Published in final edited form as:

Curr Biol. 2022 November 21; 32(22): 4941–4948.e3. doi:10.1016/j.cub.2022.09.041.

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**[3,5,*], **Dengke K. Ma**[1,4,6,7,*]

[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.

[4]Innovative Genomics Institute, University of California, Berkeley, USA.

[5]Twitter: @Daugherty_Lab

[6]Twitter: @DengkeMa

[7]Lead contact

SUMMARY

Amygdalin is a cyanogenic glycoside enriched in tissues of many edible plants, including seeds of stone fruits such as cherry (*Prunus avium*), peach (*Prunus persica*), and apple (*Malus domestica*). These plants biosynthesize amygdalin in defense against herbivore animals as amygdalin generates poisonous cyanide upon plant tissue destruction.[1–4] Poisonous to many animals, amygdalin-derived cyanide is detoxified by potent enzymes commonly found in bacteria and plants but not most animals.[5] 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*. Screen of a natural product library for hypoxia-independent regulators of HIF-1 identifies amygdalin as a potent activator of *cysl-2*, a HIF-1 transcriptional target that encodes a cyanide detoxification enzyme in *C. elegans*. As a *cysl-2* paralog similarly essential for amygdalin resistance, *cysl-1* encodes a protein homologous to cysteine biosynthetic enzymes in bacteria and plants, but functionally co-opted in *C. elegans*.

*Correspondence: dengke.ma@ucsf.edu (D.K.M.), MDDaugherty@UCSD.edu (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 and wrote the manuscript. M.D.D and D.K.M. supervised the project.

Declaration of interests

The authors declare no competing interests.

INCLUSION AND DIVERSITY

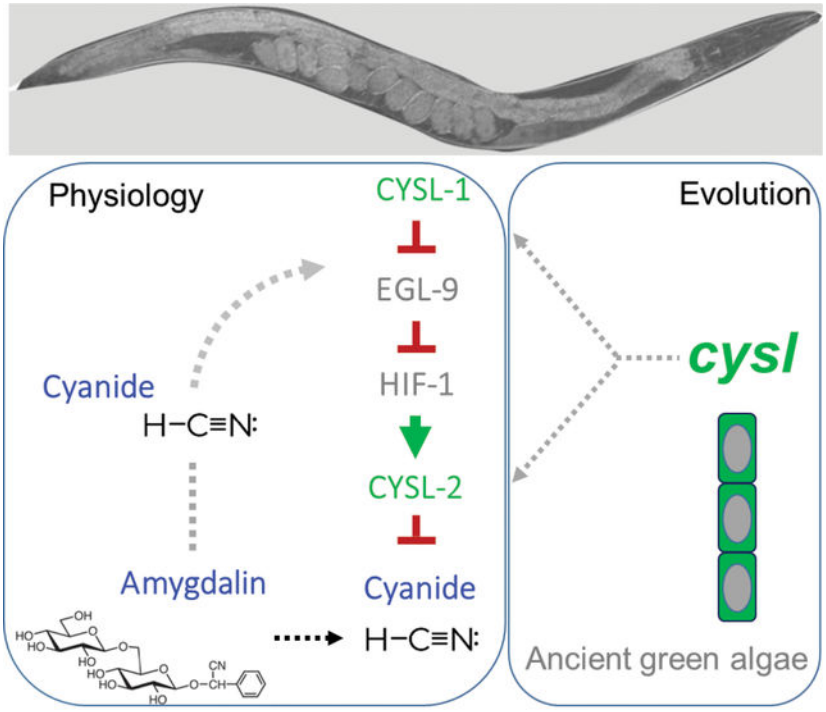
We support inclusive, diverse, and equitable conduct of research.

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 indicates that *cysl-1* and *cysl-2* were likely acquired from green algae through horizontal gene transfer (HGT) and functionally co-opted in protection against amygdalin. Since acquisition, these two genes evolved division of labor in a cellular circuit to detect and detoxify cyanide. Thus, algae-to-nematode HGT and subsequent gene function co-option events may facilitate host survival and adaptation to adverse environment stresses and biogenic toxins.

eTOC blurb

Wang et al. show how plant-derived amygdalin, a cyanogenic glycoside, is detoxified by the nematode *C. elegans* through two paralogous genes *cysl-1* and *cysl-2*. Functional and phylogenetic analyses suggest both genes originated from green algae likely by horizontal gene transfer and evolved separate roles by gene co-option for cyanide detoxification.

Graphical Abstract



RESULTS AND DISCUSSION

A chemical screen identifies amygdalin as a *cysl-2*p::GFP activator

We previously discovered several components of the HIF-1 pathway acting upstream of the oxygen-sensing HIF hydroxylase EGL-9 in *C. elegans*.[6] 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 (Figure

1A). Among the identified compounds, a glycoside named amygdalin emerged as a potent activator of *cysl-2*p::GFP. We validated the finding using amygdalin from new preparations of different sources (from 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 (Figure 1B). Activation of *cysl-2*p::GFP by amygdalin was also dose-dependent and reached peak levels at approximately 48 h post amygdalin treatment (Figures 1C and 1D).

We performed RNAseq to identify genome-wide transcriptomic changes induced by amygdalin. We treated wild-type mixed-stage *C. elegans* hermaphrodites cultivated at 20 °C with 2 mg/mL amygdalin for 48 h. After differential expression analyses of triplicate samples, we identified 16 genes that are significantly (adjusted p value < 0.05) up- or down-regulated by amygdalin treatment (Figure 1E; Data S1). 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 (Figure S1; Data S1). Gene Ontology analysis (Wormbase) indicates that amygdalin also up-regulated several genes encoding glutathione S-transferases involved in xenobiotic detoxification (Figure S1).

