Genes encoding cytochrome P450 monooxygenases and glutathione S-transferases associated with herbicide resistance evolved before the origin of land plants

Alexandra Casey[1,2], Liam Dolan[1,*]

[1] Gregor Mendel Institute, Vienna, Austria

[2] Department of Plant Sciences, University of Oxford, Oxford, Oxfordshire, UK

*Lead Contact

Correspondence: liam.dolan@gmi.oeaw.ac.at

Abstract

Cytochrome P450 (CYP) monooxygenases and glutathione S-transferases (GST) are enzymes that catalyse chemical modifications of a range of organic compounds. Herbicide tolerance is associated with higher levels of CYP and GST gene expression in some herbicide-resistant weed populations compared to sensitive populations of the same species. By comparing the protein sequences of 9 representative species of the Archaeplastida – the lineage which includes red algae, glaucophyte algae, chlorophyte algae, and streptophytes – and generating phylogenetic trees, we identified the CYP and GST proteins that existed in the common ancestor of the Archaeplastida. All CYP clans and all but one land plant GST classes present in land plants evolved before the divergence of streptophyte algae and land plants from their last common ancestor. We also demonstrate that there are more genes encoding CYP and GST proteins in land plants than in algae. The larger numbers of genes among land plants largely results from gene duplications in CYP clans 71, 72, and 85 and in the GST Phi and Tau classes. Enzymes that either chemically alter herbicides or confer herbicide resistance belong to CYP clans 71 and 72 and the GST Phi and Tau classes. These results demonstrate that the clan and class diversity in extant plant CYP and GST proteins evolved in the Proterozoic before the divergence of land plants and streptophyte algae from a last common ancestor. Then, early in embryophyte evolution during the Palaeozoic, gene duplication in four of the twelve CYP clans, and in two of the fourteen GST classes, led to the large numbers of CYP and GST proteins found in extant land plants. It is among the genes of CYP clans 71 and 72 and GST classes Phi and Tau that alleles conferring herbicide resistance evolved in the last fifty years.

Introduction

Herbicide resistance evolves in weed populations and poses a challenge in all agricultural landscapes where chemical herbicides are used for weed control. This resistance can result from two

types of mutations. Mutations in the gene targeted by the herbicide which inhibit the interaction between the two confer target site resistance (TSR). Non target site resistance (NTSR) results either from mutations that reduce the amount of herbicide chemical reaching the target or that alleviate the herbicide-induced damage [1]. Reported mechanisms of NTSR involve the reduction of herbicide uptake or translocation, chemical modification of the herbicide, or sequestration of the herbicide to a location where it cannot access the target [2–4]. Genetic changes in genes encoding enzymes that can chemically modify the herbicide, including changes such as overexpression, the expression of hyperactive forms of the enzymes, or enzymes with altered substrate specificity, can inactivate the herbicide, conferring resistance [5–7]. Mutations that result in NTSR are selected for in agricultural landscapes where chemical herbicides are used and can reach high allele frequencies in the presence of ongoing herbicide selection. While the genetic basis of NTSR is often complex and mechanistically poorly understood, the overexpression of genes encoding cytochrome P450 monooxygenases and glutathione s-transferases has been shown to confer resistance in weed populations [8–10].

Glutathione-s-transferases (GSTs) are an ancient superfamily of enzymes found in eukaryotes and prokaryotes. GSTs catalyse the conjugation of glutathione (GSH) to both endogenous and exogenous electrophilic, hydrophobic substrates to form more polar, hydrophilic compounds. GSTs also catalyse GSH-dependent peroxidase, isomerase, and deglutathionylation reactions. In plants, GSTs are active in diverse processes including abiotic and biotic detoxification pathways [11,12], ascorbic acid metabolism [13], hormone signalling such as auxin and cytokinin homeostasis [14–16], metabolism of anthocyanins and flavonoids [17,18], tyrosine catabolism [19], and in preventing apoptosis [20].

GSTs function as either monomers or dimers. Each monomer is characterised by a conserved N-terminal domain containing the active site and several GSH binding site residues (G-sites), and a less conserved C-terminal domain comprising alpha helices with class-specific substrate binding sites (H-sites) [21]. Plant GSTs are classified into groups as cytosolic, mitochondrial,or microsomal and each group is further subdivided into classes. In plants there are 12 cytosolic GST classes. These

include 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 reductase (DHAR). In contrast, there is only a single microsomal GST class, microsomal prostaglandin E-synthase type 2 (mPGES2), and a single mitochondrial GST class, Metaxin (GSTM).

Cytochrome p450 monooxygenases (CYPs) are a superfamily of membrane-bound enzymes present in plants, fungi, bacteria, and animals. They are heme-thiolate proteins that use molecular oxygen and NADPH to modify substrates with diverse chemical reactions including oxidations, hydroxylations, dealkylations, and reductions [22] 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 [22–24].

CYPs are characterised by a conserved heme-binding domain, an oxygen binding domain, two conserved motifs (X-E-X-X-R and P-E-R-F) that form what is known as the ERR triad and is involved in positioning and stabilising the heme pocket, and several highly variable substrate positioning and recognition sites [25]. The three-dimensional structure of CYPs is conserved across the family even though the amino acid sequences of individual members may be as little as 20% identical [26–30]. Previous phylogenetic analyses of CYPs grouped them into monophyletic clades termed clans, each containing one or more CYP families [31–33], with clans being named after their lowest numbered family member [34]. Clans represent the deepest clades that reproducibly appear in multiple phylogenetic trees.

Here we report the phylogenetic relationships among both the GST and CYP proteins within the Archaeplastida lineage. We discovered that those CYPs and GSTs that confer herbicide resistance among weeds are restricted to two monophyletic clans and two monophyletic classes, respectively.

These clans and classes already existed in the common ancestor of land plants, which is estimated to have existed between 980 and 473 Mya [35–37].These clans and classes diversified early in land plant evolution and now constitute the largest groups of CYP and GST proteins in extant vascular plants. This analysis suggests that natural selection caused by herbicides acts on sets of ancient genes that existed in the last common ancestor of the land plants and *K. nitens*, a streptophyte alga, and diversified in vascular plants, leading to the evolution of herbicide resistance in the agricultural landscape.

Materials and methods

Data resources

Protein sequences from *A. thaliana* were retrieved from TAIR10 [38] (https://www.arabidopsis.org/) Protein sequences from *Oryza sativa* were retrieved from the rice genome annotation project [39] (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 [40] (https://www.hornworts.uzh.ch/en/download.html). Protein sequences from the streptophyte alga *Klebsormidium nitens* were obtained from the *K. nitens* genome webpage [41] (http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/). Protein sequences from the moss *Physcomitrium patens* and the chlorophyte alga *Chlamydomonas reinhardtii* were retrieved from Phytozome 12 [42] (https://phytozome.jgi.doe.gov/pz/portal.html). Protein sequences from the red alga *Cyanidioschyzon merolae* were retrieved from the *C. merolae* genome webpage [43] (https://www.genome.jp/kegg-bin/show_organism?org=cme).

A classification of CYP genes from *A. thaliana, S. moellendorffii*, *P. patens*, *C. reinhardtii* is available on The Cytochrome P450 Homepage [44] (http://drnelson.uthsc.edu/plants/). Two other Arabidopsis CYP databases can be found on the Arabidopsis Cytochrome P450 List [45] (http://www.p450.kvl.dk/At_cyps/table.shtml) and CyPEDIA [46] (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/).

Sequence collection

CYP protein sequences from *A. thaliana* and *O. sativa* [47,48] were used to perform BLASTP searches using a minimum E value cut-off of $1e^{-10}$ against the predicted proteomes of *S. moellendorffii*, *M. polymorpha*, *A. agrestis*, *P. patens*, *K. nitens*, *C. reinhardtii*, and *C. merolae*. GST protein sequences

were retrieved by BLASTP searches using GST proteins from *A. thaliana* [49,50], *O. sativa* [51,52], and *P. patens* [53] against the predicted proteomes of *S. moellendorffii*, *M. polymorpha*, *A. agrestis*, *K. nitens*, *C. reinhardtii*, and *C. merolae.* This initial list of sequences for each species was self-blasted against the proteome of that species to retrieve additional sequences belonging to species-specific clans. Each CYP sequence was checked for the presence of the cytochrome p450 domain (PF00067, IPR00128) and each GST sequence was checked for the presence of the GST N-terminal domain (IPR004045, IPR019564, PF13409, PF17172, PF13417 and PF02798) and C-terminal domain (IPR010987, PF13410, PF00043, PF14497 and PF17171) using InterProScan 84.0 [54].

Two GST classes, mitochondrial Kappa and microsomal MAPEG, don't possess a GST N-terminal thioredoxin-like domain or GST C-terminal domain and lack the N-terminal active site found in all other GST proteins. An additional group of sequences was identified by this analysis possessing two GST N-terminal domains but lacking a C-terminal domain. Protein sequences belonging to the Kappa, MAPEG, and 2N classes were therefore not included in the phylogenetic analysis but are listed in S5 Table.

Sequence alignment

Sequences were aligned in MAFFT [55] using the FFT-NS-2 algorithm and visualised in Bioedit [56]. Sequences lacking important functional residues were removed. To trim large gaps, four approaches to alignment cleaning were undertaken. A manual approach was carried out using knowledge of the location of the functionally important CYP and GST residues. A more stringent trimming approach was also tested with the trimming software trimAl v.1.2. [57] using the three automated modes (-gappyout, -strict and -strictplus) (S2 Fig).

Phylogenetic analysis

The final alignments were subjected to a maximum-likelihood analysis conducted by PHyML 3.0 [58] 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 v1.4.4 [59] and annotated in Inkscape v1.0.2 [60].

