Co-opted Genes of Algal Origin Protect *C. elegans* against Cyanogenic Toxins

Bingying Wang[1], Taruna Pandey[1], Yong Long[2], Sofia E. Delgado-Rodriguez[3], Matthew D. Daugherty[3, *], Dengke K. Ma[1, *]

[1]Cardiovascular Research Institute and Department of Physiology, University of California San Francisco, San Francisco, USA

[2]State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China.

[3]Department of Molecular Biology, University of California San Diego, San Diego, United States.

* Correspondence should be addressed to Dengke K. Ma, Ph.D. (dengke.ma@ucsf.edu) and Matthew D. Daugherty, Ph.D. (MDDaugherty@UCSD.edu)

Abstract

Amygdalin is a cyanogenic glycoside widely used by many plants in herbivore defense. Poisonous to most animals, amygdalin-derived cyanide is detoxified by potent enzymes commonly found in bacteria and plants but not most animals. Here we show that the nematode *C. elegans* can detoxify amygdalin by a genetic pathway comprising *cysl-1*, *egl-9*, *hif-1* and *cysl-2*. Essential for amygdalin resistance, *cysl-1* encodes a protein similar to cysteine synthetic enzymes in bacteria and plants, but functionally co-opted in *C. elegans*. We identify exclusively HIF-activating *egl-9* mutations in a *cysl-1* suppressor screen and show that *cysl-1* confers amygdalin resistance by regulating HIF-1-dependent *cysl-2* transcription to protect against amygdalin toxicity. Phylogenetic analysis suggests *cysl-1* and *cysl-2* were acquired from green algae by horizontal gene transfer (HGT) and functionally co-opted by *C. elegans* in protection against amygdalin. Unlike functional HGTs from bacteria to eukaryotes that have been widely observed and characterized, our studies reveal previously unknown HGT from algae to nematodes, supporting the emerging theme that HGT-mediated evolutionary changes can facilitate host survival and adaptation to adverse environment stresses and biogenic toxins.

Introduction

In nature, many species of plants have evolved the ability to biosynthesize a group of chemically related molecules called cyanogenic glycosides as chemical toxins used in herbivore defense[1–3]. Amygdalin is a natural cyanogenic glycoside enriched in tissues of many edible plants, including seeds of stone fruits such as cherry (*Prunus avium*),

peach (*Prunus persica*), apricot (*Prunus armeniaca*), and apples (*Malus domestica*). These plant species synthesize amygdalin in defense against herbivore animals as amygdalin generates poisonous cyanide upon plant tissue destruction[1–4]. After consumption by herbivores, amygdalin is hydrolyzed by β-glucosidases and α-hydroxynitrilases, releasing hydrogen cyanide that can be highly toxic and even lethal to herbivore animals. To tolerate amygdalin, plants also evolved a family of cysteine synthase-like enzyme, β-cyanoalanine synthases, to directly detoxify cyanides[4–6].

Caenorhabditis elegans (C. elegans) is a nematode that lives with rotten fruits and bacteria in their natural ecological niches[7–9], thus may frequently encounter amygdalin. Although this commonly used genetic model organism rarely encounters amygdalin under laboratory conditions, loss-of-function mutations in the gene *egl-9* have been shown to confer resistance to high doses of hydrogen cyanides generated by a pathogenic bacterial strain of *Pseudomonas aeruginosa*[10]. EGL-9 is a prolyl hydroxylase and negative regulator of HIF-1 (hypoxia inducible factor), a transcription factor that directly activates expression of many genes involved in hypoxic adaptation, metabolic reprogramming and stress responses[11–13]. As a transcriptional target of HIF-1, CYSL-2 is a protein strongly up-regulated in *egl-9* loss-of-function mutants and can detoxify cyanides *in vitro* as a cyanoalanine synthase[14]. CYSL-1 is a paralogue of CYSL-2, both of which share sequence similarities with a family of β-cyanoalanine synthases found in bacteria and plants. We previously found that CYSL-1 regulates *cysl-2* transcription by binding to EGL-9 in regulation of *cysl-2* expression[15]. Whether this pathway is involved in response to cyanogenic amygdalin has not been characterized. While EGL-9 and HIF-1 are evolutionarily conserved in metazoans, most animals do not have proteins

homologous to CYSL-1 and CYSL-2, except certain herbivorous arthropods[16]. In addition, co-opted detoxification genes of plant origin in whiteflies can neutralize plant-derived phenolic glycoside toxins[17,18]. However, the evolutionary origin, history of domestication and co-option of *cysl* genes in *C. elegans* remain hitherto unknown.

By screening a natural product library, we identified amygdalin as a potent activator of *C. elegans cysl-2*. We showed that the CYSL-1/EGL-9/HIF-1/CYSL-2 pathway plays a crucial role in conferring resistance to amygdalin. The intriguing acquisition of amygdalin resistance in *C. elegans* through two genes (*cysl-1* and *cysl-2*) not commonly found in other animals into the conserved EGL-9/HIF-1 genetic pathway led us to further explore the evolutionary origin and path of *cysl-1* and *cysl-2* based on phylogenetic analysis. Our combined mechanistic and phylogenetic studies indicate that CYSL-1 and CYSL-2 may have originated from green algae by a horizontal gene transfer event in ancient nematodes and evolved by gene co-option into separate roles in a cellular circuit to respond to and detoxify cyanogenic cyanides, including amygdalin from plants today.

Results

We previously discovered several components of the HIF-1 pathway acting upstream of the oxygen-sensing HIF hydroxylase EGL-9 in *C. elegans*[15]. To identify small-molecule natural products that may modulate HIF-1 activity independently of oxygen-sensing in *C. elegans*, we conducted a natural product library screen for compounds that can activate a HIF-1-dependent transgenic reporter strain *cysl-2*p::GFP in liquid cultures under normoxia (Fig. 1a). Among the identified compounds, a glycoside called

amygdalin emerged as a potent activator of *cysl-2*p::GFP. We validated the finding using amygdalin from new preparations of different sources (vendor VWR or Sigma-Aldrich) to treat *C. elegans* on solid NGM media and confirmed that amygdalin can robustly activate *cysl-2*p::GFP under normoxic conditions, without affecting the transgenic co-injection marker *myo-2*p::mCherry (Fig. 1b). Activation of *cysl-2*p::GFP by amygdalin is dose-dependent and reaches peak level at approximately 48 hrs post amygdalin treatment (Fig. 1c, d).

We performed RNAseq to identify genome-wide transcriptomic changes induced by amygdalin. We treated wild-type mixed-stage *C. elegans* hermaphrodites cultivated at 20 ^{o}C with 2 mg/ml amygdalin for 48 hrs. After differential expression analyses of triplicate samples, we identified many genes that are significantly (adjusted p value < 0.05) up- or down-regulated by amygdalin treatment (Fig. 1e). As expected, *cysl-2* is among the most highly up-regulated genes ($\log_2$fold change = 2.62, adjusted p value = 3.11E-06), while expression of its upstream regulators including *hif-1*, *vhl-1* or *egl-9* remains largely unchanged (Supplementary Fig. 1 and table S1). Gene Ontology analysis (Wormbase) indicates that amygdalin also up-regulated several genes encoding glutathione S-transferases involved in the xenobiotic detoxification pathway (Supplementary Fig. 1 and table S1).