The CYSL-1/EGL-9/HIF-1/CYSL-2 pathway mediates transcriptional response and organismic tolerance to amygdalin

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.[6–9] We found that loss-of-function mutations of *hif-1* or *cysl-1* abolished *cysl-2*p::GFP induction by low-dose amygdalin (Figure 2A) (high dose is lethal to these mutants, see below). Amygdalin is a glycoside that can be metabolized sequentially to prunasin, glucose, benzaldehyde and hydrogen cyanide (Figure 2B).[1,3,10] 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 detectable effect (Figure 2C). CYSL-1 and CYSL-2 are cysteine synthase-like proteins and have been implicated in hydrogen sulfide and cyanide detoxification.[6,8,11–13] While CYSL-2 can enzymatically convert cyanide to nontoxic β-cyanoalanine,[11] 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.[6] Taking advantage of the *cysl-2*p::GFP phenotype and lethal effects of high-dose amygdalin in *cysl-1* mutants, 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 (Figure 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 (Figure 2E). These data indicate that amygdalin activates *cysl-2*p::GFP through a dis-inhibitory regulatory pathway from cyanide, CYSL-1, EGL-9 to HIF-1 (Figure 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 owing 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 (Figure 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 (Figure 3B). Loss-of-function of *cysl-2* also caused similar time-dependent sensitivity to amygdalin (Figure 3C). While loss of *egl-9* fully suppressed vulnerability of *cysl-1* mutants to amygdalin (Figure 3D), loss of either *hif-1* or *cysl-2* rendered *egl-9* mutants sensitive to amygdalin (Figures 3E and 3F). While *cysl-2; cysl-1* or *hif-1; cysl-1* double mutants exhibited strikingly similar amygdalin sensitivity profiles to *cysl-1*, *cysl-2*, or *hif-1* single mutants, loss of *rhy-1*, which leads to constitutive *cysl-2*p::GFP activation,[6] showed similar amygdalin sensitivity to loss of *egl-9*. These data indicate that amygdalin activation of *cysl-2* through the CYSL-1/EGL-9/HIF-1 pathway confers tolerance to amygdalin toxicity in *C. elegans*.

Phylogenetic analysis reveals the algal origin of *cysl-1* and *cysl-2*

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 (Figures S2A and 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 (Data S2 and S3; Figure 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 (Figure 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 (Figure 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 (Figures S2B and S3).[14] 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. We found CYSL proteins only in the *Chromadoria* clade of nematodes and not in the well sampled organisms in the *Trichinellida* (*e.g. Trichinella* species) or in genomes or transcriptomes from other non-*Chromadoria* nematodes (Figure 4C and Data S4).[15] Although we cannot rule out that it was lost in other nematodes, 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 observed two or more copies of CYSL homologs in *Tylenchina* (*e.g. Bursaphelenchus*

species) and *Rhabditina* (*e.g. Caenorhabditis* species) lineages of nematodes. Although we only detect a single CYSL homolog in the *Spirurina* nematodes (*e.g. Toxocara canis*), these homologs group with CYSL-1 proteins from other nematodes (Figure 4C). These data suggest that CYSL duplicated soon after acquisition in nematodes to CYSL-like proteins, possibly followed by loss of one paralog or other in certain lineages. 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.

Physiological and evolutionary implications of *cysl* genes

In summary, our work establishes novel physiological roles of the CYSL-1/EGL-9/HIF-1/CYSL-2 pathway in the detoxification of amygdalin in *C. elegans*. Our data indicate that CYSL-1 acts upstream of the evolutionarily conserved HIF-1 pathway as a key regulator of *cysl-2* induction upon amygdalin exposure, resulting in detoxification of cyanide. 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 functional and phylogenetic studies indicate that *cysl-1* and *cysl-2* likely originated from green algae by a HGT event in ancient nematodes and evolved separate roles since. In *C. elegans*, CYSL-1 transcriptionally regulates expression of *cysl-2* through HIF-1. As a transcriptional target of HIF-1, CYSL-2 can detoxify cyanides *in vitro* as a cyanoalanine synthase.[11] Mechanistically, CYSL-1 may regulate *cysl-2* by binding to EGL-9 and thereby promote HIF-1 activity.[6] How amygdalin or cyanide is sensed by *C. elegans* and regulates CYSL-1 and EGL-9 awaits further studies. Plausible mechanisms include mitochondrial inhibition by cyanide, posttranslational modification by S-cyanylation, and allosteric modulation of CYSL-1 and EGL-9 interaction.[6,16,17]

In nature, *C. elegans* lives with rotten fruits and bacteria,[18–21] thus may frequently encounter cyanogenic chemicals derived from its natural diets (e.g. cyanogenic plant tissues and *Pseudomonas* bacteria). After consumption, amygdalin is hydrolyzed by β-glucosidases and α-hydroxynitrilases, releasing hydrogen cyanide or cyanogenic intermediates that can inhibit mitochondrial respiration by binding to cytochrome C oxidases, thus be highly toxic and even lethal to herbivore animals.[10,17] To tolerate amygdalin, plants also evolved a family of cysteine synthase-like enzyme, β-cyanoalanine synthases, to directly detoxify cyanide.[3,5,22] Although *C. elegans* rarely encounters amygdalin under laboratory conditions, loss-of-function mutations in the gene *egl-9* have been shown to confer resistance to hydrogen cyanide generated by a pathogenic bacterial strain of *Pseudomonas aeruginosa*.[23] EGL-9 is a prolyl hydroxylase of HIF-1 (hypoxia inducible factor), a transcription factor that directly activates expression of many genes involved in hypoxic adaptation and stress responses.[24–26] 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.[14] 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 ortholog was acquired by

ancestral *Chromodoria* nematodes after the divergence from other clades of nematodes, and duplicated soon after their acquisition to generate *cysl* gene paralogs. While CYSL-2 retains cyanoalanine synthase activity to detoxify cyanide, CYSL-1 is co-opted to regulate *cysl-2* upon amygdalin/cyanide exposure, in an elegant cellular circuit that detects and subsequently defends against toxicity associated with amygdalin/cyanide derived from natural diets of *C. elegans*.