Results

1130 CYP and 358 GST sequences were identified in the genomes of 9 species of Archaeplastida

To determine the phylogenetic relationships among CYP and GST sequences in the Archaeplastida lineage, we collected sequences from online resources. CYP and GST protein-coding genes in 9 species (Table 1) representing key Archaeplastida lineages were identified as described in Methods. The resulting 1130 CYP and 358 GST sequences included sequences from the red alga *Cyanidioschyzon merolae* (5 CYP and 9 GST proteins), the chlorophyte alga *Chlamydomonas reinhardtii* (40 CYP and 19 GST proteins), the streptophyte alga *Klebsormidium nitens* (29 CYP and 24 GST proteins), the liverwort *Marchantia polymorpha* (115 CYP and 35 GST proteins), the moss *Physcomitrium patens* (69 CYP and 42 GST proteins), the hornwort *Anthoceros agrestis* (144 CYP and 26 GST proteins), the lycophyte *Selaginella moellendorffii* (199 CYP and 57 GST proteins), and the angiosperms *Oryza sativa* (291 CYP and 85 GST proteins) and *Arabidopsis thaliana* (238 CYP and 61 GST proteins) (Table 1). The *M. polymorpha* CYP sequences were named following the standard CYP nomenclature [34].

Table 1. List of species used in the analysis.

Species	Classification	Genome (Mb)	Protein-coding genes	GSTs	GSTs (% PCG)	CYPs	CYPs (% PCG)	References
Arabidopsis thaliana	Angiosperm eudicot	135	25,498	61	0.24	238	0.93	[61]
Oryza sativa	Angiosperm monocot	321	35,681	85	0.24	291	0.82	[39]
Selaginella moellendorffii	Lycophyte	212.6	22,285	57	0.26	199	0.89	[62]
Anthoceros agrestis	Hornwort	133	24,700	26	0.11	144	0.58	[40]
Physcomitrella patens	Moss	480	35,938	42	0.12	69	0.19	[63]
Marchantia polymorpha	Liverwort	225.8	19,138	35	0.18	115	0.60	[64]
Klebsormidium nitens	Streptophyte alga	117.1	16,215	24	0.15	29	0.18	[41]
Chlamydomonas reinhardtii	Chlorophyte alga	120	15,143	19	0.13	40	0.26	[65]
Cyanidioschyzon merolae	Rhodophyte alga	16.5	5,331	9	0.17	5	0.09	[43]

Including their phylum (Classification), genome size (Genome), total number of protein-coding genes (Protein-coding genes), total number of GST proteins (GSTs), GST proteins as a percentage of total protein coding genes (GSTs % PCG), total number of CYP proteins (CYPs), CYP proteins as a percentage of total protein coding genes (CYPs % PCG) and the bibliographical reference for each genome sequence.

Alignments were generated from the identified sequences and used to construct phylogenetic trees, using four alignment trimming approaches. The sequences were manually trimmed to retain the homologous domains and remove large gaps. A more stringent trimming approach was also tested with the trimming software trimAl v.1.2. [57] using the three automated modes (-gappyout, -strict and -strictplus) (S2 Fig). For both the GST and CYP phylogenetic trees, the approximate likelihood ratio test (aLRT) support values for the deepest clades of the maximum-likelihood (ML) trees resulting from the trimAl -strict and -strictplus alignments were low (0-0.23). The ML trees generated from the trimAl -gappyout alignments had correct tree topologies but had low aLRT support values for the main clades (0.05-0.23). The ML trees generated from the manually

trimmed GST and CYP alignments had the overall highest aLRT values (>0.8) for the main clades and were selected as the representative trees for further analysis (Fig 1).

Fig 1. Phylogenetic analysis of CYP and GST protein sequences in the Archaeplastida.

Unrooted cladogram of a maximum likelihood (ML) analysis of Archaeplastida CYP (A) and GST (B) proteins conducted by PHyML 3.0 [58] 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. CYP clans are indicated by light green highlighting and numbers. GST classes are indicated by light yellow highlighting and acronyms. Coloured dots indicate the presence of sequences from different species in each clan. *A. thaliana* (orange); *O. sativa ssp. japonica* (grey); *Selaginella moellendorffii* (yellow); *Physcomitrium patens* (cyan); *Anthoceros agrestis* (blue); *Marchantia polymorpha* (black); *Klebsormidium nitens* (purple); *Chlamydomonas reinhardtii* (green); *Cyanidioschyzon merolae* (red).

Plant CYP clans are ancient and two CYP clans existed in the last common ancestor of the Archaeplastida

To elucidate the evolution of CYPs in Archaeplastida, we constructed a phylogenetic tree using a maximum likelihood approach (Fig 1A). This analysis demonstrated that CYPs from the 9 representative species of Archaeplastida grouped into 17 monophyletic clans, consistent with previous analyses of plant CYP phylogeny [31–33].

CYPs encoded by the genomes of land plants *A. agrestis*, *M. polymorpha, P. patens*, S. *moellendorffii*, *O. sativa*, and *A. thaliana* corresponded to 12 of the 17 clans identified in the Archaeplastida – 51, 71, 72, 74, 85, 86, 97, 710, 711, 727, 746, and 747. Each of these 12 clans was also represented in the genome of the streptophyte alga *K. nitens.* This indicates that these clans

existed before the divergence of *K. nitens* and land plants from their last common ancestor. Members of 6 of the 12 clans – 71, 72, 74, 85, 86, and 727 – were not present in the genome of *C. reinhardtii*. This suggests that these 6 clans originated in the streptophyte lineage after the divergence of chlorophytes and streptophytes from their last common ancestor but before the divergence of *K. nitens* (Fig 2A). Members of the other 6 of the 12 CYP clans – 51, 97, 710, 711, 746, and 747 – were encoded by the *C. reinhardtii* genome indicating that they were present before the divergence of streptophytes and chlorophytic algae from the last common ancestor. Two of the clans were also present in red algae; there is one member of clan 51 and two members of clan 710 in the genome of *C. merolae*. This places the origin of clan 51 and clan 710 before the divergence of Rhodophyta and Viridiplantae (Fig 2A). We conclude that clans 51 and 710 were present in the last common ancestor of Archaeplastida and therefore constitute the most ancient Archaeplastida clans.

Fig 2. Four CYP clans and two GST classes expanded during land plant evolution.

Cladogram of Archaeplastida phylogeny showing CYP clan (A) and GST class (C) origins and losses in plants. Blue circles represent first appearance of a clan/class, black circles represent the absence of a clan or class in a particular lineage. Numbers of CYP proteins in each species showing increases in the sizes of four CYP clans (B) and two GST classes (D) during land plant evolution.

Three clans – 55, 737, and 741 – were restricted to *C. reinhardtii*. There is a single clan 55 member in *C. reinhardtii*, CrCYP55B1, which was sister to the clan 51 clade. Members of clan 55 are also present in fungi and hypothesised to have been acquired by *C. reinhardtii* from fungi through horizontal gene transfer [66]. Two *C. reinhardtii* CYP protein sequences – CrCYP741A1 and CrCYP768A1 – formed a monophyletic clade, clan 741, that was sister to the clade comprising clans 86, 97, and 747. Thirty *C. reinhardtii* CYP sequences formed a monophyletic clade – clan 737 – which

was sister to the clade containing the 86, 97, 741, and 747 clans. These data are consistent with the hypothesis that clans 737 and 741 are chlorophyte specific.

Two clans – Cm1 and Cm2 – comprised only single red algae proteins. Cm1 (CMD096C) was sister to the clade containing clans 72, 86, 97, 711, 727, 737, 746, and 747. Clan Cm1 and clans 72, 86, 97, 711, 727, 737, 746, and 747 are therefore likely derived from a protein present in the common ancestor of the red algae and the green plant lineage (chlorophytes and streptophytes). Cm2 (CMR093C) was sister to clan 710 but shares very low amino acid identity (20%) with members of 710. Cm2 is possibly an ancestral 710 protein or it could represent a red-algae specific clan. Clans Cm2 and 710 are therefore likely derived from a protein present in the common ancestor of the red algae and the green plant lineage (chlorophytes and streptophytes).

In summary, our phylogenetic analysis shows that each of the of the land plant CYP clans are also present in the genome of the streptophyte alga *K. nitens*. This indicates that the diversity of CYP sequences in plants evolved among algae in the aquatic environment before plants colonised land between 980 and 470 million years ago [35–37]. No new clans evolved among land plants after their colonisation of the land. Instead, the number of genes in each clan increased. Five CYP clans present in land plants and streptophyte algae are also present in the genome of the chlorophyte alga *C. reinhardtii*, which places their origin before the divergence of the chlorophyte and streptophyte lineages from their last common ancestor. Two clans found in land plants, streptophyte algae, and chlorophytes – 51 and 710 – are also present in the red algae. This suggests that these clans are the most ancient Archaeplastida clans and evolved before the divergence of Rhodophyta and Viridiplantae from their last common ancestor.

Plant GST classes are ancient, and 11 classes existed in the last common ancestor of the Archaeplastida

To elucidate the evolutionary history of GST classes in Archaeplastida, sequences were retrieved, aligned, and a phylogenetic tree constructed using maximum likelihood statistics (Fig 1B). The topology of the trees demonstrated that GSTs from the 9 representative species of Archaeplastida constituted 19 monophyletic classes – Ala, Alb, Alc, Cr1, DHAR, EF1B-γ, GHR, Hemerythrin, Iota, Kn1, Lambda, Metaxin, mPGES2, Phi, Tau, TCHQD, Theta, Ure2p, and Zeta. Of these 19 classes, 14 are encoded in the genomes of the land plant species *A. agrestis*, *M. polymorpha, P. patens*, S. *moellendorffii*, *O. sativa* and *A. thaliana* – DHAR, EF1B-γ, GHR, Hemerythrin, Iota, Lambda, Metaxin, mPGES2, Phi, Tau, TCHQD, Theta, Ure2p and Zeta (Fig 1B). Five of the 19 classes are novel GST classes identified in algal genomes, named Ala, Alb, Alc, Cr1, and Kn1.

