0087] GTPase Rab11/YPT3, small G protein superfamily
	Mapoly0041s0136	-14.43	$1.0\ 10^{-02}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0112s0056	-15.58	$8.4\ 10^{-03}$	[PF01814] Hemerythrin HHE cation binding domain
	Mapoly0085s0103	-16.82	$9.6\ 10^{-09}$	[KOG0767] Mitochondrial phosphate carrier protein
36 HAT	Mapoly0017s0019	14.80	$4.7\ 10^{-05}$	[PF07887] Calmodulin binding protein-like
	Mapoly0041s0136	14.62	$2.2\ 10^{-02}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0091s0081	16.97	$4.7\ 10^{-05}$	[PF00504] Chlorophyll A-B binding protein
48 HAT	Mapoly0003s0267	14.89	$3.2\ 10^{-05}$	[PF00892] EamA-like transporter family
	Mapoly0085s0103	14.87	$7.1\ 10^{-05}$	[PF00153] Mitochondrial carrier protein;
	Mapoly0018s0008	14.11	$7.0\ 10^{-04}$	[KOG0087] GTPase Rab11/YPT3, small G protein superfamily
	Mapoly0102s0020	0.58	$1.2\ 10^{-02}$	[KOG0062] ATPase component of ABC transporters
	Mapoly0046s0093	0.36	$2.1\ 10^{-02}$	[PF01084] Ribosomal protein S18
	Mapoly0018s0001	-14.36	$3.2\ 10^{-05}$	[KOG0627] Heat shock transcription factor
	Mapoly0022s0083	-15.99	$3.2\ 10^{-05}$	[PF05512] AWPM-19-like family
	Mapoly0041s0136	-16.18	$1.3\ 10^{-04}$	[PF00504] Chlorophyll A-B binding protein
	Mapoly0091s0079	-17.18	$6.2\ 10^{-05}$	[PF00504] Chlorophyll A-B binding protein

Figure S4.2. 19 genes were differentially expressed between chlorsulfuron treated and untreated target-site resistant mutant Mp*ALS*^{tsr}_7 over the 48 hour time-course. Positive Log_2(fold change) indicates the gene is up-regulated, negative log_2(fold change) indicates that the gene is down-regulated. Padj is *p*-value adjusted for multiple testing using the Benjamini-Hochberg method. Function lists PFAM, PANTHER or KOG annotation for each gene. HAT = hours after treatment.

Mp*ALS*^{tsr}_10		log_2FoldChange	padj	Function
6 HAT	Mapoly0112s0056	18.45	$5.29\ 10^{-09}$	[PF01814] Hemerythrin HHE cation binding domain
	Mapoly0041s0136	-14.55	$6.27\ 10^{-06}$	[PF00504] Chlorophyll A-B binding protein
36 HAT	Mapoly0041s0139	31.82	$8.45\ 10^{-08}$	[PF00504] Chlorophyll A-B binding protein

Figure S4.3. Three genes were differentially expressed between chlorsulfuron treated and untreated target-site resistant mutant Mp*ALS*^{tsr}_10 over the 48 hour time-course.

Positive Log_2(fold change) indicates the gene is up-regulated, negative log_2(fold change) indicates that the gene is down-regulated. Padj is *p*-value adjusted for multiple testing using the Benjamini and Hochberg method. Function includes the PFAM annotation found on Marpolbase. Function lists PFAM annotation for each gene. HAT = hours after treatment.

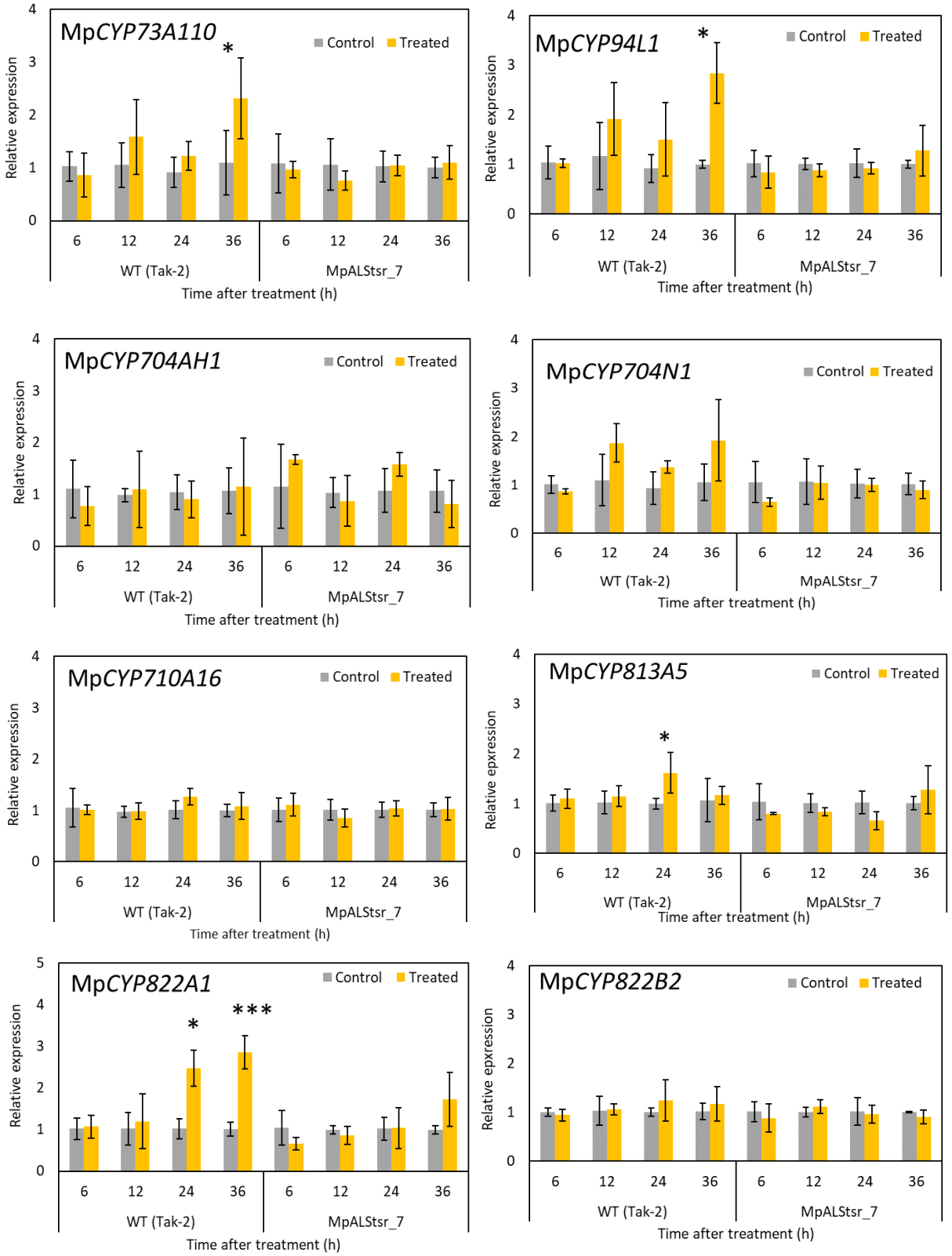


Figure S4.4. Relative expression of candidate chlorsulfuron metabolising enzymes in sensitive Tak-2 and resistant Mp*ALS*^{tsr}_7 treated and untreated plants.

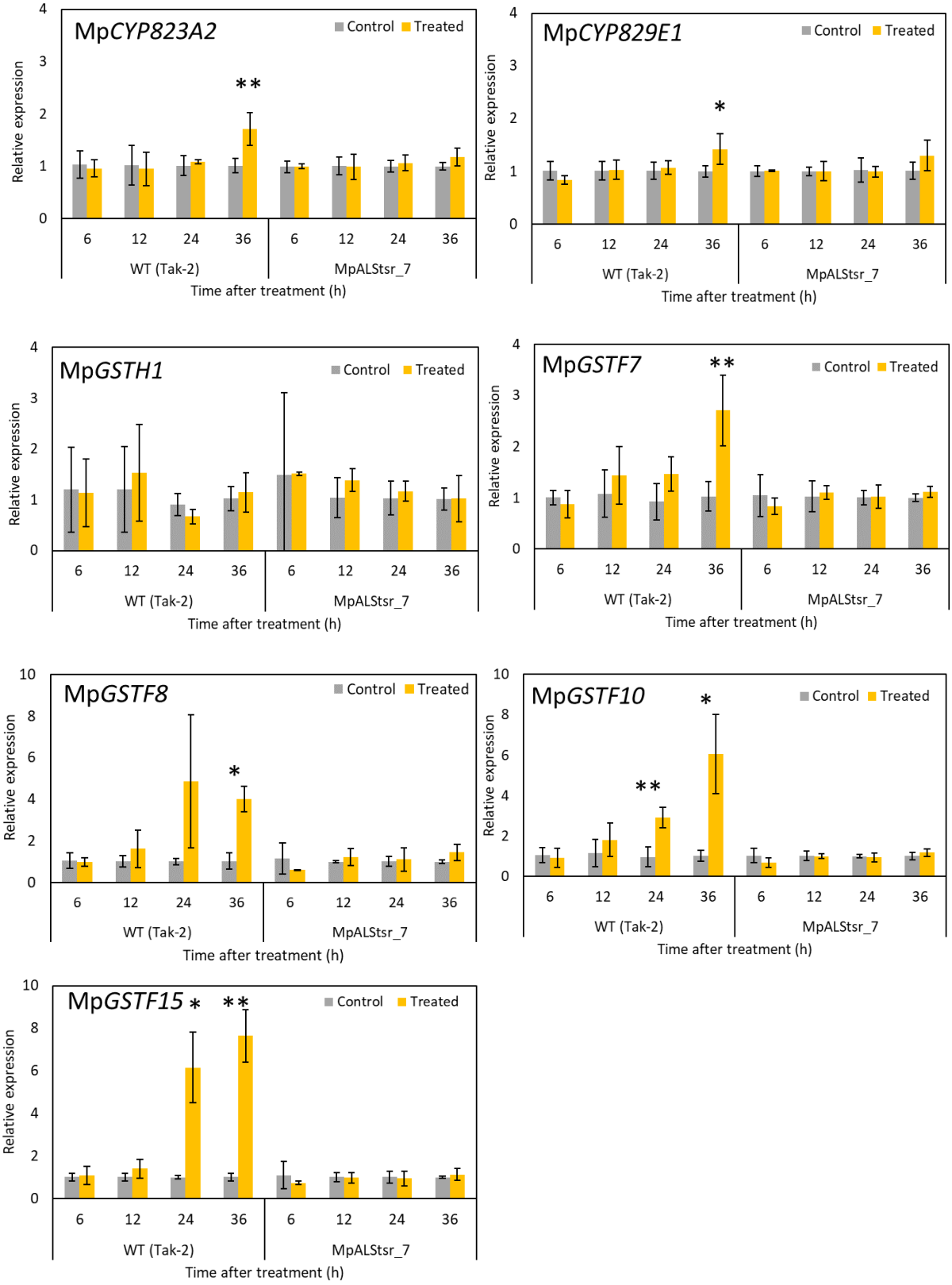


Figure S4.4. (continued) Relative expression of candidate chlorsulfuron metabolising enzymes in sensitive Tak-2 and resistant Mp*ALS*^{tsr}_7 treated and untreated plants.

4.6.2. Media and buffer recipes

Johnson's medium

Ingredient	Final concentration
KNO3	6 mM
MgSO4	1 mM
Ca(NO3)2.4H2O	4 mM
KCl	25 µM
H3BO4	10 µM
MnSO4.4H2O	1 µM
ZnSO4.7H2O	1 µM
CuSO4.5H2O	250 nM
(NH4)6Mo7O24.4H2O	250 nM
FeSO4.7H2O	25 µM
Na2EDTA	25 µM
NH4H2PO43-	0.6 mM
(NH4)2SO4	0.4 mM
inositol	0.55 mM
MES	0.533 g/l (w/v)
sucrose	1 % (w/v)
agar(Sigma-Aldrich)	0.8 % (w/v)

pH adjusted to 5.6 using 1M KOH

Gamborg medium

Ingredient	Final concentration	Half strength (1/2)
KNO3	25 mM	12.5 mM
CaCl2.2H2O	1 mM	0.5 mM
MgSO4.7H2O	10 mM	5 mM
(NH4)2SO4	1 mM	0.5 mM
NaH2PO4.H2O	1 mM	0.5 mM
KI	5 µM	2.5 µM
H3BO3	50 µM	25 µM
MnSO4.H2O	60 µM	30 µM
ZnSO4.7H2O	7 µM	3.5 µM
Na2MoO4.2H2O	1 µM	0.5 µM
CuSO4.5H2O	0.1 µM	0.05 µM
CoCl2.6H2O	105 nM	102.5 nM
Ferric-EDTA	125 µM	62.5 µM
inositol	0.55 mM	0.28 mM
nicotinic acid	8 µM	4 µM
pyridoxine. HCl	5 µM	2.5 µM
thiamine. HCl	30 µM	15 µM
kinetin	0.46 µM	0.23 µM
2,4-D	0.45 µM	0.23 µM
sucrose	2% (w/v)	1% (w/v)
agar(Sigma-Aldrich)	1.4% (w/v)	0.8% (w/v)

pH adjusted to 5.6 using 1M KOH

CTAB buffer

Ingredient	Final concentration
Tris-HCL pH 8	100 mM
NaCl	1.4 M
EDTA pH 8	20 mM
CTAB	2 %

Chapter 5

Chlorsulfuron lethality is not due to the accumulation of quinate or 2-ketobutyrate

5.1. Abstract

Chlorsulfuron is a highly potent herbicide widely used to control weeds. There is substantial evidence that chlorsulfuron binds to and blocks the activity of acetolactate synthase (ALS). However, the processes linking ALS inhibition to growth inhibition and plant death are not understood. Branched chain amino acid starvation, or the accumulation of a toxic molecule such as 2-ketobutyrate or quinate have both been implicated in the mechanism of plant death of ALS inhibitors.

I conducted experiments to investigate the cause of chlorsulfuron lethality in the liverwort *Marchantia polymorpha.* I observed that valine, leucine and isoleucine can restore growth in *M. polymorpha* Takaragaike-2 (Tak-2) plants treated with a lethal chlorsulfuron dose, while Tak-2 grown on medium with chlorsulfuron and isoleucine alone were severely growth inhibited. I generated 6 independent Mp*ALS* loss-of-function (lof) mutants (Mp*als*^{lof}) using CRISPR/Cas9 technology and measured the growth of two Mp*als*^{lof} lines and Tak-2 on medium supplemented with different combinations of branched chain amino acids. Both the complete loss of function line Mp*als*^{lofF16} and the partial loss of function line Mp*als*^{lofF12} were significantly smaller than Tak-2 on medium with all three branched chain amino acids, suggesting that a lack of branched chain amino acid production is only partially responsible for the impaired growth. I performed targeted and non-targeted metabolomics using liquid chromatography-tandem mass spectrometry (LC-MS/MS) of Tak-2 treated with a lethal chlorsulfuron dose for 24 hours. Chlorsulfuron treatment caused a significant decrease in valine and leucine but no change in isoleucine levels, and an increase in other amino acids including the aromatic amino acids tryptophan, serine and tyrosine. Dose response experiments with Tak-2 plants showed that quinate and 2-ketobutyrate are lethal at 10mM, however no increase in quinate, 2-ketobutyrate or 2-amiobutyrate levels was observed in chlorsulfuron treated plants.

I conclude that neither quinate or 2-ketobutyrate (or it's transamination product 2-amminobutyrate) are responsible for chlorsulfuron lethality in *M. polymorpha*, and a reduction of isoleucine isn't observed in chlorsulfuron-treated Tak-2. I suggest that chlorsulfuron inhibition of Mp*ALS* causes a reduction in branched chain amnio acids, which triggers an increase in other amino acids either through proteolysis or de novo biosynthesis.

5.2. Introduction

Chlorsulfuron is an extremely potent herbicide. Very low concentrations of chlorsulfuron, in the nanomolar range, are sufficient to kill *M. polymorpha*. The mode of action of chlorsulfuron was first studied in plants using excised pea roots (*Pisum sativum*) in the 80s. Thomas Ray (1984) found that supplementing valine and isoleucine to the medium reversed the growth inhibition caused by chlorsulfuron. He therefore identified acetolactate synthase, the first common enzyme in the biosynthetic pathway of valine, leucine and isoleucine, as the molecular target of chlorsulfuron, which was highly sensitive to chlorsulfuron inhibition. Subsequent studies in other plant species confirmed his findings that branched chain amino acids restored growth in plants treated with ALS-inhibiting herbicides (Rost & Reynolds, 1985; Shaner & Reider, 1986). It was therefore assumed that branched chain amino acids starvation caused plant death. However subsequent findings did not support this hypothesis, with increases or no changes in branched chain amino acids being reported despite a decline in growth in herbicide treated plants (Liu et al., 2017; Royuela et al., 2000).

ALS catalyses the condensation of two molecules of pyruvate to produce acetolactate in the pathway which leads to the biosynthesis of valine and leucine. ALS also catalyses the condensation of one molecule of pyruvate and one molecule of 2-ketobutyrate to produce acetohydroxybutyrate in the biosynthetic pathway of isoleucine. Several hypotheses have been presented for the mechanism of toxicity of ALS-inhibiting herbicides (Zhou et al., 2007). The toxic accumulation of 2-ketobutyrate has been suggested as a cause of growth inhibition in *Lemna minor* and the bacterium *Salmonella typhimurium* treated with ALS-inhibitors (Larossa et al., 1987; Rhodes et al., 1987). 2-ketobutyrate is synthesised by threonine deaminase and is an intermediate in the biosynthetic pathway of isoleucine. Isoleucine self regulates by feedback inhibition of threonine deaminase, which reduces 2-ketobutyrate production. 2-ketobutyrate accumulated in chlorsulfuron treated Salmonella and is toxic (Larossa et al., 1987). A metabolic analysis of duckweed (*Lemna minor*) treated with increasing chlorsulfuron doses found that along with lower valine, leucine and isoleucine levels an increase in 2-aminobutyrate, the transamination product of 2-ketobutyrate, was observed (Rhodes et al., 1987). Maize (*Zea mays*) treated with the ALS-inhibitor imazaquin accumulated 2-ketobutyrate and 2-aminobutyrate in shoots, and isoleucine treatment reduced both pools, but did not affect growth, therefore the growth inhibition in imazaquin-treated maize was not associated with 2-ketobutyrate or 2-aminobutyrate accumulation (Shaner & Singh, 1993). In addition, quinate, a compound synthesised in a side chain of the shikimate biosynthesis pathway, accumulated after both ALS inhibitor and glyphosate treatment in pea and was suggested to represent a shared mechanism of plant death in amino acid inhibiting herbicides (Orcaray et al., 2010; Zulet et al., 2013).

In this chapter I aim to elucidate the mechanism of chlorsulfuron lethality in the liverwort *M. polymorpha.* I propose four hypotheses for chlorsulfuron-induced plant death; the starvation of branch-chain amino acids (hypothesis 1), the accumulation of toxic levels of 2-ketobutyrate (2-KB) by lack of feedback inhibition from isoleucine (hypothesis 2), the

accumulation of toxic levels of quinate caused by a general amino acid-linked stress response (hypothesis 3), or a combination of these mechanisms (hypothesis 4). To investigate the mechanism of plant death by lack of ALS activity I generated 6 independent Mp*ALS* loss-of-function (lof) mutants (Mp*als*^{lof}) using CRISPR/Cas9 technology, which cannot grow without supplementing branch chain amino acids to the medium. Using the wild type *M. polymorpha* female accession Takaragaike-2 (Tak-2) and Mp*als*^{lof} I tested the hypothesis that the starvation of branched chain amino acids is causing plant death in chlorsulfuron treated Tak-2 and Mp*als*^{lof} by measuring the growth of Tak-2 grown on medium with chlorsulfuron and branched chain amino acids, and of Mp*als*^{lof} and Tak-2 on medium with different combinations of branched chain amino acids. I also tested the hypothesis that the feedback inhibition of threonine deaminase by isoleucine can reverse chlorsulfuron lethality in wild type plants and suppress the Mp*als*^{lof} lethal phenotype, using growth assays on of Mp*als*^{lof} on medium supplemented with isoleucine or of Tak-2 on chlorsulfuron and isoleucine. To assess whether 2-ketobutyrate or quinate are toxic in *M. polymorpha* I conducted dose response analyses of Tak-2 grown on medium with increasing concentrations of each compound. To investigate whether chlorsulfuron causes a decrease in branched chain amino acids or an accumulation of quinate or 2-ketobutyrate, as well as to determine global metabolic changes caused by chlorsulfuron treatment, I performed targeted and non-targeted metabolomic analyses of Tak-2 treated with a lethal dose of chlorsulfuron for 24h compared to untreated plants. Studying the physiological changes caused by ALS inhibition that lead to plant death will help to understand how plants overcome this stress to evolve resistance to ALS-inhibiting herbicides.