Chromodoria nematodes diverged from other clades of nematodes more than 250 million year ago, around which drastic geological and ecological changes may have exerted selection pressures on nematodes surviving in environments rife with sulfide/cyanide toxicity and ancient algal species.[27–31] 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.[31–36] Functional HGTs within prokaryotic organisms and from bacteria to eukaryotes have been widely observed and characterized.[37–40] Our functional and phylogenetic analyses reveal a previously unknown HGT from algae to nematodes, supporting the emerging theme that such events of HGT can lead to evolutionary changes and novel traits to facilitate host protection and adaptation to abiotic stress, chemical toxins and extreme environments.

STAR★Methods

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dengke Ma (dengke.ma@ucsf.edu).

Materials availability—Strains used in this study are deposited to CGC.

Data and code availability—The RNAseq read datasets were deposited in NCBI SRA (Sequence Read Archive) under the BioProject accession PRJNA843348. All other data reported in this paper will be shared by the lead contact upon request.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

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.[6,43] 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*. Genotypes of strains used are as follows: Chr. II: *cysl-2*(*ok3516*)*; rhy-1(n5500),* Chr. IV: *nIs470 [cysl-2p::GFP; myo-2p::mCherry],* Chr. V: *egl-9(dma430, 432, 437, sa307), hif-1(ia4),* Chr. X: *cysl-1(ok762).*

METHOD DETAILS

Sample preparation for RNA sequencing and data analysis—N2 control (M9 treated) and amygdalin-treated animals (2 mg/mL on NGM for 48 h) 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.[44] Three biological replicates were included for each treatment. The cleaned RNAseq reads were mapped to the genome sequence of *C. elegans* using the hisat2 tool.[45] 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.[46]

Compound screen, amygdalin sensitivity and survival assays—*C. elegans* animals carrying the integrated *cysl-2*p::GFP reporter were cultured under non-starved conditions for at least 4 generations before natural product screens using the DiscoveryProbe Bioactive Compound Library (Catalog No. L1022P, APExBIO). Synchronized L1-stage animals were grown to the L4 stage and transferred to deep 96-well plates in liquid cultures containing each compound (0.5 mM in 100 μL volume, approximately 100 animals per well) in S medium with concentrated *E. coli* OP50 as a food source. The cultures were incubated at 20 °C for 48 h before observation under an epifluorescence microscope (SMZ18). Candidate hits were subject to multiple trials and repeated with the same compounds from different providers and to determine phenotypic penetrance, fluorescence intensity and compound dose-dependency.

C. elegans 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 freshly seeded OP50 (within 24 h of seeding) 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 h 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[47] search query against the nonredundant (NR) and Reference Sequence (RefSeq) protein databases. Resulting sequences were aligned using Clustal Omega[48] 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[49] with a 0.8 sequence identity cutoff. The resulting 253 protein sequences (accession numbers and species names found in Data S2) were realigned as above and used as input for maximum likelihood phylogenetic programs. IQ-TREE[50] phylogenies were generated using the "-bb 1000 -alrt 1000" commands for generation of 1000 ultrafast bootstrap[51] and SH-aLRT support values. The best substitution model was determined by ModelFinder[52] using the "-m AUTO" command or the substitution model was specified as shown in Figure 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 Data S3.

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 Data S2, and complete IQ-TREE with support values is found in Data S3.

To generate a phylogeny that contains nematode CYSL proteins and arthropod cysteine synthases that had previously been described as an HGT event from bacteria,[14] we performed a BLASTp search against the RefSeq database using the cysteine synthase enzyme from the two spotted spider mite (*Tetranychus urticae*, accession XP_015786551.1). The top 5000 homologs from that search were reduced in number using CD-HIT with a 0.8 sequence identity cutoff. Sequences were combined with homologs from nematode CYSL described above and any redundant sequences were removed. The resulting 1040 protein sequences (accession numbers and species names found in Data S2) were realigned as above and used as input for maximum likelihood phylogenetic programs. An IQ-TREE phylogenetic tree using ModelFinder (LG+I+G4) was determined as described above. Accession numbers and species names are found in Data S2, and complete IQ-TREE with support values is found in Data S3.

To search for CYSL homologs in nematode species not found in the NR database, we used the *C. elegans* CYSL-1 protein sequence as a search query against the WGS (sequences from all genome sequencing projects) and TSA (sequences from transcriptome sequencing projects) databases using tBLASTn. Sequences with greater than 20% coverage and 40% sequence identity were retained for further analysis. Sequences from nematode clades shown in Figure 4C were removed. The remaining sequences are shown in Data S4. To identify the closest homolog of each sequence shown in Data S2, we used BLASTn from the nucleotide sequence or BLASTp from the predicted protein sequence as a query for the NR database. Annotations of resulting hits were used to determine whether the nematode sequence was (a)

a true CYSL homolog, (b) a paralogous protein (annotated as cystathionine beta-synthase) or (c) a likely contaminant in the nematode genome sequencing project.

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 h 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. To quantify the percentage of animals with *cysl-2*p::GFP expression, a randomly selected population was observed under the epifluorescence microscope (SMZ18, Nikon) with animals considered to be *cysl-2*p::GFP positive only when hypodermal *cysl-2*p::GFP fluorescence is intense at a level comparable to that activated by 2 mg/mL amygdalin that approaches the saturation of the phenotypic penetrance in the wild-type *cysl-2*p::GFP animals.

QUANTIFICATION AND STATISTICAL ANALYSIS

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 two-way ANOVA (interaction between genotype and treatment), with post-hoc Tukey HSD and Bonferroni's corrections.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 R35GM139618 (D.K.M.), R35GM133633 (M.D.D.), 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.).