We next examined how amygdalin activates *cysl-2*p::GFP. HIF-1 and CYSL-1 are two previously established positive regulators of *cysl-2p*::GFP[15,19–21]. We found that loss-of-function mutations of *hif-1* or *cysl-1* abolished *cysl-2*p::GFP induction by low-dose amygdalin (Fig. 2a). Amygdalin is a glycoside that can be metabolized sequentially to prunasin, glucose, benzaldehyde and hydrogen cyanide (Fig. 2b)[1,4,22]. We treated *cysl-*

*2*p::GFP animals with each of these amygdalin-derived molecules and found prunasin and cyanide (as in potassium cyanide) activated *cysl-2*p::GFP as amygdalin did, but others had no effect (Fig. 2c). CYSL-1 and CYSL-2 are cysteine synthase-like proteins and have been implicated in hydrogen sulfide and cyanide detoxification[14,15,20,23,24]. While CYSL-2 can enzymatically convert cyanide to nontoxic β-cyanoalanine[14], the essential role of CYSL-1 in *cysl-2*p::GFP activation by amygdalin suggests CYSL-1 acts as a transcriptional regulator rather than a cysteine synthase or an enzyme to detoxify cyanide[15]. Taking advantage of the *cysl-2*p::GFP phenotype and lethal effects of amygdalin in *cysl-1* mutants (see below), we performed a *cysl-1* suppressor screen and identified exclusively three loss-of-function mutations in *egl-9* that led to constitutive *cysl-2*p::GFP expression without amygdalin treatment (Fig. 2d). In further genetic epistasis analysis, loss of *hif-1* suppressed not only *cysl-2*p::GFP induction by amygdalin but also constitutive *cysl-2*p::GFP expression in *egl-9*; *cysl-1* mutants (Fig. 2E). These data suggest that amygdalin activates *cysl-2*p::GFP through a dis-inhibitory regulatory pathway from cyanide, CYSL-1, EGL-9 to HIF-1 (Fig. 2f).

To assess the functional importance of the CYSL-1/EGL-9/HIF-1/CYSL-2 pathway, we examined the time-dependent survival response to high-dose amygdalin (10 mg/ml on NGM; effective concentrations *in vivo* are likely much lower due to cuticle and intestinal absorption barriers) in animals with individual or combinatorial pathway gene mutations. Wild-type animals exhibited no apparent decrease in survival upon amygdalin over three days of treatment (Fig. 3a). By contrast, loss-of-function *cysl-1* mutants displayed striking sensitivity to amygdalin, with nearly complete population death at day 3 of amygdalin treatment (Fig. 3b). Loss-of-function of *cysl-2* also caused similar time-

dependent sensitivity to amygdalin (Fig. 3c). While loss of *egl-9* fully suppressed vulnerability of *cysl-1* mutants to amygdalin (Fig. 3d), loss of either *hif-1* or *cysl-2* rendered *egl-9* mutants sensitive to amygdalin (Fig. 3e and f). These data support central roles of the CYSL-1/EGL-9/HIF-1/CYSL-2 pathway in the transcriptional response and conferring resistance to amygdalin in *C. elegans*.

To investigate the evolutionary origin CYSL-1/2 in *C. elegans*, we searched for homologs of CYSL-1 and CYSL-2 proteins across all domains of life. We identified related proteins with >50% sequence identity in nematodes, green algae, land plants, and bacteria. Outside of nematode proteins, the most closely related sequences, and the only ones with sequence identities >60% were from the *Chlorophyta* clade of green algae (Supplementary Fig. 2a and Fig. 4a).

To determine the phylogenetic origin of CYSL proteins in nematodes, we performed maximum likelihood (ML) based phylogenetic analyses on CYSL homologs from a broad range of eukaryotic and prokaryotic species (Supplementary Data Files 1, 2 and Fig. 4A). Consistent with the high sequence similarity to proteins in green algae, we found strong support for nematode proteins nesting within the chlorophyte clade of green algae (Fig. 4b). Proteins from streptophytes (*Streptophyta*), which include some green algae and all land plants, are found in a distinct region of the phylogenetic tree, as are proteins from single-cell eukaryotes (*e.g.* groups including *SAR*, *Excavata*, and *Rhodophyta*) and bacteria. These inferences were robust to the ML software we used for analysis, as well as the substitution models we employed (Fig. 4b). Importantly, cysteine synthase enzymes found in some arthropods, such as mites, which are the result of a horizontal gene transfer (HGT) event from bacteria show less than 40%

sequence identity, consistent with their separate evolutionary origin[16] (Supplementary Fig. 2b). Taken together, these data support a single HGT event from a green algal species in the *Chlorophyta* that gave rise to nematode CYSL.

We further wished to analyze the distribution of CYSL proteins in nematodes to determine at what point CYSL was acquired by nematodes and at what point gene duplication occurred (Fig. 4c). We found CYSL proteins only in *Chromadoria* clade of nematodes and not in the well sampled organisms in the *Trichinellida* (*e.g. Trichinella* species). Although we cannot rule out that it was lost in the *Trichinellida*, the most parsimonious explanation for this distribution is that CYSL was acquired by an ancestral *Chromodoria* nematode after the divergence from other clades of nematodes. Within *Chromodoria*, we only detect a single CYSL homolog in the most basally branching *Spirurina* nematodes[25] (*e.g. Toxocara canis*) whereas we see two or more copies of CYSL homologs in *Tylenchina* (*e.g. Bursaphelenchus* species) and *Rhabditina* (*e.g. Caenorhabditis* species) lineages of nematodes. These data, coupled with the branching of CYSL-1 and CYSL-2 paralogs at the base of the nematode CYSL protein phylogeny (Fig. 4c), suggest that CYSL duplicated soon after acquisition in nematodes to CYSL-1 like proteins and CYSL-2 like proteins. Based on our findings of the non-overlapping functional importance of CYSL-1 and CYSL-2 in *C. elegans* for cyanide detoxification, these phylogenetic findings suggest that this network of CYSL-1 and CYSL-2 evolved soon after the acquisition of these enzymes by nematodes.

Discussion

In this study, we establish novel physiological roles of the CYSL-1/EGL-9/HIF-1/CYSL-2 pathway in detoxification of amygdalin in *C. elegans*. Our data indicate that CYSL-1 acts upstream of the evolutionarily conserved HIF pathway as a key regulator of *cysl-2* induction upon amygdalin exposure, resulting in detoxification of cyanide that derives from the natural diet of *C. elegans* and other nematodes. How amygdalin or cyanide is sensed by cells in *C. elegans* and regulates CYSL-1 and EGL-9 awaits further studies. Plausible mechanisms include mitochondrial inhibition, posttranslational modification by S-cyanylation, and allosteric modulation of CYSL-1 and EGL-9 interaction[15,26,27].

By phylogenetic analysis, we obtained evidence pointing to an intriguing evolutionary origin of *cysl-1/cysl-2* in *C. elegans*. Maximum likelihood estimation based on protein sequence similarity supports a single HGT event from a green algal species in the *Chlorophyta* that gave rise to nematode *cysl* genes. Our analysis further suggests that a *cysl* gene orthologue was acquired by ancestral *Chromodoria* nematodes after the divergence from other clades of nematodes, and duplicated soon after their acquisition to give rise to *cysl* gene paralogs. In *C. elegans*, *cysl-1* and *cysl-2* have further evolved separate functions from ancestral cysteine synthase-like activity into division of labor. While CYSL-2 retains cyanoalanine synthase enzymatic activity to detoxify cyanide, CYSL-1 is co-opted to regulate *cysl-2* expression upon amygdalin/cyanide exposure in a cellular circuit that detects and defends against amygdalin/cyanide. *Chromodoria* nematodes diverged from other clades of nematodes more than 250 million year ago, around which drastic geological and ecological changes (e.g. those leading to Permian-Triassic mass extinction) may have exerted selection pressures on nematodes surviving in environments rife with sulfide/cyanide toxicity and ancient algal species[28–32].

