Natural variation in arsenic toxicity is explained by differences in branched chain amino acid catabolism

Stefan Zdraljevic[1,2], Bennett W. Fox[3], Christine Strand[5], Oishika Panda[3,4], Francisco J. Tenjo[3,4], Shannon C. Brady[1,2], Tim A. Crombie[2], John G. Doench[5], Frank C. Schroeder[3], and Erik C. Andersen[1,2,6] *

1. Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA
2. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA
3. Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA
4. The Buck Institute for Research on Aging, Novato, CA 94945, USA
5. Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
6. Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL 60611, USA
* Corresponding author

Erik C. Andersen
Assistant Professor of Molecular Biosciences
Northwestern University
Evanston, IL 60208, USA
Tel: (847) 467-4382
Fax: (847) 491-4461
Email: Erik.Andersen@Northwestern.edu

Stefan Zdraljevic, stefanzdraljevic2018@u.northwestern.edu, ORCID 0000-0003-2883-4616
Bennett William Fox, bwf7@cornell.edu, ORCID 0000-0002-9749-3491
Christine Strand, cstrand@broadinstitute.org, ORCID 0000-0001-5783-6667
Oishika Panda, op56@cornell.edu, ORCID 0000-0002-1217-8231
Francisco J. Tenjo, fjt38@cornell.edu, 0000-0003-4848-916X
Shannon C. Brady, shannonbrady2014@u.northwestern.edu, ORCID 0000-0002-3043-1544
Tim A. Crombie, tcrombie@northwestern.edu, ORCID 0000-0002-5645-4154
John G. Doench, jdoench@broadinstitute.org, ORCID 0000-0002-3707-9889
Frank C. Schroeder, fs31@cornell.edu, ORCID 0000-0002-4420-0237
Erik Andersen, erik.andersen@northwestern.edu, ORCID 0000-0003-0229-9651

Abstract

Organisms are often exposed to the environmentally ubiquitous toxic metalloid arsenic, and genetic differences unique to individuals can cause differential susceptibility to arsenic. To understand how molecular mechanisms of arsenic toxicity vary among individuals, we used two genetic mapping approaches to show that a major source of natural differences in *Caenorhabditis elegans* responses to arsenic trioxide is caused by variation in the *dbt-1* gene. This gene encodes the E2 subunit of the branched-chain α-keto acid dehydrogenase (BCKDH) complex, a core component of branched-chain amino acid (BCAA) catabolism. We used CRISPR/Cas9-mediated genome editing to show that a single non-synonymous variant (C78S) in the highly conserved lipoyl domain of DBT-1 is the causal polymorphism underlying variation in response to arsenic trioxide. Using targeted metabolomics and chemical supplementation experiments, we demonstrate that differences in *C. elegans* responses to arsenic trioxide are caused by variation in the abundances of iso-branched chain fatty acids that serve a central role in developmental progression. We hypothesize that the presence of the additional thiol group in the sensitive DBT-1 C78 allele participates in arsenic binding, thereby more strongly inhibiting BCKDH function. We go on to show that branched chain fatty acids are affected after arsenic treatment of human cells. This finding has broad implications for arsenic toxicity and for arsenic-focused chemotherapeutics across divergent individuals in human populations. Our study implicates the BCKDH complex and BCAA metabolism in arsenic responses, demonstrating the power of using *C. elegans* natural genetic diversity in combination with comparative metabolomics to identify mechanisms by which environmental toxins affect organismal physiology.

Introduction

An estimated 100 million people are currently at risk of chronic exposure to arsenic, a toxic metalloid that can be found in the environment (Ravenscroft, Brammer, and Richards 2009). The high prevalence of environmental arsenic and the severe toxicity associated with arsenic exposure has made it the number one priority for the United States Agency for Toxic Substances and Disease Registry (https://www.atsdr.cdc.gov/SPL/). Inorganic trivalent arsenic As(III) compounds, which include arsenic trioxide (As_2O_3), are the most toxic forms of environmental arsenic (Ratnaike 2003; Mandal and Suzuki 2002). In humans, As(III) is detoxified by methylation to the least toxic dimethylarsenite (DMA) form (Khairul et al. 2017; Stýblo et al. 2002). However, this methylation process also creates the highly toxic monomethylarsenite (MMA) form, so ratios of DMA to MMA determine levels of arsenic toxicity. Both MMA and DMA are produced from As(III) via the arsenic methyltransferase (AS3MT) (Schlebusch et al. 2015). Interestingly, individuals from human subpopulations that inhabit high arsenic environments have higher DMA/MMA ratios than individuals from low-arsenic environments. The elevated DMA/MMA ratio in individuals from high-arsenic environments is associated with natural differences in the *AS3MT* gene (Chung et al. 2009; Fujihara et al. 2009; Gomez-Rubio et al. 2010), which shows signs of strong positive selection. These results suggest that a more active AS3MT enzyme in these human subpopulations makes more DMA and enables adaptation to elevated environmental arsenic levels (Schlebusch et al. 2015). Importantly, population-wide differences in responses to environmental arsenic cannot be explained solely by variation in *AS3MT*, indicating that other genes must impact arsenic toxicity.

Despite its toxicity, arsenic trioxide has been used as a therapeutic agent for hundreds of years. Most recently, it was introduced as a highly effective cancer chemotherapeutic for the treatment of acute promyelocytic leukemia (APL) (Chen et al. 1997; Antman 2001; Murgo 2001; Emi 2017). Hematopoietic differentiation and apoptosis in APL patients is blocked at the level of promyelocytes by the Promyelocytic Leukemia/Retinoic Acid Receptor alpha fusion protein caused by a t(15;17) chromosomal translocation (de Thé et al. 1990; Grignani et al. 2000). Arsenic trioxide has been shown to directly bind a cysteine-rich region of the RING-B box coiled-coil domain of PML-RARα, which causes the degradation of the oncogenic fusion protein (Zhang et al. 2010; Tomita, Kiyoi, and Naoe 2013). The success of arsenic trioxide (Trisenox®) has spurred its use in over a hundred clinical trials in the past decade (Hoonjan, Jadhav, and Bhatt 2018). Despite these successes, individual differences in the responses to arsenic-based treatments, including patient-specific dosing regimens and side effects, limit the full therapeutic benefit of this compound (Zeidan and Gore 2014). Medical practitioners require knowledge of the molecular mechanisms for how arsenic causes toxicity to provide the best individual-based therapeutic benefits.

Studies of the free-living roundworm *Caenorhabditis elegans* have greatly facilitated our understanding of basic cellular processes (Kniazeva et al. 2004; Luz et al. 2017; Spracklin et al. 2017; Watson et al. 2013), including a number of studies that show that the effects of arsenic are similar to what is observed in mammalian model systems and humans. These effects include mitochondrial toxicity (Luz and Meyer 2016; Luz et al. 2016), the generation of reactive oxygen species (ROS) (Schmeisser et al. 2013), genotoxicity (Wyatt et al. 2017), genome-wide shifts in chromatin structure (Large et al. 2016), reduced lifespan (Schmeisser et al. 2013), and

105 induction of the heat-shock response (Wang et al. 2017). However, these studies were all
106 performed in the genetic background of the standard *C. elegans* laboratory strain (N2). To date,
107 152 *C. elegans* strains have been isolated from various locations around the world (Andersen et
108 al. 2012; Cook, Zdraljevic, Tanny, et al. 2016; Cook, Zdraljevic, Roberts, et al. 2016), which
109 contain a largely unexplored pool of genetic diversity much of which could underlie adaptive
110 responses to environmental perturbations (Zdraljevic and Andersen 2017).
111 We used two quantitative genetic mapping approaches to show that a major source of
112 variation in *C. elegans* responses to arsenic trioxide is caused by natural variation in the *dbt-1*
113 gene, which encodes an essential component of the highly conserved branched-chain α-keto
114 acid dehydrogenase (BCKDH) complex (Jia et al. 2016). The BCKDH complex is a core
115 component of branched-chain amino acid (BCAA) catabolism, which has not been previously
116 implicated in arsenic responses. Furthermore, we show that a single missense variant in DBT-
117 1(S78C), located in the highly conserved lipoyl-binding domain, underlies phenotypic variation in
118 response to arsenic. Using targeted and untargeted metabolomics and chemical rescue
119 experiments, we show that differences in wild isolate responses to arsenic trioxide are caused
120 by differential synthesis of mono-methyl branched chain fatty acids (mmBCFA), metabolites with
121 a central role in development (Kniazeva et al. 2004). These results demonstrate the power of
122 using the natural genetic diversity across the *C. elegans* species to identify mechanisms by
123 which environmental toxins affect physiology.
124
125 **Results**
126
127 **Natural variation on chromosome II underlies differences in arsenic responses**
128 We quantified arsenic trioxide sensitivity in *C. elegans* using a high-throughput fitness
129 assay that utilizes the COPAS BIOSORT (Andersen et al. 2015; Zdraljevic et al. 2017). In this
130 assay, three L4 larvae from each strain were sorted into arsenic trioxide or control conditions.
131 After four days of growth, we measured the brood size and progeny lengths of these animals
132 using the BIOSORT. We defined brood size as the total number of objects detected by the
133 BIOSORT and progeny length by the median of individual progeny time-of-flight (TOF)
134 measurements. To identify an appropriate concentration of arsenic trioxide for linkage and GWA
135 mapping assays, we performed dose-response experiments on four genetically diverged
136 isolates of *C. elegans*: N2, CB4856, JU776, and DL238 (Figure S1A). When compared to
137 control conditions, all four strains produced fewer progeny at all arsenic trioxide concentrations,
138 whereby the lowest concentration at which we observed a significant reduction in brood size for
139 all strains was 1 mM. In addition to reduced brood sizes, we observed that the progeny of
140 animals exposed to arsenic trioxide were shorter in length than the progeny of animals grown in
141 control conditions, which indicates an arsenic-induced developmental delay (Figure S1B). We
142 found that CB4856 animals produced approximately 16% more offspring that were on average
143 20% larger than JU775, DL238, and N2 animals when treated with 1 mM arsenic trioxide,
144 suggesting that the CB4856 strain was more resistant to arsenic trioxide than the other three
145 strains.
146 The increased arsenic trioxide resistance of CB4856 compared to N2 motivated us to
147 perform linkage mapping experiments with a panel of recombinant inbred advanced intercross
148 lines (RIAILs) that were previously constructed through ten generations of intercrossing

between an N2 derivative (QX1430) and CB4856 (Andersen et al. 2015). To capture arsenic trioxide-induced phenotypic differences in brood size and progeny length, we exposed a panel of 252 RIAILs to 1 mM arsenic trioxide. For both traits, we corrected for growth differences among RIAILs in control conditions and assay-to-assay variability using linear regression (see Materials and Methods). Linkage mapping analysis revealed that arsenic trioxide-induced variation in RIAIL brood size and progeny length are significantly associated with genetic variation on the center of chromosome II (Figure 1A). The brood-size quantitative trait locus (QTL) explains 9.4% of the variation in brood size among strains in the RIAIL panel, and the confidence interval spans from 6.04 to 9.30 Mb on chromosome II. This QTL completely encompasses the progeny-length QTL, which has a confidence interval spanning from 6.82 to 8.93 Mb on chromosome II and explains 24.9% of the variation in progeny length among the RIAILs. Matching our observations from the dose-response experiments, RIAILs with the CB4856 genotype at the peak marker for these QTL had larger brood sizes and longer progeny in the presence of arsenic trioxide than recombinant lines with the N2 genotype (Figure S2A-B). Because the confidence intervals of these QTL overlap, variation in brood size and progeny length in the RIAIL panel could be affected by the same genetic variant(s). However, each of these QTL confidence intervals span genomic regions greater than two megabases and contain hundreds of genes that vary between N2 and CB4856 (brood size: 567, progeny length: 373). Therefore, we constructed near-isogenic lines (NILs) to isolate and narrow each QTL in a controlled genetic background.