5.3. Materials and Methods

5.3.1. Plant materials and growth conditions

The two laboratory *M. polymorpha* accessions Takaragaike-1 (Tak-1) (male) and Takaragaike-2 (Tak-2) (female) were used as wild-type plants (Ishizaki et al., 2008). For measuring growth responses to different treatments, Tak-2 was selected as the representative wild-type plant, as it tended to grow flatter on the plates than Tak-1 making it easier to image. Plants were grown on sterile petri dishes containing Johnson's medium with 0.8% agar or ½ Gamborg medium with 1% sucrose and 0.8% agar (see chapter 4.6.2.). Plants were propagated asexually from gemmae and grown in light cabinets under 30 $\mu molm^{-2}s^{-1}$ continuous white light at 23°C.

Tak-2 plants used in dose response analyses and growth assays were grown from gemmae for 14 days on 120x120mm square plates or 100x21mm round plates containing Johnson's medium with 0.8% agar supplemented with different chemicals.

Tak-2 plants used for metabolomic analyses were grown from gemmae for 11 days on 100x21mm round plates containing ½ Gamborg medium with 1% sucrose and 0.8% agar (see chapter 4.6.2) covered with a cellulose membrane (A.A. Packaging Ltd, Preston, UK). Plants were moved onto medium supplemented with 100nM chlorsulfuron or untreated medium by lifting the cellulose discs using sterile forceps. Plants were left on the chlorsulfuron-treated or untreated medium for 24h before collection.

To generate spores for transformation, two-week old Tak-1 and Tak-2 plants were grown on a 1:3 mixture of vermiculite and John Innes No.2 compost and incubated in a growth chamber under a 16h light:8h dark photoperiod at 23°C. Gametophore production was induced under far red light irradiation as described in Chiyoda et al. (2008).

5.3.2. Growth assays using chlorsulfuron and/or branched chain amino acids

Whole plant growth experiments were conducted to determine the growth responses of wild-type (Tak-2) plants to chlorsulfuron in combination with branched chain amino acids isoleucine, leucine and valine, and of Mp*als*^{lof} plants to different combinations of branched chain amino acids. Chlorsulfuron, 1-(2-chlorophenylsulfonyl)-3-(4-methyoxy-6-methyl-1,3,5-triazin-2-yl)urea, was supplied by Sigma Aldrich and diluted in milli-Q H_2O. Branched chain amino acids L-isoleucine, L-leucine, L-valine were obtained from Sigma Aldrich and dissolved in milli-Q H_2O and 1.8M NaOH. All stock solutions were filter sterilized with a 0.2uM sterile syringe filter. Control plates were treated with the equivalent volume of solvent as treatment plates. Plant area (mm^2) was measured either by ImageJ thresholding or using a fluorescence imaging system (NightOWL II LB983, Berthold Technologies, Bad Wildbad, Germany). Statistical significance was determined with the non-parametric Kruskal-Wallis test followed by the Nemenyi post hoc test.

5.3.3. Sensitivity of *M. polymorpha* to quinate and 2-ketobutyrate

Whole plant dose-response experiments were carried out to determine the sensitivity of wild-type (Tak-2) plants to quinate and 2-ketobutyrate. Quinate (1,3,4,5-tetrahydroxycyclohexanecarboxylate) and 2-Ketobutyric acid were supplied by Sigma Aldrich. Quinate and 2-ketobutyrate stocks were dissolved in milli-Q H_2O and filter sterilized with a 0.2uM sterile syringe filter. Quinate or 2-ketobutyrate was pipetted into melted medium and mixed thoroughly before pouring into plates. Plants were treated with ten 2-ketobutyrate doses: 0mM, 0.05mM, 0.1mM, 0.2 mM, 0.5mM, 1mM, 2mM, 5mM, 10mM and 20mM, and 11 quinate doses: 0mM, 0.05mM, 0.1mM, 0.3mM, 1mM, 2mM, 5mM, 10mM, 25mM, 50mM and 100mM. There were two replicate plates per dose and nine plants grown per plate. Chlorophyll autofluorescence was imaged using a fluorescence imaging system (NightOWL II LB983, Berthold Technologies, Bad Wildbad, Germany) and was used to measure plant area (mm^2) as a proxy for growth. Data were expressed as % of the average area of untreated plants, and were plotted with a non-linear fit, sigmoidal, four-parameter regression curve using GraphPad Prism 9.1.0 (GraphPad Software, Inc., San Diego, CA, USA).

5.3.4. Sample extraction for metabolomic analysis

12 day old plants were placed into 2mL Eppendorf tubes and weighed. Approximately 150mg of fresh plant tissue per sample was collected and frozen immediately in liquid N_2. 6 biological replicates were collected per treatment group. Metabolite extraction was as follows. 500uL methanol:acetonitrile:H2O (2:2:1, v/v) (stored at -20°C) was added to each sample. Plant material was then homogenised for 2 min in a tissuelyser (Qiagen) using adaptors kept at -70°C. Samples were incubated for 1h at -20°C followed by centrifugation at 14,000rpm at 4°C. The supernatant was removed and placed in a new Eppendorf tube, and stored at -20°C. 400uL of 80% methanol:water (stored at -20°C) was added to the precipitate then vortexed for 1min at 4-8°C. Samples were incubated for 1h at -20°C then centrifuged for 14,000rpm at 4°C. The supernatant from the first extraction was combined with the second extraction, and the samples were stored for 2h at -20°C. Samples were centrifuged one final time at 14,000rpm for 10 min at 4°C. The supernatant was transferred to a new Eppendorf tube, frozen in liquid N_2 then stored at -80°C.

5.3.5. Targeted metabolomic analysis using liquid chromatography-mass spectrometry

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of samples from chlorsulfuron-treated and untreated plants was performed by the Metabolomics Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria and funded by the City of Vienna through the Vienna Business Agency. For the analysis of branched chain amino acids and their precursors, 1µL of the metabolite extract was injected on a Kinetex C8 column (100 Å, 100 x 2.1 mm), employing a flow rate of 80µL/min. A 5-min-long gradient from 95% A (0.1 % formic acid in water) to 60% B (0.1 % formic acid in acetonitrile) was used for the separation. The HPLC (RSLC ultimate 3000; Thermo Fisher Scientific)

was directly coupled to a TSQ Altis mass spectrometer (Thermo Fisher Scientific) via electrospray ionization. The following SRM (selected reaction monitoring) transitions were used in the positive ion mode: m/z 118.1 to m/z 72 (valine), m/z 132.1 to m/z 86.1 (isoleucine and leucine), m/z 205.1 to m/z 188 (tryptophan). Isopropylmalate wash was measured in the negative ion mode using the transitions m/z 175.1 to m/z 115 and m/z 118.1 to m/z 113. Optimal collision energies and retention times were determined and validated for each metabolite with the respective authentic standard. Other amino acid precursor metabolites (e.g. 2-ketobutyric acid) were targeted using the respective SRMs but only detected in the mixture of authentic standards. The data was manually interpreted using Trace Finder (Thermo Fisher Scientific).

5.3.6. Non-targeted metabolomic analysis using liquid chromatography-mass spectrometry

For the non-targeted detection of metabolites in chlorsulfuron treated and untreated plants, samples were analysed by LC-MS/MS using the hydrophilic interaction liquid chromatography (HILIC) the reversed phase chromatography (RP) separation methods. 1 µL of each sample was injected independently onto two different phase systems, on a SeQuant ZIC-pHILIC HPLC column (Merck, 100 x 2.1 mm; 5 µm) and on a C18-column (Waters, ACQUITY UPLC HSS T3 150 x 2.1; 1.8 µm). 6uL of each sample was pooled into a QC sample used for normalization. Separation was performed with a flow rate of 100 µl/min employing an Ultimate 3000 HPLC system (Thermo Fisher Scientific, Germany). In HILIC (hydrophilic interaction liquid chromatography) a 25 min gradient from 10% to 80% B was used (A: acetonitrile (ACN); B: 25 mM ammonium bicarbonate in water) and in reversed phase chromatography (RP) a gradient from 1% to 90% B in (A: 0.1% formic acid (FA) in water; B: 0.1% FA in ACN). The HPLC was coupled via electrospray ionization to the mass spectrometer. Metabolites were ionized via electrospray ionization in polarity switching mode, acquiring high-resolution tandem mass spectrometry data on a Q-Exactive Focus (Thermo Fisher Scientific, Germany) in data-dependent acquisition mode. Data sets were processed by Compound Discoverer (Thermo Fisher Scientific), searching our in-house library and publicly available spectral libraries with a mass accuracy of 3 ppm for precursor masses and 10 ppm for fragment ion masses.

5.3.7. gRNA design and vector construction for the generation of Mp*als* loss of function alleles

The generation of mutations in the Mp*ALS* gene using the CRISPR/Cas9 system followed the method developed by Sugano et al. (2014). Two short guide RNAs (gRNAs) gRNA_W (GTTCTGCCTATGATTCCTGG) and gRNA_F (CTCTGAGGGGACATGCATAC) were designed to target two different locations in the Mp*ALS* DNA sequence. gRNAs must be upstream (5') of a PAM sequence (NGG) and contain a protospacer or 'target' sequence consisting of 20 nucleotides that are unique to Mp*ALS*. gRNAs were designed using the CRISPR-P website (Lei et al., 2014). Individual gRNAs were cloned in the MpU6-1pro:gRNA expression vector using BsaI overhangs. MpU6-1pro:gRNA was combined with the Cas9 expression vector, MpGE010 via LR recombination. CRISPR-Cas9 expression vectors carrying gRNA_W or gRNA_F were transformed in *Agrobacterium tumefaciens*.

5.4.8. Agrobacterium-mediated transformation of *M. polymorpha* spores

Agrobacterium-mediated spore transformation followed the method of Ishizaki et al. (2008) and described in section 4.3.10, with the additional step of supplementing 800uM isoleucine, leucine and valine to the agrobacterium and spore liquid cultures and to the medium in the selection plates.

5.4.9. Selection and genotyping of transformants

Approximately 100 transformants were grown to adult size on medium containing the hygromycin selection and 800uM isoleucine, leucine and valine. One gemma was taken from each plant and placed on medium supplemented with 800uM branched chain amino acids, while a gemma from the same gemma cup was placed on medium without branched chain amino acids. Plants from which the gemma died or grew smaller on the medium lacking branched chain amino acids were then chosen as possible Mp*ALS* mutants and genotyped. Plants were genotyped using the Phire Plant Direct PCR kit (Thermo Fisher) using a 0.5mm diameter piece of plant tissue and the manufacturer's recommended 3-step cycling protocol. Primers used for genotyping are sgRNAF_GenF2: TATTGCTATCTCGCGCTGTGT, sgRNAF_GenR2: CTGGAACGTGTAAGAGCCTG, sgRNAW_GenF: TGTCAGGGATGGACAGTATGG and sgRNAW_GenR: GGACAAGGCTACCTCTTCAATGT.

5.4. Results

5.4.1. Supplemented branched-chain amino acids restores growth of *M. polymorpha* treated with a lethal dose of chlorsulfuron, but not isoleucine alone

The molecular target of the herbicide chlorsulfuron is acetolactate synthase (ALS), which catalyses the first step in the biosynthesis of branched chain amino acids. The enzyme threonine deaminase converts threonine to 2-ketobutyrate which is used as a substrate by acetolactate synthase in isoleucine biosynthesis. Threonine deaminase is feedback inhibited by isoleucine, therefore I predict that supplementing isoleucine to plants will increase the inhibition of threonine deaminase, reducing the production of 2-ketobutyrate in the plant. If chlorsulfuron treatment causes 2-ketobutyrate to build up to toxic levels in plants, then supplementing chlorsulfuron-treated plants with isoleucine should reverse the growth inhibition seen in chlorsulfuron-treated plants, compared to untreated plants.

To investigate whether a lack of branched chain amino acids or a build-up of 2-ketobutyrate causes chlorsulfuron lethality in *M. polymorpha*, a growth assay was carried out using Tak-2 gemmae grown for 14 days on solid medium supplemented with a lethal chlorsulfuron dose, chlorsulfuron lethal dose and isoleucine, chlorsulfuron lethal dose and isoleucine, leucine and valine, and untreated (solvent only) medium. Normality was checked using quantile-quantile (Q-Q) plots. Data did not meet the assumptions of normality of a one-way analysis of variance, therefore significant differences were calculated with the Kruskal-Wallis test. Tukey-Kramer post-hoc test identified significant differences between samples.

Plants on medium with a lethal dose of chlorsulfuron did not grow, while plants grown on medium with chlorsulfuron and isoleucine, leucine and valine survived and were not significantly different in size to control plants (Figure 5.1). Thus, supplementing branched chain amino acids to the medium restored growth of chlorsulfuron-treated plants to control rates. Plants treated with chlorsulfuron and isoleucine survived but were very small (0.46mm^2 average area), 0.6% of the size of untreated plants (76mm^2 average area, p-value <0.001) and 1% the size of plants treated with chlorsulfuron and all 3 branched chain amino acids (45mm^2 average area, p-value <0.05).

In summary supplementing all 3 branched chain amino acids restored growth of Tak-2 treated with a lethal chlorsulfuron dose, while isoleucine treatment suppressed chlorsulfuron lethality but did not reduce the growth inhibitory effect of chlorsulfuron. These results suggest 2-ketobutyrate toxicity may be associated with plant death but not growth inhibition, or a lack of all 3 branched chain amino acids causes chlorsulfuron lethality. To determine whether elevated 2-ketobutyrate is lethal in *M. polymorpha* I conducted dose response experiments using Tak-2 grown on increasing concentrations 2-ketobutyrate.

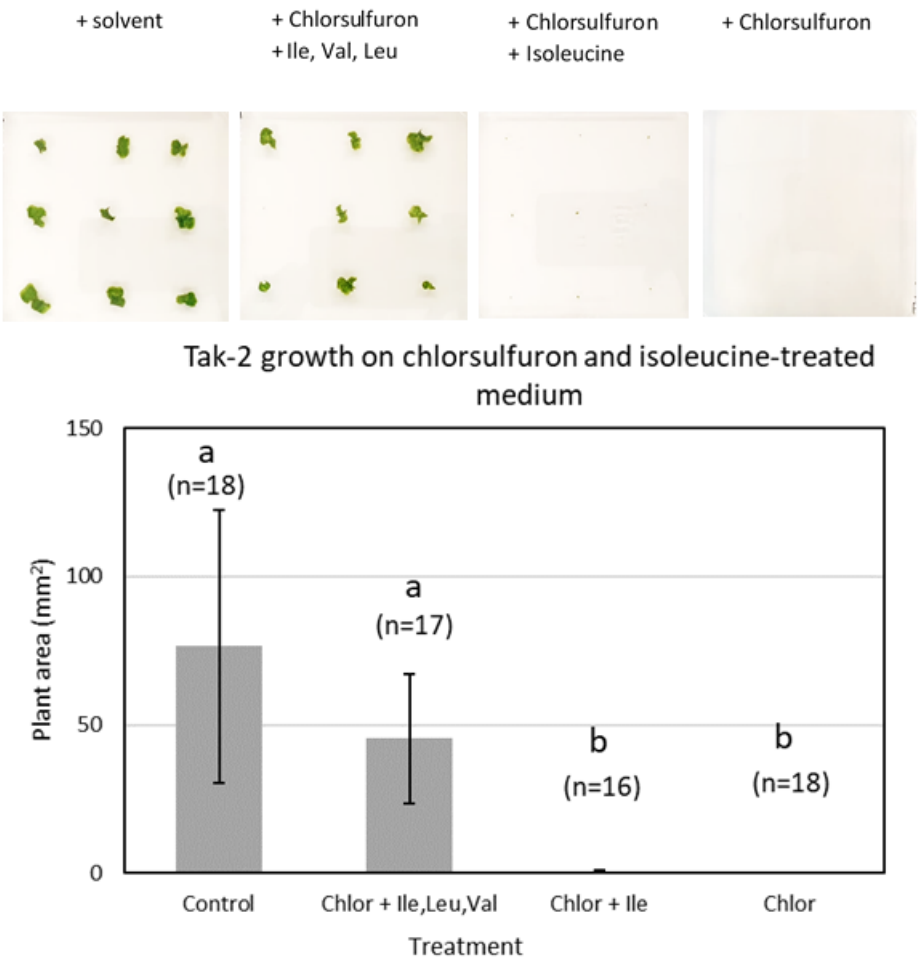


Figure 5.1. Branched chain amino acids restore growth in wild type *M. polymorpha* treated with a lethal dose of chlorsulfuron.

Growth assay of Tak-2 gemmae grown for 14 days on Johnson's medium with 100nM chlorsulfuron, 100nM chlorsulfuron and 800uM leucine, valine and isoleucine, 100nM chlorsulfuron and 800uM isoleucine or untreated (solvent only) medium. Plant area (mm^2) was measured using imageJ thresholding. Statistical significance based on Kruskal-Wallis test. Significant groups determined by Nemenyi post hoc test. Error bars are ± SD.

5.4.2. 2-ketobutyrate and quinate inhibit *M. polymorpha* growth and cause plant death

Early research in *Salmonella typhimurium* (Larossa et al., 1987) and duckweed (*Lemna minor*) (Rhodes et al., 1987) suggested that chlorsulfuron induced growth inhibition and death is caused by the accumulation of 2-ketobutyrate, a precursor of ALS produced by threonine deaminase. However a later study did not find a link between 2-ketobutyrate accumulation in maize and growth inhibition (Shaner & Singh, 1993). Quinate accumulation was also linked to defective amino acid metabolism, and growth inhibiting levels accumulated in pea plants treated with glyphosate and two ALS inhibitors (Orcaray et al., 2010).

To test whether 2-ketobutyrate or quinate cause growth inhibition or death of *M. polymorpha*, dose response analyses were conducted using the Tak-2 *M. polymorpha* accession treated with 10 2-ketobutyrate doses and 11 quinate doses (Fig. 4.2. B). Plants were grown for 14 days on treated or untreated medium before imaging using a NightOwl II LB983 fluorescence imaging system. Plant area (mm^2) was measured from chlorophyll autofluorescence and used as a proxy for growth. Growth reductions were observed from 0.3mM quinate (20% growth reduction compared to untreated plants) and 2mM 2-ketobutyrate (70% growth reduction compared to untreated plants). While quinate inhibited Tak-2 growth at lower doses than 2-ketobutyrate, the lethal dose for both 2-ketobutyrate and quinate was observed at 10mM treatment. These results indicate that high enough concentrations of quinate and 2-ketobutyrate inhibit growth and cause plant death of *M. polymorpha*.

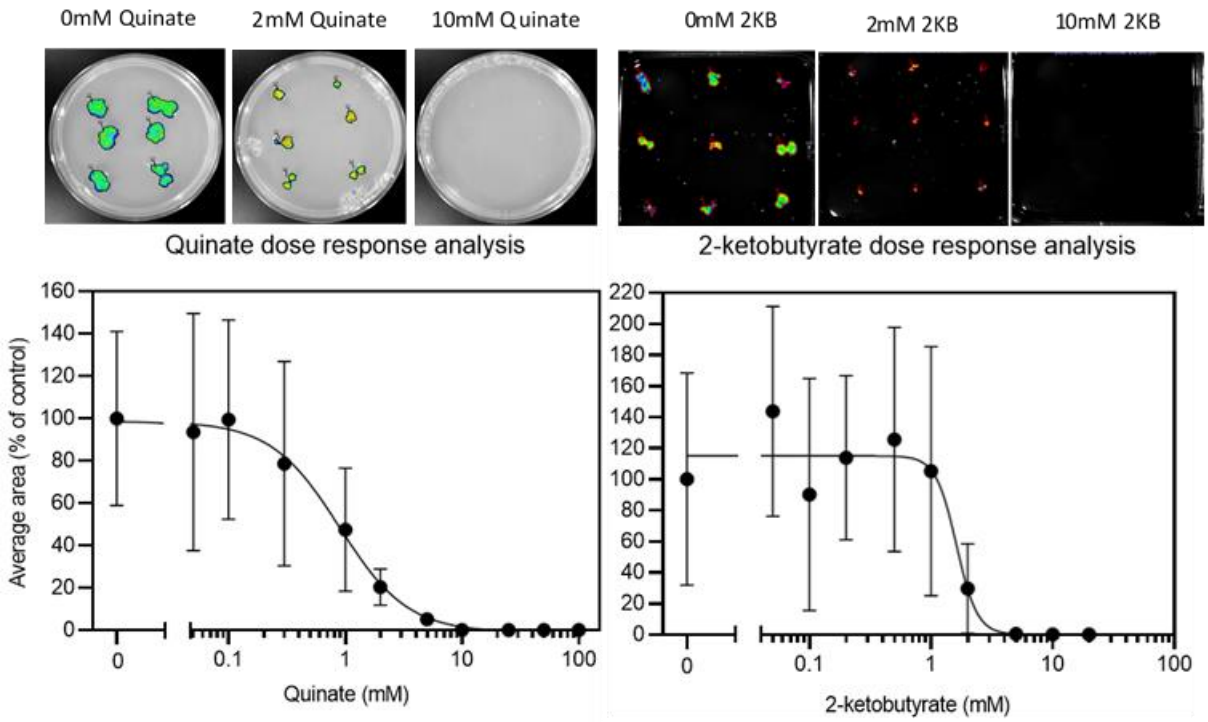


Figure 5.2. Sensitivity of wild type *M. polymorpha* to 2-ketobutyrate and quinate.