Following acquisition of *cysl* genes, nematodes may have retained and co-opted *cysl* gene functions in adaptation to new environments, including living as plant parasites or in rotten fruits associated with cyanide presence. While precise mechanisms of these evolutionary changes particularly HGT remain unclear, the evolutionary pathway leading to distinct cellular roles of *cysl-1* and *cysl-2* in *C. elegans* today exemplifies how organisms can acquire novel biological traits to survive and reproduce by HGT and gene co-option[32–37]. Unlike functional HGTs from bacteria to eukaryotes that have been widely observed and characterized[38–40], our mechanistic and phylogenetic analyses reveal previously unknown HGT from algae to nematodes, supporting the emerging principle that HGTs can lead to evolutionary changes and novel traits to facilitate host protection and adaptation to abiotic stress, chemical toxins and extreme environments.

Figures and Figure legends

Figure 1

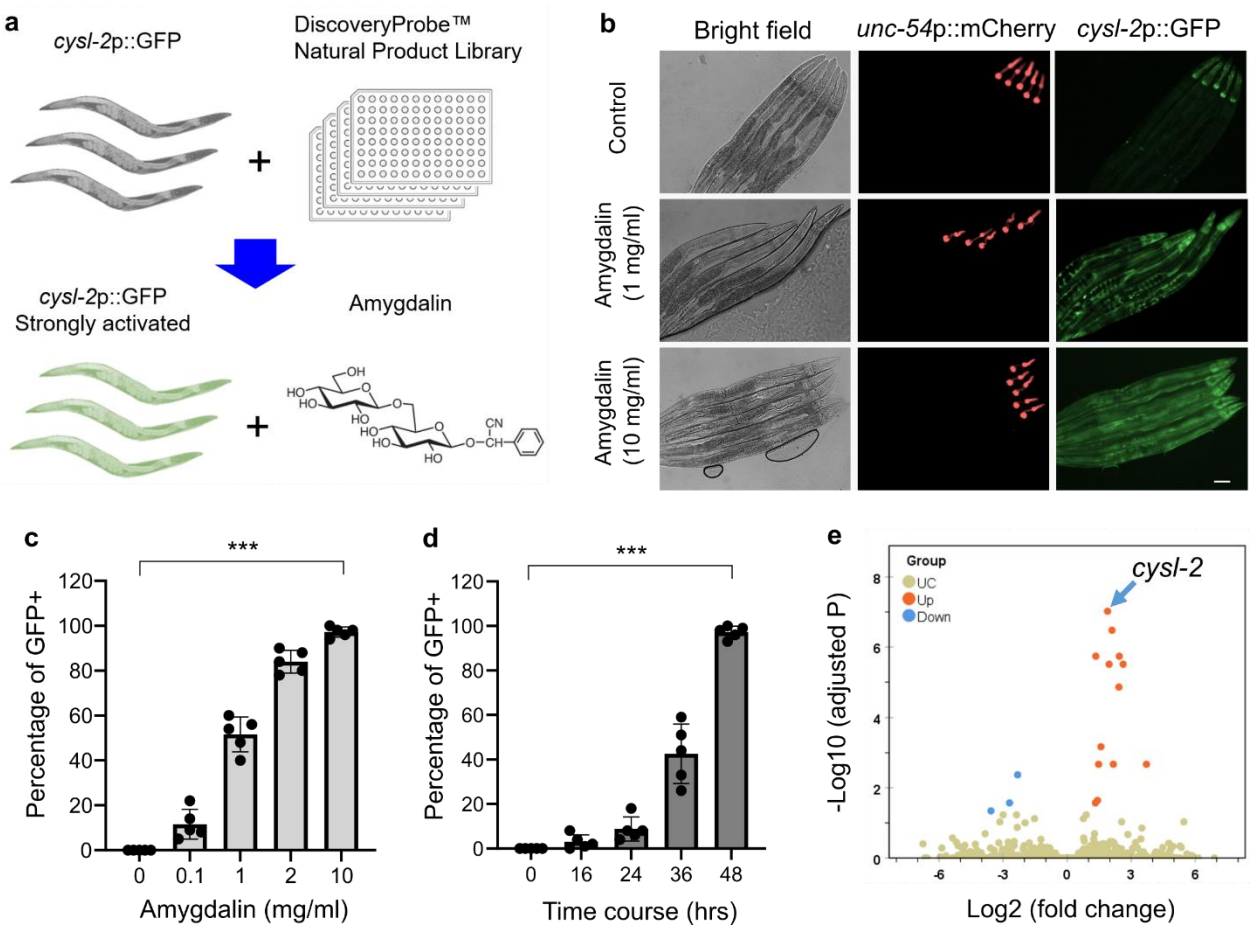
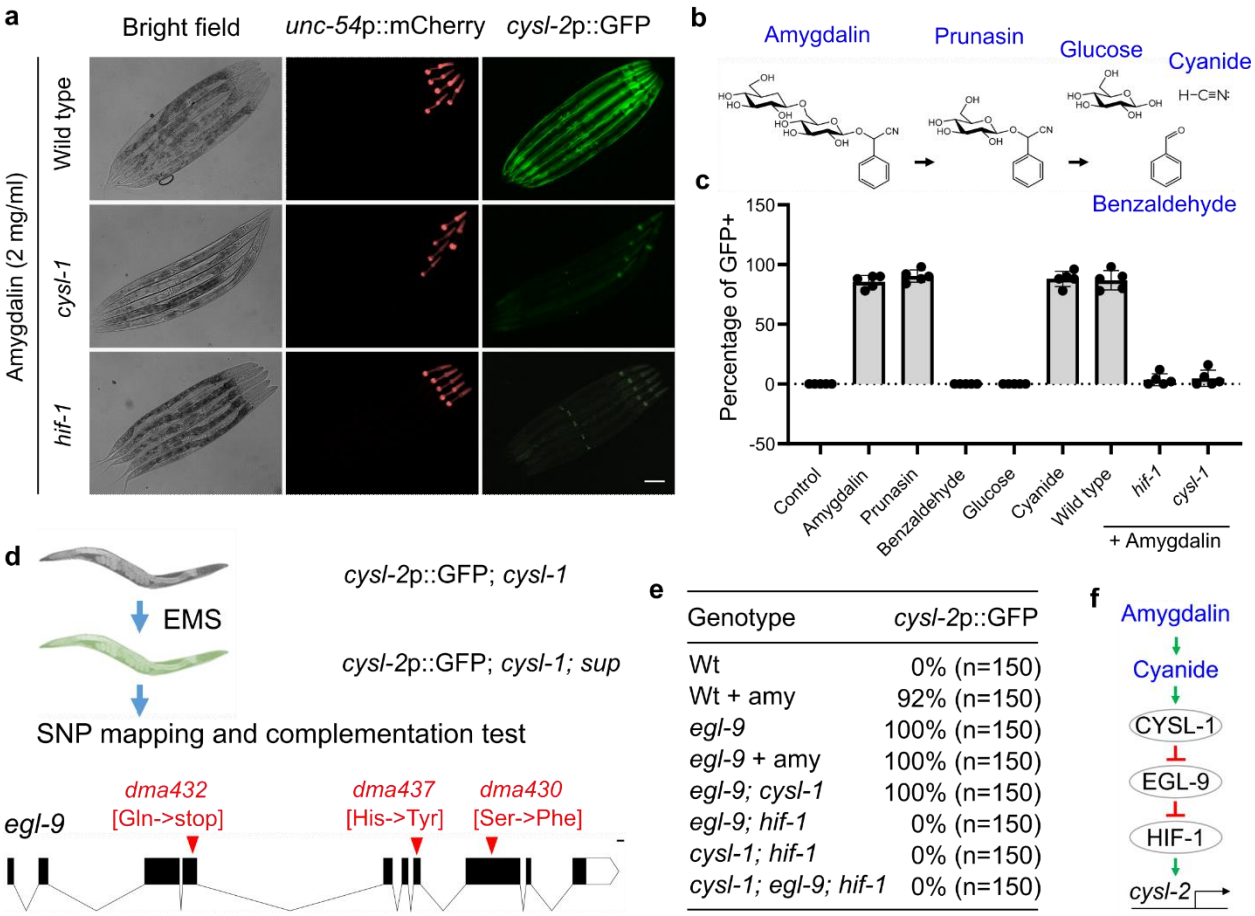