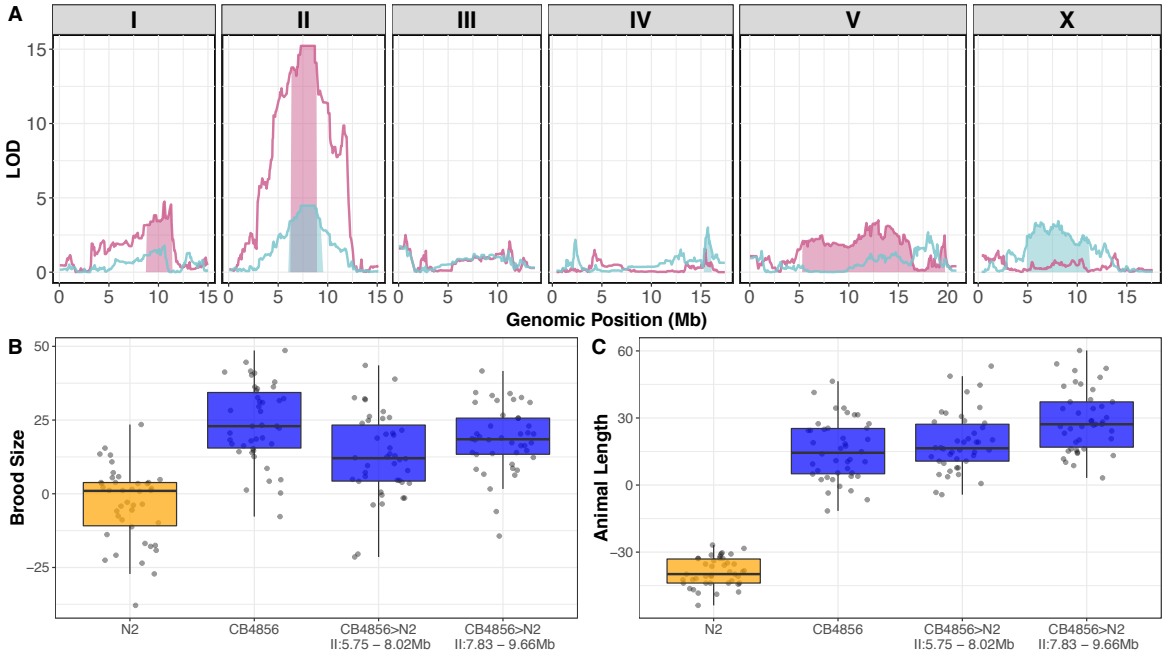


Figure 1: A large-effect QTL on the center of chromosome II explains differences in arsenic trioxide response between N2 and CB4856

(A) Linkage mapping plots for regressed brood size (teal) and median animal length (pink) in the presence of arsenic trioxide are shown. The significance values (logarithm of odds, LOD, ratio) for 1454 markers between the N2 and CB4856 strains are on the y-axis, and the genomic position (Mb) separated by chromosome is plotted on the x-axis. The associated 1.5 LOD-drop confidence intervals are represented by teal and pink boxes for brood size and median animal length, respectively. Tukey box plots of near-isogenic line (NIL) phenotype values for regressed (B) brood size and (C) median animal

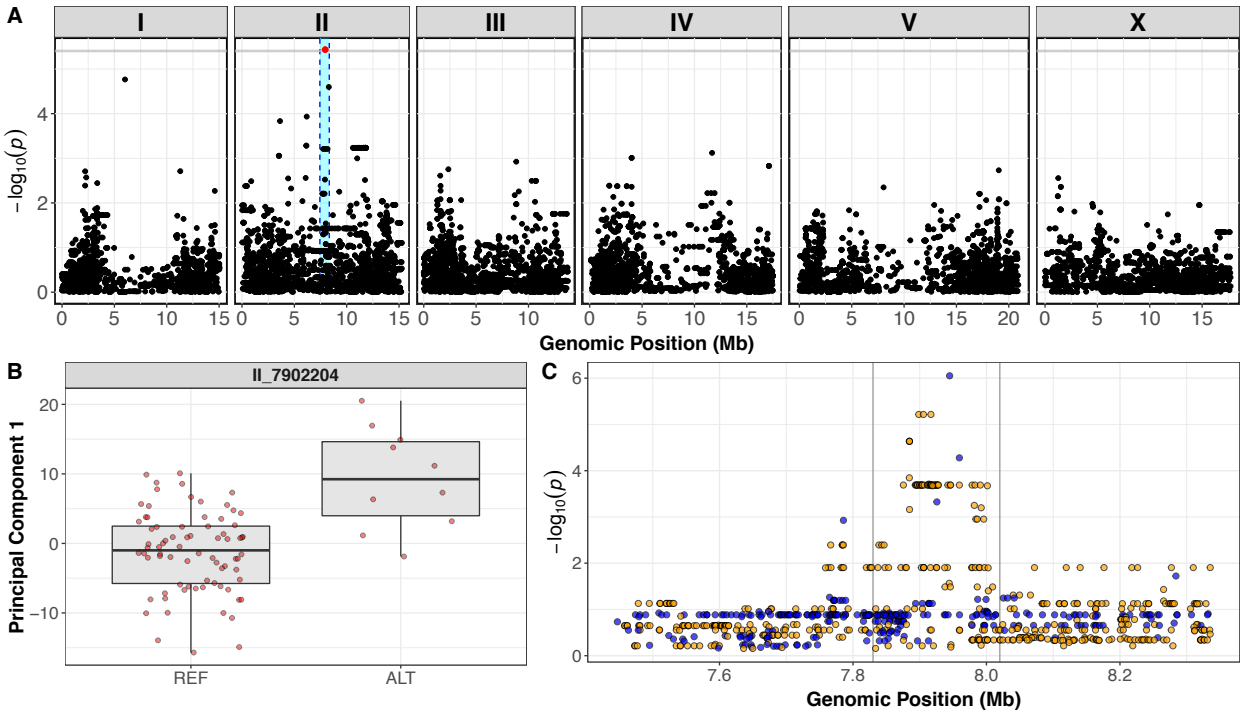
length in the presence of arsenic trioxide are shown. NIL genotypes are indicated below the plot as genomic ranges. The N2 brood sizes and progeny lengths are significantly different from all other strains (Brood size: Tukey HSD *p*-value < 1.56E-7; Animal Length: Tukey HSD *p*-value < 3.0E-14).

To split the large QTL confidence intervals in half, we introgressed genomic regions from the CB4856 strain on the left and right halves of the confidence intervals into the N2 genetic background. In the presence of arsenic trioxide, both of these NILs had similar brood sizes and progeny of similar length as the CB4856 parental strain (Figure 1B and C). The phenotypic similarity of these NILs to the CB4856 parental strain suggested that the two NILs might share an introgressed region of the CB4856 genome. To identify this shared introgressed region, we performed low-coverage sequencing of the NIL strains and defined the left and right bounds of the CB4856 genomic introgression to be from 5.75 to 8.02 Mb and 7.83 to 9.66 Mb in the left and right NILs, respectively (Supplemental table 9). Taken together, these results strongly suggested that genetic differences between N2 and CB4856 within 7.83 to 8.02 Mb on chromosome II conferred resistance to arsenic trioxide.

In parallel to the linkage-mapping approach described above, we performed a genome-wide association (GWA) mapping experiment by quantifying the responses to arsenic trioxide for 86 wild *C. elegans* strains (Andersen et al. 2012). We found no significant QTL associated with brood size nor progeny length measurements with this approach. However, the GWA experiment we performed had lower statistical power to detect QTL than the linkage mapping experiment, because we only tested 86 strains in single replicates. To circumvent this issue, we performed principal component analysis (PCA) on all traits measured by the COPAS BIOSORT and post-processed using the *easysorter* R package (Andersen et al. 2015). In total, 67 processed phenotype measurements were included in the PCA. This analysis revealed that eight principal components (PCs) explain more than 95% of the variation in the data set, and the first PC alone captures approximately 68% of the variation in the data set (Figure S3). Closer inspection of the first PC indicates that it is highly correlated with the progeny length trait, median TOF (r = 0.89, p-value = 4.7E-31) (Figure S4). We therefore considered the first PC as a progeny length-related phenotype for GWA mapping. Variation in the first PC among the wild isolates mapped to a QTL on the center of chromosome II that spans from 7.43 Mb to 8.33 Mb (Figure 2A). This QTL overlaps with the QTL identified by linkage mapping and follow-up NIL experiments, which suggested to us that the same causal genetic variant(s) underlies both QTL. Additionally, strains with the non-reference genotype at the peak SNV from the PC mapping tended to have higher PC1 values, which correlated with longer progeny lengths, than strains with the reference genotype when exposed to arsenic trioxide (Figure 2B). These results matched our observations that RIAILs with the CB4856 genotype at the QTL marker tended to be more resistant to arsenic trioxide than RIAILs with the N2 genotype at the QTL marker, further supporting that the QTL we identified by linkage and GWA mapping might be caused by the same underlying genetic variants. To explore this hypothesis further, we performed fine mapping within the PC1 QTL confidence interval identified by GWA mapping.

To fine map the PC1 QTL, we focused on variants from the *C. elegans* whole-genome variation dataset (Cook, Zdraljevic, Tanny, et al. 2016) that are shared among at least five percent of the 86 wild isolates exposed to arsenic trioxide. Under the assumption that the linkage and GWA mapping QTL are caused by the same genetic variation, we also considered variants present in the CB4856 strain. We calculated the p-value of Spearman's *rho* correlation

222 coefficients between the first PC and each single-nucleotide variant (SNV) in the QTL
223 confidence interval (Figure 2C). From this analysis, we found that a single SNV is highly
224 correlated with the arsenic trioxide response ($-log_{10}(p) = 6.05$). It affects the *dbt-1* gene, which
225 results in a cysteine to serine missense variant at position 78 (C78S) of DBT-1. This highly
226 correlated variant is also present in the NIL-defined QTL that spans from 7.83 to 8.02 Mb on
227 chromosome II.



228 **Figure 2: GWA mapping results of the first principal component**
229 A) A manhattan plot for the first principal component in the presence of arsenic trioxide is shown. Each
230 dot represents an SNV that is present in at least 5% of the assayed wild population. The genomic
231 position in Mb, separated by chromosome, is plotted on the x-axis and the $-log_{10}(p)$ for each SNV is
232 plotted on the y-axis. SNVs are colored red if they pass the genome-wide Bonferroni-corrected
233 significance threshold, which is denoted by the gray horizontal line. The genomic region of interest
234 surrounding the QTL is represented by a cyan rectangle. (B) Tukey box plots of phenotypes used for
235 association mapping in (A) are shown. Each dot corresponds to the phenotype of an individual strain,
236 which is plotted on the y-axis. Strains are grouped by their genotype at the peak QTL position (red SNV
237 from panel A, ChrII:7,902,204), where REF corresponds to the allele from the reference N2 strain. (C)
238 Fine mapping of the chromosome II region of interest, showing the calculated *p*-value associated with the
239 Spearman's *rho* statistic for every variant present in the QTL confidence interval from (A) is shown. Blue
240 dots correspond to variants that are present in the CB4856 strain, and orange dots correspond to variants
241 that are present in the rest of the wild population.
242
243 **A cysteine-to-serine variant in DBT-1 contributes to arsenic response variation**
244 The *C. elegans dbt-1* gene encodes the E2 component of the branched-chain α-keto
245 acid dehydrogenase complex (BCKDH) (Jia et al. 2016). The BCKDH complex is a core
246 component of branched-chain amino acid (BCAA) catabolism and catalyzes the irreversible
247 oxidative decarboxylation of amino acid-derived branched-chain α-ketoacids (Adeva-Andany et
248 al. 2017). The BCKDH complex belongs to a family of α-ketoacid dehydrogenases that includes

249 pyruvate dehydrogenase (PDH) and α-ketoglutarate dehydrogenase (KGDH) (Bergquist et al.
250 2009). All three of these massive enzymatic complexes include a central E2 component that is
251 lipoylated at one critical lysine residue (two residues in PDH). The function of these enzymatic
252 complexes depends on the lipoylation of these lysine residues (Bergquist et al. 2009; Reed and
253 Hackert 1990). In *C. elegans*, the putative lipoylated lysine residue is located at amino acid
254 position 71 of DBT-1, which is in close proximity to the C78S residue that we found to be highly
255 correlated with arsenic trioxide resistance.

256 To confirm that the C78S variant in DBT-1 contributes to differential arsenic trioxide
257 responses, we used CRISPR/Cas9-mediated genome editing to generate allele-replacement
258 strains by changing the C78 residue in the N2 strain to a serine and the S78 residue in the
259 CB4856 strain to a cysteine. Arsenic trioxide treatment of these strains revealed that the N2
260 DBT-1(S78) allele-replacement strain had increased resistance to arsenic trioxide, as measured
261 by brood size (Figure 3A) and progeny lengths (Figure 3B), when compared to the parental N2
262 DBT-1(C78) strain. Conversely, the CB4856 DBT-1(C78) allele-replacement strain had
263 decreased resistance to arsenic trioxide, as measured by brood size (Figure 3A) and progeny
264 lengths (Figure 3B), when compared to the parental CB4856 DBT-1(S78) strain. We obtained
265 the same result when we used the first PC from principal component analysis as a trait (Figure
266 S5). When considering brood size, the N2 DBT-1(C78S) allele-replacement strain produced an
267 intermediate number of progeny in the presence of arsenic trioxide relative to the parental N2
268 and CB4856 strains. However, the CB4856 DBT-1(S78C) allele-replacement strain produced
269 fewer offspring than both parental strains (Figure 3A). Interestingly, the N2 DBT-1(S78) strain
270 had the same level of resistance to arsenic trioxide, as measured by progeny length, as the
271 parental CB4856 strain. However, the CB4856 DBT-1(C78) strain was still more resistant to
272 arsenic trioxide than the parental N2 strain (Figure 3B). These results suggested that additional
273 genetic differences between the N2 and CB4856 strains might interact with the DBT-1(C78S)
274 allele. Nevertheless, these results functionally validate the DBT-1 C78S variant underlies
275 differences in physiological responses in the presence arsenic trioxide.