References

1. Gleadow RM, and Møller BL (2014). Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. Annu. Rev. Plant Biol. 65, 155–185. [PubMed: 24579992]
2. Møller BL (2010). Functional diversifications of cyanogenic glucosides. Curr. Opin. Plant Biol. 13, 338–347. [PubMed: 20197238]
3. Cressey P, and Reeve J (2019). Metabolism of cyanogenic glycosides: A review. Food Chem. Toxicol. Int. J. Publ. Br. Ind. Biol. Res. Assoc. 125, 225–232.
4. Mithöfer A, and Boland W (2012). Plant defense against herbivores: chemical aspects. Annu. Rev. Plant Biol. 63, 431–450. [PubMed: 22404468]
5. Machingura M, Salomon E, Jez JM, and Ebbs SD (2016). The β-cyanoalanine synthase pathway: beyond cyanide detoxification. Plant Cell Environ. 39, 2329–2341. [PubMed: 27116378]
6. Ma DK, Vozdek R, Bhatla N, and Horvitz HR (2012). CYSL-1 interacts with the O2-sensing hydroxylase EGL-9 to promote H2S-modulated hypoxia-induced behavioral plasticity in C. elegans. Neuron 73, 925–940. [PubMed: 22405203]
7. Shen C, and Powell-Coffman JA (2003). Genetic analysis of hypoxia signaling and response in C elegans. Ann. N. Y. Acad. Sci. 995, 191–199. [PubMed: 12814951]

8. Budde MW, and Roth MB (2011). The response of Caenorhabditis elegans to hydrogen sulfide and hydrogen cyanide. Genetics 189, 521–532. [PubMed: 21840852]
9. Shao Z, Zhang Y, Ye Q, Saldanha JN, and Powell-Coffman JA (2010). C. elegans SWAN-1 Binds to EGL-9 and regulates HIF-1-mediated resistance to the bacterial pathogen Pseudomonas aeruginosa PAO1. PLoS Pathog. 6, e1001075. [PubMed: 20865124]
10. He X-Y, Wu L-J, Wang W-X, Xie P-J, Chen Y-H, and Wang F (2020). Amygdalin - A pharmacological and toxicological review. J. Ethnopharmacol. 254, 112717. [PubMed: 32114166]
11. Vozdek R, Hnízda A, Krijt J, Será L, and Kožich V (2013). Biochemical properties of nematode O-acetylserine(thiol)lyase paralogs imply their distinct roles in hydrogen sulfide homeostasis. Biochim. Biophys. Acta 1834, 2691–2701. [PubMed: 24100226]
12. Topalidou I, and Miller DL (2017). Caenorhabditis elegans HIF-1 Is Broadly Required for Survival in Hydrogen Sulfide. G3 Bethesda Md 7, 3699–3704. [PubMed: 28889102]
13. Burton NO, Willis A, Fisher K, Braukmann F, Price J, Stevens L, Baugh LR, Reinke AW, and Miska EA (2021). Intergenerational adaptations to stress are evolutionarily conserved, stress-specific, and have deleterious trade-offs. eLife 10, e73425. [PubMed: 34622777]
14. Wybouw N, Dermauw W, Tirry L, Stevens C, Grbić M, Feyereisen R, and Van Leeuwen T (2014). A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. eLife 3, e02365. [PubMed: 24843024]
15. Ahmed M, Roberts NG, Adediran F, Smythe AB, Kocot KM, and Holovachov O (2022). Phylogenomic Analysis of the Phylum Nematoda: Conflicts and Congruences With Morphology, 18S rRNA, and Mitogenomes. Front. Ecol. Evol. 9.
16. García I, Arenas-Alfonseca L, Moreno I, Gotor C, and Romero LC (2019). HCN Regulates Cellular Processes through Posttranslational Modification of Proteins by S-cyanylation. Plant Physiol. 179, 107–123. [PubMed: 30377236]
17. Cooper CE, and Brown GC (2008). The inhibition of mitochondrial cytochrome oxidase by the gases carbon monoxide, nitric oxide, hydrogen cyanide and hydrogen sulfide: chemical mechanism and physiological significance. J. Bioenerg. Biomembr. 40, 533–539. [PubMed: 18839291]
18. Schulenburg H, and Félix M-A (2017). The Natural Biotic Environment of Caenorhabditis elegans. Genetics 206, 55–86. [PubMed: 28476862]
19. Frézal L, and Félix M-A (2015). C. elegans outside the Petri dish. eLife 4. 10.7554/eLife.05849.
20. Kiontke K, and Sudhaus W (2006). Ecology of Caenorhabditis species. WormBook Online Rev. C Elegans Biol, 1–14.
21. Kiontke KC, Félix M-A, Ailion M, Rockman MV, Braendle C, Pénigault J-B, and Fitch DHA (2011). A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. BMC Evol. Biol. 11, 339. [PubMed: 22103856]
22. Ohlen M. van, Herfurth A-M, and Wittstock U (2017). Herbivore Adaptations to Plant Cyanide Defenses (IntechOpen)
23. Gallagher LA, and Manoil C (2001). Pseudomonas aeruginosa PAO1 kills Caenorhabditis elegans by cyanide poisoning. J. Bacteriol. 183, 6207–6214. [PubMed: 11591663]
24. Semenza GL (2012). Hypoxia-inducible factors in physiology and medicine. Cell 148, 399–408. [PubMed: 22304911]
25. Kaelin WG, and Ratcliffe PJ (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol. Cell 30, 393–402. [PubMed: 18498744]
26. Powell-Coffman JA (2010). Hypoxia signaling and resistance in C. elegans. Trends Endocrinol. Metab. TEM 21, 435–440. [PubMed: 20335046]
27. Grice K, Cao C, Love GD, Böttcher ME, Twitchett RJ, Grosjean E, Summons RE, Turgeon SC, Dunning W, and Jin Y (2005). Photic zone euxinia during the Permian-triassic superanoxic event. Science 307, 706–709. [PubMed: 15661975]
28. Gill BC, Lyons TW, Young SA, Kump LR, Knoll AH, and Saltzman MR (2011). Geochemical evidence for widespread euxinia in the later Cambrian ocean. Nature 469, 80–83. [PubMed: 21209662]