Fig. 1. Screen of natural product library identifies amygdalin as a potent activator of *cysl-2*p::GFP. a, Schematic of compound screens that led to identification of amygdalin. **b**, Representative bright field and epifluorescence images showing dose-dependent up-regulation of *cysl-2*p::GFP by amygdalin. Scale bar: 50 µm. **c**, Quantification of percentages of animals with strong *cysl-2*p::GFP expression after 48 hrs of amygdalin treatment with increasing doses. Values are means ± S.D with ***P < 0.001 (N = 5 independent experiments, n > 50 animals per experiment). **d**, Quantification of percentages of animals with strong *cysl-2*p::GFP expression after

increasing durations of amygdalin treatment at the fixed dose of 2 mg/ml. Values are means ± S.D with ***P < 0.001 (N = 5 independent experiments, n > 50 animals per experiment). **e**, Volcano plot showing significantly (adjusted p value < 0.05) up- (green) or down- (blue) regulated genes in amygdalin-treated animals, with *cysl-2* noted (arrow).

Figure 2



Genotype	*cysl-2*p::GFP
Wt	0% (n=150)
Wt + amy	92% (n=150)
egl-9	100% (n=150)
egl-9 + amy	100% (n=150)
egl-9; cysl-1	100% (n=150)
egl-9; hif-1	0% (n=150)
cysl-1; hif-1	0% (n=150)
cysl-1; egl-9; hif-1	0% (n=150)

Fig. 2. Amygdalin activates *cysl-2*p::GFP through cyanide, CYSL-1, EGL-9 and HIF-1. a, Representative bright field and epifluorescence images showing up-regulation of *cysl-2*p::GFP by amygdalin (1 mg/ml) in wild-type but not *hif-1*(*ia4*), *cysl-1*(*ok762*) loss-of-function mutant animals. Scale bar: 50 µm. **b**, Schematic illustrating the metabolic pathway of amygdalin, leading to generation of prunasin, glucose, hydrogen cyanide and benzaldehyde. **c**, Quantification of percentages of animals with strong *cysl-2*p::GFP expression after 48 hrs of treatment with amygdalin (2 mg/ml), prunacin (2 mg/ml), glucose (2 mg/ml), potassium cyanide (0.2 mg/ml) and benzaldehyde (1 mg/ml). Values are means ± S.D with ***P < 0.001 (N = 5 independent experiments, n > 50 animals per experiment). **d**, Schematic showing *cysl-1* suppressor screens that led to

identification of three mutants, all of which are allelic to *egl-9* based on whole-genome sequencing and complementation tests. **e**, Table summary of *cysl-2*p::GFP activation phenotypes in animals with indicated genotypes and treatment conditions. Penetrance and numbers of animals examined are noted. **f**, Model illustrating the regulatory pathway that mediates transcriptional response to amygdalin.

Figure 3

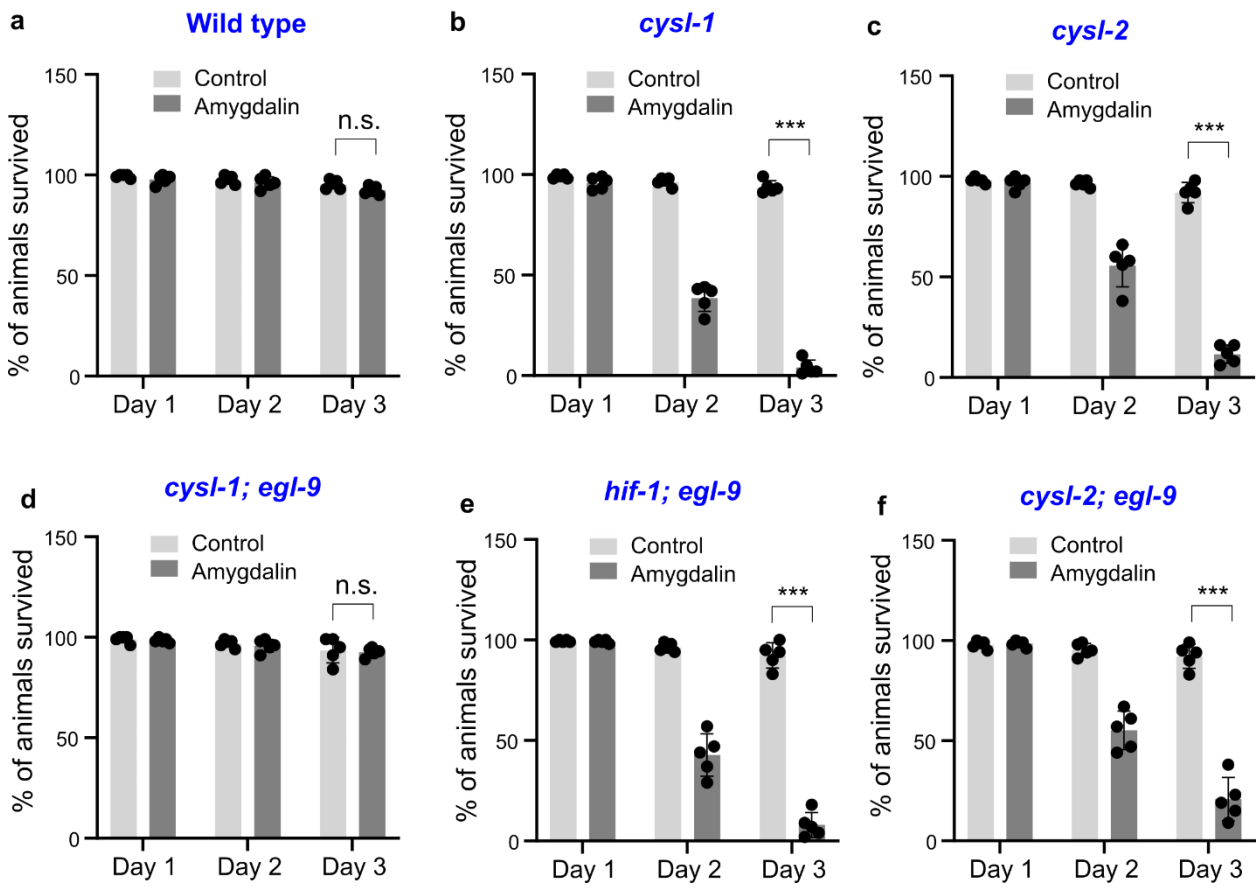
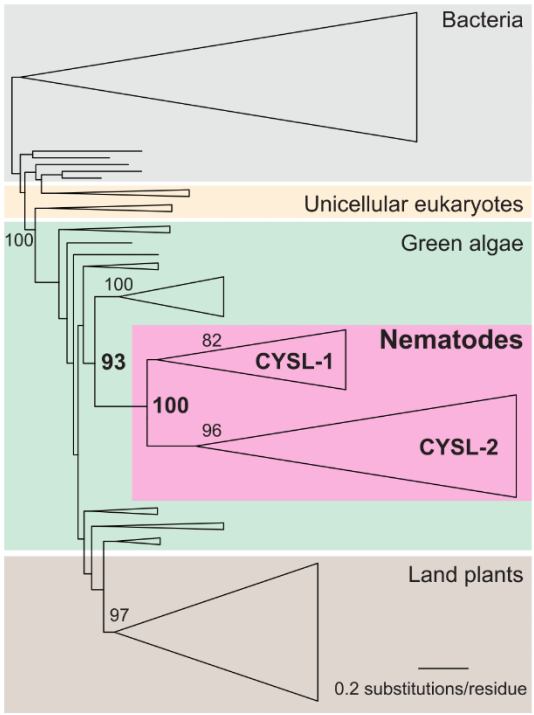