16 algal GST sequences comprised several different monophyletic clades. Three *C. reinhardtii* sequences and one *C. merolae* sequence comprised class Alc, which is a sister to the Ure2p class (Fig 1B). However, these sequences lacked a characteristic Ure2p protein domain (cd03048) and were therefore not included in the Ure2p class. Class Alb, which included one *K. nitens* sequence and one *C. merolae* sequence, is a sister to the monophyletic clade comprising both the Ure2p and Alc classes. Class Ala, comprising 7 *C. reinhardtii* sequences and a single *C. merolae* sequence, is a sister to the clade containing Phi, Theta, EFB1-γ, Ure2p, Alb, and Alc GST sequences. Ala, Alb, and Alc may represent classes that evolved in the ancestor of Archaeplastida, where Ala and Alc were lost in the common ancestor of Streptophytes, and Alb was lost in the chlorophyte lineage and in the common ancestor of land plants.

Two individual algal sequences formed two independent clades. A *C. reinhardtii* sequence (Cre12.g508850.t1) was sister to the TCHQD class. However, this sequence lacked a TCHQD protein domain (IPR044617) and was therefore designated Cr1. A *K. nitens* sequence (Kfl00304_0120_v1) was sister to the Lambda class, however there was no GST Lambda class C-terminal domain

317 (cd03203). This sequence was designated Kn1. These data suggest that Cr1 evolved in the
318 chlorophyte lineage and Kn1 evolved in the streptophyte algal lineage.

319 Of the 14 GST classes present in the genomes of the land plants *A. agrestis*, *M. polymorpha,*
320 *P. patens*, S*. moellendorffii*, *O. sativa* and *A. thaliana*, 9 classes – EF1B-γ, GHR, Metaxin, mPGES2, Phi,
321 TCHQD, Theta, Ure2p, Zeta – are also found in non-plant genomes (such as metazoans, bacteria,
322 archaea, and fungi) and therefore predate the origin of the Archaeplastida [53,67–71]. The other 5
323 GST classes – DHAR, Hemerythrin, Iota, Lambda, and Tau – have only been described from the
324 genomes of land plants and chlorophyte and streptophyte algae [49,51,53,72]. Our analysis shows
325 that Lambda and Tau members are present in the genome of the streptophyte alga *K. nitens* but not
326 in the *C. reinhardtii* and *C. merolae* genomes. This indicates that these classes evolved among the
327 streptophytes after the divergence of the red algae and chlorophytes but before the divergence of *K.*
328 *nitens* and land plants. Members of the Hemerythrin class were found in genomes of the bryophytes
329 *P. patens, M. polymorpha*, and *A. agrestis* and the lycophyte *S. moellendorffii*, but not in the
330 angiosperms or in *K. nitens, C. reinhardtii*, or *C. merolae.* This suggests that the Hemerythrin class
331 originated in the common ancestor of bryophytes and vascular plants but was lost in the common
332 ancestor of the angiosperms. There are DHAR members in the genomes of *K. nitens* and *C.*
333 *reinhardtii*. This suggests that DHAR GST proteins were present in the last common ancestor of
334 chlorophytes and streptophytes. There are Iota members in *C. merolae, C. reinhardtii*, and *K. nitens*
335 indicating that Iota class enzymes originated before the divergence of rhodophytes and chlorophytes
336 in the common ancestor of Archaeplastida (Fig 2C).

337 There are 26 GST proteins belonging to 12 classes in the genome of the hornwort
338 *Anthoceros agrestis* (S2 Table). One sequence (AagrOXF_evm.model.utg000005l.356.1) nested
339 within the monophyletic Tau GST clade and contained the conserved N- and C-terminal Tau class
340 catalytic motifs (cd03058 and cd03185). This is strong evidence that
341 AagrOXF_evm.model.utg000005l.356.1 is a Tau GST. Tau GST proteins are also present in
342 streptophyte algae, liverworts, and vascular plants but absent from mosses. This suggests that the

Tau GST class was present in the last common ancestor of the streptophyte algae and subsequently lost in the moss lineage (Fig 2C).

In summary, this analysis showed that Archaeplastida GST proteins comprise 19 classes. 11 classes – Ala, Alb, Alc, EF1B-γ, GHR, Iota, Metaxin, mPGES2, TCHQD, Theta, and Zeta – were present in the common ancestor of the Archaeplastida. 12 classes originated after the divergence of Archaeplastida from other eukaryotes. The earliest GST classes to arise in Archaeplastida were the Ala, Alb, Alc, and Iota classes, which originated before the separation of rhodophyte and chlorophyte lineages. The DHAR class originated in the common ancestor of chlorophytes and streptophytes.The Cr1 class originated in the chlorophyte lineage. Lambda, Tau, Phi, and Ure2p GSTs originated in the last common ancestor of streptophyte algae and land plants. Kn1 originated in the streptophyte algae. The most recently diverging plant GST class, the Hemerythrin class, originated in the last common ancestor of land plants.

CYP clans 71, 72, 85, and 86 and GST classes Phi and Tau GST expanded among land plants

The number of CYP genes encoded in the genomes of land plants is larger than the number encoded in the genomes of algae. We identified between 5 and 40 CYP protein genes in algae – 5 in *C. merolae*, 40 in *C. reinhardtii*, and 29 in *K. nitens*. We identified between 69 and 144 among the bryophytes – 69 in *A. agrestis*, 115 in *P. patens*, and 144 in *M. polymorpha* genomes. Among the vascular plants we identified between 199 and 291 – 199 in *S. moellendorffii*, 238 in *A. thaliana*, and 291 in *O. sativa* genomes (Table 1, S1 Table).

To determine if CYP gene numbers are correlated with the numbers of total protein coding genes in land plants, we calculated the percentage of protein-coding genes that encoded CYP proteins. CYPs represent 0.18% of the protein-coding genes in the streptophyte alga *K. nitens*, 0.19-

0.60% in bryophytes, and 0.82-0.93% in vascular plants (Table 1). These data are consistent with the hypothesis that the larger number of CYP genes in bryophytes and vascular plants compared to algae is not simply due to a general increase in gene number.

To identify the clans responsible for the increase in CYPs in land plants, clan gene numbers were compared between species. There are more genes in clans 71, 72, 85, and 86 in land plants than in streptophyte algae (Fig 2B, S1 Table), with clan 71 gene numbers differing the most between species. There are three 71 clan members in the genome of the streptophyte alga *K. nitens*. Among the bryophytes there are 59 clan 71 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 98 in the lycophyte *S. moellendorffii*, 148 in *A. thaliana*, and 163 in *O. sativa* (S1 Table). Clan 71 proteins represent 10% of all CYPs in *K. nitens* but 40-60% of all CYPs in the land plants. Together these data are consistent with the hypothesis that the expansion in the numbers of clan 71 genes contributed to the large number of CYP proteins in land plants compared to algae (non-land plant Archaeplastida). There are only a small number of genes in eight CYP clans across all streptophyte species – 51, 74, 97, 710, 711, 727, 746, and 747. Generally, there were fewer than 10 members in each of these clans in any one species (S1 Table). Thus, these clans therefore represent monophyletic groups that did not diversify among land plants.

Despite the smaller number of GST classes in land plants compared to algae, there are more GST protein coding genes in land plants than in algae. We identified 9 GST genes in the genome of *C. merolae*, 19 in *C. reinhardtii*, and 24 in *K. nitens*. Among the bryophytes we identified 35 in *M. polymorpha*, 42 in *P. patens* and 26 in *A. agrestis*. Among the vascular plants we identified 57 in *S. moellendorffii*, 85 in *O. sativa* and 61 in *A. thaliana* (Table 1, S2 Table). Genes coding for GST proteins represent 0.15% of all protein coding genes in *K. nitens*, 0.11-0.18% in bryophytes, and 0.24-0.26% in vascular plants (Table 1). These data are consistent with the hypothesis that the larger number of GST genes in vascular plants than in algae is not due to a general increase in protein number, but due to GST family expansion.

To identify the classes responsible for the increase in GSTs in vascular plants, gene numbers in each GST class were compared between species. The number of GST proteins in the Phi and Tau classes is larger in land plants than in streptophyte algae. There are 3 Phi class members in the genome of the streptophyte alga *K. nitens*. Among the bryophytes there are 18 Phi class genes in the genome of *M. polymorpha*, 10 in *P. patens* and 11 in *A. agrestis*. Only 1 Phi GST was identified in the genome of the lycophyte *S. moellendorffii*. Among the angiosperms, there are 19 Phi GST proteins in *O. sativa* and 13 in *A. thaliana*. This suggests that the Phi class expanded in the land plant lineage after the divergence of streptophyte algae and land plants from the last common ancestor but before the divergence of bryophytes and vascular plants. There are also more Tau class GST proteins in vascular plant genomes than in either the algal or bryophyte genomes (Fig 2D). 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 17 GST classes in Archaeplastida – Ala, Alb, Alc, Cr1, DHAR, EF1B-γ, GHR, Hemerythrin, Iota, Kn1, Lambda, Metaxin, mPGES2, TCHQD, Theta, Ure2p, and Zeta – gene numbers are less than 10 in each species (S2 Table), indicating these these classes have not expanded during the course of evolution.

In summary, our phylogenetic analysis shows that the 2 to 10-fold larger number of CYP genes in the genomes of land plants than in the streptophyte alga *K. nitens* results from expansions of clans 71, 72, 85, and 86. The 1.5 to 3.5-fold more GST genes in land plants than in the streptophyte alga *K. nitens* results from expansions of the Phi and Tau classes.