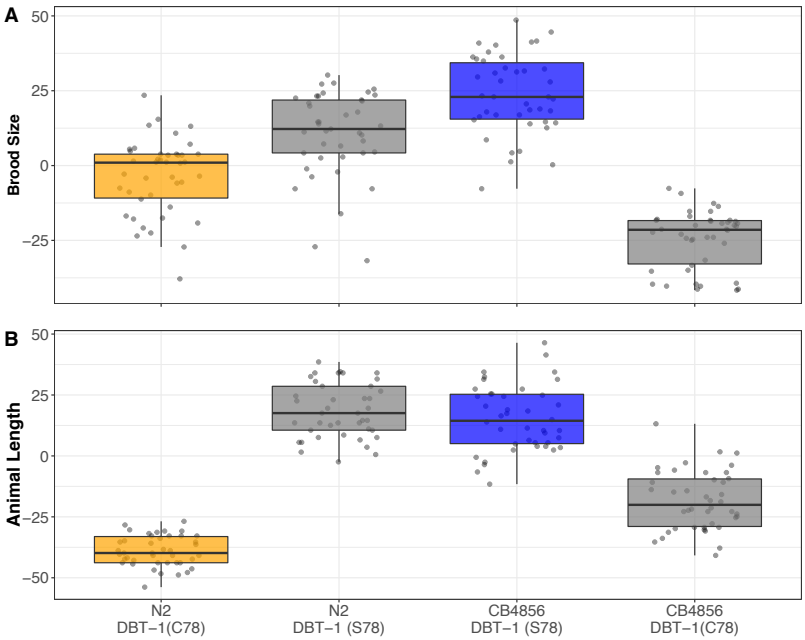


Figure 3: The DBT-1 C78S variant underlies arsenic trioxide sensitivity in *C. elegans.*
Tukey box plots of residual (A) brood size and (B) median animal length after arsenic trioxide exposure are shown (N2, orange; CB4856, blue; allele replacement strains, gray). Labels correspond to the genetic background and the corresponding residue at position 78 of DBT-1 (C for cysteine, S for serine). Every pair-wise strain comparison is significant except for the N2 DBT-1(S78) - CB4856 comparison from animal length (Brood size: Tukey's HSD *p*-value = 1.54E-5; Animal length: *p*-value < 1E-14)***.***

Arsenic trioxide inhibits the DBT-1 C78 allele

Mono-methyl branched chain fatty acids (mmBCFA) are an important class of molecules that are produced via BCAA catabolism (Jia et al. 2016; Kniazeva, Euler, and Han 2008; Kniazeva et al. 2004; Baugh 2013). The production of mmBCFA requires the BCKDH, fatty acid synthase (FASN-1), acetyl-CoA carboxylase (POD-2), fatty acyl elongases (ELO-5/6), β-ketoacyl dehydratase (LET-767), and acyl CoA synthetase (ACS-1) (Watts and Ristow 2017; Entchev et al. 2008; Kniazeva et al. 2004; Kniazeva, Euler, and Han 2008; Jia et al. 2016; Zhu et al. 2013). Strains that lack functional *elo-5*, *elo-6*, or *dbt-1* produce less C15ISO and C17ISO mmBCFAs, arrest at the L1 larval stage, and can be rescued by supplementing the growth media with C15ISO or C17ISO (Jia et al. 2016; Kniazeva, Euler, and Han 2008; Kniazeva et al. 2004) (Figure 4A).

Because DBT-1 is involved in BCAA catabolism, we hypothesized that the DBT-1(C78S)-dependent difference in progeny length between N2 and CB4856 upon arsenic trioxide treatment might be caused by depletion of downstream mmBCFAs. To test this hypothesis, we quantified the abundance of the isomer (ISO) and straight-chain forms of C15 and C17 in the N2, CB4856, and allele-swap genetic backgrounds. We measured the metabolite levels in staged L1 animals and normalized the detected amounts of C15ISO and C17ISO relative to the abundances of straight-chain C15 and C17, respectively, to mitigate confounding effects of differences in developmental speeds that could result from genetic background differences or arsenic trioxide exposure. As expected, we observed a reduction of C15ISO in arsenic-treated animals (Figure 4B). Although C15ISO was more abundant in N2 than in CB4856 animals, arsenic reduced C15ISO production more strongly in N2 than in CB4856, in agreement with our physiological observations that CB4856 was more resistant to arsenic trioxide (Figure 3). The fact that N2 animals had an approximately 1.75-fold higher ISO/straight-chain ratio in mock conditions than CB4856 animals suggested that N2 might require higher levels of mmBCFA and thus be more susceptible to reduction in these fatty acids.

We also quantified C15ISO amounts in the allele-replacement strains N2 DBT-1(S78) and CB4856 DBT-1(C78) when exposed to arsenic (Figure 4B). In both strains, arsenic exposure was correlated with reduced C15ISO production. In agreement with the predicted outcome, N2 DBT1(S78) was less impacted by arsenic than the N2 strain. By contrast, CB4856 DBT-1(C78) was more strongly affected by arsenic than the CB4856 strain. Correspondingly, the observed reduction of C15ISO by arsenic treatment appeared to be greater in CB4856 DBT-1(C78) carrying the N2 allele of DBT-1 (47% decrease; Tukey HSD *p*-value = 0.019) than in N2 DBT-1(S78) (31% decrease; Tukey HSD *p*-value = 0.26). Interestingly, production of C15ISO was significantly increased in CB4856 DBT-1(C78) compared to the parent strain CB4856, whereas C15ISO levels in N2 DBT-1(S78) were not significantly different from those measured for N2. We observed similar trends for the C17ISO/straight-chain ratio (Figure S6) and raw abundances of C15ISO and C17ISO (Figure S7). Notably, we did not detect any significant

differences between the levels of C15 and C17 straight-chain fatty acids, indicating that the effect of the DBT-1 C78S variant was specific to branched-chain fatty acids (Figure S8). These results, in combination with our physiological observations that the C78S allele, explained the majority of the physiological differences between N2 and CB4856 in the presence of arsenic trioxide (Figure 3) and suggested that the DBT-1(C78) allele was inhibited by arsenic trioxide more strongly than DBT-1(S78).

Next, we investigated if mmBCFA supplementation of the growth media used in our physiological assay could rescue the effects of arsenic trioxide toxicity. We exposed the parental and the DBT-1 allele-replacement strains to media containing arsenic trioxide alone, C15ISO alone, or a combination of arsenic trioxide and C15ISO. Median progeny length of arsenic-treated N2 and CB4856 DBT-1(S78C) strains increased by 9% and 15%, respectively, when supplemented with C15ISO (Figure 4C, Figure S9). These C15ISO-derived increases in body length correspond to a 38% and 63% rescue of the allele-specific arsenic affect in the respective genetic backgrounds. By contrast, CB4856 and N2 DBT-1(C78S) median progeny lengths were unaffected by C15ISO supplementation in arsenic trioxide media. Collectively, these data support the hypothesis that the cysteine/serine variant in DBT-1 contributes to arsenic sensitivity in *C. elegans*, and the DBT-1(C78) variant can be partially rescued by supplementation with mmBCFAs.

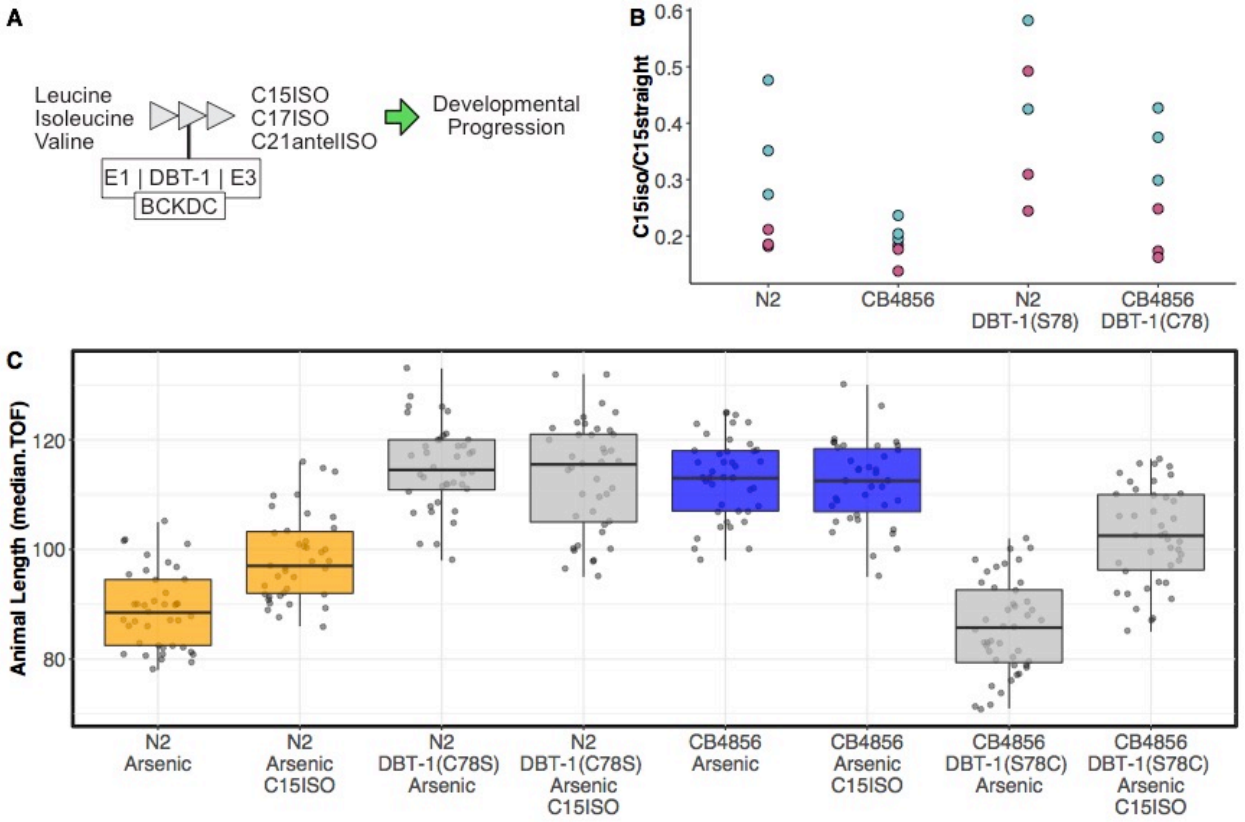


Figure 4: Differential production of mmBCFA underlies DBT-1(C78)-mediated sensitivity to arsenic trioxide.

(A) A simplified model of BCAA catabolism in *C. elegans*. The BCKDH complex, which consists of DBT-1, catalyzes the irreversible oxidative decarboxylation of branched-chain ketoacids. The products of this

344 breakdown can then serve as building blocks for the mmBCFA that are required for developmental
345 progression. (B) The ratio of C15ISO/C15straight-chain is plotted on the y-axis for three independent
346 replicates of the N2, CB4856, and allele swap strains exposed to control (teal) or 100 μM arsenic trioxide
347 (pink) conditions. The difference between N2 control and N2 arsenic trioxide conditions is statistically
348 significant (Tukey HSD *p*-value = 0.043), but the difference between CB4856 mock and arsenic
349 conditions is not (Tukey HSD *p*-value = 0.075). The difference between N2 swap control and arsenic
350 trioxide conditions is not statistically significant (Tukey HSD *p*-value = 0.26) and the difference between
351 CB4856 swap mock and arsenic conditions is significant (Tukey HSD *p*-value = 0.019). (C) Tukey box
352 plots median animal length after arsenic trioxide or arsenic trioxide and 0.64 μM C15ISO exposure are
353 shown (N2, orange; CB4856, blue; allele replacement strains, gray). Labels correspond to the genetic
354 background and the corresponding residue at position 78 of DBT-1 (C for cysteine, S for serine). Every
355 pair-wise strain comparison is significant except for the N2 DBT-1(S78) - CB4856 comparisons (Tukey's
356 HSD *p*-value < 1.43E-6).
357
358 **Arsenic exposure increases mmBCFA production and favors a cysteine allele in human**
359 **DBT1**
360 To test whether our results from *C. elegans* translate to human variation in arsenic
361 sensitivity, we tested the role of DBT1 variation on arsenic trioxide responses and mmBCFA
362 biosynthesis in human cells. The human DBT1 enzyme contains a serine at position 112 that
363 corresponds to the C78 residue in *C. elegans* DBT-1 (Figure 5A). Using CRISPR/Cas9, we
364 edited batch cultures of 293T cells to generate a subset of cells with DBT1(S112C). These cells
365 were exposed to arsenic trioxide or control conditions and we monitored changes in the fraction
366 of cells carrying the DBT1(C112) allele. We found that arsenic exposure caused a 4% increase
367 in the fraction of cells that contained the DBT1(C112) allele (Figure 5B). Though the human
368 DBT1 does not vary within the human population at S112, two residues in close spatial
369 proximity to S112 do vary among individuals in the population (R113C and W84C) (Figure 5A)
370 (Forbes et al. 2008). To test the effects of these residues on arsenic trioxide sensitivity, we
371 performed the same editing and arsenic selection procedure described above. Over the course
372 of the selection experiment, cells with DBT1(W84C) and DBT1(R113C) increased by 2% and
373 1%, respectively (Figure 5B). All three cysteine edits caused an increase in fitness in batch-
374 edited human cell cultures exposed to arsenic. To determine if branched-chain fatty acid
375 production was altered by arsenic exposure, as we found in *C. elegans*, we measured mmBCFA
376 production in unedited 293T cells in arsenic and mock-treated cultures. We found that overall
377 fatty acid production was markedly reduced in arsenic-treated cultures. In contrast to our
378 observations in *C. elegans*, straight-chain fatty acids were more drastically reduced than ISO
379 fatty acids (Supplemental data 24). These results are consistent with our observations that
380 arsenic does not have as strong of an effect on *C. elegans* strains with the DBT-1 S78 allele.
381 Additionally, the DBT1(C112) allele, similar to the *C. elegans* DBT-1(C78) allele, was associated
382 with higher production of mmBCFAs.
383
384
385
386
387
388
389