29. Meyer KM, Yu M, Jost AB, Kelley BM, and Payne JL (2011). δ13C evidence that high primary productivity delayed recovery from end-Permian mass extinction. Earth Planet. Sci. Lett. 302, 378–384.
30. Quist CW, Smant G, and Helder J (2015). Evolution of plant parasitism in the phylum Nematoda. Annu. Rev. Phytopathol. 53, 289–310. [PubMed: 26047569]
31. Coghlan A (2005). Nematode genome evolution. WormBook Online Rev. C Elegans Biol, 1–15.
32. Delaux P-M, and Schornack S (2021). Plant evolution driven by interactions with symbiotic and pathogenic microbes. Science 371, eaba6605. [PubMed: 33602828]
33. Oakley TH (2017). Furcation and fusion: The phylogenetics of evolutionary novelty. Dev. Biol. 431, 69–76. [PubMed: 28923487]
34. Zhou H, Beltrán JF, and Brito IL Functions predict horizontal gene transfer and the emergence of antibiotic resistance. Sci. Adv. 7, eabj5056. [PubMed: 34678056]
35. Hittinger CT, and Carroll SB (2007). Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449, 677–681. [PubMed: 17928853]
36. True JR, and Carroll SB (2002). Gene co-option in physiological and morphological evolution. Annu. Rev. Cell Dev. Biol. 18, 53–80. [PubMed: 12142278]
37. Husnik F, and McCutcheon JP (2018). Functional horizontal gene transfer from bacteria to eukaryotes. Nat. Rev. Microbiol. 16, 67–79. [PubMed: 29176581]
38. Keeling PJ (2009). Functional and ecological impacts of horizontal gene transfer in eukaryotes. Curr. Opin. Genet. Dev. 19, 613–619. [PubMed: 19897356]
39. Crisp A, Boschetti C, Perry M, Tunnacliffe A, and Micklem G (2015). Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biol. 16, 50. [PubMed: 25785303]
40. Sibbald SJ, Eme L, Archibald JM, and Roger AJ (2020). Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes. Trends Parasitol. 36, 927–941. [PubMed: 32828660]
41. Brenner S (1974). The genetics of Caenorhabditis elegans. Genetics 77, 71–94. [PubMed: 4366476]
42. Davis MW, Hammarlund M, Harrach T, Hullett P, Olsen S, and Jorgensen EM (2005). Rapid single nucleotide polymorphism mapping in C. elegans. BMC Genomics 6, 118. [PubMed: 16156901]
43. Ma DK, Li Z, Lu AY, Sun F, Chen S, Rothe M, Menzel R, Sun F, and Horvitz HR (2015). Acyl-CoA Dehydrogenase Drives Heat Adaptation by Sequestering Fatty Acids. Cell 161, 1152–1163. [PubMed: 25981666]
44. Jiang W, Wei Y, Long Y, Owen A, Wang B, Wu X, Luo S, Dang Y, and Ma DK (2018). A genetic program mediates cold-warming response and promotes stress-induced phenoptosis in C. elegans. eLife 7.
45. Kim D, Langmead B, and Salzberg SL (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360. [PubMed: 25751142]
46. Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. [PubMed: 25516281]
47. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410. [PubMed: 2231712]
48. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. [PubMed: 21988835]
49. Fu L, Niu B, Zhu Z, Wu S, and Li W (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinforma. Oxf. Engl. 28, 3150–3152.
50. Nguyen L-T, Schmidt HA, von Haeseler A, and Minh BQ (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. [PubMed: 25371430]
51. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, and Vinh LS (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35, 518–522. [PubMed: 29077904]

52. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, and Jermiin LS (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. [PubMed: 28481363]

Highlights

- Chemical screens identify amygdalin as a hypoxia-independent activator of HIF-1
- Amygdalin up-regulates HIF-1 target gene cysl-2 via CYSL-1 upstream of HIF-1
- CYSL-1, EGL-9, HIF-1, CYSL-2 act in a cascade to detoxify amygdalin-derived cyanide
- *cysl* genes likely originated from green algae and functionally co-opted since