Fig. 3. Amygdalin resistance requires CYSL-1, HIF-1 and CYSL-2 acting in a regulatory pathway. a, Quantification of percentages of post-L4 stage wild-type animals survived after 24, 48 or 72 hrs of treatment with amygdalin (10 mg/ml). **b**, Quantification of percentages of post-L4 stage *cysl-1*(*ok762)* loss-of-function mutant animals survived after 24, 48 or 72 hrs of treatment with amygdalin (10 mg/ml). **c**, Quantification of percentages of post-L4 stage *cysl-2*(*ok3516)* loss-of-function mutant animals survived after 24, 48 or 72 hrs of treatment with amygdalin (10 mg/ml). **d**, Quantification of percentages of post-L4 stage *cysl-1*(*ok762); egl-9(sa307)* double loss-of-function mutant animals survived after 24, 48 or 72 hrs of treatment with amygdalin (10 mg/ml). **e**, Quantification of percentages of post-L4 stage *hif-1*(*ia04); egl-9(sa307)*

double loss-of-function mutant animals survived after 24, 48 or 72 hrs of treatment with amygdalin (10 mg/ml). **f**, Quantification of percentages of post-L4 stage *cysl-2*(*ok3516); egl-9(sa307)* double loss-of-function mutant animals survived after 24, 48 or 72 hrs of treatment with amygdalin (10 mg/ml). Values are means ± S.D with ***$P < 0.001$ (N = 5 independent experiments, n > 50 animals per experiment).

Figure 4

a



b

Software / *Substitution model*	Phylogenetic clade	Branch support *Bootstrap*	*SH-aLRT*
IQ-TREE *LG + I + G4(*)*	Nematode only	100	100
	Nematode + green algae	93	88
IQ-TREE *LG*	Nematode only	100	100
	Nematode + green algae	87	89
IQ-TREE *WAG + I + G4*	Nematode only	100	100
	Nematode + green algae	97	89
IQ-TREE *WAG*	Nematode only	100	100
	Nematode + green algae	73	76
IQ-TREE *JTT + I + G*	Nematode only	100	100
	Nematode + green algae	76	71
IQ-TREE *JTT*	Nematode only	100	100
	Nematode + green algae	47	73
RAxML *LG + Gamma(*)*	Nematode only	99	
	Nematode + green algae	32	

c

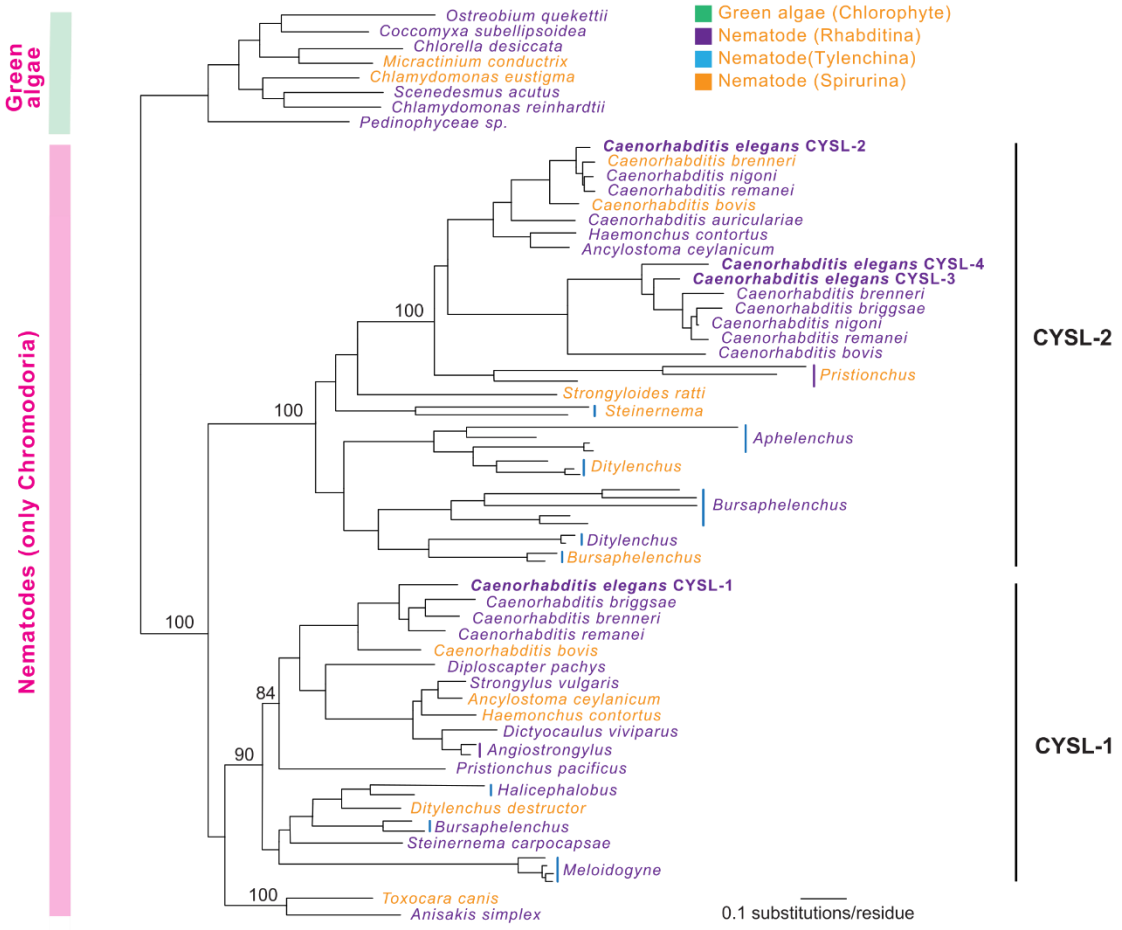
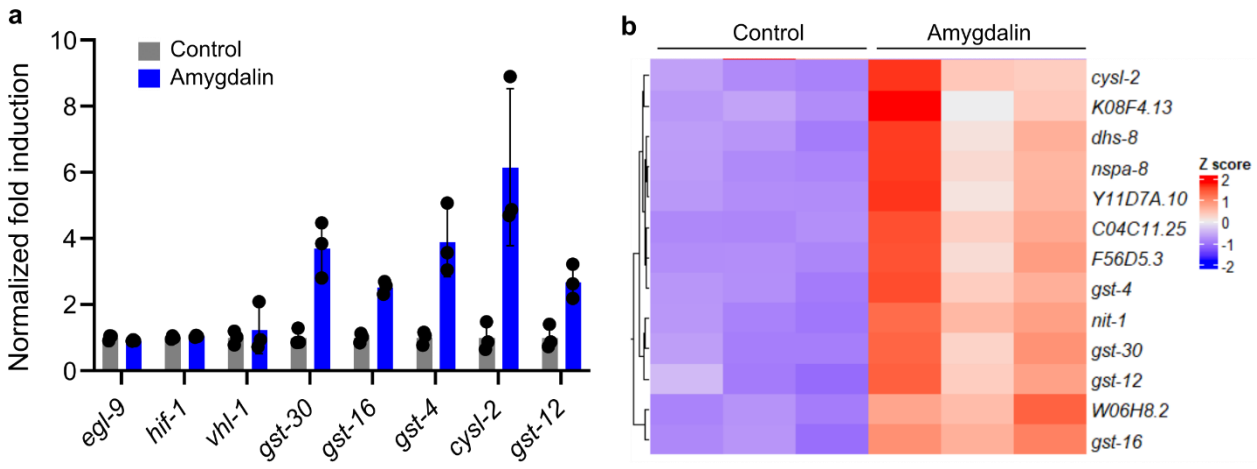