Dose response assay of Tak-2 gemmae grown for 14 days on Johnsons' medium supplemented with 2-ketobutyrate or quinate. Plant area (mm^2) based on chlorophyll autofluorescence was measured on a NightOwl II LB983 fluorescence imaging system. Data presented as % of average area of untreated plants and plotted with a non-linear fit, sigmoidal, four-parameter regression curve using GraphPad Prism. n=6-12 (quinate) or 9-18 (2KB). Error bars are ±SD.

5.4.5. Supplemented branched chain amino acids are necessary for survival of Mp*als*^{lof} mutants

Another approach to understanding the physiological effect of the inhibition of ALS activity is by generating Mp*ALS* loss of function mutants. Mp*ALS* is the molecular target of chlorsulfuron, and is first common enzyme in the biosynthetic pathway of valine, leucine and isoleucine. The activity of ALS is dependent on three cofactors, thiamine pyrophosphate (TPP), a magnesium ion (Mg2+) and flavin adenine dinucleotide (FAD). I generated Mp*ALS* loss of function (lof) mutants for use as a genetic system that mirrors the action of ALS inhibiting chemicals on wild type *M. polymorpha*.

I used the CRISPR/Cas9 system to generate the Mp*als*^{lof} lines. I designed two independent gRNAs to target different locations in the Mp*ALS* gene sequence which I cloned into separate expression vectors. One gRNA targeted the second exon of Mp*ALS* upstream of the TPP binding domain, while the other targeted the 8th or final exon of Mp*ALS* (see Figure 5.8A, B) which encodes residues that form part of the active site channel (McCourt et al., 2006). I transformed spores resulting from a cross between Tak-1 and Tak-2 accessions with a vector containing one of the two gRNAs under the expression of the U6 RNA Polymerase III promoter, the Cas9 endonuclease gene controlled by the ubiquitous Mp*EF1α* promoter, and a hygromycin resistance gene. In case mutations in Mp*ALS* were lethal, both spores and transformants were grown in liquid or on solid medium supplemented with 800uM branched chain amino acids. Transformed spores were grown for two weeks on hygromycin selection plates containing branched chain amino acids, and approximately 100 hygromycin resistant plants were tested for impaired branched chain amino acid biosynthesis. One gemma was taken from each hygromycin resistant plant and placed on medium supplemented with 800uM branched chain amino acids, while a gemma from the same gemma cup was placed on medium without branched chain amino acids. Parent plants from which the gemma died or grew smaller on the medium lacking branched chain amino acids were chosen as possible Mp*als* mutants and genotyped. 6 plants were selected for genotyping this way. I genotyped plants using primers that were 3-400bp upstream and 3-400bp downstream of the gRNA site and sequenced the PCR products.

All of the selected plants had mutations or small deletions in their Mp*ALS* gene sequence, validating this approach for identifying lethal mutations in *M. polymorpha*. Sequencing of the Mp*als* mutant lines F16, F32, F84 revealed point mutations and 1-11bp deletions at the gRNA_F binding site (Figure 5.8A,B) which are predicted to produce frameshifts in the protein sequence (Figure 5.8.C). These plants died on medium without branched chain amino acids (Figure 5.8.D). Sequencing the region around gRNA_W in Mp*als* mutant line W29 revealed a single bp deletion in Mp*ALS* which is predicted to result in a frameshift in the protein sequence at the C terminal. The Mp*als* mutant line W20 had a 4bp insertion, 4 mutations and 1bp deletion at the gRNA_W binding site, which is predicted to cause Pro to Gly and Gly to Trp substitutions and an Ile insertion in the C terminal. Mp*als*-W20 and Mp*als*-W29 plants died on medium without branched chain amino acids. This highlights that the residues in the Mp*ALS* gRNA_W binding site at the C terminal are critical for enzyme activity.

All of the above mentioned mutants have mutations that are predicted to cause frameshifts and truncations in Mp*ALS* and none survived on media without supplemented branched chain amino acids. As a result, they were determined to be complete loss of function mutants (lof) and named Mp*als*^{lofF16}, Mp*als*^{lofF32}, Mp*als*^{lofF84}, Mp*als*^{lofW20}, Mp*als*^{lofW29}. The mutant Mp*als*_F12 grew slightly smaller on media without branched chain amino acids, and sequencing revealed an in-frame 36bp insertion in the gRNA_F binding site of Mp*ALS*. This line was treated as a partial loss of function mutant named Mp*als*^{lofF12}. These results indicate that I was able to generate 6 independent Mp*als* mutant alleles with varying severity using the CRISPR/Cas9 system.

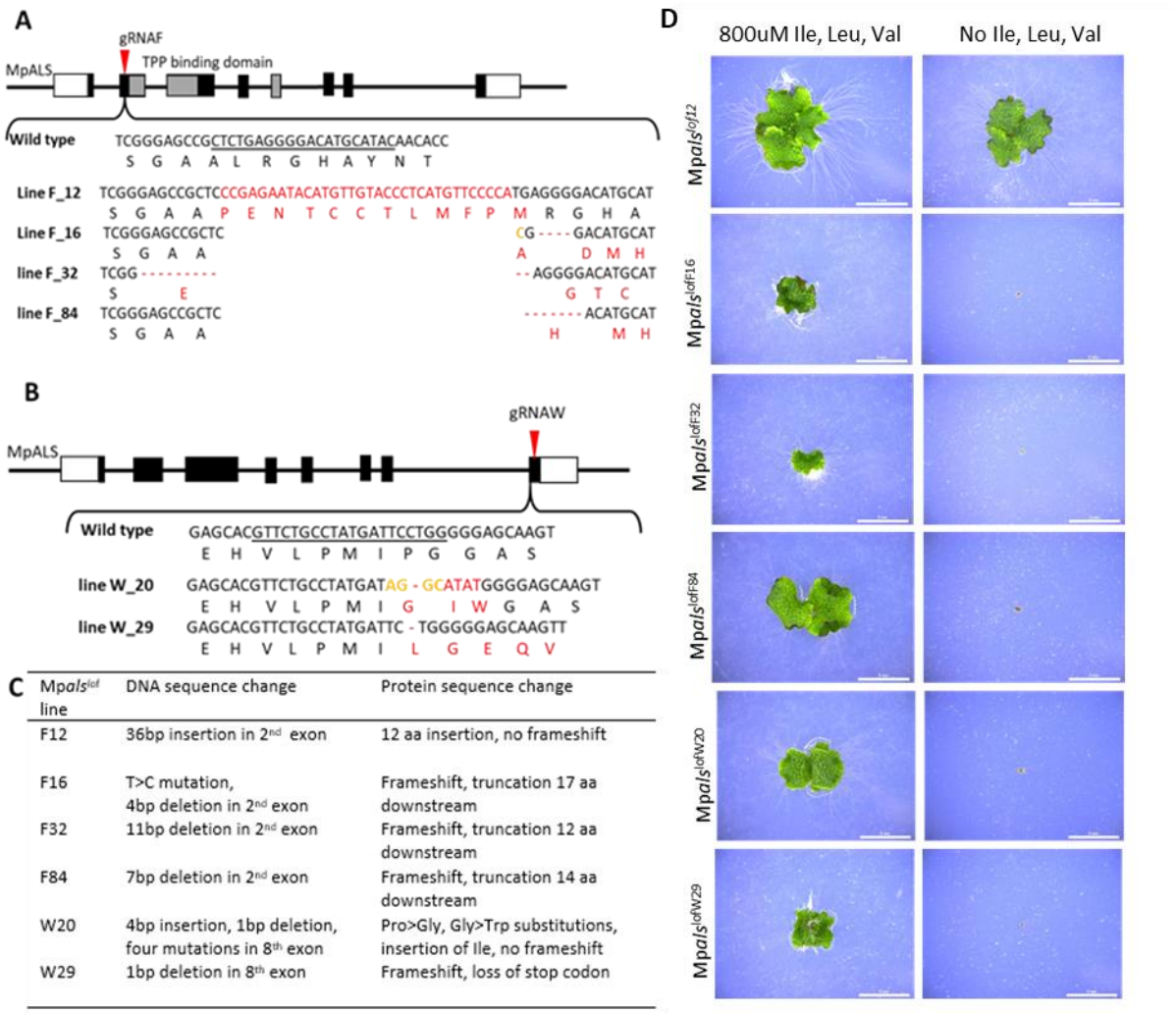


Figure 5.4. Generation of mutations and small deletions in MpALS using CRISPR/Cas9 mediated targeted mutagenesis.

Six independent Mp*als* mutant lines were generated using a gRNA targeting the 2nd exon (gRNAF) (A) or the 8th exon (gRNAW) (B) of Mp*ALS*. Exons are represented by black boxes, mutations by yellow letters, insertions by red letters, and deletions by red dashes. TPP = thiamine pyrophosphate.(C) A description of the mutations generated by CRISPR/Cas9 and predicted amino acid changes in the 6 Mp*als*^{lof} mutants. (D) Gemmae grown for 18 days on Johnsons' medium with or without supplemented branched chain amino acids (800uM). Error bar is 5mm (D).

5.4.6. Branched chain amino acids partially restore growth of Mp*als*^{lof} plants

To understand how a lack of branched chain amino acid biosynthesis causes plant death in *M. polymorpha*, two Mp*als*^{lof} lines and wild type Tak-2 plants were grown on media supplemented with different combinations of valine, leucine and isoleucine. If Mp*als*^{lof} mutants are able to grow on media supplemented with isoleucine alone, this suggests that 2-ketobutyrate accumulation may be associated with the Mp*als*^{lof} lethal phenotype. Two Mp*als*^{lof} lines were selected for characterisation; the complete loss of function line Mp*als*^{lofF16} and the partial loss of function line Mp*als*^{lofF12}. Mp*als*^{lofF16} has a 1bp mutation and 4bp deletion at the start of the Mp*ALS* coding sequence leading to a frameshift and a premature stop codon 17 amino acids downstream of the mutation site. Mp*als*^{lofF16} does not survive on media without supplemented branched chain amino acids. The partial loss of function mutant Mp*als*^{lofF12} exhibits an in-frame 36bp insertion in the start of the Mp*ALS* coding sequence and can survive on medium without branched chain amino acids.

Plants were grown for 14 days on medium supplemented with all 3 branched chain amino acids, valine, leucine or isoleucine only, or pairwise combinations of valine, leucine and isoleucine. Plates were imaged using a NightOwl II LB983 fluorescence imaging system and plant area (mm^2) was determined based on chlorophyll autofluorescence. The square root of the data was analysed to fit the assumptions of normality of a one-way analysis of variance. Normality was checked using quantile-quantile (Q-Q) plots. Tukey-Kramer post-hoc test identified significant differences between samples.

Clear differences were observed between genotypes within each treatment group. Mp*als*^{lofF16} only survived on the isoleucine and valine medium and on the medium with all 3 amino acids but grew larger on medium with all 3 amino acids. In both cases Mp*als*^{lofF16} was smaller than Mp*als*^{lofF12} and Tak-2 (Figure 5.7.). Mp*als*^{lofF12} survived on all treatment plates, and was smaller than Tak-2 on all treatments except valine only, isoleucine and valine, and isoleucine and leucine treatment plates, and was larger than Mp*als*^{lofF16} on all treatment plates.

Interestingly, Tak-2 plants were extremely small on the leucine, and leucine and valine treatments, and did not survive on the isoleucine and leucine medium. This suggests that 'excess' leucine is toxic in plants, and that both valine and isoleucine work together to regulate leucine levels. To summarise, branched chain amino acids partially restored growth of Mp*als*^{lof} mutants to wild type size. Therefore, either the concentration of branched chain amino acids that could restore growth in chlorsulfuron-treated Tak-2 was not sufficiently high to restore growth in Mp*als*^{lof} , or the smaller size of Mp*als*^{lof} is not solely due to a lack of branched chain amino acids. Moreover, Mp*als*^{lofF16} did not grow on the isoleucine treatment, suggesting that isoleucine feedback inhibition of threonine deaminase doesn't reduce the Mp*als*^{lofF16} lethal phenotype.

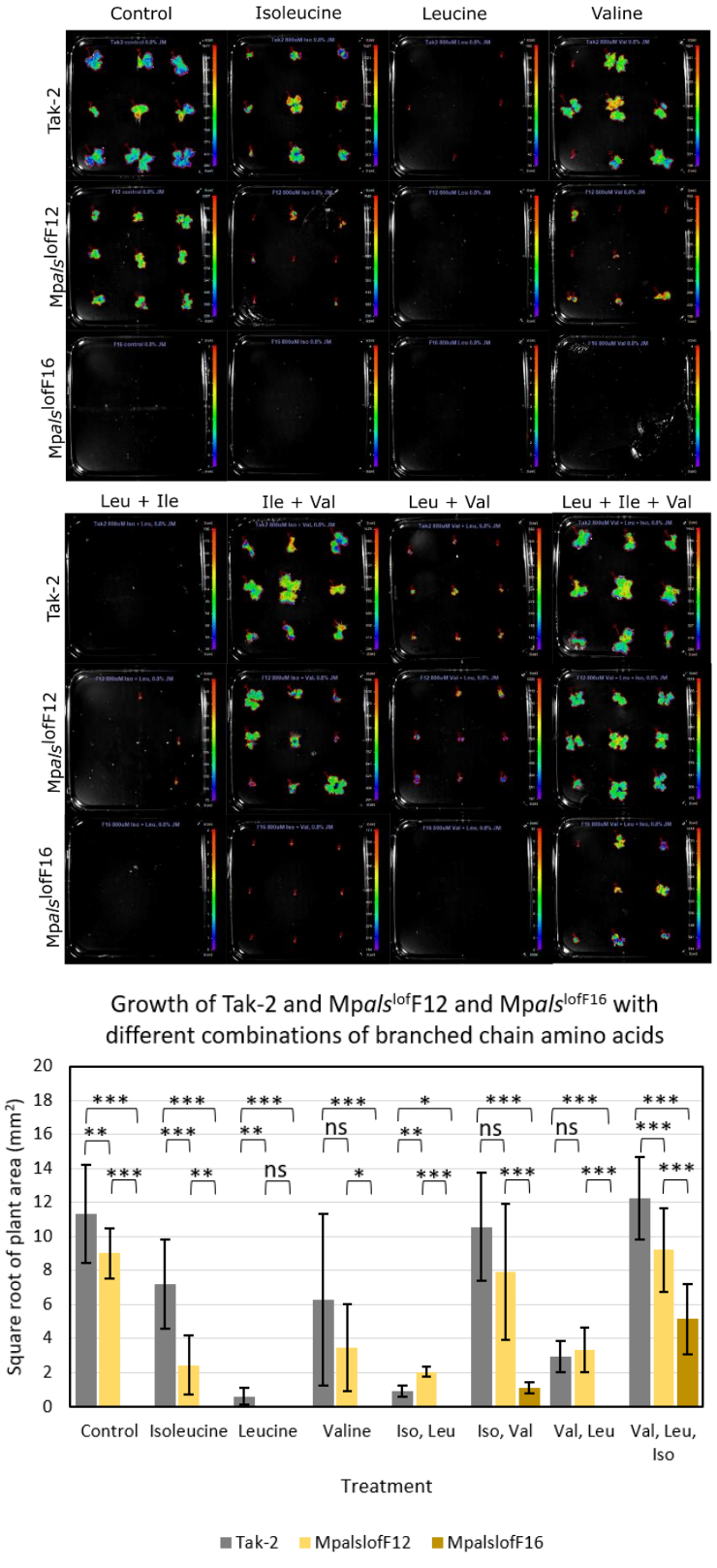


Figure 5.5. Branched chain amino acids partially restore growth in Mp*als*^{lof} but not isoleucine alone.

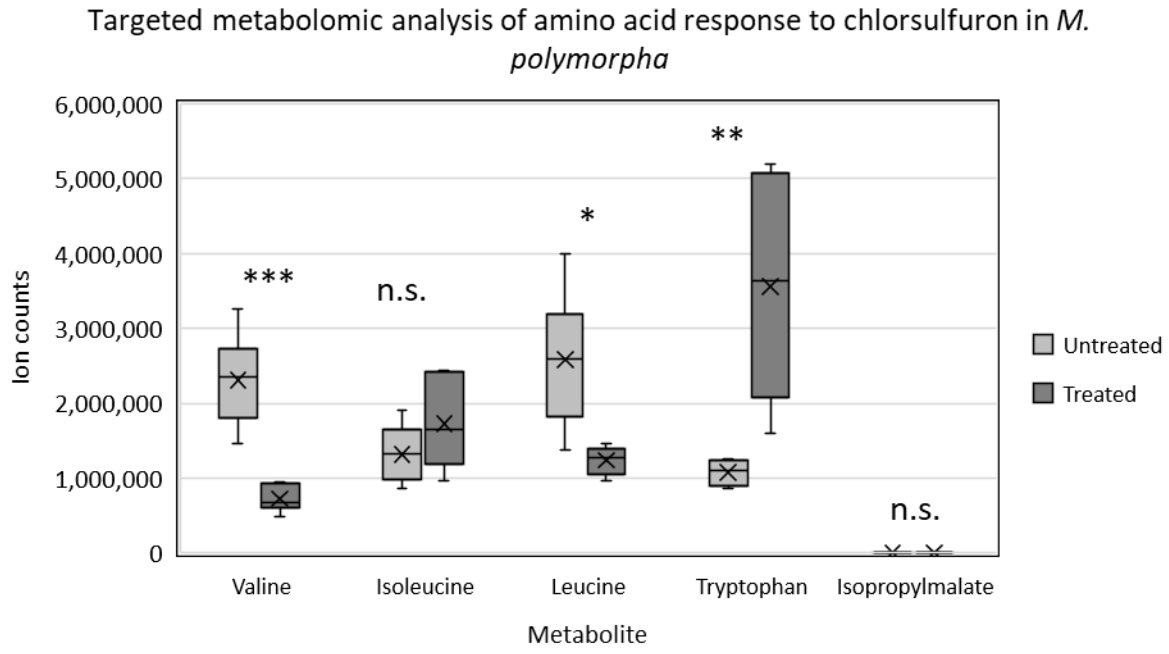
Growth assay of gemmae from Tak-2 and complete (F16) and partial (F12) Mpals^{lof} mutant lines grown for 14 days on Johnson's medium containing different combinations of branched chain amino acids (800uM) or the solvent (control). Plant area (mm^2) based on chlorophyll autofluorescence was measured on a NightOwl II LB983 fluorescence imaging system. Data was transformed to fit assumptions of normality. Statistical significance based on one-way ANOVA followed by Tukey-Kramer post-hoc test. n =3-18. Error bars are ± SD. *=p < 0.05, ** = p < 0.01, *** = p < 0.001, ns = not significant.

5.4.7. Chlorsulfuron treatment reduces the amount of valine and leucine but isoleucine levels are unchanged

To determine if chlorsulfuron alters the abundance of branched chain amino acids in plants, a targeted metabolomics approach was carried out on chlorsulfuron treated and untreated *M. polymorpha* Tak-2 plants. Plants were grown for 11 days on ½ Gamborg medium covered with a cellulose membrane, then moved onto control plates or plates treated with 100nM chlorsulfuron for 24h before collection. Metabolites from 6 untreated and 6 herbicide-treated biological replicates were extracted and analysed using liquid chromatography-tandem mass spectrometry (LC-MS/MS). The initial aim was to measure all components of the branched chain amino acid pathway; however, the abundances of these compounds were too low for detection. The metabolites that could be detected using targeted metabolomics were branched chain amino acids valine, leucine, isoleucine, the amino acid tryptophan, and isopropylmalate, a component of the leucine biosynthetic pathway. One isoleucine measurement (C4) was 20 fold higher than other C samples. It was labelled as an outlier and not included in the analysis. Normality was checked using Q-Q plots and significant differences between treatment groups for each metabolite were determined by student's t-tests.