Herbicide metabolic resistance is associated with proteins from the GST Phi and Tau classes and CYP 71 and 72 clans

GSTs and CYPs have been genetically associated with herbicide resistance in crops and weed populations [73,74]. To identify which CYP clans and GST classes are genetically and/or metabolically associated with herbicide resistance, a literature search was conducted. CYPs or GSTs reported in previous studies to increase herbicide resistance in transgenic plants or to metabolise herbicides were classified as NTSR genes (S3 and S4 Tables). CYPs and GSTs found to have increased expression in herbicide resistant weeds, but whose function was not experimentally validated, were classified as "candidate NTSR genes" and are listed in S5 and S6 Tables.

Clan 71 and clan 72 CYP proteins are associated with resistance to herbicides from 16 chemical classes

A total of thirty plant CYPs have been experimentally shown to metabolise or confer resistance to one or more herbicides in sensitivity or metabolism assays (Fig 3A, S3 Table). These CYPs were identified in the model plant Arabidopsis (*A. thaliana*) [75,76], the grass weeds barnyard grass (*Echinochloa phyllopogon*) [77–79], shortawn foxtail (*Alopecurus aequalis*) [80] and annual ryegrass (*Lolium rigidum*) [81], the gymnosperm western red cedar (*Thuja plicata*) [82], and the crops barley (*Hordeum vulgare*) [83], rice (*Oryza sativa*) [84–86], wheat (*Triticum aestivum*) [87], maize (*Zea mays*) [88], cotton (*Gossypium hirsutum*) [89], soybean (*Glycine max*) [90,91], ginseng (*Panax ginseng*) [92], Jerusalem artichoke (*Helianthus tuberosus*) [93,94] and tobacco (*Nicotiana tabacum*) [95]. These 29 CYPs metabolised or conferred resistance to diverse herbicide chemical classes, with the majority (24 of 29) metabolising phenylureas or sulfonylureas (S3 Table).

Fig 3. GST and CYP proteins associated with herbicide resistance belong to the Lambda, Phi, and Tau classes and clans 71 and 72.

(A) Number of CYP proteins associated with herbicide resistance (white bars), present in the *A. thaliana* genome (light grey bars) and in the *O. sativa* genome (dark grey bars), per clan. (B) Number of GST proteins associated with herbicide resistance (white bars), present in the *A. thaliana* genome (light grey bars) and the *O. sativa* genome (dark grey bars) per clan. (C) Number of CYP proteins associated with resistance per clan, with family membership indicated by colours. The most represented family among CYPs associated with herbicide resistance is the CYP81 family. Numbers over or within bars represent the number of proteins within that category.

All thirty of the herbicide-metabolising CYPs belong to clan 71 or 72 (Fig 3A). Twenty-six clan 71 enzymes have been shown to confer resistance to benzothiadiazinones, clomazone, DEN, DIM, FOP, isoxazolidinones, phenylureas, pyrazoles, pyridazinones, thiobenzoates, sulfonylaminocarbonyl-triazolinones, sulfonylureas, thiadiazines, triazolopyrimidines, and triketone herbicide chemicals. Clan 71 CYPs are encoded in large number in the genomes of all land species; there are 150 in *A. thaliana* and 164 in *O. sativa*. In contrast, there are much fewer clan 72 CYPs encoded in land plant genomes, with 19 in *A. thaliana* and 34 in *O. sativa.* Four clan 72 members shown to confer resistance to thiobenzoates, pelargonic acid, or sulfonylureas. Thus, all CYPs currently known to metabolise or confer resistance to herbicides belong to clans 71 and 72, which represent two of the four expanded CYP clans in land plants.

Twelve members of the clan 71 family CYP81 were shown to confer herbicide resistance. This is more than any other family or clan (Fig 3C). CYP81 enzymes catalyse hydroxylations and N-/O-demethylations of herbicide substrates [77]. The CYP81 enzymes metabolise herbicides from five chemical classes, more than any other CYP family to date.

Together these data indicate that genes encoding CYP proteins that mutate to herbicide resistance are members of clan 71 and 72. Within clan 71, more members of the CYP81 family are associated with herbicide resistance than any other family.

Phi, Tau and Lambda GST class proteins are associated with resistance to herbicides from 9 chemical classes

Thirty-three plant GSTs were found in the literature to be active towards one or more herbicides or that confer herbicide resistance (Fig 3B, S4 Table). These GST proteins were identified in the model species Arabidopsis (*Arabidopsis thaliana*) [96], moss (*P. patens*) [53], the weed species blackgrass (*Alopecurus myosuroides*) [97,98], the crops maize (*Zea mays*) [99,100,109,101–108], rice (*Oryza sativa*) [99,110,111], sorghum (*Sorghum bicolor*) [112], wheat (*Triticum aestivum*) [29,113–115] and soybean (*Glycine max*) [116–118]. These GSTs were shown to modify or confer resistance to diverse chemical classes, with most GSTs (28 of 33) modifying chloroacetanilide herbicides. Of the 33 GSTs, 11 are Phi class members, 21 are Tau class members, and one is a lambda class member (Fig 3B).

Twenty-one Tau GSTs were identified in 6 species and catalysed the GSH-conjugation of chloroacetanilide, diphenyl ether, FOP, sulfonylurea and triazine herbicide chemicals. Eleven Phi GSTs identified in 6 species catalysed the GSH-conjugation of bipyridylium, chloroacetanilide, DIM, diphenyl ether, FOP, organophosphorus, phenylurea, sulphonylurea, thiocarbamate and triazine herbicides (S4 Table).

50-70% of all GSTs encoded in vascular plant genomes are Tau or Phi class members. In *A. thaliana*, there are 41 Tau and Phi GSTs and only 20 GSTs across the other 12 classes. In *O. sativa*, there are 68 Tau and Phi GSTs and 17 in the other classes (Fig 3B). Thus, the overrepresentation of Phi and Tau class GSTs among those reported to confer herbicide resistance may simply be due to the fact that there are more genes in these classes than others. Therefore, we cannot reject the hypothesis that there is an equal probability of GST proteins from any class being able to confer

herbicide resistance. The report that overexpression of a single Lambda class GST – there are 3 Lambda class genes encoded in *A. thaliana* – in a naturally occurring herbicide tolerant weed is able to confer herbicide resistance supports this hypothesis.

In conclusion, herbicide resistance has been associated with GST proteins from Tau, Phi and Lambda classes. The probability of resistance evolving among any of those classes, is likely to be proportional to the number of genes in each class.

Discussion

Cytochrome P450 monooxygenases (CYPs) and glutathione S-transferases (GSTs) are enzymes that catalyse the chemical modification of a multitude of organic compounds in organisms from all domains of life. Overexpression of genes encoding CYPs and GSTs has been shown to confer herbicide resistance in wild weed populations subjected to herbicide selection. To classify the genes that metabolise herbicides, we carried out a phylogenetic analysis of both the CYP and GST protein families. By comparing protein sequences of 9 representative species of the Archaeplastida – the lineage that includes the red algae, glaucophyte algae, chlorophyte algae, and streptophytes – and generating phylogenetic trees, we identified the CYP and GST protein families that existed in the common ancestor of the Archaeplastida. Members of two CYP clans (clans 51 and 71) and eleven GST classes (Ala, Alb, Alc, EF1B-y, GHR, Iota, Metaxin, mPGES2, TCHQD, Theta, and Zeta) existed in the last common ancestor of the Archaeplastida. Other families evolved during the course of Archaeplastida evolution. There are more CYP and GST genes in land plants than in algae, even relative to the total number of genes, consistent with the hypothesis that these gene families expanded during Archaeplastida evolution. This expansion was largely driven by gene duplications among CYP clan 71 and 72, and among the GST Phi and Tau classes. Those CYP and GST genes that confer resistance to herbicides belong almost exclusively to these expanded CYP clans and GST classes.

In the face of intense herbicide use over the past 50 years, herbicide resistance has evolved through the selection of naturally occurring alleles that contribute to herbicide tolerance. Target site resistance can evolve as a result of mutations in the gene encoding the herbicide's target, thereby disrupting the inhibition of the target proteins by the herbicide. Non-target site resistance results from genetic changes that inhibit access of the active herbicide to its target [1]. Diverse forms of non-target site resistance have been reported which either chemically modify the herbicide, making it inactive, or sequester the herbicide to a location where it cannot access the target [2–4]. Overexpression of genes encoding CYPs and GSTs is associated with herbicide resistance in many weed populations [8–10]. Using phylogenetic trees built from protein sequences from 9 Archaeplastida species, we show that the CYP and GST proteins that confer non-target site herbicide resistance in natural weed populations belong to the expanded CYP clans 71 and 72 and the GST Phi and Tau classes.

It is unclear why mutation of genes in these CYP clans and GST classes leads to resistance in weed populations while others do not. It is possible that because these clans and classes are the largest, there is simply a greater probability of them mutating to resistance. Characterizing the cause of resistance in more resistant populations will help to resolve this question. It is also possible that the enzymatic activity of these proteins makes them more likely to metabolize herbicide compounds. Further characterization of the endogenous function of CYP clans 71 and 72 and GST Tau and Phi classes during normal plant growth and development will help to answer this question. At present, the available phylogenetic and enzymatic data do not allow us to distinguish between these alternative hypotheses.

All CYP clans and all but one land plant GST classes that are present in land plants evolved before the divergence of streptophyte algae and land plants from their last common ancestor. These results demonstrate that the clan and class diversity in extant plant CYP and GST proteins, respectively, evolved in the Proterozoic, before the divergence of land plants and streptophyte algae from a last common ancestor. Then, early in embryophyte evolution during the Palaeozoic,

538 expansion of four of the twelve CYP clans and two of the fourteen GST classes resulted in the large
539 number of CYP and GST proteins found in extant land plants. It is among these expanded groups that
540 herbicide resistance genes are found. Thus, resistance depends on the deregulation of the
541 expression of genes that evolved in the Proterozoic, with original functions unrelated to herbicide
542 resistance. This is an unusual example of exaptation – whereby traits [or genes] that evolved in
543 response to one selection pressure – in this case probably metabolic biosynthesis – are selected for
544 in response to an entirely different selection pressure – here herbicide resistance [119]. Exaptation is
545 likely to be a general principle underpinning the evolution of herbicide resistance mechanisms
546 among weeds in the agricultural landscape.