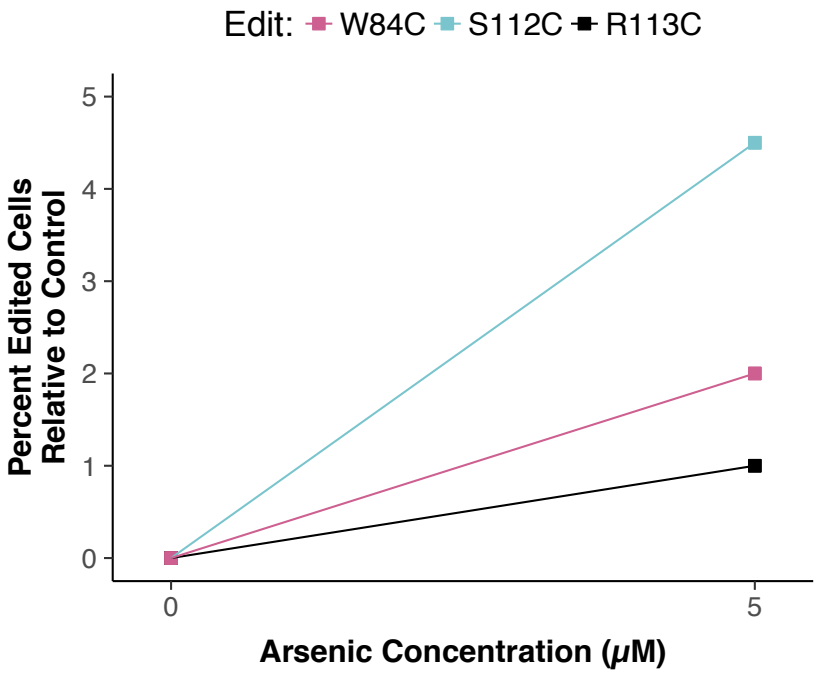


Figure 5: Protective effect of cysteine residues in human DBT1.
A) Alignment of *C. elegans* DBT-1 and *H. sapiens* DBT1. The residues tested for an arsenic-specific effect are indicated with arrows - W84C (pink), S112C (blue), and R113C (black). The lysine that is post-translationally modified with a lipoid acid is highlighted in red. B) The fraction of human cells that contain the W84C, S112C, or R113C amino acid change in DBT1 in the presence of 0 or 5 μM arsenic trioxide are shown. The fraction of edited cells is normalized to the fraction edited in the 0 μM arsenic condition. Each dot corresponds to the mean of two replicate experiments.

Discussion

In this study, we characterized the effects of *C. elegans* natural genetic variation on physiological responses to the pervasive environmental toxin arsenic trioxide. Though the effects of this toxin have been extensively studied in a variety of systems (Bergquist et al. 2009; Ratnaike 2003; Mandal and Suzuki 2002; Paul et al. 2014; Shen et al. 2013), recent evidence from human population studies have revealed local adaptations within region-specific subpopulations (Schlebusch et al. 2015; Fujihara et al. 2009; Gomez-Rubio et al. 2010; Li et al. 2017). Our investigation into the natural variation in *C. elegans* responses to arsenic trioxide led to the discovery of a novel mechanism by which this compound could elicit toxicity. We show that arsenic trioxide differentially inhibits two natural alleles of the E2 domain of the BCKDH complex, which is encoded by the *dbt-1* gene. Specifically, strains with the DBT-1(C78) allele are more sensitive to arsenic trioxide than strains carrying the DBT-1(S78) allele, as measured by progeny production and growth (Figure 3). Furthermore, we show that the increased sensitivity of the DBT-1(C78) allele is largely explained by differences in the production of mmBCFAs (Figure 4B-C), which are critical molecules for developmental progression beyond the first larval stage. Arsenic is thought to inhibit the activity of both the pyruvate dehydrogenase

416 (PDH) and the α-ketoglutarate (KGDH) dehydrogenase complexes through interactions with the
417 reduced form of lipoate (Bergquist et al. 2009), which is the cofactor for the E2 domain of these
418 complexes. Like the PDH and KGDH complexes, the E2 domain of BCKDH complex requires
419 the cofactor lipoate to perform its enzymatic function (Pettit, Yeaman, and Reed 1978;
420 Heffelfinger, Sewell, and Danner 1983; Yeaman 1989). The inhibition of DBT-1 by arsenic
421 trioxide could involve three-point coordination of arsenic by the C78 thiol and the reduced thiol
422 groups of the nearby lipoate. However, based on the crystal structure (PDB:1Y8N), the atomic
423 distance between the C78 thiol group and the thiol groups from the lipoylated lysine is ~32 Å,
424 which might be too large a distance for coordinating arsenic (Kato et al. 2005) (Figure S10).
425 Alternatively, arsenic trioxide could inhibit DBT-1(C78) through coordination between the thiol
426 groups of C78 and C65 (~8.5 Å) (Figure S10). Analogous thiol-dependent mechanisms have
427 been proposed for the inhibition of other enzymes by arsenic (Shen et al. 2013). Despite
428 structural similarities and a shared cofactor, no evidence in the literature indicates that BCKDH
429 is inhibited by arsenic trioxide, so these results demonstrate the first connection of arsenic
430 toxicity to BCKDH E2 subunit inhibition.
431 Multiple sequence alignments show that cysteine residues C65 and C78 of *C. elegans*
432 DBT-1 correspond to residues S112 and C99 of human DBT1 (Figure 5A). Though DBT1 does
433 not vary at position 112 within the human population, two residues (R113C and W84C) in close
434 spatial proximity do (Forbes et al. 2008). We hypothesized that cysteine variants in DBT1 would
435 sensitize human cells to arsenic trioxide. Interestingly, we found that the cysteine variants
436 (W84C, S112C, and R113C) proliferated more rapidly than the parental cells in the presence of
437 arsenic. In agreement with these results, a growing body of evidence suggests that certain
438 cancer cells upregulate components involved in BCAA, and this upregulation promotes tumor
439 growth (Burrage et al. 2014; Tönjes et al. 2013). Perhaps the increased proliferation of human
440 cell lines that contain the DBT1 C112 allele (Figure 5) is caused by increased activity of the
441 BCKDH complex. In agreement, we found that *C. elegans* strains with the DBT-1 C78 allele
442 produce more mmBCFAs than strains with the S78 allele (Figure 4b and Figure S6). It is worth
443 noting that human cell lines grown in culture do not have the same strict requirements for
444 mmBCFA, and the requirements for different fatty acids are variable among diverse
445 immortalized cell lines (Hughes-Fulford, Chen, and Tjandrawinata 2001; Agostini et al. 2004).
446 Furthermore, in *C. elegans*, the developmental defects associated with *dbt-1* loss-of-function
447 can be rescued by neuronal-specific expression of *dbt-1* (Jia et al. 2016), suggesting that the
448 physiological requirements of mmBCFA in *C. elegans* depend on the coordination of multiple
449 tissues that cannot be recapitulated with cell-culture experiments. These results further highlight
450 the complexity of arsenic toxicity as well as the physiological requirements of BCAA within and
451 across species and could explain the discrepancy between the physiological effects we
452 observed in *C. elegans* and human cell-line experiments. Given that arsenic trioxide has
453 become the standard-of-care for treating AML (Coombs, Tavakkoli, and Tallman 2015) and is
454 gaining traction in treating other leukemias, our results indicate that further exploration of the
455 effects of arsenic and BCAA catabolism on cancer growth is essential.
456 The C78 allele of DBT-1 is likely the derived allele in the *C. elegans* population because
457 all other organisms have a serine at the corresponding position. The loss of the serine allele in
458 the *C. elegans* population might have been caused by relaxed selection at this locus. It is
459 hypothesized that the *C. elegans* species originated from the Pacific Rim and that the ancestral

460 state more closely resembles the CB5846 strain than the N2 strain (Andersen et al. 2012;
461 Thompson et al. 2015). The CB4856 strain was isolated from the Hawaiian island of O'ahu
462 (Hodgkin and Doniach 1997). The Hawaiian islands have elevated levels of arsenic in the soil
463 from volcanic activity, the farming of cane sugar, former construction material (canec)
464 production facilities, or wood treatment plants (Hawaii.gov). It is possible that as the *C. elegans*
465 species spread across the globe into areas with lower levels of arsenic in the soil and water, the
466 selective pressure to maintain high arsenic tolerance was relaxed and the cysteine allele
467 appeared. Alternatively, higher mmBCFA levels in strains with the DBT-1(C78) allele (Figure
468 S7) might cause faster development in certain conditions, although we did not observe allele-
469 specific growth differences in laboratory conditions. Despite these clues suggesting selection in
470 local environments, the genomic region surrounding the *dbt-1* locus does not appear to have a
471 strong signature of selection as measured by Tajima's D (Tajima 1989) (Figure S11), and the
472 strains with the DBT-1 S78 allele show no signs of geographic clustering (Figure S12). In the
473 future, it will be important to sample additional *C. elegans* wild strains and measure arsenic
474 levels in the natural substrates to further explore this possible connection. Our study suggests
475 that *C. elegans* is a powerful model to investigate the molecular mechanisms for how
476 populations respond to environmental toxins.
477
478 **Materials and methods**
479
480 **Strains**
481 Animals were fed the bacterial strain OP50 and grown at 20°C on modified nematode growth
482 medium (NGM), containing 1% agar and 0.7% agarose to prevent burrowing of the wild isolates
483 (Boyd, Smith, and Freedman 2012). For each assay, strains were grown for five generations
484 with no strain entering starvation or encountering dauer-inducing conditions (Andersen et al.
485 2014). Wild *C. elegans* isolates used for genome-wide association and recombinant inbred
486 advanced intercross lines (RIAILs) used for linkage mapping have been described previously
487 (Cook, Zdraljevic, Tanny, et al. 2016; Cook, Zdraljevic, Roberts, et al. 2016; Andersen et al.
488 2015). Strains constructed for this manuscript are listed in Supplemental table 1.
489
490 **High-throughput fitness assay**
491 We used the high-throughput fitness assay (HTA) described previously (Andersen et al. 2015).
492 In short, strains are passaged for four generations before bleach-synchronization and aliquoted
493 to 96-well microtiter plates at approximately one embryo per microliter in K medium (Boyd,
494 Smith, and Freedman 2012). The final concentration of NaCl in the K medium for the genome-
495 wide association (GWA) and linkage mapping assays was 51 mM. For all subsequent
496 experiments the final NaCl concentration was 10.2 mM. The following day, hatched and
497 synchronized L1 animals were fed HB101 bacterial lysate (Pennsylvania State University
498 Shared Fermentation Facility, State College, PA, (García-González et al. 2017)) at a final
499 concentration of 5 mg/ml and grown for two days at 20ºC. Next, three L4 larvae were sorted
500 using a large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston, MA) into
501 microtiter plates that contain HB101 lysate at 10 mg/ml, K medium, 31.25 µM kanamycin, and
502 either drug dissolved in 1% water or 1% water alone. The animals were then grown for four