The average ion counts for each metabolite in the 6 herbicide treated and 6 untreated samples are shown in Figure 5.6. Valine and leucine were detected at significantly lower abundances in chlorsulfuron treated plants compared to untreated plants. Valine was on average 70% less abundant in chlorsulfuron treated plants (0.72±0.60 10^{60} ions) compared to untreated plants (2.32±0.18 10^{06} ions) (*p*-value =0.0001), and leucine was on average 52% less abundant in chlorsulfuron treated plants (1.24±0.17 10^{06} ions) compared to untreated plants (2.58±0.89 10^{06} ions) (*p*-value=0.013). Isopropylmalate was 24% less abundant in chlorsulfuron treated plants (4.08±1.8 10^{03} ions) than untreated plants (6.20±2.3 10^{03} ions) on average, although the difference was not significant. By contrast, isoleucine abundance was not significantly different between chlorsulfuron-treated (1.73±0.60 10^{06} ions) and untreated plants (1.32±0.39 10^{06} ions). The reduced levels of valine and leucine are consistent with an inhibition of branched chain amino acid synthesis, and the lack of change in isoleucine, which is synthesised on a parallel pathway to leucine and valine, suggests that there are different impacts on the 3 branched chain amino acids after chlorsulfuron treatment.

Tryptophan levels increased more than 3-fold in chlorsulfuron treated plants (3.55±1.46 10^{06} ions) compared to untreated plants (1.08±0.17 10^{06} ions) (*p*-value=0.009). Tryptophan is not associated with the branched chain amino acid pathway, and the increase in tryptophan abundance in chlorsulfuron-treated plants suggests that chlorsulfuron treatment impacts multiple different amino acid biosynthesis pathways. Next, a non-targeted metabolomic analysis was conducted to detect if any other metabolites are significantly differential abundance chlorsulfuron treated and untreated *M. polymorpha* plants.



	Valine	*Isoleucine*	*Leucine*	*Tryptophan*	*Isopropylmalate*
Untreated (U)	2.31±0.60 10^{06}	1.32±0.39 10^{06}	2.58±0.89 10^{06}	1.08±0.17 10^{06}	6.20±2.3 10^{03}
Chlorsulfuron (T)	0.72±0.18 10^{06}	1.73±0.60 10^{06}	1.24±0.17 10^{06}	3.55±1.46 10^{06}	4.08±1.8 10^{03}
Log_2(fold change) (T/U)	-1.68	0.40	-1.06	1.72	-0.60

Figure 5.6. Chlorsulfuron treatment reduces valine and leucine levels but isoleucine is unaffected in *M. polymorpha* Tak-2 plants.

Targeted metabolic analysis showing the accumulation of valine, leucine and isoleucine, tryptophan, and isopropylmalate in untreated Tak-2 plants and in plants treated with 100nM chlorsulfuron. Plants were grown for 11 days, transferred to treatment plates for 24h, then extracted. Extracts were analysed using a targeted approach by LC-MS. Statistical significance based on student's t-test between treatment groups for each metabolite. *=$p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, n.s. = not significant. N=6 except in isoleucine untreated group n=5 (one outlier was removed). Box and whisker plot where whiskers are first and fourth quartiles, boxes represent second and third quartiles, and x is the mean.

5.4.8. Chlorsulfuron treatment causes an increase in other amino acids but no increase in quinate or 2-ketobutyrate

To understand the global metabolic changes caused by chlorsulfuron treatment, the same 12 samples used for the targeted metabolomic analysis were subjected to a non-targeted metabolomic analysis. Samples were analysed twice by LC-MS/MS using two separation methods: hydrophilic interaction liquid chromatography (HILIC) and reversed phase chromatography (RP). The two separation methods have slightly different metabolite coverage, and some compounds can only be detected using one of the methods, therefore using both allows the detection of a wider range of metabolites. Differences in metabolite levels were compared between chlorsulfuron treated and untreated samples. Samples had highly variable metabolic profiles, and one control sample, C4, was more similar to the herbicide-treated samples than to the other control samples (Figure 5.7A and B). Despite the large variability between biological replicates, a clear chlorsulfuron-induced metabolic signature was visible.

A 2-4.9-fold increase in amino acid levels was detected in herbicide treated plants (Figure 5.7C, Supplementary Table S5.1). Consistent with the targeted metabolomic analysis, aromatic amino acid tryptophan was more abundant as well as phenylalanine and tyrosine in chlorsulfuron treated plants. In addition, amino acids threonine, citrulline, glutamine and its analogue gamma-glutamylglutamine (gamma-glu-gln), alanine, serine, arginine and methionine were more abundant in chlorsulfuron-treated plants relative to untreated plants. Citrulline was the most differentially abundant metabolite with a $\log_2$(fold change) of 2.3 (*p*-value = 0.002), equivalent to a 4.9 fold increase in treated compared to untreated plants. 5 other compounds accumulated in chlorsulfuron treated plants. N-Nitrosoguvacoline, glycerophosphoglycerol, 2-hydroxycinnamic acid, agmatine and hydroxy-1,4-benzoquinone and 1-(sn-glycero-3-phospho)-1D-myo-inositol were 2.3-4 fold more abundant in chlorsulfuron treated plants (Figure 5.7C, Supplementary Table S5.1). Gluconic acid (*p*-value = 0.009) and leucine (*p*-value = 0.001) were significantly decreased following chlorsulfuron treatment (Figure 5.7C, Supplementary Table S5.1).

2-ketobutyrate and quinate levels were not significantly different in chlorsulfuron treated and untreated plants (Figure 5.7C, Supplementary Table S5.1). This indicates that quinate or 2-ketobutyrate are not accumulating in chlorsulfuron-treated plants and therefore are unlikely to be responsible for chlorsulfuron toxicity in *M. polymorpha*. One study in *Lemna minor* showed the accumulation of the transamination product of 2-ketobutyrate, 2-amiobutyrate, in herbicide treated plants (Rhodes et al., 1987), however 2-aminobutyrate was also not found to accumulate in chlorsulfuron-treated plants. All together with the targeted analysis data, these results indicate that the main metabolic effect of chlorsulfuron stress on wild type plants is an accumulation of a range of amino acids, combined with a decrease in abundance in valine and leucine.

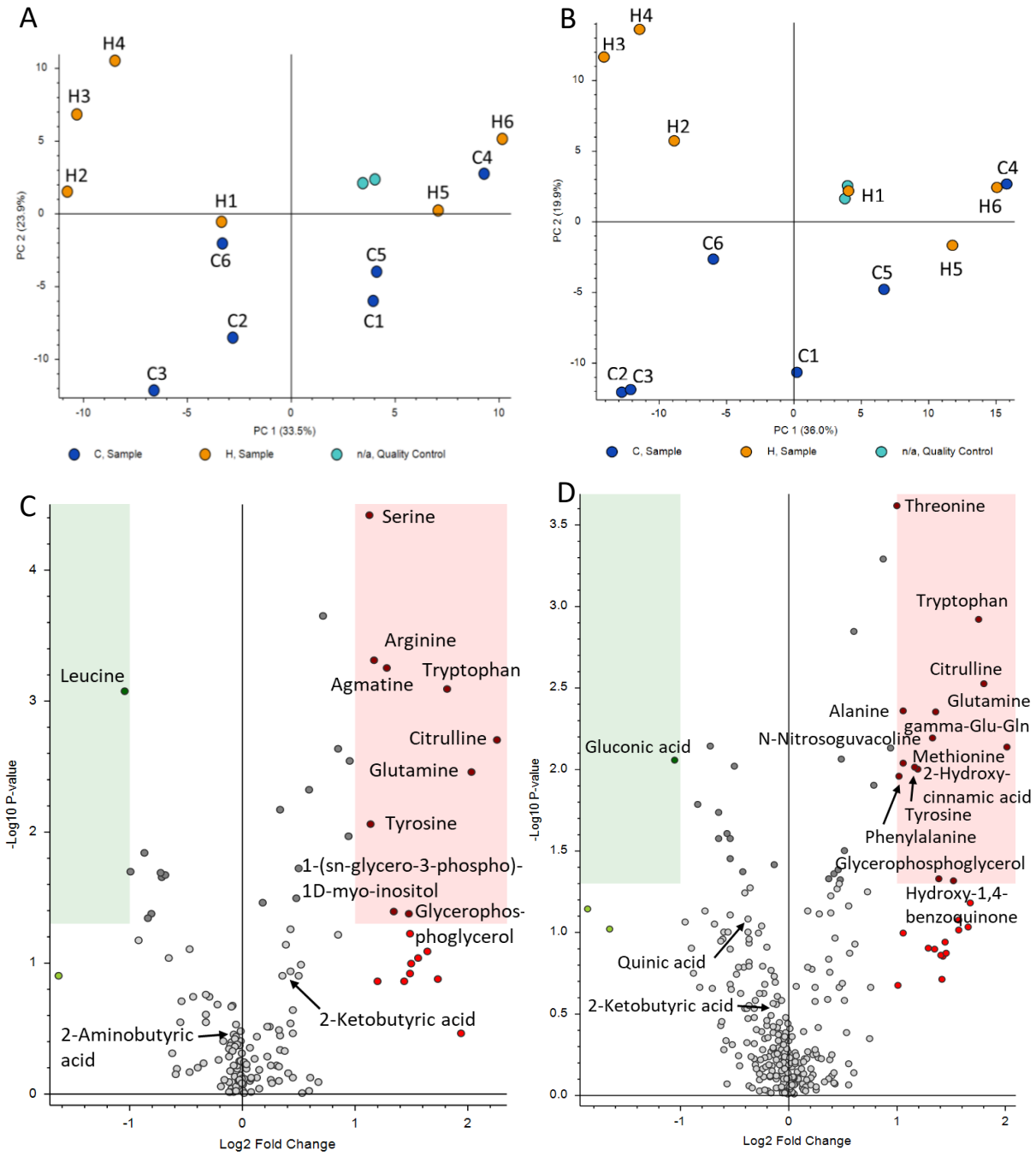


Figure 5.7. Non-targeted metabolic analysis of 24h chlorsulfuron treatment of *M. polymorpha* Tak-2 plants.

Samples were analysed by LC-MS/MS using reversed phase chromatography (RP) (A, C) and hydrophilic interaction liquid chromatography (HILIC) (B, D) separation methods. Principle component analyses of control (C1-6) and herbicide-treated (H1-6) samples and QC samples using RP (A) and HILIC (B). Volcano plots of differentially abundant metabolites in chlorsulfuron-treated vs. untreated plants using RP (C) and HILIC (D). Green box indicates significantly lower levels of metabolites and red box indicates significantly higher levels of metabolites in chlorsulfuron-treated compared to untreated Tak-2. Significance based on one-way ANOVA followed by Tukey HSD post-hoc test.

5.5. Discussion

In this chapter I aimed to understand the mechanism of chlorsulfuron lethality in *M. polymorpha*. Here I showed that wild type Tak-2 plants grown on medium with a lethal dose of chlorsulfuron and all 3 branched chain amino acids grew to the size of untreated plants, confirming that supplementing branched chain amino acids reverses chlorsulfuron induced plant death and growth inhibition. Isoleucine treatment did not restore the growth of chlorsulfuron-treated Tak-2. Complete deletion of Mp*als* is lethal, and supplemented branched chain amino acids in the medium reversed the Mp*als*^{lof} lethal phenotype. I showed that the complete Mp*als*^{lof} line Mp*als*^{lofF16} did not survive on medium with isoleucine, indicating that the feedback inhibition of threonine deaminase by isoleucine did not suppress growth inhibition in chlorsulfuron-treated plants and in Mp*als*^{lof} plants. Using targeted and non-targeted metabolomic analyses I observed that chlorsulfuron treatment caused a decrease in valine, leucine, and gluconic acid, and an increase in many other amino acids. There was no difference in quinate, 2-ketobutyrate or 2-aminobutyrate abundance between chlorsulfuron treated and untreated Tak-2.

I generated 6 independent Mp*als*^{lof} lines using the CRISPR/Cas9 system and characterised the growth of 2 of these lines. All Mp*als*^{lof} lines die - or in the case of the partial Mp*als*^{lofF12} line, grow smaller than Tak-2 - on media without supplemented branched chain amino acids, indicating that they lack or have impaired branched chain amino acid biosynthesis. This is the first report of a complete knock-out ALS mutant in plants. Because *als*^{lof} plants need continuous branched chain amino acid treatment to survive, producing homozygous *als*^{lof} mutants in a diploid organism that grows on soil such as *A. thaliana* would be challenging. This validates the use of *M. polymorpha* as a system for researching the physiological impacts of lethal mutations in plants.

The results in this study do not support the hypothesis that the accumulation of quinate or 2-ketobutyrate is the mechanism of toxicity in chlorsulfuron-treated or Mp*als*^{lof} plants. Tak-2 plants had severely inhibited growth on medium with isoleucine and a lethal dose of chlorsulfuron, while complete Mp*als*^{lof} plants did not survive on medium supplemented with isoleucine, indicating that reducing 2-ketobutyrate production by isoleucine mediated feedback inhibition of threonine deaminase didn't improve growth of chlorsulfuron treated Tak-2 and Mp*als*^{lof}. Moreover, the levels of 2-ketobutyrate, it's transamination product 2-amiobutyrate, and quinate in chlorsulfuron-treated Tak-2 were not significantly different from untreated plants. Therefore, although dose response curves indicted that quinate and 2-ketobutyrate killed Tak-2 at high concentrations, these data do not support the scenarios of quinate or 2-ketobutyrate being responsible for growth inhibition or plant death in Mp*als*^{lof} and chlorsulfuron-treated plants. This is consistent with research in imazaquin-treated maize (Shaner & Singh, 1993) in which supplemented isoleucine decreased 2-ketobutyrate and 2-aminobutyrate levels but this did not coincide with improved growth.

Valine, leucine and isoleucine supplemented to medium with a lethal chlorsulfuron dose restored growth of Tak-2 to untreated levels, indicating that a lack of branched chain amino acids in chlorsulfuron-treated plants leads to growth

inhibition. The same concentration of branched chain amino acids only partially restored growth in Mp*als*^{lof}. This could be explained by Mp*ALS* still being active in chlorsulfuron treated Tak-2 at a much-reduced rate. Therefore, the concentration of branched chain amino acids that restores growth in chlorsulfuron treated Tak-2 may not be high enough to increase growth of Mp*als*^{lof} to Tak-2 levels.

If branched chain amino acid starvation is the mechanism of plant death I would predict a decrease in all 3 branched chain amino acids in chlorsulfuron treated plants. However, chlorsulfuron treatment caused a decrease in valine and leucine but no change in isoleucine. Decreased valine and leucine and increased isoleucine levels was observed in common vetch (*Vicia sativa*) treated with the ALS inhibitor imazamox (García-Garijo et al., 2013) and in imazapyr treated pea plants (Zabalza et al., 2011). ALS requires two pyruvate molecules for leucine and valine biosynthesis and one pyruvate and one 2-ketobutyrate molecule for isoleucine biosynthesis, so it was suggested that ALS inhibition causes an imbalance of pyruvate/2-ketobutyrate with pyruvate being diverted to fermentative metabolism, contributing to growth inhibition and plant death (Gaston et al., 2002; Zabalza et al., 2005). However, pyruvate wasn't significantly reduced by chlorsulfuron treatment in this study. Instead, methionine and threonine were more abundant in chlorsulfuron treated plants. These amino acids act as substrates for isoleucine synthesis, therefore their accumulation may explain the higher isoleucine levels compared to valine and leucine in chlorsulfuron-treated plants in this study.

A significant increase in other amino acids occurred in chlorsulfuron treated plants. Higher pools of free amino acids are often reported in plants treated with ALS inhibitors (Gaston et al., 2002; Orcaray et al., 2010; Rhodes et al., 1987; Trenkamp et al., 2009). The aromatic amino acids tryptophan, phenylalanine and tyrosine, and the amino acids methionine, glutamine, citrulline, threonine, alanine, and serine were all significantly more abundant in chlorsulfuron-treated plants. This is consistent with a study by Hikosaka et al. (2021) who identified alanine, aspartic acid, serine, and threonine as markers of ALS inhibition in *Schoenoplectiella juncoides* using three ALS inhibitors. Citrulline was the most differentially abundant metabolite, with a 4.6 fold increase in chlorsulfuron treated compared to untreated plants. Citrulline is a precursor for arginine biosynthesis, and also acts as a hydroxyl radical scavenger (Akashi et al., 2001). Therefore some of the accumulating amino acids may help to reduce oxidative stress caused by chlorsulfuron treatment.

Two hypotheses would explain the observed increase in free amino acids after ALS inhibition. The inhibition of branched chain amino acid biosynthesis may cause general protein breakdown releasing previously bound amino acids. Maize plants showed a 40% reduction in soluble protein levels and 32% increase in levels of free amino acids after 24h of ALS inhibitor treatment (Shaner & Reider, 1986). Therefore, the decrease in branched chain amino acid pools may be masked by the increase in free amino acids due to proteolysis. Another explanation is the disruption of the branched chain amino acid pathway causes an increase in flux to other amino acid pathways. The branched chain amino acid biosynthetic pathway is closely tied to other amino acid pathways including the shikimate pathway. Glyphosate treated pea plants also

showed increased free amino acid content including an accumulation of branched chain amino acids and decrease in aromatic amino acids (Orcaray et al., 2010). Thus, herbicides that inhibit amino acid pathways may cause similar metabolic changes, where inhibiting one amino acid pathway increases fluxes to other amino acid pathways. Amino acids may have a key role in plant abiotic stress response by acting as osmolytes, nitrogen storage proteins or precursors for secondary metabolism (Hildebrandt, 2018).

A handful of non-amino acid metabolites accumulated in chlorsulfuron treated plants. N-nitrosoguvacoline, glycerophosphoglycerol, 2-hydroxycinnamic acid, agmatine, hydroxy-1,4-benzoquinone and 1-(sn-glycero-3-phospho)-1D-myo-inositol were all significantly more abundant in chlorsulfuron treated plants. Hydroxycinnamic acid amides are synthesised from hydroxycinnamic acids and agmatine and have roles in plant defence against pathogens (Muroi et al., 2009). N-nitrosoguvacoline is a potential toxic metabolite with mutagenic activity in Salmonella (Wang & Peng, 1996) Quinones are defence compounds in plants with cytotoxic activity (Farkas et al., 1960), and glycerophosphoglycerol, a hydrolytic product of phospholipids, is associated with abiotic stress response (Yuan et al., 2018) but there is no evidence for its toxicity. Therefore two potentially toxic metabolites, N-nitrosoguvacoline and hydroxy-1,4-benzoquinone and one stress response metabolite accumulated after chlorsulfuron treatment.

In conclusion, the data doesn't support the hypotheses that 2-ketobutyrate (hypothesis 2) or quinate (hypothesis 3) accumulation in chlorsulfuron-treated plants leads to plant death. One caveat of this study is that the plants were extracted after 24h on herbicide treated medium, therefore their accumulation may occur after this timepoint. However significant increases in 2-ketobutyrate in *Salmonella typhimurium*, 2-aminobutyrate in duckweed (*Lemna minor)* and quinate in pea (*Pisum sativum*) were observed within 20 minutes to 24h after herbicide treatment (Larossa et al., 1987; Orcaray et al., 2010; Rhodes et al., 1987). The partial restoration of Mp*als*^{lof} growth with branched chain amino acids suggests that a lack of branched chain amino acids isn't the only mechanism of plant death in chlorsulfuron-treated plants (hypothesis 1). Based on these results I propose a 4th hypothesis for chlorsulfuron lethality. I propose that chlorsulfuron kills plants by blocking the branched chain amino acid pathway which destabilises other amino acid pathways and causes increases in potentially toxic metabolites. I predict that all together these metabolite changes disrupt essential plant processes leading to plant death. If NTSR arises from genes transcriptionally induced by chlorsulfuron treatment (as suggested in chapter 4), understanding the physiological and metabolic changes caused by chlorsulfuron treatment will help to understand how plants evolve resistance mechanisms to overcome the herbicide stress. For example NTSR mechanisms may include the compartmentalisation or detoxification of the herbicide and of toxic metabolites, as well as mechanisms to maintain amino acid homeostasis.