Fig. 4. Nematode CYSL genes were acquired from green algae by horizontal gene transfer. a, IQ-TREE generated maximum likelihood (ML) phylogenetic tree of homologs of nematode CYSL across a broad range of species. Species groups are shaded. Clades of similar proteins are collapsed and shown as triangles. Bootstrap support values for relevant branches are shown. Bolded values indicate the support for a clade that only contains CYSL-like proteins from nematodes (100% bootstrap support) and a well-supported clade that contains only CYSL-like proteins from nematodes and a group of green algae in the *Chlorophyta* lineage (93% bootstrap support). Within the nematodes, two clear, well-supported clades of CYSL proteins exist, one that contains *C. elegans* CYSL-1 and one that contains *C. elegans* CYSL-2. **b**, ML phylogenetic analyses were performed with multiple programs and multiple amino acid substitution models as indicated. Support values (bootstrap and SH-aLRT when available) are shown for a clade containing only nematode proteins and a clade containing nematode and chlorophyte green algae proteins as are bolded in part A. Asterisks next to model names indicate that best fitting model as described in Methods. **c**, Expanded phylogenetic tree of CYSL proteins from nematodes and their nearest homologs in green algae. Font colors denote major groups of nematode species as found in [25]. Bootstrap support values are shown at relevant nodes. For panels A-C, a complete list of all sequences used is found in Supplementary Data file 1 and all trees, including support values, are found in Supplementary Data file 2.

Supplementary Figure 1



Supplementary Fig. 1. Amygdalin-independent and regulated genes identified by RNAseq. a, Normalized fold induction by amygdalin (2 mg/ml on NGM for 48 hrs) of genes acting upstream of *cysl-2* (*egl-9*, *hif-1*, *vhl-1*) and GST-encoding genes involved in xenobiotic detoxification. **b**, Heat map showing hierarchical clustering of genes differentially up-regulated by amygdalin (2 mg/ml on NGM for 48 hrs) from RNAseq experiments.

Supplementary Figure 2

a

***C. elegans* CYSL-1 alignment to top hit green algae (*Raphidocelis subcapitata*)**

Blast score: 392 bit Identities: 63% Similarity: 79%

```
C. elegans CYSL-1 [NP_001369978.1]        4    RNSIGENAAALIGNTPMVYINKLTKGLPGTVAVKIEYMNPAGSVKDRIGAAMLAAAEKDG  63
                                               + +I ++   LIGNTPMVY+ ++ +GLPG +A K+E M P  SVKDRIG +M++AAE DG
R. subcapitata (greenalgae) [GBF88368.1]  36   KMNIAKDVTELIGNTPMVYLKRVNEGLPGKIACKLELMEPCCSVKDRIGYSMISAAEADG  95

C. elegans CYSL-1 [NP_001369978.1]        64   TVIPGVTTLIEPTSGNTGIALAFVAAAKGYRCIVTMPASMSGERRTLLKAYGSEVVLTDP  123
                                                +  G TTL+EPTSGNTGI LAF+AAAKGY+ ++TMPASMS ERR LL+A+G+E+VLTDP
R. subcapitata (greenalgae) [GBF88368.1]  96   KITAGKTTLVEPTSGNTGIGLAFIAAAKGYKLVLTMPASMSLERRILLRAFGAELVLTDP  155

C. elegans CYSL-1 [NP_001369978.1]        124  AKGMKGAIDMANQLKENIPGSIILAQFDNPNNPLVHYQTTGPEIWRQTKGTVDAVVFGVG  183
                                               AKGMKGA+  A ++  N P S IL QF+NPNNP VHY+TTGPEIW+ T G +D +V GVG
R. subcapitata (greenalgae) [GBF88368.1]  156  AKGMKGAVLKAQEIAANTPDSYILQQFENPNNPKVHYETTGPEIWKATDGKIDFLVSGVG  215

C. elegans CYSL-1 [NP_001369978.1]        184  TGGTITGVGRYLQEQNPGVRVFAVEPEESAILSGRPAGPHKIQGIGAGFAPAVLDTKIYE  243
                                               TGGTITG GRYL+EQNP V++ AVEP ES +LSG   GPHKIQGIGAGF P VLDT + +
R. subcapitata (greenalgae) [GBF88368.1]  216  TGGTITGAGRYLKEQNPNVKLVAVEPSESPVLSGGKPGPHKIQGIGAGFVPGVLDTHLID  275

C. elegans CYSL-1 [NP_001369978.1]        244  DVIRIHSDEAIVMAQRLSYEEGLLGGISSGANVAAALQLAARPEMAGKLIVTCLPSCGER  303
                                               +VI++ SD+A+ MA+RL+ EEGLL GISSGA   AA+Q+AARPE AGKLI   LPS GER
R. subcapitata (greenalgae) [GBF88368.1]  276  EVIQVSSDDAVSMARRLATEEGLLTGISSGAATVAAIQVAARPENAGKLIAVILPSFGER  335

C. elegans CYSL-1 [NP_001369978.1]        304  YMTSPLYTDIRESAMALA  321
                                               Y++S L+ +IR+ A A++
R. subcapitata (greenalgae) [GBF88368.1]  336  YLSSVLFQNIRDEAEAMS  353
```

b

***C. elegans* CYSL-1 alignment to *Tetranychus urticae* (mite) cysteine synthase**

Blast score: 211 bit Identities: 39% Similarity: 56%

```
C. elegans CYSL-1 [NP_001369978.1]        3    DR-NSIGENAAALIGNTPMVYINKLTKGLPGTVAVKIEYMNPAGSVKDRIGAAMLAAAEK  61
                                               DR N I  +A  LIGNTP++ +++L  G PG +  K E++NP  S+KDR    M+A A++
T. urticae (mite) [XP_015786551.1]        7    DRINGITPSALDLIGNTPLIALDRLWPG-PGRLLAKCEFLNPTASLKDRSSYYMIAKAKE  65

C. elegans CYSL-1 [NP_001369978.1]        62   DGTVIPGVTTLIEPTSGNTGIALAFVAAAKGYRCIVTMPASMSGERRTLLKAYGSEVVLT  121
                                                G +  G  ++IE TSGN G  +A V A  G+   VTM    S +R  ++ A G+ V+L
T. urticae (mite) [XP_015786551.1]        66   SGQLKDG-ESVIEVTSGNQGGGIACVTAVMGHPFTVTMSKGNSPQRAIMMNALGANVILV  124

C. elegans CYSL-1 [NP_001369978.1]        122  DPAKGMKGAI---------DMANQLKENIPGSIILAQFDNPNNPLVHYQTTGPEIWRQTK  172
                                               D   G  G +         + A +++E    +  + QF+NP N L HY+TTGPEIWRQT
T. urticae (mite) [XP_015786551.1]        125  DQVTGKPGNVTADDVAAAEETAMKIREET-NAYYVDQFNNPTNCLAHYETTGPEIWRQTN  183

C. elegans CYSL-1 [NP_001369978.1]        173  GTVDAVVFGVGTGGTITGVGRYLQEQNPGVRVFAVEPEESAILSG----RPAGPHKIQGI  228
                                               G +DA + G GTGG   G  ++L+E+NP VR F VEPE    ++G    +P   H +QG
T. urticae (mite) [XP_015786551.1]        184  GRIDAFLVGCGTGGCFVGTSKFLKEKNPNVRCFVVEPEGCQPIAGCTITKPL--HLLQGS  241

C. elegans CYSL-1 [NP_001369978.1]        229  GAGFAPAVLDTKIYEDVIRIHSDEAIVMAQRLSYEEGLLGGISSGANVAAALQLAARPEM  288
                                               G G  P + D K+Y D I +  +EAI   + L  +EGL  G ++G N+AAA++L    ++
T. urticae (mite) [XP_015786551.1]        242  GYGCVPTLFDKKVYNDSISVSDEEAIEYRKLLGQKEGLFCGFTTGGNIAAAIKLLKSGQL  301

C. elegans CYSL-1 [NP_001369978.1]        289  -AGKLIVTCLPSCGERY  304
                                                    +VT L   G +Y
T. urticae (mite) [XP_015786551.1]        302  PKDAWVVTILCDSGLKY  318
```

Supplementary Fig. 2. Protein sequence alignments of CYSL-1 with orthologs from green algae (a) and mite (b). Blast scores, a.a. identities and similarities are shown.