547

548

549 Acknowledgements

550 The authors would like to thank Professor David Nelson for the nomenclature of *Marchantia*
551 *polymorpha* CYPs in this article, and Dr Sandy Hetherington for his input and advice on the methods.

552

553 Funding

554 This research was supported by a European Research Council (ERC) Advanced Grants EVO500 project
555 number 250284 and De Novo-P (project number 787613) to LD from the European Commission. AC
556 was supported by a British Biological Sciences Research Council (BBSRC) Scholarship through a
557 doctoral training partnership (BB/XXXX /X). The funders had no role in study design, data collection
558 and analysis, decision to publish, or preparation of the manuscript.

559

References

1. Powles SB, Yu Q. Evolution in action: plants resistant to herbicides. Annu Rev Plant Biol. 2010;61(1):317–47.

2. Ghanizadeh H, Kerry &, Harrington CC. Non-target Site Mechanisms of Resistance to Herbicides. CRC Crit Rev Plant Sci. 2017;36(1):24–34.

3. Délye C. Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: A major challenge for weed science in the forthcoming decade. Pest Manag Sci. 2013;69(2):176–87.

4. Délye C, Duhoux A, Pernin F, Riggins CW, Tranel PJ. Molecular mechanisms of herbicide resistance. Weed Sci. 2015;63(SP1):91–115.

5. Dixon DP, McEwen AG, Lapthorn AJ, Edwards R. Forced evolution of a herbicide detoxifying glutathione transferase. J Biol Chem. 2003 Jul 27;278(26):23930–5.

6. Reade JPH, Milner LJ, Cobb AH. A role for glutathione-S-transferases in resistance to herbicides in grasses. Weed Sci. 2004 May 2;52(3):468–74.

7. Yu Q, Powles S. Metabolism-Based Herbicide Resistance and Cross-Resistance in Crop Weeds: A Threat to Herbicide Sustainability and Global Crop Production. Plant Physiol. 2014;166:1106–18.

8. Yuan JS, Tranel PJ, Stewart CN. Non-target-site herbicide resistance: a family business. Trends Plant Sci. 2007 Jan 1;12(1):6–13.

9. Cummins I, Dixon DP, Freitag-Pohl S, Skipsey M, Edwards R. Multiple roles for plant glutathione transferases in xenobiotic detoxification. Drug Metab Rev. 2011;43(2):266–80.

10. Dimaano NG, Iwakami S. Cytochrome P450-mediated herbicide metabolism in plants: current understanding and prospects. Pest Manag Sci. 2021 Jan 31;77(1):22–32.

11. Kumar S, Asif MH, Chakrabarty D, Tripathi RD, Dubey RS, Trivedi PK. Expression of a rice Lambda class of glutathione S-transferase, OsGSTL2, in Arabidopsis provides tolerance to

heavy metal and other abiotic stresses. J Hazard Mater. 2013 Mar 5;248–249(1):228–37.

12. Riechers DE, Vaughn KC, Molin WT. The role of plant glutathione S-transferases in herbicide metabolism. ACS Symp Ser. 2005;899:216–32.

13. Dixon D, Davis B, Edwards R. Functional divergence in the glutathione transferase superfamily in plants: Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. J Biol Chem. 2002 Aug 23;277(34):30859–69.

14. Gonneau M, Mornet R, Laloue M. A *Nicotiana plumbaginifolia* protein labeled with an azido cytokinin agonist is a glutathione S-transferase. Physiol Plant. 1998 May 1;103(1):114–24.

15. Zettl R, Schell J, Palme K. Photoaffinity labeling of Arabidopsis thaliana plasma membrane vesicles by 5-azido-[7-3H]indole-3-acetic acid: Identification of a glutathione S- transferase. Proc Natl Acad Sci U S A. 1994;91(2):689–93.

16. Bilang J, Sturm A. Cloning and characterization of a glutathione S-transferase that can be photolabeled with 5-azido-indole-3-acetic acid. Plant Physiol. 1995 Sep 1;109(1):253–60.

17. Dixon DP, Edwards R. Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. J Biol Chem. 2010 Nov 19;285(47):36322–9.

18. Liu Y, Jiang H, Zhao Y, Li X, Dai X, Zhuang J, et al. Three Camellia sinensis glutathione S-transferases are involved in the storage of anthocyanins, flavonols, and proanthocyanidins. Planta. 2019 Oct 1;250(4):1163–75.

19. Fernández-Cañón JM, Baetscher MW, Finegold M, Burlingame T, Gibson KM, Grompe M. Maleylacetoacetate Isomerase (MAAI/GSTZ)-Deficient Mice Reveal a Glutathione-Dependent Nonenzymatic Bypass in Tyrosine Catabolism. Mol Cell Biol. 2002 Jul 1;22(13):4943–51.

20. Kampranis SC, Damianova R, Atallah M, Toby G, Kondi G, Tsichlis PN, et al. A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. J Biol Chem. 2000 Sep 22;275(38):29207–16.

21. Edwards R, Dixon DP, Walbot V. Plant glutathione S-transferases: enzymes with multiple

functions in sickness and in health. Trends Plant Sci. 2000 May 1;5(5):193–8.

22. Mizutani M, Sato F. Unusual P450 reactions in plant secondary metabolism. Vol. 507, Archives of Biochemistry and Biophysics. Academic Press Inc.; 2011. p. 194–203.

23. Wakabayashi T, Hamana M, Mori A, Akiyama R, Ueno K, Osakabe K, et al. Direct conversion of carlactonoic acid to orobanchol by cytochrome P450 CYP722C in strigolactone biosynthesis. Sci Adv. 2019 Dec 18;5(12):9067–85.

24. Mizutani M, Ohta D. Diversification of P450 genes during land plant evolution. Annu Rev Plant Biol. 2010 Jun 2;61:291–315.

25. Rupasinghe S, Schuler MA. Homology modeling of plant cytochrome P450s. Vol. 5, Phytochemistry Reviews. Springer; 2006. p. 473–505.

26. Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J. Structure and function of cytochromes P450:a comparative analysis of three crystal structures. Structure. 1995 Jan 1;3(1):41–62.

27. Neuefeind T, Huber R, Reinemer P, Knäblein J, Prade L, Mann K, et al. 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. J Mol Biol. 1997 Dec 12;274(4):577–87.

28. Reinemer P, Prade L, Hof P, Neuefeind T, Huber R, Zettl R, et al. 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. J Mol Biol. 1996;255(2):289–309.

29. Thom R, Cummins I, Dixon DP, Edwards R, Cole DJ, Lapthorn AJ. Structure of a Tau Class Glutathione S-Transferase from Wheat Active in Herbicide Detoxification. Biochemistry. 2002;41:7008–20.

30. Neuefeind T, Huber R, Dasenbrock H, Prade L, Bieseler B. Crystal structrure of herbicide-detoxifying maize glutathione S-transferase-I in complex with lactoylglutathione: Evidence for an induced-fit mechanism. Vol. 274, Journal of Molecular Biology. Academic Press; 1997. p.

446–53.

31. Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. Plant cytochrome P450 plasticity and evolution. Mol Plant. 2021 Aug 2;14(8):1244–65.

32. Nelson DR. Plant cytochrome P450s from moss to poplar. Vol. 5, Phytochemistry Reviews. Springer; 2006. p. 193–204.

33. Nelson DR. Cytochrome P450 diversity in the tree of life. Biochim Biophys Acta - Proteins Proteomics. 2018 Jan 1;1866(1):141–54.

34. Nelson D, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, et al. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics. 1996 Feb;6(1):1–42.

35. Su D, Yang L, Shi X, Ma X, Zhou X, Hedges SB, et al. Large-Scale Phylogenomic Analyses Reveal the Monophyly of Bryophytes and Neoproterozoic Origin of Land Plants. Mol Biol Evol. 2021 Aug 1;38(8):3332.

36. Nie Y, Foster CSP, Zhu T, Yao R, Duchêne DA, Ho SYW, et al. Accounting for Uncertainty in the Evolutionary Timescale of Green Plants Through Clock-Partitioning and Fossil Calibration Strategies. Syst Biol. 2020 Jan 1;69(1):1–16.

37. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, et al. The timescale of early land plant evolution. Proc Natl Acad Sci U S A. 2018 Mar 6;115(10):E2274–83.

38. Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, et al. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 2001 Jan 1;29(1):102–5.

39. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, Mccombie WR, Ouyang S, et al. Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. Rice. 2013 Feb 6;6(1):3–10.

40. Li F-W, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z, et al. Anthoceros genomes

illuminate the origin of land plants and the unique biology of hornworts. Nat Plants. 2020 Mar 13;6(3):259–72.

41. Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, et al. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. Nat Commun. 2014 May 28;5.

42. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 2012 Jan 1;40(D1):D1178–86.

43. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, et al. Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. Nature. 2004 Apr 8;428(6983):653–7.

44. Nelson DR. The cytochrome P450 homepage. Hum Genomics. 2009 Oct 1;4(1):59–65.

45. Paquette SM, Bak S, Feyereisen R. Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. DNA Cell Biol. 2000;19(5):307–17.

46. Ehlting J, Sauveplane V, Olry A, Ginglinger JF, Provart NJ, Werck-Reichhart D. An extensive (co-)expression analysis tool for the cytochrome P450 superfamily in Arabidopsis thaliana. BMC Plant Biol. 2008 Apr 23;8(1):47.

47. Nelson D, Schuler M, Paquette SM, Werck-Reichhart D, Bak S. Comparative Genomics of Rice and Arabidopsis. Analysis of 727 Cytochrome P450 Genes and Pseudogenes from a Monocot and a Dicot. Plant Physiol. 2004;1355:756–72.