503 days at 20ºC. The resulting populations were treated with sodium azide (50 mM) prior to being
504 measured with the BIOSORT.
505
506 **Calculation of fitness traits for genetic mappings**
507 Phenotype data generated using the BIOSORT were processed using the R package
508 *easysorter*, which was specifically developed for processing this type of data (Shimko and
509 Andersen 2014). Briefly, the function *read_data*, reads in raw phenotype data and runs a
510 support vector machine to identify and eliminate bubbles. Next, the *remove_contamination*
511 function eliminates any wells that were identified as contaminated prior to scoring population
512 parameters. The *sumplate* function is then used to generate summary statistics of the measured
513 parameters for each animal in each well. These summary statistics include the 10th, 25th, 50th,
514 75th, and 90th quantiles for time of flight (TOF), which corresponds to animal length. Measured
515 brood sizes (n) are normalized by the number of animals that were originally sorted into each
516 well (norm.n). After summary statistics for each well are calculated, we accounted for
517 differences between assays using the the *regress(assay=TRUE)* function in the *easysorter*
518 package. Outliers in the GWA and linkage mapping experiments were identified and eliminated
519 using the *bamf_prune* function in *easysorter*. For follow-up experiments that contained multiple
520 replicates for each strain, we eliminated strain replicates that were more than two standard
521 deviations from the strain mean for each condition tested. Finally, drug-specific effects were
522 calculated using the *regress(assay=FALSE)* function from *easysorter*, which accounts for strain-
523 specific differences in growth parameters present in control conditions.
524
525 **Arsenic trioxide dose-response assays**
526 All dose-response experiments were performed on four genetically diverged strains (N2,
527 CB4856, DL238, and JU258) in technical quadruplicates prior to performing GWA and linkage
528 mapping experiments (Supplemental data 1). Animals were assayed using the HTA, and
529 phenotypic analyses were performed as described above. The arsenic trioxide concentration for
530 GWA and linkage mapping experiments was chosen based on an observable effect for animal
531 length and brood size phenotypes in the presence of arsenic.
532
533 **Linkage mapping**
534 A total of 262 RIAILs were phenotyped in the HTA described previously for control and arsenic
535 trioxide conditions (Andersen et al. 2015; Zdraljevic et al. 2017). The phenotype and genotype
536 data were entered into R and scaled to have a mean of zero and a variance of one for linkage
537 analysis (Supplemental data 2-4). Quantitative trait loci (QTL) were detected by calculating
538 logarithm of odds (LOD) scores for each marker and each trait as $-n(ln(1-r^2)/2ln(10))$,
539 where r is the Pearson correlation coefficient between RIAIL genotypes at the marker and
540 phenotype trait values (Bloom et al. 2013). The maximum LOD score for each chromosome for
541 each trait was retained from three iterations of linkage mappings (Supplemental data 5). We
542 randomly permuted the phenotype values of each RIAIL while maintaining correlation structure
543 among phenotypes 1000 times to estimate the significance threshold empirically. The
544 significance threshold was set using a genome-wide error rate of 5%. Confidence intervals were
545 defined as the regions contained within a 1.5 LOD drop from the maximum LOD score (Broman
546 et al. 2003).

547
548 **Near-isogenic line (NIL) generation**
549 NILs were generated by crossing N2xCB4856 RIAILs to each parental genotype. For each NIL,
550 eight crosses were performed followed by six generations of selfing to homozygose the
551 genome. Reagents used to generate NILs are detailed in Supplemental table 1 and 2. The NILs
552 responses to 1000 μM arsenic trioxide were quantified using the HTA described above
553 (Supplemental data 6-8). NIL whole-genome sequencing and analysis was performed as
554 described previously (Brady et al. 2018) (Supplemental data 9).
555
556 **Genome-wide association mapping**
557 Genome-wide association (GWA) mapping was performed using phenotype data from 86
558 *C. elegans* isotypes (Supplemental data 10-13). We used the *cegwas* R package for association
559 mapping and post-mapping processing (Cook, Zdraljevic, Roberts, et al. 2016). This package
560 uses the EMMA algorithm for performing association mapping and correcting for population
561 structure (Kang et al. 2008), which is implemented by the *GWAS* function in the *rrBLUP*
562 package (Endelman 2011). Specifically, the *GWAS* function in the *rrBLUP* package was called
563 with the following command: *rrBLUP::GWAS(pheno = ph, geno = y, K = kin, min.MAF = 0.05,*
564 *n.core = 1, P3D = FALSE, plot = FALSE)*. The kinship matrix used for association mapping was
565 generated using a whole-genome high-quality single-nucleotide variant (SNV) set from CeNDR
566 release 20160408 (Cook, Zdraljevic, Tanny, et al. 2016; Zdraljevic et al. 2017; Evans et al.
567 2017) and the *A.mat* function from the *rrBLUP* package. SNVs previously identified using RAD-
568 seq (Andersen et al. 2012) that had at least 5% minor allele frequency in the 86 isotype set
569 were used for performing GWA mappings. Association mappings that contained at least one
570 SNV that had a value greater than the Bonferroni-corrected value were processed further using
571 fine mapping. Brood size and animal length significance scores can be found in Supplemental
572 data 14.
573
574 **Principal component analysis**
575 The COPAS BIOSORT quantifies individual animal length (TOF), optical density (EXT),
576 fluorescence (green, yellow, and red), and total animals in a well normalized by the number of
577 animals initially sorted into the well (brood size). All of these measurements were then
578 summarized using the *easysorter* package to generate various summary statistics of each raw
579 BIOSORT parameter, including five distribution quantiles and measures of dispersion (Andersen
580 et al. 2015). Prior to principal component analysis (PCA), HTA phenotypes were scaled to have
581 a mean of zero and a standard deviation of one using the *scale* function in R. PCA was
582 performed using the *prcomp* function in R (Team 2017) (Supplemental data 15). The variance
583 explained for each principal components is located in Supplemental data 16. Genome-wide
584 mapping results using the first principal component are available in Supplemental data 17.
585
586 **Fine mapping**
587 We previously described the identification of QTL confidence intervals and fine-mapping
588 (Zdraljevic et al. 2017). In short, we used the *process_correlations()* and the
589 *variant_correlation()* functions in the *cegwas* package. These functions calculate the *p*-value
590 associated with the Spearman's correlation coefficient between each variant in the QTL

591 confidence interval and the kinship-corrected phenotype used in the GWA (Supplemental data
592 18).
593
594 **Generation of *dbt-1* allele replacement strains**
595 Allele replacement strains were generated using CRISPR/Cas9-mediated genome editing, using
596 the co-CRISPR approach (Kim et al. 2014) with Cas9 ribonucleoprotein delivery (Paix et al.
597 2015). Alt-R^{TM} crRNA and tracrRNA (Supplemental table 2) were purchased from IDT (Skokie,
598 IL). tracrRNA (IDT, 1072532) was injected at a concentration of 13.6 μM. The *dpy-10* and the
599 *dbt-1* crRNAs were injected at 4 μM and 9.6 μM, respectively. The *dpy-10* and the *dbt-1* single-
600 stranded oligodeoxynucleotides (ssODN) repair templates were injected at 1.34 μM and 4 μM,
601 respectively. Cas9 protein (IDT, 1074182) was injected at 23 μM. To generate injection mixes,
602 the tracrRNA and crRNAs were incubated at 95ºC for five minutes and 10ºC for 10 minutes.
603 Next, Cas9 protein was added and incubated for five minutes at room temperature. Finally,
604 repair templates and nuclease-free water were added to the mixtures and loaded into pulled
605 injection needles (1B100F-4, World Precision Instruments, Sarasota, FL). Individual injected P_0
606 animals were transferred to new 6 cm NGM plates approximately 18 hours after injections.
607 Individual F_1 rollers were then transferred to new 6 cm plates to generate self-progeny. The
608 region surrounding the desired S78C (or C78S) edit was then amplified from F_1 rollers using
609 primers oECA1163 and oECA1165. The PCR products were digested using the *Sfc*I restriction
610 enzyme (R0561S, New England Biolabs, Ipswich, MA). Differential band patterns signified
611 successfully edited strains because the N2 S78C, which is encoded by the CAG codon, creates
612 an additional *Sfc*I cut site. Non-Dpy, non-Rol progeny from homozygous edited F_1 animals were
613 propagated. If no homozygous edits were obtained, heterozygous F_1 progeny were propagated
614 and screened for the presence of the homozygous edits. F_1 and F_2 progeny were then Sanger
615 sequenced to verify the presence of the proper edited sequence. The phenotypes of allele swap
616 strains in control and arsenic trioxide conditions were measured using the HTA described above
617 (Supplemental data 6-8). PCA phenotypes for allele-swap strains were generated the same way
618 as described above for GWA mapping traits and are located in Supplemental data 19.
619
620 **Rescue with 13-methyltetradecanoic acid**
621 Strains were grown as described for a standard HTA experiment. In addition to adding arsenic
622 trioixde to experimental wells, we also added a range of C15iso (13-methyltetradecanoic acid,
623 Matreya Catalog # 1605) concentrations to assay rescue of arsenic effects (Supplemental data
624 20-21).
625
626 **Growth conditions for metabolite profiling**
627 Chunks (~1 cm) were taken from starved plates and placed on multiple fresh 10 cm plates. Prior
628 to starvation, animals were washed off of the plates using M9, and embryos were prepared by
629 bleach-synchronization. Approximately 40,000 embryos were resuspended in 25 ml of K
630 medium and allowed to hatch overnight at 20ºC. L1 larvae were fed 15 mg/ml of HB101 lysate
631 the following morning and allowed to grow at 20ºC for 72 hours. We harvested 100,000
632 embryos from gravid adults by bleaching. These embryos were hatched overnight in 50 ml of K
633 medium in a 125 ml flask. The following day, we added arsenic trioxide to a final concentration
634 of 100 μM and incubated the cultures for 24 hours. After 24 hours, we added HB101 bacterial

635 lysate (2 mg/ml final) to each culture. Finally, we transferred the cultures to 50 ml conical tubes,
636 centrifuged the cultures at 3000 RPM for three minutes, and separated the pellet and
637 supernatant. The supernatant and pellets from the cultures were frozen at -80ºC and prepared
638 for analysis.
639
640 **Nematode extraction**
641 Pellets were lyophilized 18-24 hours using a VirTis BenchTop 4K Freeze Dryer until a chalky
642 consistency was achieved. Dried pellets were transferred to 1.5 mL microfuge tubes and dry
643 pellet weight recorded. Pellets were disrupted in a Spex 1600 MiniG tissue grinder after the
644 addition of three stainless steel grinding balls to each sample. Microfuge tubes were placed in a
645 Cryoblock (Model 1660) cooled in liquid nitrogen, and samples were disrupted at 1100 RPM for
646 two cycles of 30 seconds. At this point, each sample was individually dragged across a
647 microfuge tube rack eight times, inverted, and flicked five times to prevent clumping. This
648 process was repeated two additional rounds for a total of six disruptions. Pellets were
649 transferred to 4 mL glass vials in 3 mL 100% ethanol. Samples were sonicated for 20 minutes
650 (on/off pulse cycles of two seconds at power 90 A) using a Qsonica Ultrasonic Processor
651 (Model Q700) with a water bath cup horn adaptor (Model 431C2). Following sonication, glass
652 vials were centrifuged at 2750 RCF for five minutes in an Eppendorf 5702 Centrifuge using rotor
653 F-35-30-17. The resulting supernatant was transferred to a clean 4 mL glass vial and
654 concentrated to dryness in an SC250EXP Speedvac Concentrator coupled to an RVT5105
655 Refrigerated Vapor Trap (Thermo Scientific). The resulting powder was suspended in 100%
656 ethanol according to its original dry pellet weight: 0.01 mL 100% ethanol per mg of material. The
657 suspension was sonicated for 10 minutes (pulse cycles of two seconds on and three seconds
658 off at power 90 A) followed by centrifugation at 20,817 RCF in a refrigerated Eppendorf
659 centrifuge 5417R at 4°C. The resulting supernatant was transferred to an HPLC vial containing
660 a Phenomenex insert (cat #AR0-4521-12) and centrifuged at 2750 RCF for five minutes in an
661 Eppendorf 5702 centrifuge. The resulting supernatant was transferred to a clean HPLC vial
662 insert and stored at -20°C or analyzed immediately.
663
664 **Mass spectrometric analysis**
665 Reversed-phase chromatography was performed using a Dionex Ultimate 3000 Series LC
666 system (HPG-3400 RS High Pressure pump, TCC-3000RS column compartment, WPS-
667 3000TRS autosampler, DAD-3000 Diode Array Detector) controlled by Chromeleon Software
668 (ThermoFisher Scientific) and coupled to an Orbitrap Q-Exactive mass spectrometer controlled
669 by Xcalibur software (ThermoFisher Scientific). Metabolites were separated on a Kinetex EVO
670 C18 column, 150 mm x 2.1 mm, particle size 1.7 µm, maintained at 40°C with a flow rate of 0.5
671 mL/min. Solvent A: 0.1% ammonium acetate in water; solvent B: acetonitrile (ACN). A/B
672 gradient started at 5% B for 30 seconds, followed by a linear gradient to 95% B over 13.5
673 minutes, then a linear gradient to 100% B over three minutes. 100% B was maintained for one
674 minute. Column was washed after each run with 5:1 isopropanol:ACN, flow rate of 0.12 mL/min
675 for five minutes, followed by 100% ACN for 2.9 minutes, a linear gradient to 95:5 water:ACN
676 over 0.1 minutes, and then 95:5 water:ACN for two minutes with a flow rate of 0.5 mL/min. A
677 heated electrospray ionization source (HESI-II) was used for the ionization with the following
678 mass spectrometer parameters: spray voltage: 3 kV; capillary temperature: 320°C; probe heater