Methods and Materials

C. elegans

C. elegans strains were maintained under standard laboratory conditions unless otherwise specified. The N2 Bristol strain was used as the reference wild type, and the polymorphic Hawaiian strain CB4856 was used for genetic linkage mapping and SNP analysis [41,42]. Forward genetic screens for *cysl-2*p::GFP; *cysl-1* suppressing mutants after ethyl methanesulfonate (EMS)-induced random mutagenesis were performed as described previously [15,43]. Identification of mutations by whole-genome sequencing and complementation tests by crossing EMS mutants with *egl-9(sa307)* heterozygous males were used to determine *dma430, dma432* and *dma437* as new loss-of-function alleles of *egl-9*. Transgenic strains were generated by germline transformation as described [44]. Transgenic constructs were co-injected (at 10 - 50 ng/µl) with dominant *unc-54p*::mCherry, and stable extrachromosomal lines of fluorescent animals were established. Genotypes of strains used are as follows: Chr. II: *cysl-2*(*ok3516*)*;* Chr. IV: *nIs470 [cysl-2p::GFP; unc-54p::mCherry],* Chr. V: *egl-9(dma430, 432, 437, sa307), hif-1(ia4),* Chr. X: *cysl-1(ok762).*

Sample preparation for RNA sequencing and data analysis

N2 control (M9 treated) and amygdalin-treated animals (2 mg/ml on NGM for 48 hrs) were maintained at 20 °C and washed down from NGM plates using M9 solution and subjected to RNA extraction using the RNeasy Mini Kit from Qiagen. RNA-seq library preparation and data analysis were performed as previously described [45]. Three biological replicates were included for each treatment. The cleaned RNAseq reads were mapped to the genome sequence of *C. elegans* using hisat2 [46]. Abundance of genes

was expressed as FPKM (Reads per kilobase per million mapped reads). Identification of differentially expressed genes was performed using the DESeq2 package [47].

C. elegans amygdalin sensitivity and survival assays

Animals were cultured under non-starved conditions for at least 4 generations before amygdalin sensitivity assays. NGM plates spread with equal agar thickness seeded with equal amounts of OP50 were used for M9 or amygdalin (dissolved in M9, TCA0443-010G, VWR or equivalent A6005, Sigma-Aldrich) supplementation. Fresh preparations of fully dissolved amygdalin in M9 were added (250 µl/ 60 cm plate) and spread evenly on NGM plates with pre-seeded OP50. Once the plates with amygdalin were briefly air dried, synchronized-stage (L4) animals were placed and monitored for survival over 24, 48 and 72 hrs post treatment. Animals were scored as dead if they showed no pumping and movement upon light touch with the body necrosis subsequently confirmed.

Phylogenetic analysis

C. elegans CYSL-1 (accession NP_001369978.1) was used as a BLASTp [48] search query against the nonredundant (NR) and Reference Sequence (RefSeq) protein databases. Resulting sequences were aligned using Clustal Omega [49] using two iterations of refinement and duplicate sequences were removed. Non-aligning termini and alignment positions in which >95% of all sequences had a gap were manually removed from subsequent analyses. To generate phylogenetic trees encompassing a broad range of CYSL homologs, sequences with >80% sequence identity were reduced to a single unique sequence using CD-HIT [50] with a 0.8 sequence identity cutoff. The resulting 253 protein sequences (accession numbers and species names found in

Supplemental file 1) were realigned as above and used as input for maximum likelihood phylogenetic programs. IQ-TREE [51] phylogenies were generated using the "-bb 1000 -alrt 1000" commands for generation of 1000 ultrafast bootstrap [52] and SH-aLRT support values. The best substitution model was determined by ModelFinder [53] using the "-m AUTO" command or the substitution model was specified as shown in Fig. 4B using the "-m" command. RAxML phylogenies were generated using "-f a" with 500 bootstrap replicates and "-m PROTGAMMAAUTO" to determine the best substitution model with gamma distribution. All phylogenetic trees generated from IQ-TREE and RAxML, with support values, can be found in Supplemental file 2.

To generate a phylogeny of nematode CYSLs, the relevant subtree containing nematodes and the nearest green algae was extracted for the larger phylogeny generated above. Additional nematode sequences that had been removed using the CD-HIT filtering described above were added back in. Sequences were realigned and degapped as described above with a CDHIT similarity cutoff of 98% followed by additional manual removal of poorly aligning sequences. An IQ-TREE phylogenetic tree using ModelFinder (LG+I+G) was determined as described above. Accession numbers and species names are found in Supplemental file 1, and complete IQ-TREE with support values is found in Supplemental file 2.

Confocal and epifluorescence microscopic imaging

SPE epifluorescence compound microscopes (Leica) were used to capture fluorescence images. Animals of different genotypes were randomly picked at the same young adult stage (24 hrs post L4) and treated with 1 mM Levamisole sodium Azide in M9 solution (31,742-250MG, Sigma-Aldrich), aligned on an 4% agar pad on slides for

imaging. Identical setting and conditions were used to compare genotypes, treatment experimental groups with control.

Statistical analysis (wet lab data)

Data were analyzed using GraphPad Prism 9.2.0 Software (Graphpad, San Diego, CA) and presented as means ± S.D. unless otherwise specified, with *P* values calculated by unpaired two-tailed t-tests (comparisons between two groups), one-way ANOVA (comparisons across more than two groups) and adjusted with Bonferroni's corrections.

Data availability

The RNAseq read datasets were deposited in NCBI SRA (Sequence Read Archive) under the BioProject accession PRJNA843348. All other data generated for this study are included in this article.

Acknowledgments

Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). The work was supported by NIH grant 1R35GM139618 (D.K.M.), Packard Fellowships for Science and Engineering (D.K.M.) and the Pew Scholars Program in the Biomedical Sciences (D.K.M. and M.D.D.).

Author contributions

B.W., T.P. and D.K.M. conceptualized, performed and analyzed the wet experiments and wrote the manuscript. Y.L. performed RNAseq bioinformatic analysis. S.D.R. and M.D.D. performed phylogenetic analysis. M.D.D and D.K.M. supervised the project.

Competing interests

The authors declare no competing interests.

Correspondence

Correspondence should be addressed to Dengke K. Ma, Ph.D. (dengke.ma@ucsf.edu) and Matthew D. Daugherty, Ph.D. (MDDaugherty@UCSD.edu)

References

1. Gleadow, R. M. & Møller, B. L. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu. Rev. Plant Biol.* **65**, 155–185 (2014).
2. Mithöfer, A. & Boland, W. Plant defense against herbivores: chemical aspects. *Annu. Rev. Plant Biol.* **63**, 431–450 (2012).
3. Møller, B. L. Functional diversifications of cyanogenic glucosides. *Curr. Opin. Plant Biol.* **13**, 338–347 (2010).
4. Cressey, P. & Reeve, J. Metabolism of cyanogenic glycosides: A review. *Food Chem. Toxicol. Int. J. Publ. Br. Ind. Biol. Res. Assoc.* **125**, 225–232 (2019).
5. Ohlen, M. van, Herfurth, A.-M. & Wittstock, U. *Herbivore Adaptations to Plant Cyanide Defenses*. *Herbivores* (IntechOpen, 2017). doi:10.5772/66277.
6. Machingura, M., Salomon, E., Jez, J. M. & Ebbs, S. D. The β-cyanoalanine synthase pathway: beyond cyanide detoxification. *Plant Cell Environ.* **39**, 2329–2341 (2016).
7. Schulenburg, H. & Félix, M.-A. The Natural Biotic Environment of Caenorhabditis elegans. *Genetics* **206**, 55–86 (2017).
8. Frézal, L. & Félix, M.-A. C. elegans outside the Petri dish. *eLife* **4**, (2015).