Appendix

Compound	RT (min)	Method of detection	*P*-value (T/U)	Fold change (T/U)	Regulation
1-(sn-glycero-3-phospho)-1D-myo-inositol	3.83	RP	$4.12\ 10^{-02}$	1.3	Up
2-Aminobutyric acid/γ-Aminobutyric acid	4.36	RP	$3.72\ 10^{-01}$	0.1	No change
2-Hydroxycinnamic acid	12.89	HILIC	$9.75\ 10^{-03}$	1.2	Up
2-Ketobutyric acid	16.72	HILIC	$5.54\ 10^{-01}$	-0.1	No change
Agmatine	3.34	RP	$5.62\ 10^{-04}$	1.3	Up
Alanine	14.89	HILIC	$4.40\ 10^{-03}$	1.1	Up
Arginine	3.61	RP	$4.88\ 10^{-04}$	1.2	Up
Citrulline	16.82	HILIC	$3.01\ 10^{-03}$	1.8	Up
	4.12	RP	$2.01\ 10^{-03}$	2.3	Up
gamma-Glu-Gln	15.43	HILIC	$7.33\ 10^{-03}$	2.0	Up
Gluconic acid	12.94	HILIC	$8.83\ 10^{-03}$	-1.1	Down
Glutamine	15.71	HILIC	$4.43\ 10^{-03}$	1.4	Up
	3.86	RP	$3.53\ 10^{-03}$	2.0	Up
Glycerophosphoglycerol	11.91	HILIC	$4.71\ 10^{-02}$	1.4	Up
	4.09	RP	$4.21\ 10^{-02}$	1.5	Up
Leucine	10.43	RP	$8.49\ 10^{-04}$	-1.0	Down
Methionine	11.00	HILIC	$9.25\ 10^{-03}$	1.1	Up
N-Nitrosoguvacoline	15.71	HILIC	$6.45\ 10^{-03}$	1.3	Up
Phenylalanine	9.50	HILIC	$1.11\ 10^{-02}$	1.0	Up
Quinic acid	10.44	HILIC	$7.70\ 10^{-01}$	-0.2	No change
Serine	3.73	RP	$3.85\ 10^{-05}$	1.1	Up
Threonine	14.24	HILIC	$2.41\ 10^{-04}$	1.0	Up
Tryptophan	13.88	RP	$8.16\ 10^{-04}$	1.8	Up
	11.49	HILIC	$1.21\ 10^{-03}$	1.8	Up
Tyrosine	12.89	RP	$1.01\ 10^{-02}$	1.2	Up

Figure S5.1. Non-targeted metabolic analysis summary table of metabolites significantly differentially abundant between chlorsulfuron treated and untreated *M. polymorpha*

Tak-2 plants were grown for 11 days then moved to chlorsulfuron medium for 24h before extraction. Samples were analysed by LC-MS/MS using reversed phase chromatography (RP) and hydrophilic interaction liquid chromatography (HILIC) separation methods. RT is retention time.
p-value is calculated with a Tukey-HSD post-hoc test after an analysis of variance.

Chapter 6

General Discussion

The continuous use of herbicides has imposed a selection pressure on weed populations to evolve herbicide metabolic resistance. Herbicide metabolic resistance is a form of non-target site herbicide resistance (NTSR) which involves the enhanced detoxification of herbicides most frequently by glutathione s-transferases (GSTs) and cytochrome P450s (CYPs). In chapters 2 and 3 I show that herbicide metabolising CYPs and GSTs belong to the two largest GST classes (Phi and Tau) and two largest CYP clans (71 and 72) in vascular plants. In chapter 2 I show that 71 and 72 clans originated in the streptophyte algal lineage and underwent gene duplications and expansions in two phases, once in embryophytes, and once in the ancestor of vascular plants. Based on these findings I propose two hypotheses: CYPs that have the ability to detoxify herbicides are only found in the 71 and 72 clans (hypothesis 1), or there is an equal probability that a CYP from any clan can metabolise herbicides, and the reason that herbicide metabolising CYPs have only been identified in 71 and 72 clans is because the majority of CYPs in weeds belong to these clans (hypothesis 2). Further functional experiments of CYPs from non-71 and 72 clans are needed to clarify this question. In chapter 3 I show that GSTs that directly metabolise herbicides by glutathione (GSH) conjugation belong to the Phi and Tau classes. The Tau class evolved in the common ancestor of streptophytes, while the Phi class is also present in bacteria and fungi, and first appeared in Archaeplastida in the common ancestor of streptophytes. The Phi class expanded in early land plant evolution before the divergence of vascular plants and bryophytes, while the Tau class underwent gene duplications in the common ancestor of vascular plants.

In contrast to CYPs, experimental evidence indicates that only GSTs belonging to 5 of the 22 Archaeplastida classes, namely Tau, Phi, Theta, Zeta and TCHQD, can catalyse GSH conjugation reactions. Therefore the hypothesis that there is an equal likelihood that a GST from any class can detoxify herbicides can be rejected (hypothesis 2). However while hypothesis 2 doesn't hold true in GSTs, it may still be correct for herbicide metabolising enzymes in the CYP family, where there is more functional overlap between genes in different clans.

In addition the phylogenetic analyses of the CYP and GST enzyme families in Archaeplastida indicated that the clans and classes that contain herbicide metabolising enzymes are present in the genome of *M. polymorpha*, therefore enzymes belonging to the 71 and 72 CYP clans and Tau and Phi GSTs classes in *M. polymorpha* have the potential to metabolise herbicides such as chlorsulfuron.

The genes involved in herbicide detoxification have been described in mutant resistant backgrounds, either alone or combined with herbicide treated sensitive plants. In this case, I used a sensitive plant to investigate genes that may function in herbicide detoxification. In chapter 4 I identified a CYP, Mp*CYP822A1*, that was up-regulated by chlorsulfuron treatment in sensitive plants. The over-expression of Mp*CYP822A1* increased plant growth under a non-lethal chlorsulfuron dose compared to control plants. This is the first report of a gene conferring NTSR which was identified exclusively through transcriptomic analyses of a sensitive plant, and is the first report of a NTSR gene in *M. polymorpha*.

Although metabolomic experiments need to confirm whether Mp*CYP822A1* is directly metabolising chlorsulfuron or a toxic by-product, this result suggests that NTSR arises from mutations in genes that are induced by the herbicide treatment. In addition, plants with a point mutation in Mp*ALS* (Mp*ALS*^{TSR}) which are resistant to chlorsulfuron did not up-regulate any genes that could confer metabolic resistance in response to chlorsulfuron treatment. Altogether the transcriptomic data suggests that chlorsulfuron metabolic resistance isn't conferred by genes that directly detect the xenobiotic chemical, but by genes induced by Mp*ALS* inhibition. Mp*CYP822A1* over-expressing plants were twice as large than control plants when treated with a growth inhibiting dose of chlorsulfuron, but did not survive on a lethal chlorsulfuron dose, indicating that the over-expression of Mp*CYP822A1* only confers a low level of resistance to chlorsulfuron. Therefore, the high level of resistance observed in NTSR weeds is likely to evolve from mutations in multiple genes each conferring a small increase in resistance. This is important because it shows how mutations in genes may evolve in sensitive weeds under low herbicide doses to confer herbicide NTSR.

Chlorsulfuron kills plants at very low concentrations, however the mechanism of chlorsulfuron lethality is unclear. I show that chlorsulfuron treatment at a nanomolar concentration led to the de-regulation of thousands of genes in sensitive plants, which represented 30% of expressed genes in the samples or 15% of the total genes in the genome. By contrast chlorsulfuron-treated Mp*ALS*^{TSR} plants differentially expressed less than 20 genes in response to chlorsulfuron treatment. These two results suggest that the transcriptional response to chlorsulfuron is caused by the inhibition of the ALS enzyme, and not by the detection of the chlorsulfuron xenobiotic. Therefore I hypothesise that chlorsulfuron lethality is due to an inactive ALS enzyme.

To independently test this hypothesis I generated loss of function (LOF) mutations in the Mp*ALS* gene. Although mutations in ALS have been generated in bacteria (Wiersma et al., 1990; Zhao et al., 2021) and the down-regulation of ALS genes in tomato was achieved using RNA interference (Gao et al., 2014), these are the first loss of function *als* mutants ever achieved in a plant. I've exploited the resources of *M. polymorpha* to do this, by growing Mp*als*^{lof} lines on medium supplemented with branched chain amino acids. To determine whether they had a mutation in Mp*ALS*, I grew gemmae from each transformant on medium with and without branched chain amino acids. The parents of any plants that died without branched chain amino acids were selected for sequencing, and also used for experiments. Due to the haploid nature of *M. polymorpha* there was no requirement to grow plants on soil or cross them to obtain homozygous Mp*als*^{lof} mutants. While it may be technically possible to generate homozygous *als*^{lof} plants in a diploid seed plant such as *A. thaliana*, it would be challenging and time consuming, due to the difficulty of supplementing branched chain amino acids to soil-grown plants, and the need to generate homozygous lines to obtain a visible At*als*^{lof} phenotype.

In chapter 5 I show that Mp*ALS* deletion is lethal, supporting the hypothesis that a lack of Mp*ALS* activity causes plant death. Mp*als*^{lof} mutants survive on medium supplemented with branched chain amino acids, but are smaller than

Tak-2 plants. Therefore branched chain amino acids only partially restored growth of Mp*als*^{lof}. This suggests that a lack of branched chain amino acids in plants may only partially explain lethality caused by a defective Mp*ALS* gene. In addition, chlorsulfuron kills wild type plants at nanomolar concentrations, while 10^4 fold higher concentrations of branched chain amino acids are required to restore growth to the rate of untreated plants. These results support the hypothesis that a lack of branched chain amino acids in plants may only be partially responsible for chlorsulfuron lethality. Metabolomic analyses revealed that chlorsulfuron-treated plants have reduced abundances of valine and leucine, no change in isoleucine levels, and higher levels of other amino acids. This confirms that chlorsulfuron causes decreases in branched chain amino acids, and the higher levels of methionine and threonine which act as substrates for isoleucine biosynthesis may explain the lack of decrease in isoleucine abundance following chlorsulfuron treatment. The higher levels of other amino acids could be explained by an increase in protein degradation or increased amino acid biosynthesis due to higher fluxes to other amino acid pathways. If increased proteolysis is occurring this would mask the decrease in branched chain amino acids due to ALS inhibition, and would explain why branched chain amino acids are still detectable in chlorsulfuron-treated plants. The physiological implications of increased amino acids are not known. Amino acids are known for their roles as signalling molecules or potential roles as osmolytes to relief osmotic stress, therefore they may help the plant tolerate the herbicide induced stress. N-nitrosoguvacoline and hydroxy-1,4-benzoquinone were more abundant in herbicide treated plants. The accumulation of these compounds could be toxic, suggesting another possible mechanism of chlorsulfuron lethality. Further experiments are required to test their toxicity in plants.

To conclude, the results in this thesis are consistent with the hypothesis that chlorsulfuron lethality is due to the inhibition of Mp*ALS* which leads to a reduction in branched chain amino acids, triggering a large de-regulation of gene expression and eventually plant death. I predict that NTSR to chlorsulfuron arises from mutations in genes that are induced by the chlorsulfuron inhibition of Mp*ALS*, and belong to the GST classes and CYP clans that diversified on land. Therefore I predict that NTSR evolves from the herbicide stress response, and genes conferring NTSR are most likely to be genes with endogenous functions in stress response. Future studies to understand how the transcriptional response to chlorsulfuron is activated and how it leads to plant death would test this hypothesis and further our understanding of the action of the highly potent ALS inhibiting herbicides.

References

Akashi, K., Miyake, C., & Yokota, A. (2001). Citrulline, a novel compatible solute in drought-tolerant wild watermelon leaves, is an efficient hydroxyl radical scavenger. *FEBS Letters*, *508*(3), 438–442. https://doi.org/10.1016/S0014-5793(01)03123-4

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), 1–12. https://doi.org/10.1186/gb-2010-11-10-r106

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'donovan, C., Redaschi, N., & Yeh, L.-S. L. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*, *32*, D115–D119. https://doi.org/10.1093/nar/gkh131

Aravind, L., & Koonin, E. V. (2000). SAP - A putative DNA-binding motif involved in chromosomal organization. *Trends in Biochemical Sciences*, *25*(3), 112–114. https://doi.org/10.1016/S0968-0004(99)01537-6

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*(SUPPL. 2), W202–W208. https://doi.org/10.1093/nar/gkp335

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., & Werck-Reichhart, D. (2011). Cytochromes p450. *The Arabidopsis Book*, *9*, e0144. https://doi.org/10.1199/tab.0144

Batard, Y., LeRet, M., Schalk, M., Robineau, T., Durst, F., & Werck-Reichhart, D. (1998). Molecular cloning and functional expression in yeast of CYP76B1, a xenobiotic-inducible 7-ethoxycoumarin *O* -de-ethylase from *Helianthus tuberosus*. *The Plant Journal*, *14*(1), 111–120. https://doi.org/10.1046/j.1365-313X.1998.00099.x

Bateman, R. M., Crane, P. R., DiMichele, W. A., Kenrick, P. R., Rowe, N. P., Speck, T., & Stein, W. E. (1998). EARLY EVOLUTION OF LAND PLANTS: Phylogeny, Physiology, and Ecology of the Primary Terrestrial Radiation. *Annual Review of Ecology and Systematics*, *29*(1), 263–292. https://doi.org/10.1146/annurev.ecolsys.29.1.263

Bellinder, R. R., Gummesson, G., & Karlsson, C. (1994). Percentage-Driven Government Mandates for Pesticide Reduction: The Swedish Model. *Weed Technology*, *8*(2), 350–359. https://doi.org/10.1017/s0890037x00038914

Benekos, K., Kissoudis, C., Nianiou-Obeidat, I., Labrou, N., Madesis, P., Kalamaki, M., Makris, A., & Tsaftaris, A. (2010). Overexpression of a specific soybean GmGSTU4 isoenzyme improves diphenyl ether and chloroacetanilide herbicide tolerance of transgenic tobacco plants. *Journal of Biotechnology*, *150*(1), 195–201. https://doi.org/10.1016/j.jbiotec.2010.07.011

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. In *Journal of the Royal Statistical Society* (Vol. 57, Issue 1). https://doi.org/10.2307/2346101

Betts, H. C., Puttick, M. N., Clark, J. W., Williams, T. A., Donoghue, P. C. J., & Pisani, D. (2018). Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nature Ecology and Evolution*, *2*(10), 1556–1562. https://doi.org/10.1038/s41559-018-0644-x

Bilang, J., & Sturm, A. (1995). Cloning and characterization of a glutathione S-transferase that can be photolabeled with 5-azido-indole-3-acetic acid. *Plant Physiology*, *109*(1), 253–260. https://doi.org/10.1104/pp.109.1.253

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., … Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354. https://doi.org/10.1093/nar/gkaa977

Board, P. G., Baker, R. T., Chelvanayagam, G., & Jermiin, L. S. (1997). Zeta, a novel class of glutathione transferases in a range of species from plants to humans. *Biochemical Journal*, *328*(3), 929–935. https://doi.org/10.1042/bj3280929

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S. S., Althoff, F., Araki, T., Arteaga-Vazquez, M. A., Balasubrmanian, S., Barry, K., Bauer, D., Boehm, C. R., … Schmutz, J. (2017). Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell*, *171*(2), 287-304.e15. https://doi.org/10.1016/J.CELL.2017.09.030

Bravo, J. M., Campo, S., Murillo, I., Coca, M., & San Segundo, B. (2003). Fungus- and wound-induced accumulation of mRNA containing a class II chitinase of the pathogenesis-related protein 4 (PR-4) family of maize. *Plant Molecular Biology*, *52*(4), 745–759. https://doi.org/10.1023/A:1025016416951

Brazier, M., Cole, D. J., & Edwards, R. (2002). O-Glucosyltransferase activities toward phenolic natural products and xenobiotics in wheat and herbicide-resistant and herbicide-susceptible black-grass (*Alopecurus myosuroides*). *Phytochemistry*, *59*(2), 149–156. https://doi.org/10.1016/S0031-9422(01)00458-7

Brionne, A., Juanchich, A., & Hennequet-Antier, C. (2019). ViSEAGO: A Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity. *BioData Mining*, *12*(1). https://doi.org/10.1186/s13040-019-0204-1

Burnet, Michael; Hart, Quentin; Holtum, Josheph; Powles, S. B. (1994). Resistance to nine herbicide classes in a population of rigid ryegrass (*Lolium rigidum*). *Weed Science*, *42*, 369–377. https://doi.org/10.1007/s13398-014-0173-7.2

Burnet, M. W. M., Loveys, B. R., Holtum, J. A. M., & Powles, S. B. (1993). Increased Detoxification Is a Mechanism of Simazine Resistance in *Lolium rigidum*. *Pesticide Biochemistry and Physiology*, *46*(3), 207–218. https://doi.org/10.1006/PEST.1993.1052

Burns, E. E., Keith, B. K., Talbert, L. E., & Dyer, W. E. (2018). Non-target site resistance to flucarbazone, imazamethabenz and pinoxaden is controlled by three linked genes in *Avena fatua*. *Weed Research*, *58*(1), 8–16. https://doi.org/10.1111/wre.12279

Burton, J. D., Gronwald, J. W., Somers, D. A., Connelly, J. A., Gengenbach, B. G., & Wyse, D. L. (1987). Inhibition of plant acetyl coenzyme A carboxylase by the herbicides sethoxydim and haloxyfop. *Biochemical and Biophysical Research Communications*, *148*, 1039–1044.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *BIOINFORMATICS APPLICATIONS NOTE*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chen, J.-H., Jiang, H.-W., Hsieh, E.-J., Chen, H.-Y., Chien, C.-T., Hsieh, H.-L., & Lin, T.-P. (2012). Drought and Salt Stress Tolerance of an Arabidopsis Glutathione *S* -Transferase U17 Knockout Mutant Are Attributed to the Combined Effect of Glutathione and Abscisic Acid . *Plant Physiology*, *158*(1), 340–351. https://doi.org/10.1104/pp.111.181875

Chiyoda, S., Ishizaki, K., Kataoka, H., Yamato, K. T., & Kohchi, T. (2008). Direct transformation of the liverwort Marchantia polymorpha L. by particle bombardment using immature thalli developing from spores. *Plant Cell Reports*, *27*(9), 1467–1473. https://doi.org/10.1007/s00299-008-0570-5

Choi, J. H., Lou, W., & Vancura, A. (1998). A novel membrane-bound glutathione S-transferase functions in the stationary phase of the yeast Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *273*(45), 29915–29922. https://doi.org/10.1074/jbc.273.45.29915

Christopher, J. T., Powles, S. B., Liljegren, D. R., & Holtum, J. A. M. (1991). Cross-Resistance to Herbicides in Annual Ryegrass (*Lolium rigidum*) II. Chlorsulfuron Resistance Involves a Wheat-Like Detoxification System. *Plant Physiology*, *95*, 1036–1043. http://www.plantphysiol.org/content/plantphysiol/95/4/1036.full.pdf

Cole, D. J., & John Owen, W. (1987). Influence of monooxygenase inhibitors on the metabolism of the herbicides chlortoluron and metolachlor in cell suspension cultures. *Plant Science*, *50*(1), 13–20. https://doi.org/10.1016/0168-9452(87)90025-2

Coschigano, P. W., & Magasanik, B. (1991). The URE2 gene product of Saccharomyces cerevisiae plays an important role in the cellular response to the nitrogen source and has homology to glutathione s-transferases. *Molecular and Cellular Biology*, *11*(2), 822–832. https://doi.org/10.1128/mcb.11.2.822

Cummins, I., Cole, D. J., & Edwards, R. (1999). A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in blackgrass. *Plant Journal*, *18*(3), 285–292. https://doi.org/10.1046/j.1365-313X.1999.00452.x

Cummins, I., Dixon, D. P., Freitag-Pohl, S., Skipsey, M., & Edwards, R. (2011). Multiple roles for plant glutathione transferases in xenobiotic detoxification. *Drug Metabolism Reviews*, *43*(2), 266–280. https://doi.org/10.3109/03602532.2011.552910

Cummins, I., O'Hagan, D., Jablonkai, I., Cole, D. J., Hehn, A., Werck-Reichhart, D., & Edwards, R. (2003). Cloning, characterization and regulation of a family of phi class glutathione transferases from wheat. *Plant Molecular Biology*, *52*(3), 591–603. https://doi.org/10.1023/A:1024858218804

Cummins, I., Wortley, D. J., Sabbadin, F., He, Z., Coxon, C. R., Straker, H. E., Sellars, J. D., Knight, K., Edwards, L., Hughes, D., Kaundun, S. S., Hutchings, S.-J., Steel, P. G., Edwards, R., & Dixon, R. A. (2013). Key role for a glutathione transferase in multiple-herbicide resistance in grass weeds. *Proceedings of the National Academy of Science*, *110*(15). https://doi.org/10.1073/pnas.1221179110

Danielson, P. B. (2002). The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans.