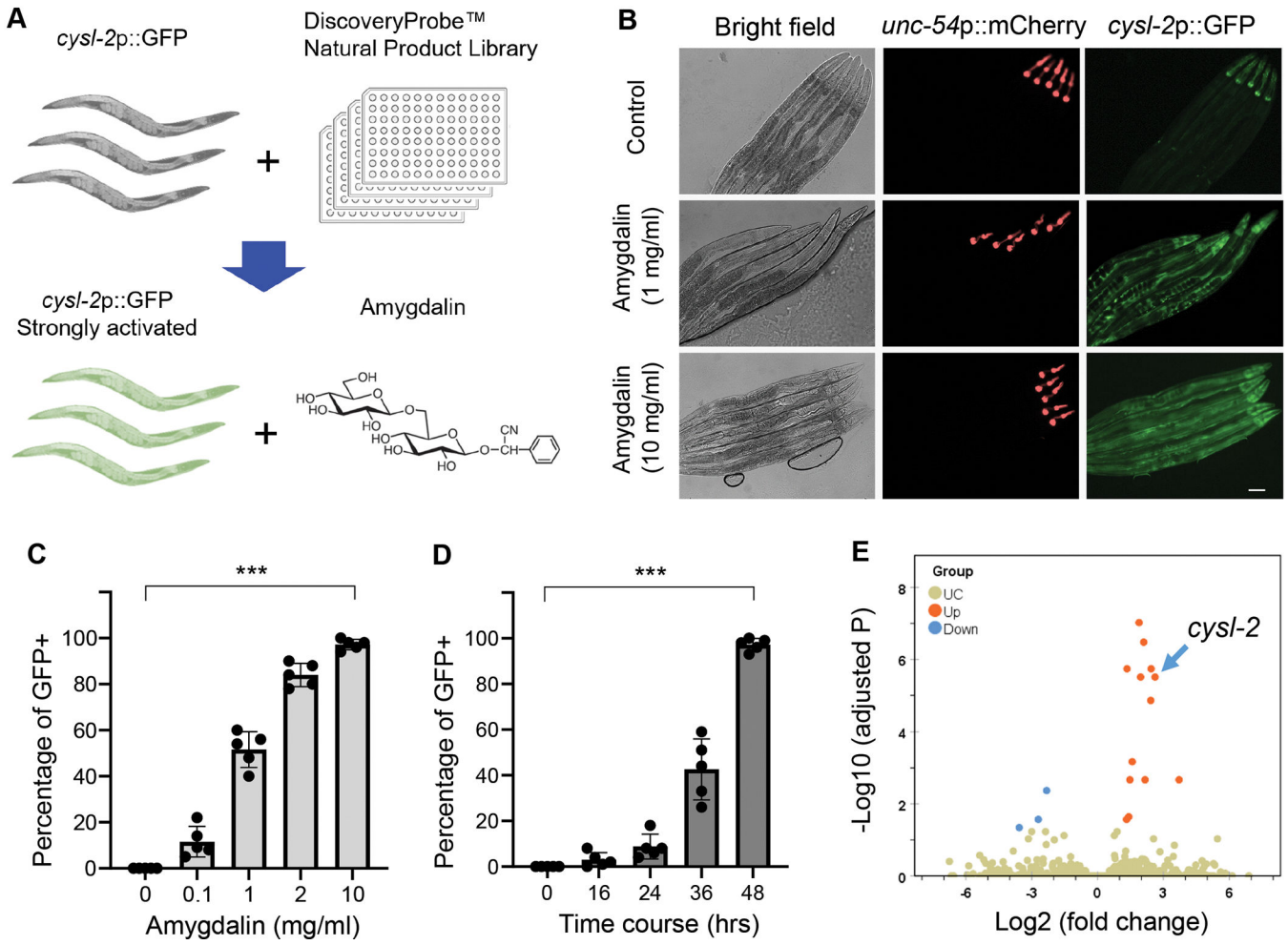
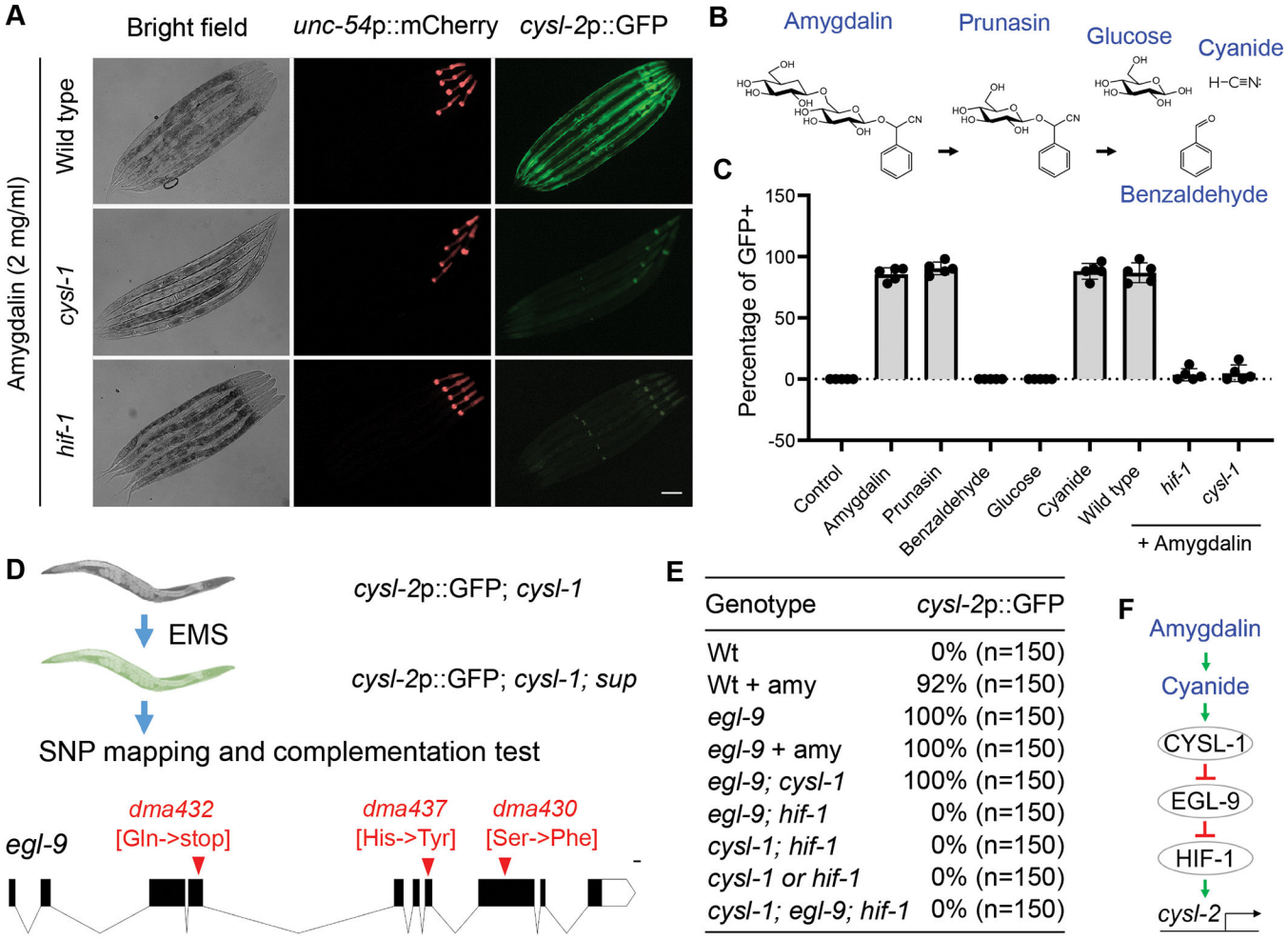


Figure 1. A compound screen of natural product library identifies amygdalin as a potent activator of *cysl-2*p::GFP.

(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 *cysl-2*p::GFP expression after 48 h of amygdalin treatment with increasing doses. Values are means ± S.D with ***P < 0.001 (one-way ANOVA with post-hoc Tukey HSD, N = 5 independent experiments, n > 50 animals per experiment). **(D)**, Quantification of percentages of animals with *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 (one-way ANOVA with post-hoc Tukey HSD, 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). A list of all genes with expression values and statistics are in Data S1.