679 temperature: 300°C; sheath gas: 70 AU; auxiliary gas flow: 2 AU; resolution: 240,000 FWHM at
680 m/z 200; AGC target: 5e6; maximum injection time: 300 ms. Each sample was analyzed in
681 negative and positive modes with m/z range 200-800. Fatty acids and most ascarosides were
682 detected as $[M-H]^-$ ions in negative ionization mode. Peaks of known abundant ascarosides and
683 fatty acids were used to monitor mass accuracy, chromatographic peak shape, and instrument
684 sensitivity for each sample. Processed metabolite measures can be found in Supplemental data
685 22.
686
687 **Statistical analyses**
688 All *p*-values testing the differences of strain phenotypes in the NIL, allele-replacement, and
689 C15ISO experiments were performed in R using the *TukeyHSD* function with an ANOVA model
690 with the formula (*phenotype ~ strain*). *p*-values of individual pairwise strain comparisons are
691 reported in each figure legend.
692
693 **CRISPR-Cas9 gene editing in human cells**
694 Gene-editing experiments were performed in a single parallel culture experiment using human
695 293T cells (ATCC) grown in DMEM with 10% FBS. On day zero, 300,000 cells were seeded per
696 well in a six-well plate format. The following day, two master mixes were prepared: a) LT-1
697 transfection reagent (Mirus) was diluted 1:10 in Opti-MEM and incubated for five minutes; b) a
698 DNA mix of 500 ng Cas9-sgRNA plasmid with 250 pmol repair template oligonucleotide
699 (Supplemental table 2) was diluted in Opti-MEM in a final volume of 100 µL. 250 µL of the lipid
700 mix was added to each of the DNA mixes and incubated at room temperature for 25 minutes.
701 Following incubation, the full 350 µL volume of DNA and lipid mix was added dropwise to the
702 cells. These six-well plates were then centrifuged at 1000 x g for 30 minutes. After six hours, the
703 media on the cells was replaced. For the next six days, cells were expanded and passaged as
704 needed. On day seven, one million cells were taken from each set of edited and unedited cells
705 and placed into separate T75s with either media-only or 5 µM arsenic-containing media. Days
706 seven to fourteen, arsenic and media-only conditions were maintained at healthy cell densities.
707 Days fourteen to eighteen, arsenic exposed cell populations were maintained off arsenic to
708 allow the populations to recover prior to sequencing. Media-only conditions were maintained in
709 parallel. On day eighteen, all arsenic and media-only conditions were pelleted for genomic DNA
710 extraction.
711
712 **Analysis of CRISPR-Cas9 editing in human cells**
713 Genomic DNA was extracted from cell pellets using the QIAGEN (QIAGEN, Hilden, Germany)
714 Midi or Mini Kits based on the size of the cell pellet (51183, 51104) according to the
715 manufacturer's recommendations. DBT1 loci were first amplified with 17 cycles of PCR using a
716 touchdown protocol and the NEBnext 2x master mix (New England Biolabs M0541). The
717 resulting product served as input to a second PCR, using primers that appended a sample-
718 specific barcode and the necessary adaptors for Illumina sequencing. The resulting DNA was
719 pooled, purified with SPRI beads (A63880, Beckman Coulter, Brea, CA), and sequenced on an
720 Illumina MiSeq with a 300-nucleotide single-end read with an eight nucleotide index read. For
721 each sample, the number of reads exactly matching the wild-type and edited DBT1 sequence
722 were determined (Supplemental data 23)

723
724 **Preparing human cells for Mass Spectroscopy**
725 Mass spectroscopy experiments used human 293T cells (ATCC) grown in DMEM with 10%
726 FBS. On day zero, 150,000 cells were seeded into 15 cm tissue cultures dishes with 15 mL of
727 either 2.5 μM arsenic or no arsenic media. Each condition had five replicates. On day three, the
728 no arsenic cells were approaching confluence and required passaging. Arsenic conditions were
729 at ~30% confluence and received a media change. On day seven, both conditions were near
730 confluence, media was removed, and plates were rinsed with ice cold PBS, remaining liquid
731 removed. Plates were frozen at -80ºC before processing for mass spectrometric analysis.
732 Metabolite measurements can be found in Supplemental data 24.
733
734 **Tajima's D calculation**
735 We used the VCF corresponding to CeNDR release 20160408 (Supplemental data 25a-b) to
736 calculate Tajima's D. Tajima's D was calculated using the *tajimas_d* function in the *cegwas*
737 package using default parameters (window size = 500 SNVs, sliding window distance = 50
738 SNVs, outgroup = N2) (Supplemental data 26). Isolation locations of strains can be found in
739 Supplemental data 27.
740
741 **Author Contributions:**
742 SZ and ECA conceptualized and designed the *C. elegans* experiments. CS and JGD designed
743 the human cell-line experiments. SZ performed *C. elegans* experiments. TAC helped with
744 preparation of *C. elegans* for metabolite extraction. CS performed the human cell-line
745 experiments. BWF processed *C. elegans* samples for metabolite extraction and performed and
746 analyzed the metabolite profiling experiments. OP and FJT prepared and analyzed metabolite
747 experiments during initial optimization of the growth conditions. SCB helped construct an allele-
748 replacement strain. Formal data curation and analysis of *C. elegans* experiments, excluding
749 metabolite profiling, was performed by SZ. SZ wrote the manuscript and processed and
750 analyzed the data for the manuscript figures. All authors contributed to the editing of the
751 manuscript. Supervision of the project was performed by ECA, FCS, and JGD.
752
753 **Funding:**
754 This work was supported by a National Institutes of Health R01 subcontract to ECA
755 (GM107227), the Chicago Biomedical Consortium with support from the Searle Funds at the
756 Chicago Community Trust, a Sherman-Fairchild Cancer Innovation Award to ECA, and an
757 American Cancer Society Research Scholar Grant to ECA (127313-RSG-15-135-01-DD), along
758 with support from the Cell and Molecular Basis of Disease training grant (T32GM008061) to SZ.
759 JGD is a Merkin Institute Fellow and is supported by the Next Generation Fund at the Broad
760 Institute of MIT and Harvard. FCS, BWF, OP, and FJT were supported by R01 GM088290. The
761 funders had no role in study design, data collection and analysis, decision to publish, or
762 preparation of the manuscript.
763
764 **Acknowledgments:**
765 The authors would like to thank Samuel Rosenberg for assistance on early mappings of drug
766 sensitivities, Mudra Hegde of the Broad Institute for assistance with sequence analysis, and
767 members of the Andersen laboratory for critical reading of this manuscript.

768 **References**

769 Adeva-Andany, María M., Laura López-Maside, Cristóbal Donapetry-García, Carlos Fernández-
770 Fernández, and Cristina Sixto-Leal. 2017. "Enzymes Involved in Branched-Chain Amino
771 Acid Metabolism in Humans." *Amino Acids* 49 (6): 1005–28.
772 https://doi.org/10.1007/s00726-017-2412-7.
773 Agostini, Michelle, Sabrina D. Silva, Karina G. Zecchin, Ricardo D. Coletta, Jacks Jorge,
774 Massimo Loda, and Edgard Graner. 2004. "Fatty Acid Synthase Is Required for the
775 Proliferation of Human Oral Squamous Carcinoma Cells." *Oral Oncology* 40 (7): 728–35.
776 https://doi.org/10.1016/j.oraloncology.2004.01.011.
777 Andersen, Erik C., Joshua S. Bloom, Justin P. Gerke, and Leonid Kruglyak. 2014. "A Variant in
778 the Neuropeptide Receptor Npr-1 Is a Major Determinant of *Caenorhabditis Elegans*
779 Growth and Physiology." 10 (2): e1004156. https://doi.org/10.1371/journal.pgen.1004156.
780 Andersen, Erik C., Joshua S. Bloom, Leonid Kruglyak, Marie-Anne Félix, Rajarshi Ghosh, Justin
781 P. Gerke, Joshua A. Shapiro, and Jonathan R. Crissman. 2012. "Chromosome-Scale
782 Selective Sweeps Shape *Caenorhabditis Elegans* Genomic Diversity." *Nature Genetics* 44
783 (3): 285–90. https://doi.org/10.1038/ng.1050.
784 Andersen, Erik C., Tyler C. Shimko, Jonathan R. Crissman, Rajarshi Ghosh, Joshua S. Bloom,
785 Hannah S. Seidel, Justin P. Gerke, and Leonid Kruglyak. 2015. "A Powerful New
786 Quantitative Genetics Platform, Combining *Caenorhabditis Elegans* High-Throughput
787 Fitness Assays with a Large Collection of Recombinant Strains." *G3* 5 (5): g3.115.017178–
788 920. https://doi.org/10.1534/g3.115.017178.
789 Antman, K. H. 2001. "Introduction: The History of Arsenic Trioxide in Cancer Therapy." *The*
790 *Oncologist* 6 Suppl 2: 1–2. https://doi.org/10.1634/theoncologist.6-suppl_2-1.
791 Baugh, L. R. 2013. "To Grow or Not to Grow: Nutritional Control of Development During
792 Caenorhabditis Elegans L1 Arrest." *Genetics* 194 (3): 539–55.
793 https://doi.org/10.1534/genetics.113.150847.
794 Bergquist, Erik R., Robert J. Fischer, Kent D. Sugden, and Brooke D. Martin. 2009. "Inhibition
795 by Methylated Organo-Arsenicals of the Respiratory 2-Oxo-Acid Dehydrogenases." *Journal*
796 *of Organometallic Chemistry* 694 (6): 973–80.
797 https://doi.org/10.1016/j.jorganchem.2008.12.028.
798 Bloom, Joshua S., Ian M. Ehrenreich, Wesley T. Loo, Thúy-Lan Võ Lite, and Leonid Kruglyak.
799 2013. "Finding the Sources of Missing Heritability in a Yeast Cross." *Nature* 494 (7436): 1–
800 6. https://doi.org/10.1038/nature11867.
801 Boyd, Windy A., Marjolein V. Smith, and Jonathan H. Freedman. 2012. "Caenorhabditis Elegans
802 as a Model in Developmental Toxicology." *Methods in Molecular Biology* 889 (Chapter 3):
803 15–24. https://doi.org/10.1007/978-1-61779-867-2_3.
804 Brady, Shannon, Kathryn Evans, Joshua Bloom, Robyn Tanny, Daniel Cook, Sarah Giuliani,
805 Stephen Hippleheuser, Mostafa Zamanian, and Erik Andersen. 2018. "Common Loci
806 Underlie Natural Variation in Diverse Toxin Responses." *bioRxiv*.
807 https://doi.org/10.1101/325399.
808 Broman, Karl W., Hao Wu, Saunak Sen, and Gary A. Churchill. 2003. "R/qtl: QTL Mapping in
809 Experimental Crosses." *Bioinformatics* 19 (7): 889–90.
810 http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12724300&retmod
811 e=ref&cmd=prlinks.
812 Burrage, Lindsay C., Sandesh C. S. Nagamani, Philippe M. Campeau, and Brendan H. Lee.
813 2014. "Branched-Chain Amino Acid Metabolism: From Rare Mendelian Diseases to More
814 Common Disorders." *Human Molecular Genetics* 23 (R1): R1–8.
815 https://doi.org/10.1093/hmg/ddu123.
816 Chen, G. Q., X. G. Shi, W. Tang, S. M. Xiong, J. Zhu, X. Cai, Z. G. Han, et al. 1997. "Use of
817 Arsenic Trioxide (As2O3) in the Treatment of Acute Promyelocytic Leukemia (APL): I.