Current Drug Metabolism, *3*(6), 561–597. http://www.ncbi.nlm.nih.gov/pubmed/12369887

Das, M., Reichman, J. R., Haberer, G., Welzl, G., Aceituno, F. F., Mader, M. T., Watrud, L. S., Pfleeger, T. G., Gutiérrez, R. A., Schäffner, A. R., & Olszyk, D. M. (2010). A composite transcriptional signature differentiates responses towards closely related herbicides in Arabidopsis thaliana and Brassica napus. *Plant Molecular Biology*, *72*(4), 545–556. https://doi.org/10.1007/s11103-009-9590-y

Délye, C., Gardin, J. A. C., Boucansaud, K., Chauvel, B., & Petit, C. (2011). Non-target-site-based resistance should be the centre of attention for herbicide resistance research: Alopecurus myosuroides as an illustration. *Weed Research*, *51*(5), 433–437. https://doi.org/10.1111/j.1365-3180.2011.00864.x

Délye, C, Duhoux, A., Gardin, J. A. C., Gouzy, J., & Carrère, S. (2018). High conservation of the transcriptional response to acetolactate-synthase-inhibiting herbicides across plant species. *Weed Research*, *58*(1), 2–7. https://doi.org/10.1111/wre.12276

Délye, Christophe. (2013). Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: A major challenge for weed science in the forthcoming decade. *Pest Management Science*, *69*(2), 176–187. https://doi.org/10.1002/ps.3318

Didierjean, L., Gondet, L., Perkins, R., Cindy Lau, S.-M., Schaller, H., O, D. P., & Werck-Reichhart, D. (2002). Engineering Herbicide Metabolism in Tobacco and Arabidopsis with CYP76B1, a Cytochrome P450 Enzyme from Jerusalem Artichoke. *Plant Physiology*, *130*, 179–189. https://doi.org/10.1104/pp.005801

Dimaano, N. G., & Iwakami, S. (2021). Cytochrome <scp>P450</scp> -mediated herbicide metabolism in plants: current understanding and prospects. *Pest Management Science*, *77*(1), 22–32. https://doi.org/10.1002/ps.6040

Dimaano, N. G., Yamaguchi, T., Fukunishi, K., Tominaga, T., & Iwakami, S. (2020). Functional characterization of cytochrome P450 CYP81A subfamily to disclose the pattern of cross-resistance in Echinochloa phyllopogon. *Plant Molecular Biology*, *102*(4–5), 403–416. https://doi.org/10.1007/s11103-019-00954-3

Dixon, D., Davis, B., & Edwards, R. (2002). Functional divergence in the glutathione transferase superfamily in plants: Identification of two classes with putative functions in redox homeostasis in Arabidopsis thaliana. *Journal of Biological Chemistry*, *277*(34), 30859–30869. https://doi.org/10.1074/jbc.M202919200

Dixon, D., Lapthorn, A., & Edwards, R. (2002). Plant glutathione transferases. *Genome Biology*, *3*(3), 1–10. https://doi.org/10.1186/gb-2002-3-3-reviews3004

Dixon, D. P., & Edwards, R. (2009). Selective binding of glutathione conjugates of fatty acid derivatives by plant glutathione transferases. *Journal of Biological Chemistry*, *284*(32), 21249–21256. https://doi.org/10.1074/jbc.M109.020107

Dixon, D. P., & Edwards, R. (2010a). Glutathione Transferases. *The Arabidopsis Book*, *8*, e0131. https://doi.org/10.1199/tab.0131

Dixon, D. P., & Edwards, R. (2010b). Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. *Journal of Biological Chemistry*, *285*(47), 36322–36329. https://doi.org/10.1074/jbc.M110.164806

Dixon, D. P., Hawkins, T., Hussey, P. J., & Edwards, R. (2009). Enzyme activities and subcellular localization of members of the arabidopsis glutathione transferase superfamily. *Journal of Experimental Botany*, *60*(4), 1207–1218. https://doi.org/10.1093/jxb/ern365

Dixon, D. P., Lapthorn, A., Madesis, P., Mudd, E. A., Day, A., & Edwards, R. (2008). Binding and glutathione conjugation of porphyrinogens by plant glutathione transferases. *Journal of Biological Chemistry*, *283*(29), 20268–20276. https://doi.org/10.1074/jbc.M802026200

Droog, F. (1997). Plant glutathione S-transferases, a tale of theta and tau. *Journal of Plant Growth Regulation*, *16*(2), 95–107. https://doi.org/10.1007/PL00006984

Du, H., Ran, F., Dong, H.-L., Wen, J., Li, J.-N., & Liang, Z. (2016). Genome-Wide Analysis, Classification, Evolution, and Expression Analysis of the Cytochrome P450 93 Family in Land Plants. *PLOS ONE*, *11*(10), e0165020. https://doi.org/10.1371/journal.pone.0165020

Duhoux, A., Carrère, S., Duhoux, A., & Délye, C. (2017). Transcriptional markers enable identification of rye-grass (*Lolium* sp.) plants with non-target-site-based resistance to herbicides inhibiting acetolactate-synthase. *Plant Science*, *257*, 22–36. https://doi.org/10.1016/J.PLANTSCI.2017.01.009

Duhoux, A., Carrère, S., Gouzy, J., Bonin, L., & Délye, C. (2015). RNA-Seq analysis of rye-grass transcriptomic response to an herbicide inhibiting acetolactate-synthase identifies transcripts linked to non-target-site-based resistance. *Plant Molecular Biology*, *87*(4–5), 473–487. https://doi.org/10.1007/s11103-015-0292-3

Duke, S. O. (2012). Why have no new herbicide modes of action appeared in recent years? *Pest Management Science*, *68*(4), 505–512. https://doi.org/10.1002/ps.2333

Edwards, R., & Dixon, D. P. (2005). Plant glutathione transferases. In *Methods in Enzymology* (Vol. 401, pp. 169–186). Academic Press Inc. https://doi.org/10.1016/S0076-6879(05)01011-6

Edwards, R., Dixon, D. P., & Walbot, V. (2000). Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends in Plant Science*, *5*(5), 193–198. https://doi.org/10.1016/S1360-1385(00)01601-0

Ehlting, J., Sauveplane, V., Olry, A., Ginglinger, J. F., Provart, N. J., & Werck-Reichhart, D. (2008). An extensive (co-)expression analysis tool for the cytochrome P450 superfamily in Arabidopsis thaliana. *BMC Plant Biology*, *8*(1), 47. https://doi.org/10.1186/1471-2229-8-47

Ezaki, B., Higashi, A., Nanba, N., & Nishiuchi, T. (2016). An S-adenosyl methionine synthetase (SAMS) gene from Andropogon virginicus L. confers aluminum stress tolerance and facilitates epigenetic gene regulation in Arabidopsis thaliana. *Frontiers in Plant Science*, *7*(November 2016), 1627. https://doi.org/10.3389/fpls.2016.01627

Fahey, R. C., & Sundquist, A. R. (1991). Evolution of glutathione metabolism. In *Advances in enzymology and related areas of molecular biology* (Vol. 64, pp. 1–53). Adv Enzymol Relat Areas Mol Biol. https://doi.org/10.1002/9780470123102.ch1

Fang, X., An, Y., Zheng, J., Shangguan, L., & Wang, L. (2020). Genome-wide identification and comparative analysis of GST gene family in apple (Malus domestica) and their expressions under ALA treatment. *3 Biotech*, *10*(7), 307. https://doi.org/10.1007/s13205-020-02299-x

Farkas, G. L., Király, Z., & Solymosy, F. (1960). Role of oxidative metabolism in the localization of plant viruses. *Virology*, *12*(3), 408–421. https://doi.org/10.1016/0042-6822(60)90163-X

Fernández-Cañón, J. M., Baetscher, M. W., Finegold, M., Burlingame, T., Gibson, K. M., & Grompe, M. (2002). Maleylacetoacetate Isomerase (MAAI/GSTZ)-Deficient Mice Reveal a Glutathione-Dependent Nonenzymatic Bypass in Tyrosine Catabolism. *Molecular and Cellular Biology*, *22*(13), 4943–4951. https://doi.org/10.1128/mcb.22.13.4943-4951.2002

Fonné-Pfister, R., Gaudin, J., Kreuz, K., Ramsteiner, K., & Ebert, E. (1990). Hydroxylation of primisulfuron by an inducible cytochrome P450-dependent monooxygenase system from maize. *Pesticide Biochemistry and Physiology*, *37Fonné-Pf*(2), 165–173. https://doi.org/10.1016/0048-3575(90)90122-I

Frear, D. S., & Swanson, H. R. (1970). Biosynthesis of S-(4-ethylamino-6-isopropylamino- 2-s-triazino) glutathione: Partial purification and properties of a glutathione S-transferase from corn. *Phytochemistry*, *9*(10), 2123–2132. https://doi.org/10.1016/S0031-9422(00)85377-7

Frear, D. S., Swanson, H. R., & Tanaka, F. S. (1969). N-demethylation of substituted 3-(phenyl)-1-methylureas: Isolation and characterization of a microsomal mixed function oxidase from cotton. *Phytochemistry*, *8*(11), 2157–2169. https://doi.org/10.1016/S0031-9422(00)88175-3

Frear, D. S., Swanson, H. R., & Thalacker, F. W. (1991). Induced Microsomal Oxidation of Diclofop, Triasulfuron, Chlorsulfuron, and Linuron in Wheat. *PESTICIDE BIOCHEMISTRY AND PHYSIOLOGY*, *41*, 274–287. https://ac.els-cdn.com/004835759190082W/1-s2.0-004835759190082W-main.pdf?_tid=ec12a440-4333-499d-bd2c-9c4659615961&acdnat=1528242800_17e8ff15d0c355f969c211f5495cee5a

Frova, C. (2006). Glutathione transferases in the genomics era: New insights and perspectives. In *Biomolecular Engineering* (Vol. 23, Issue 4, pp. 149–169). Elsevier. https://doi.org/10.1016/j.bioeng.2006.05.020

Gábrišová, D., Klubicová, K., Danchenko, M., Gömöry, D., Berezhna, V. V., Skultety, L., Miernyk, J. A., Rashydov, N., & Hajduch, M. (2016). Do cupins have a function beyond being seed storage proteins? *Frontiers in Plant Science*, *6*(JAN2016), 1215. https://doi.org/10.3389/fpls.2015.01215

Gaines, T. A., Lorentz, L., Figge, A., Herrmann, J., Maiwald, F., Ott, M.-C., Han, H., Busi, R., Yu, Q., Powles, S. B., & Beffa, R. (2014). RNA-Seq transcriptome analysis to identify genes involved in metabolism-based diclofop resistance in *Lolium rigidum*. *The Plant Journal*, *78*(5), 865–876. https://doi.org/10.1111/tpj.12514

Gao, D., Huibers, R. P., Loonen, A. E. H. M., Visser, R. G. F., Wolters, A. M. A., & Bai, Y. (2014). Down-regulation of acetolactate synthase compromises Ol-1- mediated resistance to powdery mildew in tomato. *BMC Plant Biology*, *14*(1), 1–11. https://doi.org/10.1186/1471-2229-14-32

Gao, J., Chen, B., Lin, H., Liu, Y., Wei, Y., Chen, F., & Li, W. (2020). Identification and characterization of the glutathione S-Transferase (GST) family in radish reveals a likely role in anthocyanin biosynthesis and heavy metal stress tolerance. *Gene*, *743*, 144484. https://doi.org/10.1016/j.gene.2020.144484

García-Garijo, A., Palma, F., Lluch, C., & Tejera, N. A. (2013). Physiological and biochemical responses of common vetch to

the imazamox accumulation. *Plant Physiology and Biochemistry*, *73*, 321–325. https://doi.org/10.1016/j.plaphy.2013.10.013

Gardin, J. A. C., Gouzy, J., Carrère, S., & Délye, C. (2015). ALOMYbase, a resource to investigate non-target-site-based resistance to herbicides inhibiting acetolactate-synthase (ALS) in the major grass weed Alopecurus myosuroides (black-grass). *BMC Genomics*, *16*(590). https://doi.org/10.1186/s12864-015-1804-x

Gaston, S., Zabalza, A., Gonzalez, E. M., Arrese-Igor, C., Aparicio-Tejo, P. M., & Royuela, M. (2002). Imazethapyr, an inhibitor of the branched-chain amino acid biosynthesis, induces aerobic fermentation in pea plants. *Physiologia Plantarum*, *114*(4), 524–532. https://doi.org/10.1034/j.1399-3054.2002.1140404.x

Ghangal, R., Singh Rajkumar, M., Garg, R., & Jain, · Mukesh. (2020). Genome-wide analysis of glutathione S-transferase gene family in chickpea suggests its role during seed development and abiotic stress. *Molecular Biology Reports*, *47*, 2749–2761. https://doi.org/10.1007/s11033-020-05377-8

Glombitza, S., Dubuis, P. H., Thulke, O., Welzl, G., Bovet, L., Götz, M., Affenzeller, M., Geist, B., Hehn, A., Asnaghi, C., Ernst, D., Seidlitz, H. K., Gundlach, H., Mayer, K. F., Martinoia, E., Werck-Reichhart, D., Mauch, F., & Schäffner, A. R. (2004). Crosstalk and differential response to abiotic and biotic stressors reflected at the transcriptional level of effector genes from secondary metabolism. *Plant Molecular Biology*, *54*(6), 817–835. https://doi.org/10.1007/s11103-004-0274-3

Gonneau, M., Mornet, R., & Laloue, M. (1998). A *Nicotiana plumbaginifolia* protein labeled with an azido cytokinin agonist is a glutathione S-transferase. *Physiologia Plantarum*, *103*(1), 114–124. https://doi.org/10.1034/j.1399-3054.1998.1030114.x

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Guindon, S., Ois Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Guo, F., Iwakami, S., Yamaguchi, T., Uchino, A., Sunohara, Y., & Matsumoto, H. (2019). Role of CYP81A cytochrome P450s in clomazone metabolism in Echinochloa phyllopogon. *Plant Science*, *283*, 321–328. https://doi.org/10.1016/j.plantsci.2019.02.010

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Han, H., Yu, Q., Beffa, R., González, S., Maiwald, F., Wang, J., & Powles, S. B. (2021). Cytochrome P450 CYP81A10v7 in *Lolium rigidum* confers metabolic resistance to herbicides across at least five modes of action. *The Plant Journal*, *105*(1), 79–92. https://doi.org/10.1111/tpj.15040

Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A., & Deisenhofer, J. (1995). Structure and function of cytochromes P450:a comparative analysis of three crystal structures. *Structure*, *3*(1), 41–62. https://doi.org/10.1016/S0969-2126(01)00134-4

Hatzios, K. (1997). Regulation of Enzymatic Systems Detoxifying Xenobiotics in Plants: A Brief Overview and Directions for Future Research Publishers. In KK Hatzios (Ed.), *Regulation of Enzymatic Systems Detoxifying Xenobiotics in Plants*. Kluwer Academic Publishers.

He, G., Guan, C. N., Chen, Q. X., Gou, X. J., Liu, W., Zeng, Q. Y., & Lan, T. (2016). Genome-wide analysis of the glutathione S-transferase gene family in Capsella rubella: Identification, expression, and biochemical functions. *Frontiers in Plant Science*, *7*(AUG2016). https://doi.org/10.3389/fpls.2016.01325

Heap, I. (2021). *The International Survey of Herbicide Resistant Weeds*. www.weedscience.org

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., & Vandesompele, J. (2008). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology*, *8*(2), R19. https://doi.org/10.1186/gb-2007-8-2-r19

Helvig, C., Tardif, F. J., Seyer, A., Powles, S. B., Mioskowski, C., Durst, F., & Sala, J.-P. (1996). Selective Inhibition of a Cytochrome P450 Enzyme in Wheat That Oxidizes Both the Natural Substrate Lauric Acid and the Synthetic Herbicide Diclofop. *Pesticide Biochemistry and Physiology*, *54*, 161–171. https://ac.els-cdn.com/S0048357596900206/1-s2.0-S0048357596900206-main.pdf?_tid=11785e82-5fa6-47e9-96e5-acadc1a4f6b8&acdnat=1528301067_5263073b00e076f556caa2227b2e925a

Hervé, C., De Franco, P. O., Groisillier, A., Tonon, T., & Boyen, C. (2008). New members of the glutathione transferase

family discovered in red and brown algae. *Biochemical Journal*, *412*(3), 535–544. https://doi.org/10.1042/BJ20071464

Hikosaka, M., Iwahashi, F., & Yamato, S. (2021). Metabolomic analysis of Schoenoplectus juncoides reveals common markers of acetolactate synthase inhibition among paddy weeds. *Pesticide Biochemistry and Physiology*, *174*, 104827. https://doi.org/10.1016/j.pestbp.2021.104827

Hildebrandt, T. M. (2018). Synthesis versus degradation: directions of amino acid metabolism during Arabidopsis abiotic stress response. *Plant Molecular Biology*, *98*(1–2), 121–135. https://doi.org/10.1007/s11103-018-0767-0

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., Moriyama, T., Ikeuchi, M., Watanabe, M., Wada, H., Kobayashi, K., Saito, M., Masuda, T., Sasaki-Sekimoto, Y., Mashiguchi, K., … Ohta, H. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4978

Hu, T., Qv, X., Xiao, G., & Huang, X. (2009). Enhanced tolerance to herbicide of rice plants by over-expression of a glutathione S-transferase. *Molecular Breeding*, *24*(4), 409–418. https://doi.org/10.1007/s11032-009-9302-y

Huala, E., Dickerman, A. W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L. A., Bhattacharyya, D., Bhaya, D., Sobral, B. W., Beavis, W., Meinke, D. W., Town, C. D., Somerville, C., & Rhee, S. Y. (2001). The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Research*, *29*(1), 102–105. https://doi.org/10.1093/nar/29.1.102

Ishizaki, K., Chiyoda, S., Yamato, K. T., & Kohchi, T. (2008). Agrobacterium-mediated transformation of the haploid liverwort Marchantia polymorpha L., an emerging model for plant biology. *Plant and Cell Physiology*, *49*(7), 1084–1091. https://doi.org/10.1093/pcp/pcn085

Ishizaki, K., Nishihama, R., Ueda, M., Inoue, K., Ishida, S., Nishimura, Y., Shikanai, T., & Kohchi, T. (2015). Development of Gateway Binary Vector Series with Four Different Selection Markers for the Liverwort Marchantia polymorpha. *PLOS ONE*, *10*(9), e0138876. https://doi.org/10.1371/journal.pone.0138876

Ishizaki, K., Nishihama, R., Yamato, K. T., & Kohchi, T. (2016). Molecular Genetic Tools and Techniques for *Marchantia polymorpha* Research. *Plant and Cell Physiology*, *57*(2), 262–270. https://doi.org/10.1093/pcp/pcv097

Iwakami, S., Endo, M., Saika, H., Okuno, J., Nakamura, N., Yokoyama, M., Watanabe, H., Toki, S., Uchino, A., & Inamura, T. (2014). Cytochrome P450 CYP81A12 and CYP81A21 are associated with resistance to two acetolactate synthase inhibitors in Echinochloa phyllopogon. *Plant Physiology*, *165*(2), 618–629. https://doi.org/10.1104/pp.113.232843

Iwakami, S., Kamidate, Y., Yamaguchi, T., Ishizaka, M., Endo, M., Suda, H., Nagai, K., Sunohara, Y., Toki, S., Uchino, A., Tominaga, T., & Matsumoto, H. (2019). CYP81A P450s are involved in concomitant cross-resistance to acetolactate synthase and acetyl-CoA carboxylase herbicides in Echinochloa phyllopogon. *New Phytologist*, *221*(4), 2112–2122. https://doi.org/10.1111/nph.15552

Iwata, S., Nakayama, N., Nakagawara, S., Ohta, Y., Tanaka, T., Wakabayashi, K., & Boger, P. (1992). Response of Liverwort Cells to Peroxidizing Herbicide. *Zeitschrift Fur Naturforschung - Section C Journal of Biosciences*, *47*(5–6), 394–399. https://doi.org/10.1515/znc-1992-0613

Jain, M., Ghanashyam, C., & Bhattacharjee, A. (2010). Comprehensive expression analysis suggests overlapping and specific roles of rice glutathione S-transferase genes during development and stress responses. *BMC Genomics*, *11*(1), 73. https://doi.org/10.1186/1471-2164-11-73

Kampranis, S. C., Damianova, R., Atallah, M., Toby, G., Kondi, G., Tsichlis, P. N., & Makris, A. M. (2000). A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. *Journal of Biological Chemistry*, *275*(38), 29207–29216. https://doi.org/10.1074/jbc.M002359200