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 or hif-1	0% (n=150)
cysl-1; egl-9; hif-1	0% (n=150)

Figure 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 *cysl-2*p::GFP expression after 48 h 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. **(D)**, Schematic showing *cysl-1* suppressor screens resulting in 3 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 in animals with indicated genotypes and treatment conditions. Penetrance and numbers of animals examined are noted. **(F)**, Model illustrating the dis-inhibitory regulatory pathway that mediates the transcriptional response to amygdalin.

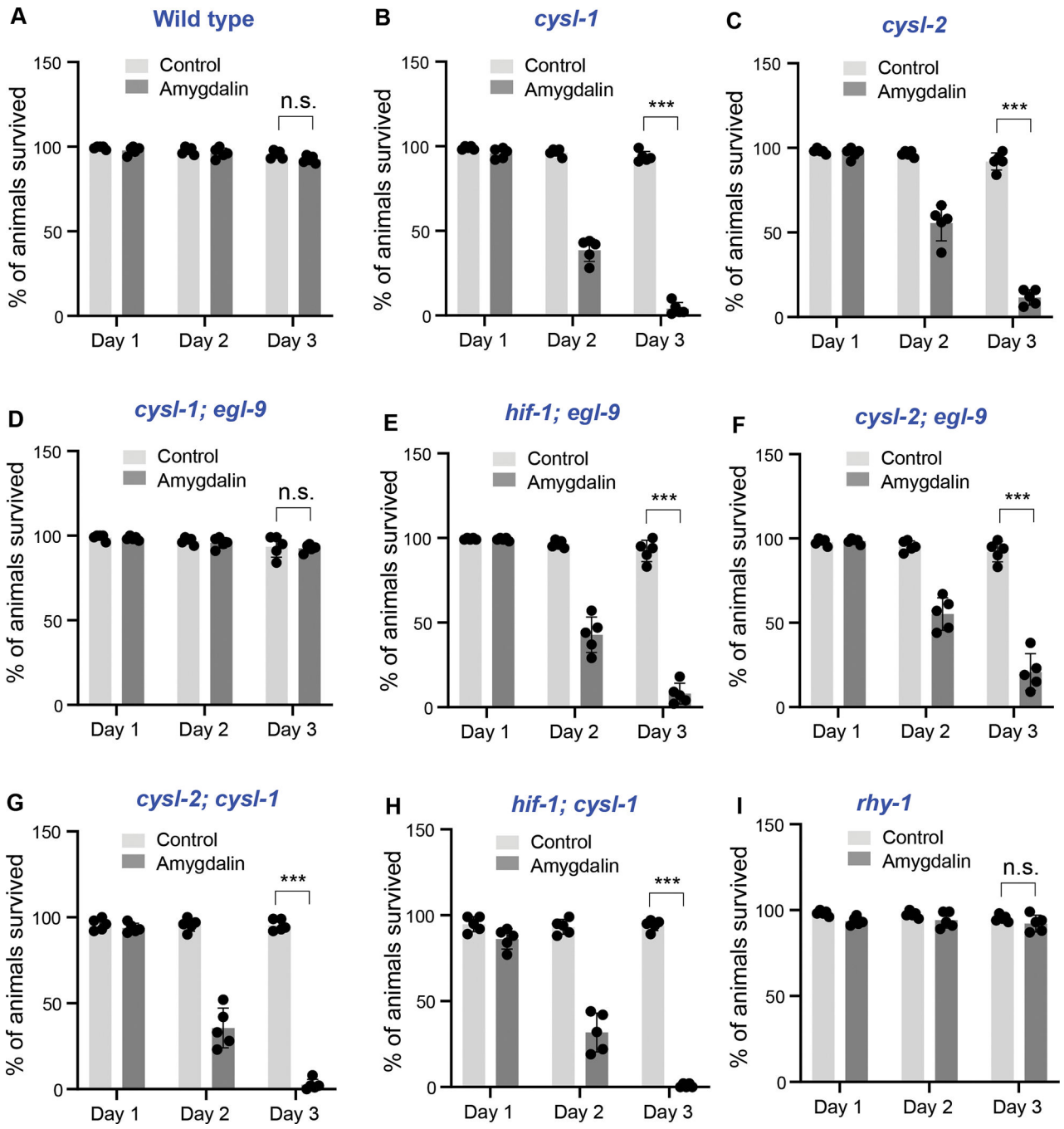


Figure 3. Amygdalin resistance requires CYSL-1, HIF-1 and CYSL-2 acting in a regulatory pathway.

(A), Quantification of percentages of post-L4 stage wild-type animals survived after 24, 48 or 72 h 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 h 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 h 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 h 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 h 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 h of treatment with amygdalin (10 mg/mL). **(G)**, Quantification of percentages of post-L4 stage *cysl-2*(*ok3516); cysl-1*(*ok762)* double loss-of-function mutant animals survived after 24, 48 or 72 h of treatment with amygdalin (10 mg/mL). **(H)**, Quantification of percentages of post-L4 stage *hif-1*(*ia04); cysl-1*(*ok762)* double loss-of-function mutant animals survived after 24, 48 or 72 h of treatment with amygdalin (10 mg/mL). **(I)**, Quantification of percentages of post-L4 stage *rhy-1*(*n5500)* loss-of-function mutant animals survived after 24, 48 or 72 h of treatment with amygdalin (10 mg/mL). All strains carry *nIs470 [cysl-2p::GFP + myo-2p::mCherry]*. Values are means ± S.D. with ***P < 0.001 (one-way ANOVA with post-hoc Tukey HSD, N = 5 independent experiments, n > 50 animals per experiment). n.s., non-significant.