48. Rice Cyt P450 Database [Internet]. 2015. Available from: https://ricephylogenomics.ucdavis.edu/p450/genInfo.shtml

49. Wagner U, Edwards R, Dixon DP, Mauch F. Probing the Diversity of the Arabidopsis glutathione S-Transferase Gene Family. Vol. 49, Plant Molecular Biology. 2002.

50. Dixon DP, Edwards R. Glutathione Transferases. Arab B. 2010 Jan;8:e0131.

689 51. Soranzo N, Sari Gorla M, Mizzi L, De Toma G, Frova C. Organisation and structural evolution of
690 the rice glutathione S-transferase gene family. Vol. 271, Molecular Genetics and Genomics.
691 2004. p. 511–21.
692 52. Jain M, Ghanashyam C, Bhattacharjee A. Comprehensive expression analysis suggests
693 overlapping and specific roles of rice glutathione S-transferase genes during development
694 and stress responses. BMC Genomics. 2010 Jan 29;11(1):73.
695 53. Liu Y-J, Han X-M, Ren L-L, Yang H-L, Zeng Q-Y. Functional Divergence of the Glutathione S-
696 Transferase Supergene Family in Physcomitrella patens Reveals Complex Patterns of Large
697 Gene Family Evolution in Land Plants. Plant Physiol. 2013 Feb 1;161(2):773–86.
698 54. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro
699 protein families and domains database: 20 years on. Nucleic Acids Res. 2021 Jan
700 8;49(D1):D344–54.
701 55. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
702 Improvements in Performance and Usability. Mol Biol Evol. 2013;30:772–80.
703 56. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for
704 windows 95/98/NT. Nucleic Acids Symp Ser. 1999;41:95–8.
705 57. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment
706 trimming in large-scale phylogenetic analyses. Bioinforma Appl NOTE. 2009;25(15):1972–3.
707 58. Guindon S, Ois Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms
708 and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of
709 PhyML 3.0. Syst Biol. 2010;59(3):307–21.
710 59. FigTree [Internet]. 2022. Available from: http://tree.bio.ed.ac.uk/software/figtree/
711 60. Inkscape. Inkscape Project [Internet]. 2022. Available from: https://inkscape.org
712 61. Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, et al. Analysis of the genome sequence
713 of the flowering plant *Arabidopsis thaliana*. Nature. 2000 Dec 14;408(6814):796–815.
714 62. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, DePamphilis C, et al. The

selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science (80-). 2011 May 20;332(6032):960–3.

63. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science (80-). 2008 Jan 4;319(5859):64–9.

64. Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, et al. Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. Cell. 2017 Oct 5;171(2):287-304.e15.

65. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science (80-). 2007 Oct 12;318(5848):245–51.

66. Nelson D, Werck-Reichhart D. A P450-centric view of plant evolution. Plant J. 2011 Apr;66(1):194–211.

67. Board PG, Baker RT, Chelvanayagam G, Jermiin LS. Zeta, a novel class of glutathione transferases in a range of species from plants to humans. Biochem J. 1997 Dec 15;328(3):929–35.

68. Koonin E V., Tatusov RL, Altschul SF, Bryant SH, Mushegian AR, Bork P, et al. 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 Sci. 1994 Nov 1;3(11):2045–55.

69. Lallement P-A, Meux E, Gualberto JM, Dumarcay S, Favier F, Didierjean C, et al. Glutathionyl-hydroquinone reductases from poplar are plastidial proteins that deglutathionylate both reduced and oxidized glutathionylated quinones. FEBS Lett. 2015 Jan 2;589(1):37–44.

70. Lister R, Carrie C, Duncan O, Ho LHM, Howell KA, Murcha MW, et al. Functional Definition of Outer Membrane Proteins Involved in Preprotein Import into Mitochondria. Plant Cell. 2007;19:3739–59.

741 71. Yamada T, Komoto J, Watanabe K, Ohmiya Y, Takusagawa F. Crystal structure and possible
742 catalytic mechanism of microsomal prostaglandin E synthase type 2 (mPGES-2). J Mol Biol.
743 2005 May 20;348(5):1163–76.
744 72. Lallement PA, Brouwer B, Keech O, Hecker A, Rouhier N. The still mysterious roles of cysteine-
745 containing glutathione transferases in plants. Vol. 5 AUG, Frontiers in Pharmacology.
746 Frontiers Research Foundation; 2014. p. 192.
747 73. Batard Y, LeRet M, Schalk M, Robineau T, Durst F, Werck-Reichhart D. Molecular cloning and
748 functional expression in yeast of CYP76B1, a xenobiotic-inducible 7-ethoxycoumarin *O*
749 -de-ethylase from *Helianthus tuberosus*. Plant J. 1998 Apr 5;14(1):111–20.
750 74. Gaines TA, Lorentz L, Figge A, Herrmann J, Maiwald F, Ott M-C, et al. RNA-Seq transcriptome
751 analysis to identify genes involved in metabolism-based diclofop resistance in *Lolium rigidum*.
752 Plant J. 2014 Jun;78(5):865–76.
753 75. Hayashi E, Fuzimoto K, Imaishi H. Expression of Arabidopsis thaliana cytochrome P450
754 monooxygenase, CYP71A12, in yeast catalyzes the metabolism of herbicide pyrazoxyfen.
755 Plant Biotechnol. 2007;24(4):393–6.
756 76. Höfer R, Boachon B, Renault H, Gavira C, Miesch L, Iglesias J, et al. Dual function of the
757 cytochrome p450 cyp76 family from arabidopsis thaliana in the metabolism of
758 monoterpenols and phenylurea herbicides. Plant Physiol. 2014 Nov 1;166(3):1149–61.
759 77. Dimaano NG, Yamaguchi T, Fukunishi K, Tominaga T, Iwakami S. Functional characterization
760 of cytochrome P450 CYP81A subfamily to disclose the pattern of cross-resistance in
761 *Echinochloa phyllopogon*. Plant Mol Biol. 2020 Mar 1;102(4–5):403–16.
762 78. Guo F, Iwakami S, Yamaguchi T, Uchino A, Sunohara Y, Matsumoto H. Role of CYP81A
763 cytochrome P450s in clomazone metabolism in *Echinochloa phyllopogon*. Plant Sci. 2019 Jun
764 1;283:321–8.
765 79. Iwakami S, Endo M, Saika H, Okuno J, Nakamura N, Yokoyama M, et al. Cytochrome P450
766 CYP81A12 and CYP81A21 are associated with resistance to two acetolactate synthase

767 inhibitors in *Echinochloa phyllopogon*. Plant Physiol. 2014 Jun 1;165(2):618–29.

768 80. Zhao N, Yan Y, Liu W, Wang J. Cytochrome P450 CYP709C56 metabolizing mesosulfuron-
769 methyl confers herbicide resistance in Alopecurus aequalis. Cell Mol Life Sci. 2022 Apr
770 1;79(4):1–14.

771 81. Han H, Yu Q, Beffa R, González S, Maiwald F, Wang J, et al. Cytochrome P450 CYP81A10v7 in
772 Lolium rigidum confers metabolic resistance to herbicides across at least five modes of
773 action. Plant J. 2021 Jan 27;105(1):79–92.

774 82. Gesell A, Blaukopf M, Madilao L, Yuen MMS, Withers SG, Mattsson J, et al. The gymnosperm
775 cytochrome P450 CYP750B1 catalyzes stereospecific monoterpene hydroxylation of (+)-
776 sabinene in thujone biosynthesis in western redcedar. Plant Physiol. 2015;168(1):94–106.

777 83. Iwakami S, Kamidate Y, Yamaguchi T, Ishizaka M, Endo M, Suda H, et al. CYP81A P450s are
778 involved in concomitant cross-resistance to acetolactate synthase and acetyl-CoA carboxylase
779 herbicides in *Echinochloa phyllopogon*. New Phytol. 2019 Mar 1;221(4):2112–22.

780 84. Imaishi H, Matumoto S. Isolation and functional characterization in yeast of CYP72A18, a rice
781 cytochrome P450 that catalyzes (ω-1)-hydroxylation of the herbicide pelargonic acid. Pestic
782 Biochem Physiol. 2007 May 1;88(1):71–7.

783 85. Saika H, Horita J, Taguchi-Shiobara F, Nonaka S, Nishizawa-Yokoi A, Iwakami S, et al. A novel
784 rice cytochrome P450 gene, CYP72A31, confers tolerance to acetolactate synthase-inhibiting
785 herbicides in rice and Arabidopsis. Plant Physiol. 2014 Nov;166(3):1232–40.

786 86. Zhang L, Lu Q, Chen H, Pan G, Xiao S, Dai Y, et al. Identification of a cytochrome P450
787 hydroxylase, CYP81A6, as the candidate for the bentazon and sulfonylurea herbicide
788 resistance gene, Bel, in rice. Mol Breed. 2006 Dec 27;19(1):59–68.

789 87. Xiang W, Wang X, Ren T. Expression of a wheat cytochrome P450 monooxygenase cDNA in
790 yeast catalyzes the metabolism of sulfonylurea herbicides. Pestic Biochem Physiol.
791 2005;85(1):1–6.

792 88. Li J, Yu H, Zhang F, Lin C, Gao J, Fang J, et al. A Built-In Strategy to Mitigate Transgene

Spreading from Genetically Modified Corn. Alvarez ML, editor. PLoS One. 2013 Dec 6;8(12):e81645.

89. Thyssen GN, Naoumkina M, McCarty JC, Jenkins JN, Florane C, Li P, et al. The P450 gene CYP749A16 is required for tolerance to the sulfonylurea herbicide trifloxysulfuron sodium in cotton (*Gossypium hirsutum* L.). BMC Plant Biol. 2018 Dec 10;18(1):186.