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., & Martinoia, E. (2011). Plant ABC Transporters. *Arabidopsis Book*, *9*, e0153. https://doi.org/10.1199/tab.0153

Karavangeli, M., Labrou, N. E., Clonis, Y. D., & Tsaftaris, A. (2005). Development of transgenic tobacco plants overexpressing maize glutathione S-transferase I for chloroacetanilide herbicides phytoremediation. *Biomolecular Engineering*, *22*(4), 121–128. https://doi.org/10.1016/J.BIOENG.2005.03.001

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., Mccombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S., Childs, K. L., Davidson, R. M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., … Matsumoto, T. (2013). Improvement of the oryza sativa nipponbare reference genome using next

generation sequence and optical map data. *Rice*, *6*(1), 3–10. https://doi.org/10.1186/1939-8433-6-4

Kim, H. B., Schaller, H., Goh, C. H., Kwon, M., Choe, S., An, C. S., Durst, F., Feldmann, K. A., & Feyereisen, R. (2005). Arabidopsis cyp51 mutant shows postembryonic seedling lethality associated with lack of membrane integrity. *Plant Physiology*, *138*(4), 2033–2047. https://doi.org/10.1104/pp.105.061598

Kim, M. C., Lee, S. H., Kim, J. K., Chun, H. J., Choi, M. S., Chung, W. S., Moon, B. C., Kang, C. H., Park, C. Y., Yoo, J. H., Kang, Y. H., Koo, S. C., Koo, Y. D., Jung, J. C., Kim, S. T., Schulze-Lefert, P., Lee, S. Y., & Cho, M. J. (2002). Mlo, a modulator of plant defense and cell death, is a novel calmodulin-binding protein. Isolation and characterization of a rice Mlo homologue. *Journal of Biological Chemistry*, *277*(22), 19304–19314. https://doi.org/10.1074/jbc.M108478200

Kitamura, S., Shikazono, N., & Tanaka, A. (2004). *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *The Plant Journal*, *37*(1), 104–114. https://doi.org/10.1046/j.1365-313X.2003.01943.x

Klein, M., Burla, B., & Martinoia, E. (2006). The multidrug resistance-associated protein (MRP/ABCC) subfamily of ATP-binding cassette transporters in plants. *FEBS Letters*, *580*(4), 1112–1122. https://doi.org/10.1016/j.febslet.2005.11.056

Koonin, E. V., Tatusov, R. L., Altschul, S. F., Bryant, S. H., Mushegian, A. R., Bork, P., & Valencia, A. (1994). Eukaryotic translation elongation factor 1γ contains a glutathione transferase domain-Study of a diverse, ancient protein super family using motif search and structural modeling. *Protein Science*, *3*(11), 2045–2055. https://doi.org/10.1002/pro.5560031117

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Kreiner, J. M., Giacomini, D. A., Bemm, F., Waithaka, B., Regalado, J., Lanz, C., Hildebrandt, J., Sikkema, P. H., Tranel, P. J., Weigel, D., Stinchcombe, J. R., & Wright, S. I. (2019). Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(42), 21076–21084. https://doi.org/10.1073/pnas.1900870116

Kreuz, K., Gaudin, J., Stingelin, J., & Ebert, E. (1991). Metabolism of the Aryloxyphenoxypropanoate Herbicide, CGA 184927, in Wheat, Barley and Maize: Differential Effects of the Safener, CGA 185072. *Zeitschrift Fur Naturforschung - Section C Journal of Biosciences*, *46*(9–10), 901–905. https://doi.org/10.1515/znc-1991-9-1030

Kumar, S., Asif, M. H., Chakrabarty, D., Tripathi, R. D., Dubey, R. S., & Trivedi, P. K. (2013). Expression of a rice Lambda class of glutathione S-transferase, OsGSTL2, in Arabidopsis provides tolerance to heavy metal and other abiotic stresses. *Journal of Hazardous Materials*, *248–249*(1), 228–237. https://doi.org/10.1016/j.jhazmat.2013.01.004

Kumar, S., & Trivedi, P. K. (2018). Glutathione S-Transferases: Role in Combating Abiotic Stresses Including Arsenic Detoxification in Plants. *Frontiers in Plant Science*, *9*, 751. https://doi.org/10.3389/fpls.2018.00751

Lallement, P.-A., Meux, E., Gualberto, J. M., Dumarcay, S., Favier, F., Didierjean, C., Saul, F., Haouz, A., Morel-Rouhier, M., Gelhaye, E., Rouhier, N., & Hecker, A. (2015). Glutathionyl-hydroquinone reductases from poplar are plastidial proteins that deglutathionylate both reduced and oxidized glutathionylated quinones. *FEBS Letters*, *589*(1), 37–44. https://doi.org/10.1016/j.febslet.2014.11.021

Lallement, P. A., Brouwer, B., Keech, O., Hecker, A., & Rouhier, N. (2014). The still mysterious roles of cysteine-containing glutathione transferases in plants. In *Frontiers in Pharmacology: Vol. 5 AUG* (p. 192). Frontiers Research Foundation. https://doi.org/10.3389/fphar.2014.00192

Lan, T., Yang, Z.-L., Yang, X., Liu, Y.-J., Wang, X.-R., & Zeng, Q.-Y. (2010). Extensive Functional Diversification of the *Populus* Glutathione *S* -Transferase Supergene Family . *The Plant Cell*, *21*(12), 3749–3766. https://doi.org/10.1105/tpc.109.070219

Larossa, R. A., Van Dyk, T. K., & Smulski, D. R. (1987). Toxic Accumulation of a-Ketobutyrate Caused by Inhibition of the Branched-Chain Amino Acid Biosynthetic Enzyme Acetolactate Synthase in Salmonella typhimurium. In *JOURNAL OF BACTERIOLOGY* (Vol. 169, Issue 4). http://jb.asm.org/

Lei, Y., Lu, L., Liu, H. Y., Li, S., Xing, F., & Chen, L. L. (2014). CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. In *Molecular Plant* (Vol. 7, Issue 9, pp. 1494–1496). Cell Press. https://doi.org/10.1093/mp/ssu044

Letouzé, A., & Gasquez, J. (2003). Enhanced activity of several herbicide-degrading enzymes: a suggested mechanism responsible for multiple resistance in blackgrass (*Alopecurus myosuroides* Huds.). *Agronomie*, *23*(7), 601–608. https://doi.org/10.1051/agro:2003036

Levitt, G., Ploeg, H. L., Weigel, R. C., & Fitzgerald, D. J. (1981). 2-Chloro-N-[(4-methoxy-6-methyl-1,3,5-triazin-2-

yl)aminocarbonyl]benzenesulfonamide, a new herbicide. *Journal of Agricultural and Food Chemistry*, *29*(2), 416–418. https://doi.org/10.1021/jf00104a049

Li, D., Gao, Q., Xu, L., Pang, S., Liu, Z., Wang, C., & Tan, W. (2016). *Characterization of glutathione S-transferases in the detoxification of metolachlor in two maize cultivars of differing herbicide tolerance*. https://doi.org/10.1016/j.pestbp.2016.12.003

Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N., Barker, M. S., Bennett, T., Blázquez, M. A., Cheng, S., Cuming, A. C., de Vries, J., de Vries, S., Delaux, P.-M., Diop, I. S., Harrison, C. J., Hauser, D., Hernández-García, J., … Szövényi, P. (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants*, *6*(3), 259–272. https://doi.org/10.1038/s41477-020-0618-2

Li, J., Yu, H., Zhang, F., Lin, C., Gao, J., Fang, J., Ding, X., Shen, Z., & Xu, X. (2013). A Built-In Strategy to Mitigate Transgene Spreading from Genetically Modified Corn. *PLoS ONE*, *8*(12), e81645. https://doi.org/10.1371/journal.pone.0081645

Li, X., Schuler, M. A., & Berenbaum, M. R. (2007). Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics. *Annual Review of Entomology*, *52*(1), 231–253. https://doi.org/10.1146/annurev.ento.51.110104.151104

Li, Y., & Wei, K. (2020). Comparative functional genomics analysis of cytochrome P450 gene superfamily in wheat and maize. *BMC Plant Biology*, *20*(1), 1–22. https://doi.org/10.1186/s12870-020-2288-7

Lister, R., Carrie, C., Duncan, O., Ho, L. H. M., Howell, K. A., Murcha, M. W., & Whelan, J. (2007). Functional Definition of Outer Membrane Proteins Involved in Preprotein Import into Mitochondria. *The Plant Cell*, *19*, 3739–3759. https://doi.org/10.1105/tpc.107.050534

Liu, W., Bai, S., Zhao, N., Jia, S., Li, W., Zhang, L., & Wang, J. (2018). Non-target site-based resistance to tribenuron-methyl and essential involved genes in Myosoton aquaticum (L.). *BMC Plant Biology*, *18*(1), 225. https://doi.org/10.1186/s12870-018-1451-x

Liu, X.-Q., Yu, C.-Y., Dong, J.-G., Hu, S.-W., & Xu, A.-X. (2017). Acetolactate Synthase-Inhibiting Gametocide Amidosulfuron Causes Chloroplast Destruction, Tissue Autophagy, and Elevation of Ethylene Release in Rapeseed. *Frontiers in Plant Science*, *8*, 1625. https://doi.org/10.3389/fpls.2017.01625

Liu, Y.-J., Han, X.-M., Ren, L.-L., Yang, H.-L., & Zeng, Q.-Y. (2013). Functional Divergence of the Glutathione S-Transferase Supergene Family in Physcomitrella patens Reveals Complex Patterns of Large Gene Family Evolution in Land Plants. *Plant Physiology*, *161*(2), 773–786. https://doi.org/10.1104/PP.112.205815

Liu, Y., Jiang, H., Zhao, Y., Li, X., Dai, X., Zhuang, J., Zhu, M., Jiang, X., Wang, P., Gao, L., & Xia, T. (2019). Three Camellia sinensis glutathione S-transferases are involved in the storage of anthocyanins, flavonols, and proanthocyanidins. *Planta*, *250*(4), 1163–1175. https://doi.org/10.1007/s00425-019-03206-2

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Manabe, Y., Tinker, N., Colville, A., & Miki, B. (2007). CSR1, the Sole Target of Imidazolinone Herbicide in Arabidopsis thaliana. *Plant and Cell Physiology*, *48*(9), 1340–1358. https://doi.org/10.1093/pcp/pcm105

Marshall, R., Hanley, S. J., Hull, R., & Moss, S. R. (2013). The presence of two different target-site resistance mechanisms in individual plants of *Alopecurus myosuroides* Huds., identified using a quick molecular test for the characterisation of six ALS and seven ACCase SNPs. *Pest Management Science*, *69*(6), 727–737. https://doi.org/10.1002/ps.3429

Martin, J. L. (1995). Thioredoxin -a fold for all reasons. *Structure*, *3*(3), 245–250. https://doi.org/10.1016/S0969-2126(01)00154-X

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S. Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., … Kuroiwa, T. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature*, *428*(6983), 653–657. https://doi.org/10.1038/nature02398

Matzrafi, M., Seiwert, B., Reemtsma, T., Rubin, B., & Peleg, Z. (2016). Climate change increases the risk of herbicide-resistant weeds due to enhanced detoxification. *Planta*, *244*(6), 1217–1227. https://doi.org/10.1007/s00425-016-2577-4

McCourt, J. A., Pang, S. S., King-Scott, J., Guddat, L. W., & Duggleby, R. G. (2006). Herbicide-binding sites revealed in the structure of plant acetohydroxyacid synthase. *Proceedings of the National Academy of Sciences*, *103*(3), 569–573. https://doi.org/10.1073/pnas.0508701103

McGonigle, B., Keeler, S. J., Lau, S. M. C., Koeppe, M. K., & O'Keefe, D. P. (2000). A genomics approach to the comprehensive analysis of the glutathione S-transferase gene family in soybean and maize. *Plant Physiology*, *124*(3),

1105–1120. https://doi.org/10.1104/pp.124.3.1105

Mizutani, M., & Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annual Review of Plant Biology*, *61*, 291–315. https://doi.org/10.1146/annurev-arplant-042809-112305

Mizutani, M., & Sato, F. (2011). Unusual P450 reactions in plant secondary metabolism. In *Archives of Biochemistry and Biophysics* (Vol. 507, Issue 1, pp. 194–203). Academic Press Inc. https://doi.org/10.1016/j.abb.2010.09.026

Monticolo, F., Colantuono, C., & Chiusano, M. L. (2017). Shaping the evolutionary tree of green plants: evidence from the GST family. *Scientific Reports*, *7*, Article n. 14363. https://doi.org/10.1038/s41598-017-14316-w

Morikawa, T., Mizutani, M., Aoki, N., Watanabe, B., Saga, H., Saito, S., Oikawa, A., Suzuki, H., Sakurai, N., Shibata, D., Wadano, A., Sakata, K., & Ohta, D. (2006). Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato. *Plant Cell*, *18*(4), 1008–1022. https://doi.org/10.1105/tpc.105.036012

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H., & Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), E2274–E2283. https://doi.org/10.1073/pnas.1719588115

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. In *ChemBioChem* (Vol. 22, Issue 2, pp. 264–287). Wiley-VCH Verlag. https://doi.org/10.1002/cbic.202000354

Mougin, C., Polge, N., Scalla, R., & Cabanne, F. (1991). Interactions of various agrochemicals with cytochrome P450-dependent monooxygenases of wheat cells. *Pesticide Biochemistry and Physiology*, *40*(1), 1–11. https://doi.org/10.1016/0048-3575(91)90044-M

Munyampundu, J.-P., Xu, Y.-P., & Cai, X.-Z. (2016). Phi Class of Glutathione S-transferase Gene Superfamily Widely Exists in Nonplant Taxonomic Groups. *Evolutionary Bioinformatics Online*, *12*, 59–71. https://doi.org/10.4137/EBO.S35909

Muroi, A., Ishihara, A., Tanaka, C., Ishizuka, A., Takabayashi, J., Miyoshi, H., & Nishioka, T. (2009). Accumulation of hydroxycinnamic acid amides induced by pathogen infection and identification of agmatine coumaroyltransferase in Arabidopsis thaliana. *Planta*, *230*(3), 517–527. https://doi.org/10.1007/s00425-009-0960-0

Nebert, D. W., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper, B., Levin, W., Phillips, I. R., Sato, R., & Waterman, M. R. (1987). The P450 Gene Superfamily: Recommended Nomenclature. *DNA*, *6*(1), 1–11. https://doi.org/10.1089/dna.1987.6.1

Nelson, D., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., & Nebert, D. W. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, *6*(1), 1–42. http://www.ncbi.nlm.nih.gov/pubmed/8845856

Nelson, D. R. (1998). Metazoan cytochrome P450 evolution. *Comparative Biochemistry and Physiology - C Pharmacology Toxicology and Endocrinology*, *121*(1–3), 15–22. https://doi.org/10.1016/S0742-8413(98)10027-0

Nelson, D. R. (2006). Plant cytochrome P450s from moss to poplar. In *Phytochemistry Reviews* (Vol. 5, Issues 2–3, pp. 193–204). Springer. https://doi.org/10.1007/s11101-006-9015-3

Nelson, D. R. (2009). The cytochrome P450 homepage. *Human Genomics*, *4*(1), 59–65. https://doi.org/10.1186/1479-7364-4-1-59

Nelson, D. R. (2018). Cytochrome P450 diversity in the tree of life. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1866*(1), 141–154. https://doi.org/10.1016/j.bbapap.2017.05.003

Nelson, D. R., Ming, R., Alam, M., & Schuler, M. A. (2008). Comparison of Cytochrome P450 Genes from Six Plant Genomes. *Tropical Plant Biology*, *1*(3–4), 216–235. https://doi.org/10.1007/s12042-008-9022-1

Nelson, D., Schuler, M., Paquette, S. M., Werck-Reichhart, D., & Bak, S. (2004). Comparative Genomics of Rice and Arabidopsis. Analysis of 727 Cytochrome P450 Genes and Pseudogenes from a Monocot and a Dicot. *Plant Physiology*, *1355*, 756–772. https://doi.org/10.1104/pp.104.039826

Nelson, D., & Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *The Plant Journal*, *66*(1), 194–211. https://doi.org/10.1111/j.1365-313X.2011.04529.x

Neuefeind, T., Huber, R., Dasenbrock, H., Prade, L., & Bieseler, B. (1997). Crystal structrure of herbicide-detoxifying maize glutathione S-transferase-I in complex with lactoylglutathione: Evidence for an induced-fit mechanism. In *Journal of Molecular Biology* (Vol. 274, Issue 4, pp. 446–453). Academic Press. https://doi.org/10.1006/jmbi.1997.1402

Neuefeind, T., Huber, R., Reinemer, P., Knäblein, J., Prade, L., Mann, K., & Bieseler, B. (1997). Cloning, sequencing,

crystallization and x-ray structure of glutathione S-transferase-III from Zea mays var. mutin: A leading enzyme in detoxification of maize herbicides. *Journal of Molecular Biology*, *274*(4), 577–587. https://doi.org/10.1006/jmbi.1997.1401

Neve, P., & Powles, S. (2005). Recurrent selection with reduced herbicide rates results in the rapid evolution of herbicide resistance in Lolium rigidum. *Theoretical and Applied Genetics*, *110*(6), 1154–1166. https://doi.org/10.1007/s00122-005-1947-2

Nikolenko, S. I., Korobeynikov, A. I., & Alekseyev, M. A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, *14*(Suppl 1), S7. https://doi.org/10.1186/1471-2164-14-S1-S7

Oerke, E. C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, *144*(01), 31. https://doi.org/10.1017/S0021859605005708

Ohkawa, H., Tsujii, H., & Ohkawa, Y. (1999). The use of cytochrome P450 genes to introduce herbicide tolerance in crops: A review. In *Pesticide Science* (Vol. 55, Issue 9, pp. 867–874). https://doi.org/10.1002/(SICI)1096-9063(199909)55:9<867::AID-PS31>3.0.CO;2-S

Orcaray, L., Igal, M., Marino, D., Zabalza, A., & Royuela, M. (2010). The possible role of quinate in the mode of action of glyphosate and acetolactate synthase inhibitors. *Pest Management Science*, *66*(3), 262–269. https://doi.org/10.1002/ps.1868

Pan, G., Zhang, X., Liu, K., Zhang, J., Wu, X., Zhu, J., & Tu, J. (2006). Map-based cloning of a novel rice cytochrome P450 gene CYP81A6 that confers resistance to two different classes of herbicides. *Plant Molecular Biology*, *61*(6), 933–943. https://doi.org/10.1007/s11103-006-0058-z

Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V., & Jugulam, M. (2020). Role of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants*, *9*(5), 454. https://doi.org/10.3390/antiox9050454

Paquette, S. M., Bak, S., & Feyereisen, R. (2000). Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA and Cell Biology*, *19*(5), 307–317. http://online.liebertpub.com/doi/pdf/10.1089/10445490050021221

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. https://doi.org/10.1038/nmeth.4197

Persans, M. W., Wang, J., & Schuler, M. A. (2001). Characterization of maize cytochrome P450 monooxygenases induced in response to safeners and bacterial pathogens. *Plant Physiology*, *109*(4), 1483–1490. https://doi.org/10.1104/pp.125.2.1126

Pessina, S., Lenzi, L., Perazzolli, M., Campa, M., Dalla Costa, L., Urso, S., Valè, G., Salamini, F., Velasco, R., & Malnoy, M. (2016). Knockdown of MLO genes reduces susceptibility to powdery mildew in grapevine. *Horticulture Research*, *3*(1), 1–9. https://doi.org/10.1038/hortres.2016.16

Pfister, K., & Arntzen, C. J. (1979). The Mode of Action of Photosystem II-Specific Inhibitors in Herbicide-Resistant Weed Biotypes. *Zeitschrift Für Naturforschung*, *34*, 996–1009. https://www.degruyter.com/downloadpdf/j/znc.1979.34.issue-11/znc-1979-1123/znc-1979-1123.pdf

Piślewska-Bednarek, M., Nakano, R. T., Hiruma, K., Pastorczyk, M., Sanchez-Vallet, A., Singkaravanit-Ogawa, S., Ciesiołka, D., Takano, Y., Molina, A., Schulze-Lefert, P., & Bednarek, P. (2018). Glutathione transferase U13 functions in pathogen-triggered glucosinolate metabolism. *Plant Physiology*, *176*(1), 538–551. https://doi.org/10.1104/pp.17.01455

Porebski, S., Bailey, L. G., & Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, *15*(1), 8–15. https://doi.org/10.1007/BF02772108

Preston, C. (2004). Herbicide resistance in weeds endowed by enhanced detoxification: complications for management. *Weed Science*, *52*, 448–453.