398 9. Kiontke, K. & Sudhaus, W. Ecology of Caenorhabditis species. *WormBook Online Rev. C Elegans Biol.*
399 1–14 (2006) doi:10.1895/wormbook.1.37.1.
400 10. Gallagher, L. A. & Manoil, C. Pseudomonas aeruginosa PAO1 kills Caenorhabditis elegans by cyanide
401 poisoning. *J. Bacteriol.* **183**, 6207–6214 (2001).
402 11. Semenza, G. L. Hypoxia-inducible factors in physiology and medicine. *Cell* **148**, 399–408 (2012).
403 12. Kaelin, W. G. & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase
404 pathway. *Mol. Cell* **30**, 393–402 (2008).
405 13. Powell-Coffman, J. A. Hypoxia signaling and resistance in C. elegans. *Trends Endocrinol. Metab. TEM*
406 **21**, 435–440 (2010).
407 14. Vozdek, R., Hnízda, A., Krijt, J., Será, L. & Kožich, V. Biochemical properties of nematode O-
408 acetylserine(thiol)lyase paralogs imply their distinct roles in hydrogen sulfide homeostasis. *Biochim.*
409 *Biophys. Acta* **1834**, 2691–2701 (2013).
410 15. Ma, D. K., Vozdek, R., Bhatla, N. & Horvitz, H. R. CYSL-1 interacts with the O2-sensing hydroxylase
411 EGL-9 to promote H2S-modulated hypoxia-induced behavioral plasticity in C. elegans. *Neuron* **73**,
412 925–940 (2012).
413 16. Wybouw, N. *et al.* A gene horizontally transferred from bacteria protects arthropods from host plant
414 cyanide poisoning. *eLife* **3**, e02365 (2014).
415 17. Xia, J. *et al.* Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell* **184**, 3588
416 (2021).
417 18. Lapadula, W. J., Mascotti, M. L. & Juri Ayub, M. Whitefly genomes contain ribotoxin coding genes
418 acquired from plants. *Sci. Rep.* **10**, 15503 (2020).
419 19. Shen, C. & Powell-Coffman, J. A. Genetic analysis of hypoxia signaling and response in C elegans.
420 *Ann. N. Y. Acad. Sci.* **995**, 191–199 (2003).

421 20. Budde, M. W. & Roth, M. B. The response of Caenorhabditis elegans to hydrogen sulfide and
422 hydrogen cyanide. *Genetics* **189**, 521–532 (2011).
423 21. Shao, Z., Zhang, Y., Ye, Q., Saldanha, J. N. & Powell-Coffman, J. A. C. elegans SWAN-1 Binds to EGL-9
424 and regulates HIF-1-mediated resistance to the bacterial pathogen Pseudomonas aeruginosa PAO1.
425 *PLoS Pathog.* **6**, e1001075 (2010).
426 22. He, X.-Y. *et al.* Amygdalin - A pharmacological and toxicological review. *J. Ethnopharmacol.* **254**,
427 112717 (2020).
428 23. Topalidou, I. & Miller, D. L. Caenorhabditis elegans HIF-1 Is Broadly Required for Survival in
429 Hydrogen Sulfide. *G3 Bethesda Md* **7**, 3699–3704 (2017).
430 24. Burton, N. O. *et al.* Intergenerational adaptations to stress are evolutionarily conserved, stress-
431 specific, and have deleterious trade-offs. *eLife* **10**, e73425 (2021).
432 25. Smythe, A. B., Holovachov, O. & Kocot, K. M. Improved phylogenomic sampling of free-living
433 nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evol. Biol.* **19**, 121
434 (2019).
435 26. García, I., Arenas-Alfonseca, L., Moreno, I., Gotor, C. & Romero, L. C. HCN Regulates Cellular
436 Processes through Posttranslational Modification of Proteins by S-cyanylation. *Plant Physiol.* **179**,
437 107–123 (2019).
438 27. Cooper, C. E. & Brown, G. C. The inhibition of mitochondrial cytochrome oxidase by the gases
439 carbon monoxide, nitric oxide, hydrogen cyanide and hydrogen sulfide: chemical mechanism and
440 physiological significance. *J. Bioenerg. Biomembr.* **40**, 533–539 (2008).
441 28. Grice, K. *et al.* Photic zone euxinia during the Permian-triassic superanoxic event. *Science* **307**, 706–
442 709 (2005).
443 29. Gill, B. C. *et al.* Geochemical evidence for widespread euxinia in the later Cambrian ocean. *Nature*
444 **469**, 80–83 (2011).

30. Meyer, K. M., Yu, M., Jost, A. B., Kelley, B. M. & Payne, J. L. δ13C evidence that high primary productivity delayed recovery from end-Permian mass extinction. *Earth Planet. Sci. Lett.* **302**, 378–384 (2011).

31. Quist, C. W., Smant, G. & Helder, J. Evolution of plant parasitism in the phylum Nematoda. *Annu. Rev. Phytopathol.* **53**, 289–310 (2015).

32. Coghlan, A. Nematode genome evolution. *WormBook Online Rev. C Elegans Biol.* 1–15 (2005) doi:10.1895/wormbook.1.15.1.

33. Delaux, P.-M. & Schornack, S. Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science* **371**, eaba6605 (2021).

34. Oakley, T. H. Furcation and fusion: The phylogenetics of evolutionary novelty. *Dev. Biol.* **431**, 69–76 (2017).

35. Zhou, H., Beltrán, J. F. & Brito, I. L. Functions predict horizontal gene transfer and the emergence of antibiotic resistance. *Sci. Adv.* **7**, eabj5056.

36. Hittinger, C. T. & Carroll, S. B. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**, 677–681 (2007).

37. True, J. R. & Carroll, S. B. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **18**, 53–80 (2002).

38. Husnik, F. & McCutcheon, J. P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* **16**, 67–79 (2018).

39. Keeling, P. J. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr. Opin. Genet. Dev.* **19**, 613–619 (2009).

40. Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A. & Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **16**, 50 (2015).

469 41. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71–94 (1974).

470 42. Davis, M. W. *et al.* Rapid single nucleotide polymorphism mapping in C. elegans. *BMC Genomics* **6**,
471 118 (2005).

472 43. Ma, D. K. *et al.* Acyl-CoA Dehydrogenase Drives Heat Adaptation by Sequestering Fatty Acids. *Cell*
473 **161**, 1152–1163 (2015).

474 44. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in C.elegans:
475 extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–
476 3970 (1991).

477 45. Jiang, W. *et al.* A genetic program mediates cold-warming response and promotes stress-induced
478 phenoptosis in C. elegans. *eLife* **7**, (2018).

479 46. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements.
480 *Nat. Methods* **12**, 357–360 (2015).

481 47. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq
482 data with DESeq2. *Genome Biol.* **15**, 550 (2014).

483 48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J.*
484 *Mol. Biol.* **215**, 403–410 (1990).

485 49. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using
486 Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

487 50. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation
488 sequencing data. *Bioinforma. Oxf. Engl.* **28**, 3150–3152 (2012).

489 51. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic
490 algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

491 52. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the
492 Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

53. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).