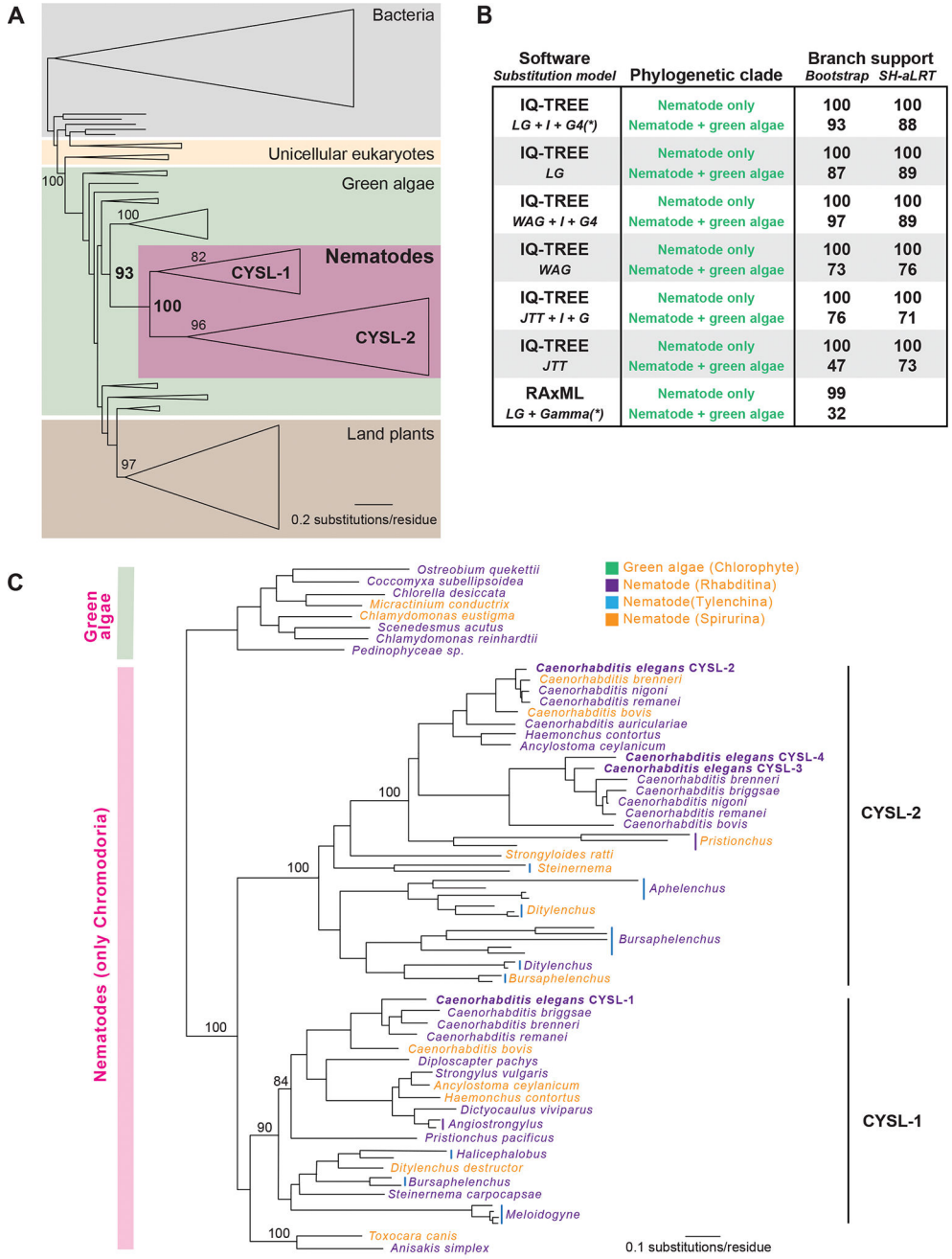


Figure 4. Nematode *cysl* genes were likely acquired from green algae by horizontal gene transfer.

(A), IQ-TREE generated maximum likelihood (ML) phylogenetic tree of homologs of nematode CYSL proteins 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 Bootstrap support values are shown at relevant nodes. For panels A-C, a complete list of all sequences used is in Data S2 and all trees, including support values, are found in Data S3. CYSL homologs in nematode species from the WGS (sequences from all genome sequencing projects) and TSA (sequences from transcriptome sequencing projects) databases using tBLASTn searches except those shown in Figure 4C are listed in Data S4.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Amygdalin	VWR	TCA0443-010G
Amygdalin	SIGMA	A6005-5G
Prunasin	CAYMAN	15959
Benzaldehyde	SIGMA	09143
Potassium cyanide	FISHER SCIENTIFIC	12136-30
Ethyl methanesulfonate	OAKWOOD PRODUCTS	001758-25G
DiscoveryProbe Bioactive Compound Library	APExBIO	L1022P
Critical Commercial Assays		
RNeasy Mini Kit	Qiagen	74004
Deposited Data		
RNAseq read datasets	NCBI SRA (Sequence Read Archive)	BioProject accession PRJNA843348
Experimental Models: Organisms/Strains		
C. elegans strain	CAENORHABDITIS GENETICS CENTER (CGC)	*cysl-2(ok3516)* II
C. elegans strain	CAENORHABDITIS GENETICS CENTER (CGC)	*nIs470 [cysl-2p::GFP + myo-2p::mCherry]* IV.
C. elegans strain	CAENORHABDITIS GENETICS CENTER (CGC) and this work	*egl-9(dma430, 432, 437, sa307)*
C. elegans strain	CAENORHABDITIS GENETICS CENTER (CGC)	*hif-1(ia4)*
C. elegans strain	CAENORHABDITIS GENETICS CENTER (CGC)	*cysl-1(ok762)*
Software and Algorithms		
hisat2	[45]	http://daehwankimlab.github.io/hisat2/
DESeq2	[46]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Clustal Omega	[48]	https://www.ebi.ac.uk/Tools/msa/clustalo/
CD-HIT	[49]	http://weizhong-cluster.ucsd.edu/cd-hit/
IQ-TREE	[50]	http://www.iqtree.org/
UFBoot2	[51]	https://academic.oup.com/mbe/article/35/2/518/4565479
ModelFinder	[52]	https://www.nature.com/articles/nmeth.4285