Preston, Christopher, Tardif, F. J., Christopher, J. T., & Powles, S. B. (1996). Multiple resistance to dissimilar herbicide chemistries in a biotype of *Lolium rigidum* due to enhanced activity of several herbicide degrading enzymes. *Pesticide Biochemistry and Physiology*, *54*(2), 123–134. https://doi.org/10.1006/PEST.1996.0016

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L. L., Eddy, S. R., Bateman, A., & Finn, R. D. (2012). The Pfam protein families database. *Nucleic Acids Research*, *40*(D1), D290–D301. https://doi.org/10.1093/nar/gkr1065

Ray, T. B. (1982). The mode of action of chlorsulfuron: A new herbicide for cereals. *Pesticide Biochemistry and Physiology*, *17*(1), 10–17. https://doi.org/10.1016/0048-3575(82)90120-1

Ray, T. B. (1984). Site of Action of Chlorsulfuron. *Plant Physiology*, *75*(3), 827–831. https://doi.org/10.1104/pp.75.3.827

Reade, J. P. H., Milner, L. J., & Cobb, A. H. (2004). A role for glutathione-S-transferases in resistance to herbicides in grasses. *Weed Science*, *52*(3), 468–474. https://doi.org/10.1614/P2002-168D

Reinemer, P., Prade, L., Hof, P., Neuefeind, T., Huber, R., Zettl, R., Palme, K., Schell, J., Koelln, I., Bartunik, H. D., & Bieseler, B. (1996). Three-dimensional structure of glutathione S-transferase from Arabidopsis thaliana at 2.2 Å resolution: Structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. *Journal of Molecular Biology*, *255*(2), 289–309. https://doi.org/10.1006/JMBI.1996.0024

Rhodes, D., Hogan, A. L., Deal, L., Jamieson, G. C., & Haworth, P. (1987). Amino Acid Metabolism of Lemna minor L. *Plant Physiology*, *84*(3), 775–780. https://doi.org/10.1104/pp.84.3.775

Riechers, D. E., Vaughn, K. C., & Molin, W. T. (2005). The role of plant glutathione S-transferases in herbicide metabolism. *ACS Symposium Series*, *899*, 216–232. https://doi.org/10.1021/bk-2005-0899.ch019

Rost, T. L., & Reynolds, T. (1985). Reversal of Chlorsulfuron-Induced Inhibition of Mitotic Entry by Isoleucine and Valine. *Plant Physiology*, *77*(2), 481–482. https://doi.org/10.1104/pp.77.2.481

Royuela, M., Gonzalez, A., Gonzalez, E. M., Arrese-Igor, C., Aparicio-Tejo, P. M., & Gonzalez-Murua, C. (2000). Physiological consequences of continuous, sublethal imazethapyr supply to pea plants. *Journal of Plant Physiology*, *157*(3), 345–354. https://doi.org/10.1016/S0176-1617(00)80057-7

Rupasinghe, S., & Schuler, M. A. (2006). Homology modeling of plant cytochrome P450s. In *Phytochemistry Reviews* (Vol. 5, Issues 2–3, pp. 473–505). Springer. https://doi.org/10.1007/s11101-006-9028-y

Saika, H., Horita, J., Taguchi-Shiobara, F., Nonaka, S., Nishizawa-Yokoi, A., Iwakami, S., Hori, K., Matsumoto, T., Tanaka, T., Itoh, T., Yano, M., Kaku, K., Shimizu, T., & Toki, S. (2014). A novel rice cytochrome P450 gene, CYP72A31, confers tolerance to acetolactate synthase-inhibiting herbicides in rice and Arabidopsis. *Plant Physiology*, *166*(3), 1232–1240. https://doi.org/10.1104/pp.113.231266

Saint-Marcoux, D., Proust, H., Dolan, L., & Langdale, J. A. (2015). Identification of Reference Genes for Real-Time Quantitative PCR Experiments in the Liverwort Marchantia polymorpha. *PLOS ONE*, *10*(3), e0118678. https://doi.org/10.1371/journal.pone.0118678

Sappl, P. G., Carroll, A. J., Clifton, R., Lister, R., Whelan, J., Harvey Millar, A., & Singh, K. B. (2009). The Arabidopsis glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress. *Plant Journal*, *58*(1), 53–68. https://doi.org/10.1111/j.1365-313X.2008.03761.x

Scarabel, L., Pernin, F., & Délye, C. (2015). Occurrence, genetic control and evolution of non-target-site based resistance to herbicides inhibiting acetolactate synthase (ALS) in the dicot weed Papaver rhoeas. *Plant Science*, *238*, 158–169. https://doi.org/10.1016/J.PLANTSCI.2015.06.008

Schuler, M. A., Hui, A., Ae, D., Bilgin, M., & Ali, S. (2006). Arabidopsis cytochrome P450s through the looking glass: a window on plant biochemistry. *Phytochemistry Reviews*, *5*(2–3), 205–237. https://doi.org/10.1007/s11101-006-9035-z

Shah, D. M., Hironaka, C. M., Wiegand, R. C., Harding, E. I., Krivi, G. G., & Tiemeier, D. C. (1986). Structural analysis of a maize gene coding for glutathione-S-transferase involved in herbicide detoxification. *Plant Molecular Biology*, *6*, 203–211. https://link.springer.com/content/pdf/10.1007%2FBF00015226.pdf

Shaner, D. L., & Reider, M. L. (1986). Physiological responses of corn (Zea mays) to AC 243,997 in combination with valine, leucine, and isoleucine. *Pesticide Biochemistry and Physiology*, *25*(2), 248–257. https://doi.org/10.1016/0048-3575(86)90051-9

Shaner, D. L., & Singh, B. K. (1993). Phytotoxicity of Acetohydroxyacid Synthase Inhibitors Is Not Due to Accumulation of 2-Ketobutyrate and/or 2-Aminobutyrate. *Plant Physiology*, *103*(4), 1221–1226. https://doi.org/10.1104/pp.103.4.1221

Sheehan, D., Meade, G., Foley, V. M., & Dowd, C. A. (2001). Structure, function and evolution of glutathione transferases : implications for classification of non-mammalian members of an ancient enzyme superfamily. In *Biochem. J* (Vol. 360).

Shimabukuro, R. H., Walsh, W. C., & Jacobson, A. (1987). Aryl-O-glucoside of diclofop: a detoxication product in wheat shoots and wild oat cell suspension culture. *Journal of Agricultural and Food Chemistry*, *35*(3), 393–397. https://doi.org/10.1021/jf00075a027

Shimamura, M. (2016). Marchantia polymorpha: Taxonomy, phylogeny and morphology of a model system. *Plant and Cell Physiology*, *57*(2), 230–256. https://doi.org/10.1093/pcp/pcv192

Siminszky, B, Corbin, F. T., Ward, E. R., Fleischmann, T. J., & Dewey, R. E. (1999). Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(4), 1750–1755. https://doi.org/10.1073/PNAS.96.4.1750

Siminszky, Balazs. (2006). Plant cytochrome P450-mediated herbicide metabolism. In *Phytochemistry Reviews* (Vol. 5, Issues 2–3, pp. 445–458). https://doi.org/10.1007/s11101-006-9011-7

Soranzo, N., Sari Gorla, M., Mizzi, L., De Toma, G., & Frova, C. (2004). Organisation and structural evolution of the rice glutathione S-transferase gene family. In *Molecular Genetics and Genomics* (Vol. 271, Issue 5, pp. 511–521). https://doi.org/10.1007/s00438-004-1006-8

Sugano, S. S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T., Hara-Nishimura, I., & Kohchi, T. (2014a). CRISPR/Cas9-mediated targeted mutagenesis in the liverwort Marchantia polymorpha L. *Plant and Cell Physiology*, *55*(3), 475–481. https://doi.org/10.1093/pcp/pcu014

Sugano, S. S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T., Hara-Nishimura, I., & Kohchi, T. (2014b). CRISPR/Cas9-Mediated Targeted Mutagenesis in the Liverwort *Marchantia polymorpha* L. *Plant and Cell Physiology*, *55*(3), 475–481. https://doi.org/10.1093/pcp/pcu014

Sweetser, P. B., Schow, G. S., & Hutchison, J. M. (1982). Metabolism of chlorsulfuron by plants: Biological basis for selectivity of a new herbicide for cereals. *Pesticide Biochemistry and Physiology*, *17*(1), 18–23. https://doi.org/10.1016/0048-3575(82)90121-3

Sylvestre-Gonon, E., Law, S. R., Schwartz, M., Robe, K., Keech, O., Didierjean, C., Dubos, C., Rouhier, N., & Hecker, A. (2019). Functional, structural and biochemical features of plant serinyl-glutathione transferases. In *Frontiers in Plant Science* (Vol. 10, p. 608). Frontiers Media S.A. https://doi.org/10.3389/fpls.2019.00608

Tanikawa, N., Ohmiya, Y., Kojima, M., Ohkubo, H., Hashimoto, K., Kangawa, K., Ito, S., & Watanabe, K. (2002). Identification and characterization of a novel type of membrane-associated prostaglandin E synthase. *Biochemical and Biophysical Research Communications*, *291*(4), 884–889. https://doi.org/10.1006/bbrc.2002.6531

Thom, R., Cummins, I., Dixon, D. P., Edwards, R., Cole, D. J., & Lapthorn, A. J. (2002). Structure of a Tau Class Glutathione S-Transferase from Wheat Active in Herbicide Detoxification. *Biochemistry*, *41*, 7008–7020. https://doi.org/10.1021/bi015964x

Thyssen, G. N., Naoumkina, M., McCarty, J. C., Jenkins, J. N., Florane, C., Li, P., & Fang, D. D. (2018). The P450 gene CYP749A16 is required for tolerance to the sulfonylurea herbicide trifloxysulfuron sodium in cotton (*Gossypium hirsutum* L.). *BMC Plant Biology*, *18*(1), 186. https://doi.org/10.1186/s12870-018-1414-2

Tranel, P. J., & Wright, T. R. (2002). Resistance of weeds to ALS-inhibiting herbicides: what have we learned? *Weed Science*, *50*, 700–712. http://www.open-source-biology.com/synthese_ALS_fichiers/TRANEL WRIGHT ALS-inhiting herbicides 2002.pdf

Trenkamp, S., Eckes, P., Busch, M., & Fernie, A. R. (2009). Temporally resolved GC-MS-based metabolic profiling of herbicide treated plants treated reveals that changes in polar primary metabolites alone can distinguish herbicides of differing mode of action. *Metabolomics*, *5*(3), 277–291. https://doi.org/10.1007/s11306-008-0149-8

Tsuboyama-Tanaka, S., & Kodama, Y. (2015). AgarTrap-mediated genetic transformation using intact gemmae/gemmalings of the liverwort Marchantia polymorpha L. *Journal of Plant Research*, *128*(2), 337–344. https://doi.org/10.1007/s10265-014-0695-2

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*(7), research0034.1. https://doi.org/10.1186/gb-2002-3-7-research0034

Wagner, U., Edwards, R., Dixon, D. P., & Mauch, F. (2002). Probing the Diversity of the Arabidopsis glutathione S-Transferase Gene Family. In *Plant Molecular Biology* (Vol. 49). http://www.isrec.isb-sib.ch/software/

Wakabayashi, T., Hamana, M., Mori, A., Akiyama, R., Ueno, K., Osakabe, K., Osakabe, Y., Suzuki, H., Takikawa, H., Mizutani, M., & Sugimoto, Y. (2019). Direct conversion of carlactonoic acid to orobanchol by cytochrome P450 CYP722C in strigolactone biosynthesis. *Science Advances*, *5*(12), 9067–9085. https://doi.org/10.1126/sciadv.aax9067

Walsh, M. J., Powles, S. B., Beard, B. R., Parkin, B. T., & Porter, S. A. (2004). Multiple-herbicide resistance across four modes of action in wild radish (*Raphanus raphanistrum*). *Weed Science*, *52*(01), 8–13. https://doi.org/10.1614/WS-03-016R

Wang, C. K., & Peng, C. H. (1996). The mutagenicities of alkaloids and N-nitrosoguvacoline from betel quid. *Mutation Research - Environmental Mutagenesis and Related Subjects*, *360*(3), 165–171. https://doi.org/10.1016/S0165-1161(96)90013-8

Wang, J., Zhang, Z., Wu, J., Han, X., Wang-Pruski, G., & Zhang, Z. (2020). Genome-wide identification, characterization, and expression analysis related to autotoxicity of the GST gene family in Cucumis melo L. *Plant Physiology and Biochemistry*, *155*, 59–69. https://doi.org/10.1016/j.plaphy.2020.06.046

Warner, J. R., & Copley, S. D. (2007). Mechanism of the severe inhibition of tetrachlorohydroquinone dehalogenase by its aromatic substrates. *Biochemistry*, *46*(14), 4438–4447. https://doi.org/10.1021/bi0620104

Wei, K., & Chen, H. (2018). Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice. *BMC Genomics*, *19*(1), 35. https://doi.org/10.1186/s12864-017-4425-8

Werck-Reichhart, D., & Feyereisen, R. (2000). Cytochromes P450: a success story. In *Genome biology* (Vol. 1, Issue 6, p. reviews3003.1). BioMed Central. https://doi.org/10.1186/gb-2000-1-6-reviews3003

Werck-Reichhart, Danièle, Hehn, A., & Didierjean, L. (2000). Cytochromes P450 for engineering herbicide tolerance. *Trends in Plant Science Reviews*, *5*(3). https://ac.els-cdn.com/S1360138500015673/1-s2.0-S1360138500015673-main.pdf?_tid=0f15b41b-c4ea-4100-918c-be7d790c3ebd&acdnat=1528281314_4bee649e7d340085b2ce02862e0a667b

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., Ruhfel, B. R., Wafula, E., Der, J. P., Graham, S. W., Mathews, S., Melkonian, M., Soltis, D. E., Soltis, P. S., Miles, N. W., … Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), E4859–E4868. https://doi.org/10.1073/pnas.1323926111

Wiersma, P. A., Hachey, J. E., Crosby, W. L., & Moloney, M. M. (1990). Specific truncations of an acetolactate synthase gene from Brassica napus efficiently complement ilvB/ilvG mutants of Salmonella typhimurium. *MGG Molecular & General Genetics*, *224*(1), 155–159. https://doi.org/10.1007/BF00259463

Wittenbach, V. A., Koeppe, M. K., Lichtner, F. T., Zimmerman, W. T., & Reiser, R. W. (1994). Basis of Selectivity of Triflusulfuron Methyl in Sugar Beets (Beta vulgaris). *Pesticide Biochemistry and Physiology*, *49*(1), 72–81. https://doi.org/10.1006/PEST.1994.1035

Xiang, W., Wang, X., & Ren, T. (2005). Expression of a wheat cytochrome P450 monooxygenase cDNA in yeast catalyzes the metabolism of sulfonylurea herbicides. *Pesticide Biochemistry and Physiology*, *85*(1), 1–6. https://doi.org/10.1016/j.pestbp.2005.09.001

Xu, J., Wang, X., & Guo, W. (2015). The cytochrome P450 superfamily: Key players in plant development and defense. *Journal of Integrative Agriculture*, *14*(9), 1673–1686. https://doi.org/10.1016/S2095-3119(14)60980-1

Yamada, T., Komoto, J., Watanabe, K., Ohmiya, Y., & Takusagawa, F. (2005). Crystal structure and possible catalytic mechanism of microsomal prostaglandin E synthase type 2 (mPGES-2). *Journal of Molecular Biology*, *348*(5), 1163–1176. https://doi.org/10.1016/j.jmb.2005.03.035

Yang, Q., Deng, W., Li, X., Yu, Q., Bai, L., & Zheng, M. (2016). Target-site and non-target-site based resistance to the herbicide tribenuron-methyl in flixweed (Descurainia sophia L.). *BMC Genomics*, *17*(1), 17:551. https://doi.org/10.1186/s12864-016-2915-8

Yu, J. G., Lee, G. H., & Park, Y. D. (2012). Physiological role of endogenous S-adenosyl-L-methionine synthetase in Chinese cabbage. *Horticulture Environment and Biotechnology*, *53*(3), 247–255. https://doi.org/10.1007/s13580-012-0021-7

Yu, Q., Abdallah, I., Han, H., Owen, M., & Powles, S. (2009). Distinct non-target site mechanisms endow resistance to glyphosate, ACCase and ALS-inhibiting herbicides in multiple herbicide-resistant *Lolium rigidum*. *Planta*, *230*, 713–723. https://doi.org/10.1007/s00425-009-0981-8

Yu, Q., & Powles, S. (2014a). Metabolism-Based Herbicide Resistance and Cross-Resistance in Crop Weeds: A Threat to Herbicide Sustainability and Global Crop Production. *Plant Physiology*, *166*, 1106–1118. https://doi.org/10.1104/pp.114.242750

Yu, Q., & Powles, S. B. (2014b). Resistance to AHAS inhibitor herbicides: current understanding. *Pest Management Science*, *70*(9), 1340–1350. https://doi.org/10.1002/ps.3710

Yu, Q., & Powles, S. B. (2014c). Metabolism-based herbicide resistance and cross-resistance in crop weeds: a threat to herbicide sustainability and global crop production. *Plant Physiology*, *166*(3), 1106–1118. https://doi.org/10.1104/pp.114.242750

Yuan, H., Zeng, X., Shi, J., Xu, Q., Wang, Y., Jabu, D., Sang, Z., & Nyima, T. (2018). Time-course comparative metabolite profiling under osmotic stress in tolerant and sensitive tibetan hulless barley. *BioMed Research International*, *2018*. https://doi.org/10.1155/2018/9415409

Yuan, J. S., Tranel, P. J., & Stewart, C. N. (2007). Non-target-site herbicide resistance: a family business. *Trends in Plant*

Science, *12*(1), 6–13. https://doi.org/10.1016/J.TPLANTS.2006.11.001

Yuan, S., Chan, H. C. S., Filipek, S., & Vogel, H. (2016). PyMOL and Inkscape Bridge the Data and the Data Visualization. In *Structure* (Vol. 24, Issue 12, pp. 2041–2042). Cell Press. https://doi.org/10.1016/j.str.2016.11.012

Zabalza, A., González, E. M., Arrese-Igor, C., & Royuela, M. (2005). Fermentative metabolism is induced by inhibiting different enzymes of the branched-chain amino acid biosynthesis pathway in pea plants. *Journal of Agricultural and Food Chemistry*, *53*(19), 7486–7493. https://doi.org/10.1021/jf050654x

Zabalza, A., Orcaray, L., Igal, M., Schauer, N., Fernie, A. R., Geigenberger, P., van Dongen, J. T., & Royuela, M. (2011). Unraveling the role of fermentation in the mode of action of acetolactate synthase inhibitors by metabolic profiling. *Journal of Plant Physiology*, *168*(13), 1568–1575. https://doi.org/10.1016/j.jplph.2011.02.015

Zettl, R., Schell, J., & Palme, K. (1994). Photoaffinity labeling of Arabidopsis thaliana plasma membrane vesicles by 5-azido-[7-3H]indole-3-acetic acid: Identification of a glutathione S- transferase. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(2), 689–693. https://doi.org/10.1073/pnas.91.2.689

Zhang, L., Lu, Q., Chen, H., Pan, G., Xiao, S., Dai, Y., Li, Q., Zhang, J., Wu, X., Wu, J., Tu, J., & Liu, K. (2006). Identification of a cytochrome P450 hydroxylase, CYP81A6, as the candidate for the bentazon and sulfonylurea herbicide resistance gene, Bel, in rice. *Molecular Breeding*, *19*(1), 59–68. https://doi.org/10.1007/s11032-006-9044-z

Zhao, J., Meng, Z., Ma, X., Zhao, S., An, Y., & Xiao, Z. (2021). Characterization and Regulation of the Acetolactate Synthase Genes Involved in Acetoin Biosynthesis in Acetobacter pasteurianus. *Foods*, *10*(5), 1013. https://doi.org/10.3390/foods10051013

Zhao, N., Li, W., Bai, S., Guo, W., Yuan, G., Wang, F., Liu, W., & Wang, J. (2017). Transcriptome profiling to identify genes involved in mesosulfuron-methyl resistance in *Alopecurus aequalis*. *Frontiers in Plant Science*, *8*, 1391. https://doi.org/10.3389/fpls.2017.01391

Zhou, Q., Liu, W., Zhang, Y., & Liu, K. K. (2007). *Action mechanisms of acetolactate synthase-inhibiting herbicides*. https://doi.org/10.1016/j.pestbp.2007.04.004

Zulet, A., Zabalza, A., & Royuela, M. (2013). Phytotoxic and Metabolic Effects of Exogenous Quinate on Pisum sativum L. *Journal of Plant Growth Regulation*, *32*(4), 779–788. https://doi.org/10.1007/s00344-013-9345-5