90. Kato S, Yokota Y, Suzuki R, Fujisawa Y, Sayama T, Kaga A, et al. Identification of a cytochrome P450 hydroxylase, CYP81E22, as a causative gene for the high sensitivity of soybean to herbicide bentazon. Theor Appl Genet. 2020 Jul 1;133(7):2105–15.

91. Siminszky B, Corbin FT, Ward ER, Fleischmann TJ, Dewey RE. Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. Proc Natl Acad Sci U S A. 1999 Feb 16;96(4):1750–5.

92. Khanom S, Jang J, Lee OR. Overexpression of ginseng cytochrome P450 CYP736A12 alters plant growth and confers phenylurea herbicide tolerance in Arabidopsis. J Ginseng Res. 2019 Oct 1;43(4):645–53.

93. Robineau T, Batard Y, Nedelkina S, Cabello-Hurtado F, LeRet M, Sorokine O, et al. The chemically inducible plant cytochrome P450 CYP76B1 actively metabolizes phenylureas and other xenobiotics. Plant Physiol. 1998 Nov 1;118(3):1049–56.

94. Pierrel MA, Batard Y, Kazmaier M, Mignotte-Vieux C, Durst F, Werck-Reichhart D. Catalytic Properties of the Plant Cytochrome P450 CYP73 Expressed in Yeast. Substrate Specificity of a Cinnamate Hydroxylase. Eur J Biochem. 1994 Sep 1;224(3):835–44.

95. Yamada T, Kambara Y, Imaishi H, Ohkawa H. Molecular cloning of novel cytochrome P450 species induced by chemical treatments in cultured tobacco cells. Pestic Biochem Physiol. 2000 Sep 1;68(1):11–25.

96. Deridder BP, Dixon DP, Beussman DJ, Edwards R, Goldsbrough PB. Induction of Glutathione S-Transferases in Arabidopsis by Herbicide Safeners. Plant Physiol. 2002;130:1497–505.

97. Cummins I, Cole DJ, Edwards R. A role for glutathione transferases functioning as glutathione

peroxidases in resistance to multiple herbicides in blackgrass. Plant J. 1999 May 1;18(3):285–92.

98. Cummins I, Wortley DJ, Sabbadin F, He Z, Coxon CR, Straker HE, et al. Key role for a glutathione transferase in multiple-herbicide resistance in grass weeds. Proc Natl Acad Sci. 2013;110(15).

99. Cho HY, Kong KH. Molecular cloning, expression, and characterization of a phi-type glutathione S-transferase from *Oryza sativa*. Pestic Biochem Physiol. 2005 Sep 1;83(1):29–36.

100. Karavangeli M, Labrou NE, Clonis YD, Tsaftaris A. Development of transgenic tobacco plants overexpressing maize glutathione S-transferase I for chloroacetanilide herbicides phytoremediation. Biomol Eng. 2005 Oct 1;22(4):121–8.

101. Mozer TJ, Tiemeier DC, Jaworski EG. Purification and characterization of corn glutathione S-transferase. Biochemistry. 1983;22(5):1068–72.

102. Shah DM, Hironaka CM, Wiegand RC, Harding EI, Krivi GG, Tiemeier DC. Structural analysis of a maize gene coding for glutathione-S-transferase involved in herbicide detoxification. Plant Mol Biol. 1986;6:203–11.

103. Dixon D, Cole DJ, Edwards R. Characterisation of Multiple Glutathione Transferases Containing the GST I Subunit with Activities toward Herbicide Substrates in Maize (*Zea mays*). Pestic Sci. 1997 May 1;50(1):72–82.

104. Fuerst EP, Irzyk GP, Miller KD, Townson J, Edwards R. Partial characterization of glutathione S-transferase isozymes induced by the herbicide safener benoxacor in maize. Plant Physiol. 1993 Jul 1;102(3):795–802.

105. Milligan AS, Daly A, Parry MAJ, Lazzeri PA, Jepson I. The expression of a maize glutathione S-transferase gene in transgenic wheat confers herbicide tolerance, both in planta and in vitro. Mol Breed. 2001;7(4):301–15.

106. Sommer A, Böger P. Characterization of recombinant corn glutathione S-transferase isoforms I, II, III, and IV. Pestic Biochem Physiol. 1999 Mar 1;63(3):127–38.

845 107. Irzyk GP, Fuerst EP. Purification and characterization of a glutathione S-transferase from
846 benoxacor-treated maize (*Zea mays*). Plant Physiol. 1993 Jul 1;102(3):803–10.
847 108. David P. Dixon, Cole DJ, Edwards R. Purification, regulation and cloning of a glutathione
848 transferase (GST) from maize resembling the auxin-inducible type-III GSTs. Vol. 36, Plant
849 Molecular Biology. 1998.
850 109. Dixon DP, Cole DJ, Edwards R. Dimerisation of maize glutathione transferases in recombinant
851 bacteria. Plant Mol Biol. 1999 Aug;40(6):997–1008.
852 110. Hu T, Qv X, Xiao G, Huang X. Enhanced tolerance to herbicide of rice plants by over-
853 expression of a glutathione S-transferase. Mol Breed. 2009 Nov 20;24(4):409–18.
854 111. Jo HJ, Lee JJ, Kong KH. A plant-specific tau class glutathione S-transferase from *Oryza sativa*
855 with very high activity against 1-chloro-2,4-dinitrobenzene and chloroacetanilide herbicides.
856 Pestic Biochem Physiol. 2011 Nov 1;101(3):265–9.
857 112. Gronwald JW, Plaisance KL. Isolation and characterization of glutathione S-transferase
858 isozymes from sorghum. Plant Physiol. 1998 Jul 1;117(3):877–92.
859 113. Cummins I, Cole DJ, Edwards R. Purification of Multiple Glutathione Transferases Involved in
860 Herbicide Detoxification from Wheat (*Triticum aestivum* L.) Treated with the Safener
861 Fenchlorazole-ethyl. Pestic Biochem Physiol. 1997;59:35–49.
862 114. Cummins I, O'Hagan D, Jablonkai I, Cole DJ, Hehn A, Werck-Reichhart D, et al. Cloning,
863 characterization and regulation of a family of phi class glutathione transferases from wheat.
864 Plant Mol Biol. 2003 Jun;52(3):591–603.
865 115. Pascal S, Scalla R. Purification and characterization of a safener-induced glutathione S-
866 transferase from wheat (*Triticum aestivum*). Physiol Plant. 1999 May 1;106(1):17–27.
867 116. Benekos K, Kissoudis C, Nianiou-Obeidat I, Labrou N, Madesis P, Kalamaki M, et al.
868 Overexpression of a specific soybean GmGSTU4 isoenzyme improves diphenyl ether and
869 chloroacetanilide herbicide tolerance of transgenic tobacco plants. J Biotechnol. 2010 Oct
870 1;150(1):195–201.

871 117. McGonigle B, Keeler SJ, Lau SMC, Koeppe MK, O'Keefe DP. A genomics approach to the
872 comprehensive analysis of the glutathione S-transferase gene family in soybean and maize.
873 Plant Physiol. 2000 Nov 1;124(3):1105–20.
874 118. Skipsey M, Cummins I, Andrews CJ, Jepson I, Edwards R. Manipulation of plant tolerance to
875 herbicides through co-ordinated metabolic engineering of a detoxifying glutathione
876 transferase and thiol cosubstrate. Plant Biotechnol J. 2005 Jul 1;3(4):409–20.
877 119. Gould SJ, Vrba ES. Exaptation-a missing term in the science of form. Paleobiology.
878 1982;8(1):4–15.
879 120. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. Vol. 1, Genome
880 Biology. BioMed Central; 2000. p. 1–9.
881 121. Pégeot H, Koh CS, Petre B, Mathiot S, Duplessis S, Hecker A, et al. The poplar Phi class
882 glutathione transferase: expression, activity and structure of GSTF1. Front Plant Sci. 2014 Dec
883 23;5:712.
884 122. Lee JJ, Jo HJ, Kong KH. A plant-specific tau class glutathione s-transferase from *Oryza sativa*
885 having significant detoxification activity towards chloroacetanilide herbicides. Bull Korean
886 Chem Soc. 2011 Oct 20;32(10):3756–9.
887 123. Gardin JAC, Gouzy J, Carrère S, Délye C. ALOMYbase, a resource to investigate non-target-
888 site-based resistance to herbicides inhibiting acetolactate-synthase (ALS) in the major grass
889 weed *Alopecurus myosuroides* (black-grass). BMC Genomics. 2015;16(590).
890 124. Piasecki C, Yang Y, Benemann DP, Kremer FS, Galli V, Millwood RJ, et al. Transcriptomic
891 analysis identifies new non-target site glyphosate-resistance genes in *Conyza bonariensis*.
892 Plants. 2019 Jun 7;8(6):157.
893 125. Liu XM, Xu X, Li BH, Yao XX, Zhang HH, Wang GQ, et al. Genomic and transcriptomic insights
894 into cytochrome P450 monooxygenase genes involved in nicosulfuron tolerance in maize (*Zea*
895 *mays* L.). J Integr Agric. 2018 Aug 1;17(8):1790–9.
896 126. Bai S, Zhao Y, Zhou Y, Wang M, Li Y, Luo X, et al. Identification and expression of main genes

involved in non-target site resistance mechanisms to fenoxaprop-p-ethyl in *Beckmannia syzigachne*. Pest Manag Sci. 2020 Aug 12;76(8):2619–26.

127. Iwakami S, Uchino A, Kataoka Y, Shibaike H, Watanabe H, Inamura T. Cytochrome P450 genes induced by bispyribac-sodium treatment in a multiple-herbicide-resistant biotype of *Echinochloa phyllopogon*. Pest Manag Sci. 2014 Apr 1;70(4):549–58.

128. Bai S, Liu W, Wang H, Zhao N, Jia S, Zou N, et al. Enhanced herbicide metabolism and metabolic resistance genes identified in tribenuron-methyl resistant *Myosoton aquaticum* L. J Agric Food Chem. 2018;66:9850–7.