818 As2O3 Exerts Dose-Dependent Dual Effects on APL Cells." *Blood* 89 (9): 3345–53.
819 https://www.ncbi.nlm.nih.gov/pubmed/9129041.
820 Chung, Chi-Jung, Yu-Mei Hsueh, Chyi-Huey Bai, Yung-Kai Huang, Ya-Li Huang, Mo-Hsiung
821 Yang, and Chien-Jen Chen. 2009. "Polymorphisms in Arsenic Metabolism Genes, Urinary
822 Arsenic Methylation Profile and Cancer." *Cancer Causes & Control: CCC* 20 (9): 1653–61.
823 https://doi.org/10.1007/s10552-009-9413-0.
824 Cook, Daniel E., Stefan Zdraljevic, Joshua P. Roberts, and Erik C. Andersen. 2016. "CeNDR,
825 the Caenorhabditis Elegans Natural Diversity Resource." *Nucleic Acids Research*, October.
826 https://doi.org/10.1093/nar/gkw893.
827 Cook, Daniel E., Stefan Zdraljevic, Robyn E. Tanny, Beomseok Seo, David D. Riccardi, Luke M.
828 Noble, Matthew V. Rockman, et al. 2016. "The Genetic Basis of Natural Variation in
829 Caenorhabditis Elegans Telomere Length." *Genetics* 204 (1): 371–83.
830 https://doi.org/10.1534/genetics.116.191148.
831 Coombs, C. C., M. Tavakkoli, and M. S. Tallman. 2015. "Acute Promyelocytic Leukemia: Where
832 Did We Start, Where Are We Now, and the Future." *Blood Cancer Journal* 5 (April): e304.
833 https://doi.org/10.1038/bcj.2015.25.
834 Emi, Nobuhiko. 2017. "Arsenic Trioxide: Clinical Pharmacology and Therapeutic Results." In
835 *Chemotherapy for Leukemia*, 221–38. Springer, Singapore. https://doi.org/10.1007/978-
836 981-10-3332-2_13.
837 Endelman, Jeffrey B. 2011. "Ridge Regression and Other Kernels for Genomic Selection with R
838 Package rrBLUP." *The Plant Genome Journal* 4 (3): 250–56.
839 https://doi.org/10.3835/plantgenome2011.08.0024.
840 Entchev, Eugeni V., Dominik Schwudke, Vyacheslav Zagoriy, Vitali Matyash, Aliona
841 Bogdanova, Bianca Habermann, Lin Zhu, Andrej Shevchenko, and Teymuras V.
842 Kurzchalia. 2008. "LET-767 Is Required for the Production of Branched Chain and Long
843 Chain Fatty Acids in Caenorhabditis Elegans." *The Journal of Biological Chemistry* 283
844 (25): 17550–60. https://doi.org/10.1074/jbc.M800965200.
845 Evans, Kathryn S., Yuehui Zhao, Shannon C. Brady, Lijiang Long, Patrick T. McGrath, and Erik
846 C. Andersen. 2017. "Correlations of Genotype with Climate Parameters Suggest
847 Caenorhabditis Elegans Niche Adaptations." *G3* 7 (1): 289–98.
848 https://doi.org/10.1534/g3.116.035162.
849 Forbes, S. A., G. Bhamra, S. Bamford, E. Dawson, C. Kok, J. Clements, A. Menzies, J. W.
850 Teague, P. A. Futreal, and M. R. Stratton. 2008. "The Catalogue of Somatic Mutations in
851 Cancer (COSMIC)." *Current Protocols in Human Genetics / Editorial Board, Jonathan L.*
852 *Haines ... [et Al.]* Chapter 10 (April): Unit 10.11.
853 https://doi.org/10.1002/0471142905.hg1011s57.
854 Fujihara, Junko, Yoshimi Fujii, Tetsuro Agusa, Takashi Kunito, Toshihiro Yasuda, Tamami
855 Moritani, and Haruo Takeshita. 2009. "Ethnic Differences in Five Intronic Polymorphisms
856 Associated with Arsenic Metabolism within Human Arsenic (+3 Oxidation State)
857 Methyltransferase (AS3MT) Gene." *Toxicology and Applied Pharmacology* 234 (1): 41–46.
858 https://doi.org/10.1016/j.taap.2008.09.026.
859 García-González, Aurian P., Ashlyn D. Ritter, Shaleen Shrestha, Erik C. Andersen, L. Safak
860 Yilmaz, and Albertha J. M. Walhout. 2017. "Bacterial Metabolism Affects the C. elegans
861 Response to Cancer Chemotherapeutics." *Cell* 169 (3): 431–41.e8.
862 https://doi.org/10.1016/j.cell.2017.03.046.
863 Gomez-Rubio, Paulina, Maria M. Meza-Montenegro, Ernesto Cantu-Soto, and Walter T.
864 Klimecki. 2010. "Genetic Association between Intronic Variants in AS3MT and Arsenic
865 Methylation Efficiency Is Focused on a Large Linkage Disequilibrium Cluster in
866 Chromosome 10." *Journal of Applied Toxicology: JAT* 30 (3): 260–70.
867 https://doi.org/10.1002/jat.1492.
868 Grignani, F., M. Valtieri, M. Gabbianelli, V. Gelmetti, R. Botta, L. Luchetti, B. Masella, et al.

869 2000. “PML/RAR Alpha Fusion Protein Expression in Normal Human Hematopoietic
870 Progenitors Dictates Myeloid Commitment and the Promyelocytic Phenotype.” *Blood* 96 (4):
871 1531–37. https://www.ncbi.nlm.nih.gov/pubmed/10942402.
872 Heffelfinger, S. C., E. T. Sewell, and D. J. Danner. 1983. “Identification of Specific Subunits of
873 Highly Purified Bovine Liver Branched-Chain Ketoacid Dehydrogenase.” *Biochemistry* 22
874 (24): 5519–22. https://doi.org/10.1021/bi00293a011.
875 Hodgkin, J., and T. Doniach. 1997. “Natural Variation and Copulatory Plug Formation in
876 Caenorhabditis Elegans.” *Genetics* 146 (1): 149–64.
877 https://www.ncbi.nlm.nih.gov/pubmed/9136008.
878 Hoonjan, Maneka, Vaibhav Jadhav, and Purvi Bhatt. 2018. “Arsenic Trioxide: Insights into Its
879 Evolution to an Anticancer Agent.” *Journal of Biological Inorganic Chemistry: JBIC: A*
880 *Publication of the Society of Biological Inorganic Chemistry* 23 (3): 313–29.
881 https://doi.org/10.1007/s00775-018-1537-9.
882 Hughes-Fulford, M., Y. Chen, and R. R. Tjandrawinata. 2001. “Fatty Acid Regulates Gene
883 Expression and Growth of Human Prostate Cancer PC-3 Cells.” *Carcinogenesis* 22 (5):
884 701–7. https://doi.org/10.1093/carcin/22.5.701.
885 Jia, Fan, Mingxue Cui, Minh T. Than, and Min Han. 2016. “Developmental Defects of
886 Caenorhabditis Elegans Lacking Branched-Chain α-Ketoacid Dehydrogenase Are Mainly
887 Caused by Monomethyl Branched-Chain Fatty Acid Deficiency.” *The Journal of Biological*
888 *Chemistry* 291 (6): 2967–73. https://doi.org/10.1074/jbc.M115.676650.
889 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. 2008.
890 “Efficient Control of Population Structure in Model Organism Association Mapping.”
891 *Genetics* 178 (3): 1709–23. https://doi.org/10.1534/genetics.107.080101.
892 Kato, Masato, Jacinta L. Chuang, Shih-Chia Tso, R. Max Wynn, and David T. Chuang. 2005.
893 “Crystal Structure of Pyruvate Dehydrogenase Kinase 3 Bound to Lipoyl Domain 2 of
894 Human Pyruvate Dehydrogenase Complex.” *The EMBO Journal* 24 (10): 1763–74.
895 https://doi.org/10.1038/sj.emboj.7600663.
896 Khairul, Islam, Qian Qian Wang, Yu Han Jiang, Chao Wang, and Hua Naranmandura. 2017.
897 “Metabolism, Toxicity and Anticancer Activities of Arsenic Compounds.” *Oncotarget* 8 (14):
898 23905–26. https://doi.org/10.18632/oncotarget.14733.
899 Kim, Heesun, Takao Ishidate, Krishna S. Ghanta, Meetu Seth, Darryl Conte, Masaki Shirayama,
900 and Craig C. Mello. 2014. “A Co-CRISPR Strategy for Efficient Genome Editing in
901 Caenorhabditis Elegans.” *Genetics* 197 (4): 1069–80.
902 https://doi.org/10.1534/genetics.114.166389.
903 Kniazeva, Marina, Quinn T. Crawford, Matt Seiber, Cun-Yu Wang, and Min Han. 2004.
904 “Monomethyl Branched-Chain Fatty Acids Play an Essential Role in Caenorhabditis
905 Elegans Development.” *PLoS Biology* 2 (9): E257.
906 https://doi.org/10.1371/journal.pbio.0020257.
907 Kniazeva, Marina, Tanya Euler, and Min Han. 2008. “A Branched-Chain Fatty Acid Is Involved
908 in Post-Embryonic Growth Control in Parallel to the Insulin Receptor Pathway and Its
909 Biosynthesis Is Feedback-Regulated in C. elegans.” *Genes & Development* 22 (15): 2102–
910 10. https://doi.org/10.1101/gad.1692008.
911 Large, Edward E., Wen Xu, Yuehui Zhao, Shannon C. Brady, Lijiang Long, Rebecca A. Butcher,
912 Erik C. Andersen, and Patrick T. McGrath. 2016. “Selection on a Subunit of the NURF
913 Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of
914 Caenorhabditis Elegans.” *PLoS Genetics* 12 (7): e1006219.
915 https://doi.org/10.1371/journal.pgen.1006219.
916 Li, Jiaojiao, Charles Packianathan, Toby G. Rossman, and Barry P. Rosen. 2017.
917 “Nonsynonymous Polymorphisms in the Human AS3MT Arsenic Methylation Gene:
918 Implications for Arsenic Toxicity.” *Chemical Research in Toxicology* 30 (7): 1481–91.
919 https://doi.org/10.1021/acs.chemrestox.7b00113.

920 Luz, Anthony L., Tewodros R. Godebo, Dhaval P. Bhatt, Olga R. Ilkayeva, Laura L. Maurer,
921 Matthew D. Hirschey, and Joel N. Meyer. 2016. “From the Cover: Arsenite Uncouples
922 Mitochondrial Respiration and Induces a Warburg-like Effect in Caenorhabditis Elegans.”
923 *Toxicological Sciences: An Official Journal of the Society of Toxicology* 152 (2): 349–62.
924 https://doi.org/10.1093/toxsci/kfw093.
925 Luz, Anthony L., Tewodros R. Godebo, Latasha L. Smith, Tess C. Leuthner, Laura L. Maurer,
926 and Joel N. Meyer. 2017. “Deficiencies in Mitochondrial Dynamics Sensitize Caenorhabditis
927 Elegans to Arsenite and Other Mitochondrial Toxicants by Reducing Mitochondrial
928 Adaptability.” *Toxicology* 387 (July): 81–94. https://doi.org/10.1016/j.tox.2017.05.018.
929 Luz, Anthony L., and Joel N. Meyer. 2016. “Effects of Reduced Mitochondrial DNA Content on
930 Secondary Mitochondrial Toxicant Exposure in Caenorhabditis Elegans.” *Mitochondrion* 30
931 (September): 255–64. https://doi.org/10.1016/j.mito.2016.08.014.
932 Mandal, Badal Kumar, and Kazuo T. Suzuki. 2002. “Arsenic Round the World: A Review.”
933 *Talanta* 58 (1): 201–35. https://doi.org/10.1016/S0039-9140(02)00268-0.
934 Murgo, A. J. 2001. “Clinical Trials of Arsenic Trioxide in Hematologic and Solid Tumors:
935 Overview of the National Cancer Institute Cooperative Research and Development
936 Studies.” *The Oncologist* 6 Suppl 2: 22–28. https://doi.org/10.1634/theoncologist.6-suppl_2-
937 22.
938 Paix, Alexandre, Andrew Folkmann, Dominique Rasoloson, and Geraldine Seydoux. 2015.
939 “High Efficiency, Homology-Directed Genome Editing in Caenorhabditis Elegans Using
940 CRISPR-Cas9 Ribonucleoprotein Complexes.” *Genetics* 201 (1): 47–54.
941 https://doi.org/10.1534/genetics.115.179382.
942 Paul, Somnath, Nilanjana Banerjee, Aditi Chatterjee, Tanmoy J. Sau, Jayanta K. Das, Prafulla
943 K. Mishra, Partha Chakrabarti, Arun Bandyopadhyay, and Ashok K. Giri. 2014. “Arsenic-
944 Induced Promoter Hypomethylation and over-Expression of ERCC2 Reduces DNA Repair
945 Capacity in Humans by Non-Disjunction of the ERCC2-Cdk7 Complex.” *Metallomics:*
946 *Integrated Biometal Science* 6 (4): 864–73. https://doi.org/10.1039/c3mt00328k.
947 Pettit, F. H., S. J. Yeaman, and L. J. Reed. 1978. “Purification and Characterization of Branched
948 Chain Alpha-Keto Acid Dehydrogenase Complex of Bovine Kidney.” *Proceedings of the*
949 *National Academy of Sciences of the United States of America* 75 (10): 4881–85.
950 https://www.ncbi.nlm.nih.gov/pubmed/283398.
951 Ratnaike, R. N. 2003. “Acute and Chronic Arsenic Toxicity.” *Postgraduate Medical Journal* 79
952 (933): 391–96. https://doi.org/10.1136/pmj.79.933.391.
953 Ravenscroft, Peter, Hugh Brammer, and Keith Richards. 2009. *Arsenic Pollution : A Global*
954 *Synthesis*. Chichester, U.K.: Wiley-Blackwell.
955 http://www.worldcat.org/oclc/214285927?referer=xid.
956 Reed, L. J., and M. L. Hackert. 1990. “Structure-Function Relationships in Dihydrolipoamide
957 Acyltransferases.” *The Journal of Biological Chemistry* 265 (16): 8971–74.
958 https://www.ncbi.nlm.nih.gov/pubmed/2188967.
959 Schlebusch, Carina M., Lucie M. Gattepaille, Karin Engström, Marie Vahter, Mattias Jakobsson,
960 and Karin Broberg. 2015. “Human Adaptation to Arsenic-Rich Environments.” *Molecular*
961 *Biology and Evolution* 32 (6): 1544–55. https://doi.org/10.1093/molbev/msv046.
962 Schmeisser, Sebastian, Kathrin Schmeisser, Sandra Weimer, Marco Groth, Steffen Priebe,
963 Eugen Fazius, Doreen Kuhlow, et al. 2013. “Mitochondrial Hormesis Links Low-Dose
964 Arsenite Exposure to Lifespan Extension.” *Aging Cell* 12 (3): 508–17.
965 https://doi.org/10.1111/acel.12076.
966 Shen, Shengwen, Xing-Fang Li, William R. Cullen, Michael Weinfeld, and X. Chris Le. 2013.
967 “Arsenic Binding to Proteins.” *Chemical Reviews* 113 (10): 7769–92.
968 https://doi.org/10.1021/cr300015c.
969 Shimko, Tyler C., and Erik C. Andersen. 2014. “COPASutils: An R Package for Reading,
970 Processing, and Visualizing Data from COPAS Large-Particle Flow Cytometers.” *PloS One*