129. Liu W, Bai S, Zhao N, Jia S, Li W, Zhang L, et al. Non-target site-based resistance to tribenuron-methyl and essential involved genes in *Myosoton aquaticum* (L.). BMC Plant Biol. 2018 Oct 11;18(1):225.

130. Franco-Ortega S, Goldberg-Cavalleri A, Walker A, Brazier-Hicks M, Onkokesung N, Edwards R. Non-target site herbicide resistance is conferred by two distinct mechanisms in black-grass (*Alopecurus myosuroides*). Front Plant Sci. 2021 Mar 3;12:636652.

131. Duhoux A, Carrère S, Duhoux A, Délye C. Transcriptional markers enable identification of rye-grass (*Lolium* sp.) plants with non-target-site-based resistance to herbicides inhibiting acetolactate-synthase. Plant Sci. 2017 Apr 1;257:22–36.

132. Leslie T, Baucom RS. De novo assembly and annotation of the transcriptome of the agricultural weed *Ipomoea purpurea* uncovers gene expression changes associated with herbicide resistance. G3 Genes, Genomes, Genet. 2014 Oct 1;4(10):2035–47.

133. Salas-Perez RA, Saski CA, Noorai RE, Srivastava SK, Lawton-Rauh AL, Nichols RL, et al. RNA-Seq transcriptome analysis of *Amaranthus palmeri* with differential tolerance to glufosinate herbicide. PLoS One. 2018;13:1–33.

134. Matzrafi M, Shaar-Moshe L, Rubin B, Peleg Z. Unraveling the Transcriptional Basis of Temperature-Dependent Pinoxaden Resistance in *Brachypodium hybridum*. Front Plant Sci. 2017 Jun 21;8:1064.

923 135. Pan L, Gao H, Xia W, Zhang T, Dong L. Establishing a herbicide-metabolizing enzyme library in
924 *Beckmannia syzigachne* to identify genes associated with metabolic resistance. J Exp Bot.
925 2016 Mar 1;67(6):1745–57.
926 136. Wang J, Chen J, Li X, Cui H. RNA-Seq transcriptome analysis to identify candidate genes
927 involved in non-target site-based mesosulfuron-methyl resistance in *Beckmannia syzigachne*.
928 Pestic Biochem Physiol. 2021 Jan 1;171:104738.
929 137. Zhao N, Li W, Bai S, Guo W, Yuan G, Wang F, et al. Transcriptome profiling to identify genes
930 involved in mesosulfuron-methyl resistance in *Alopecurus aequalis*. Front Plant Sci. 2017 Aug
931 9;8:1391.
932 138. Yang Q, Deng W, Li X, Yu Q, Bai L, Zheng M. Target-site and non-target-site based resistance
933 to the herbicide tribenuron-methyl in flixweed (*Descurainia sophia* L.). BMC Genomics.
934 2016;17(1):17:551.
935 139. Taylor VL, Cummins I, Brazier-Hicks M, Edwards R. Protective responses induced by herbicide
936 safeners in wheat. Environ Exp Bot. 2013 Apr;88:93–9.
937 140. Li D, Gao Q, Xu L, Pang S, Liu Z, Wang C, et al. Characterization of glutathione S-transferases in
938 the detoxification of metolachlor in two maize cultivars of differing herbicide tolerance.
939 Pestic Biochem Physiol. 2016;143:265–71.
940 141. Tétard-Jones C, Sabbadin F, Moss S, Hull R, Neve P, Edwards R. Changes in the proteome of
941 the problem weed blackgrass correlating with multiple-herbicide resistance. Plant J. 2018
942 May 1;94(4):709–20.
943 142. Wu J, Cramer CL, Hatzios KK. Characterization of two cDNAs encoding glutathione *S*-
944 transferases in rice and induction of their transcripts by the herbicide safener fenclorim.
945 Physiol Plant. 2002 Jan 4;105(1):102–8.
946 143. Li D, Xu L, Pang S, Liu Z, Wang K, Wang C. Variable levels of Glutathione S-Transferases are
947 responsible for the differential tolerance to metolachlor between Maize (*Zea mays*) shoots &
948 roots. J Agric Food Chem. 2017 Jan 11;65(1):39–44.

144. Li G, Wu SG, Yu RX, Cang T, Chen LP, Zhao XP, et al. Identification and expression pattern of a glutathione S-transferase in *Echinochloa crus-galli*. Zwerger P, editor. Weed Res. 2013 Oct 1;53(5):314–21.

Supporting information

S1 Fig. Overview of cytochrome P450 protein features in plants. 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.

S2 Fig. Plant CYP and GST phylogenetic analysis using automatic and manual trimming approaches. Unrooted cladograms of maximum likelihood (ML) analysis conducted by PHyML 3.0 [58] 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. CYP (A) and GST (B) 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 or GST classes. aLRT Support values for some of the clades are shown for comparison.

S1 Table. Cytochrome P450 clans and gene numbers in green plants and red algae. Numbers of CYP proteins in each clan, excluding pseudogenes. *At, Arabidopsis thaliana; Os, Oryza sativa; Sm, Selaginella moellendorffii; Aa, Anthoceros agrestis; Pp, Physcomitrium patens; Mp, Marchantia*

974 *polymorpha; Kn, Klebsormidium nitens; Cr, Chlamydomonas reinhardtii; Cm, Cyanidioschyzon*
975 *merolae.*
976
977 **S2 Table. Glutathione-S-transferase classes and gene numbers in green plants and red algae.**
978 Numbers of GST proteins in each clan, excluding pseudogenes. *At, Arabidopsis thaliana; Os, Oryza*
979 *sativa; Sm, Selaginella moellendorffii; Aa, Anthoceros agrestis; Pp, Physcomitrium patens; Mp,*
980 *Marchantia polymorpha; Kn, Klebsormidium nitens; Cr, Chlamydomonas reinhardtii; Cm,*
981 *Cyanidioschyzon merolae.*
982
983 **S3 Table. Plant CYPs that metabolise or confer resistance to herbicides are found within clans 71**
984 **and 72.** Table adapted from [10].
985
986 **S4 Table. Plant GST proteins that conjugate or confer resistance to herbicides are members of**
987 **classes Phi, Tau, and Lambda.** Maize and *A. thaliana* genes were renamed according to current
988 nomenclature [21].
989
990 **S5 Table. Candidate NTSR CYPs are found within clans 71, 72, 85 and 86.**
991
992 **S6 Table. Candidate NTSR genes belong to several GST classes.** The organophosphorus class refers
993 to the herbicide glyphosate.
994
995 **S7 Table. Number of GST proteins identified from classes 2N, Kappa, and MAPEG in green plants**
996 **and red algae.** Sequences from these classes were not included in the phylogenetic analysis due to
997 lack of classical N-terminal and C-terminal GST domains. 2N GST sequences have two N-terminal
998 domains and lack a C-terminal domain. Kappa GST proteins lack the two classical GST domains and

instead have one single thioredoxin-like kappa GST domain (InterPro domain IPR014440). MAPEG GST proteins lack both C and N-terminal GST domains and have instead a single 'MAPEG' domain.

S3 Fig. 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. Sequences were aligned in MAFFT using the FFT-NS-i algorithm. The locations of the substrate recognition sites are based on those identified in Arabidopsis CYPs in [25]. The absolutely conserved cysteine that binds the heme within the heme-binding domain is marked with an asterisk.

S4 Fig. Amino acid alignment of representative plant GST proteins showing the location of conserved GST domains. Sequences were aligned in MAFFT using the FFT-NS-I algorithm. Four representative sequences from different species are shown for each GST class. The location of the putative catalytic residue is indicated with an asterisk. Predicted GSH-binding sites (G-sites) based on the crystal structure of TaGSTU4 [29] are indicated in solid pink, and G-sites based on the crystal structure of PtGSTF1 [121] are indicated by pink boxes. Predicted substrate-binding sites (H-sites) based on the crystal structure of TaGSTU4 [29] are indicated in solid green, and H-sites based on the crystal structure of PtGSTF1 [121] are indicated by green boxes. Residues conserved in at least 80% of samples are indicated by black 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.

S1 Text. Untrimmed alignment of all CYP sequences used in the phylogenetic analysis.

S2 Text. Manually trimmed alignment of all CYP sequences used in the phylogenetic analysis.

S3 Text. Trimmed alignment of all CYP sequences used in the phylogenetic analysis using the trimAl -gappyout automated setting.

S4 Text. Trimmed alignment of all CYP sequences used in the phylogenetic analysis using the trimAl -strict automated setting.

S5 Text. Trimmed alignment of all CYP sequences used in the phylogenetic analysis using the trimAl -strictplus automated setting.

S6 Text. Untrimmed alignment of all GST sequences used in the phylogenetic analysis.

S7 Text. Manually trimmed alignment of all GST sequences used in the phylogenetic analysis.

S8 Text. Trimmed alignment of all GST sequences used in the phylogenetic analysis using the trimAl -gappyout automated setting.

S9 Text. Trimmed alignment of all GST sequences used in the phylogenetic analysis using the trimAl -strict automated setting.

S10 Text. Trimmed alignment of all GST sequences used in the phylogenetic analysis using the trimAl -strictplus automated setting.

A

clan 711
clan 727
Cm1
clan 746
clan 51
clan 72
clan 55
clan 737
clan 741
clan 97
clan 85
clan 747
clan 86
clan 710
Cm2
clan 74
clan 71

B

EF1Bγ
URE2p
GSTT
Alc
Alb
GSTF
Ala
TCQHD
Cr1
GHR
mPGES2
GSTI
Kn1
GSTL
DHAR
GSTM
GSTH
GSTZ
GSTU

Dicots
Monocots
Lycophytes
Mosses
Hornworts
Liverworts
Streptophytes
Chlorophytes
Rhodophytes