971 9 (10): e111090. https://doi.org/10.1371/journal.pone.0111090.
972 Spracklin, George, Brandon Fields, Gang Wan, Diveena Becker, Ashley Wallig, Aditi Shukla,
973 and Scott Kennedy. 2017. “The RNAi Inheritance Machinery of Caenorhabditis Elegans.”
974 *Genetics* 206 (3): 1403–16. https://doi.org/10.1534/genetics.116.198812.
975 Stýblo, Miroslav, Zuzana Drobná, Ilona Jaspers, Shan Lin, and David J. Thomas. 2002. “The
976 Role of Biomethylation in Toxicity and Carcinogenicity of Arsenic: A Research Update.”
977 *Environmental Health Perspectives* 110 Suppl 5 (October): 767–71.
978 https://www.ncbi.nlm.nih.gov/pubmed/12426129.
979 Tajima, F. 1989. “Statistical Method for Testing the Neutral Mutation Hypothesis by DNA
980 Polymorphism.” *Genetics* 123 (3): 585–95. https://www.ncbi.nlm.nih.gov/pubmed/2513255.
981 Team, R. Core. 2017. “R: A Language and Environment for Statistical Computing [Internet].
982 Vienna, Austria; 2014.”
983 Thé, H. de, C. Chomienne, M. Lanotte, L. Degos, and A. Dejean. 1990. “The t(15;17)
984 Translocation of Acute Promyelocytic Leukaemia Fuses the Retinoic Acid Receptor Alpha
985 Gene to a Novel Transcribed Locus.” *Nature* 347 (6293): 558–61.
986 https://doi.org/10.1038/347558a0.
987 Thompson, Owen A., L. Basten Snoek, Harm Nijveen, Mark G. Sterken, Rita J. M. Volkers,
988 Rachel Brenchley, Arjen Van’t Hof, et al. 2015. “Remarkably Divergent Regions Punctuate
989 the Genome Assembly of the Caenorhabditis Elegans Hawaiian Strain CB4856.” *Genetics*
990 200 (3): 975–89. https://doi.org/10.1534/genetics.115.175950.
991 Tomita, Akihiro, Hitoshi Kiyoi, and Tomoki Naoe. 2013. “Mechanisms of Action and Resistance
992 to All-Trans Retinoic Acid (ATRA) and Arsenic Trioxide (As2O 3) in Acute Promyelocytic
993 Leukemia.” *International Journal of Hematology* 97 (6): 717–25.
994 https://doi.org/10.1007/s12185-013-1354-4.
995 Tönjes, Martje, Sebastian Barbus, Yoon Jung Park, Wei Wang, Magdalena Schlotter, Anders M.
996 Lindroth, Sabrina V. Pleier, et al. 2013. “BCAT1 Promotes Cell Proliferation through Amino
997 Acid Catabolism in Gliomas Carrying Wild-Type IDH1.” *Nature Medicine* 19 (7): 901–8.
998 https://doi.org/10.1038/nm.3217.
999 Wang, Yunbiao, Anastasia N. Ezemaduka, Zhuheng Li, Zhanyan Chen, and Chuantao Song.
1000 2017. “Joint Toxicity of Arsenic, Copper and Glyphosate on Behavior, Reproduction and
1001 Heat Shock Protein Response in Caenorhabditis Elegans.” *Bulletin of Environmental*
1002 *Contamination and Toxicology* 98 (4): 465–71. https://doi.org/10.1007/s00128-017-2042-5.
1003 Watson, Emma, Lesley T. MacNeil, H. Efsun Arda, Lihua Julie Zhu, and Albertha J. M. Walhout.
1004 2013. “Integration of Metabolic and Gene Regulatory Networks Modulates the C. elegans
1005 Dietary Response.” *Cell* 153 (1): 253–66. https://doi.org/10.1016/j.cell.2013.02.050.
1006 Watts, Jennifer L., and Michael Ristow. 2017. “Lipid and Carbohydrate Metabolism in
1007 Caenorhabditis Elegans.” *Genetics* 207 (2): 413–46.
1008 https://doi.org/10.1534/genetics.117.300106.
1009 Wyatt, Lauren H., Anthony L. Luz, Xiou Cao, Laura L. Maurer, Ashley M. Blawas, Alejandro
1010 Aballay, William K. Y. Pan, and Joel N. Meyer. 2017. “Effects of Methyl and Inorganic
1011 Mercury Exposure on Genome Homeostasis and Mitochondrial Function in Caenorhabditis
1012 Elegans.” *DNA Repair* 52 (April): 31–48. https://doi.org/10.1016/j.dnarep.2017.02.005.
1013 Yeaman, S. J. 1989. “The 2-Oxo Acid Dehydrogenase Complexes: Recent Advances.”
1014 *Biochemical Journal* 257 (3): 625–32. https://www.ncbi.nlm.nih.gov/pubmed/2649080.
1015 Zdraljevic, Stefan, and Erik C. Andersen. 2017. “Natural Diversity Facilitates the Discovery of
1016 Conserved Chemotherapeutic Response Mechanisms.” *Current Opinion in Genetics &*
1017 *Development* 47 (September): 41–47. https://doi.org/10.1016/j.gde.2017.08.002.
1018 Zdraljevic, Stefan, Christine Strand, Hannah S. Seidel, Daniel E. Cook, John G. Doench, and
1019 Erik C. Andersen. 2017. “Natural Variation in a Single Amino Acid Substitution Underlies
1020 Physiological Responses to Topoisomerase II Poisons.” *PLoS Genetics* 13 (7): e1006891.
1021 https://doi.org/10.1371/journal.pgen.1006891.

1022 Zeidan, Amer M., and Steven D. Gore. 2014. “New Strategies in Acute Promyelocytic Leukemia:
1023 Moving to an Entirely Oral, Chemotherapy-Free Upfront Management Approach.” *Clinical*
1024 *Cancer Research: An Official Journal of the American Association for Cancer Research* 20
1025 (19): 4985–93. https://doi.org/10.1158/1078-0432.CCR-13-2725.
1026 Zhang, Xiao-Wei, Xiao-Jing Yan, Zi-Ren Zhou, Fei-Fei Yang, Zi-Yu Wu, Hong-Bin Sun, Wen-
1027 Xue Liang, et al. 2010. “Arsenic Trioxide Controls the Fate of the PML-RARalpha
1028 Oncoprotein by Directly Binding PML.” *Science* 328 (5975): 240–43.
1029 https://doi.org/10.1126/science.1183424.
1030 Zhu, Huanhu, Huali Shen, Aileen K. Sewell, Marina Kniazeva, and Min Han. 2013. “A Novel
1031 Sphingolipid-TORC1 Pathway Critically Promotes Postembryonic Development in
1032 Caenorhabditis Elegans.” *eLife* 2 (May): e00429. https://doi.org/10.7554/eLife.00429.

Supplemental Data:

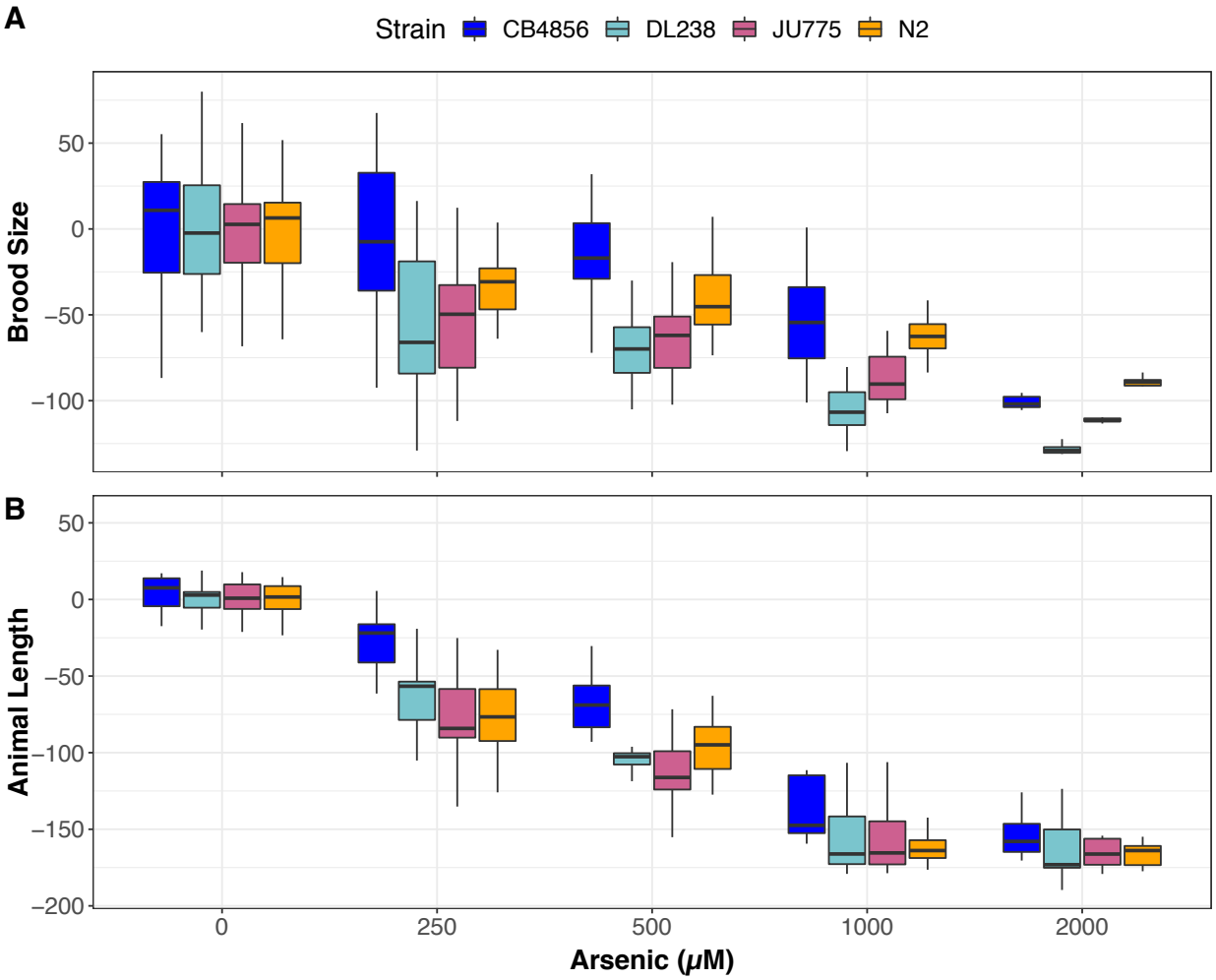
Supplemental Data 1: Phenotype measurements for arsenic trioxide dose response data for four diverged *C. elegans* strains.
Supplemental Data 2: Raw RIAIL phenotype data used for linkage mapping
Supplemental Data 3: Assay-regressed RIAIL phenotype data used for linkage mapping
Supplemental Data 4: Control-regressed RIAIL phenotype data used for linkage mapping
Supplemental Data 5: Genome-wide LOD scores for animal length and brood size measurements from linkage mapping experiments.
Supplemental Data 6: Raw NIL and DBT-1 swap phenotype data.
Supplemental Data 7: Summarized NIL and DBT-1 swap phenotype data.
Supplemental Data 8: Control-regressed NIL and DBT-1 swap phenotype data.
Supplemental Data 9: NIL genotype data.
Supplemental Data 10: Raw wild-isolate data used for GWA mapping.
Supplemental Data 11: Summarized wild-isolate data used for GWA mapping.
Supplemental Data 12: Assay-regressed wild-isolate data used for GWA mapping.
Supplemental Data 13: Control-regressed wild-isolate data used for GWA mapping.
Supplemental Data 14: Processed brood size and animal length GWA mappings.
Supplemental Data 15: PCA traits for wild isolates used for GWA mappings.
Supplemental Data 16: Variance explained of principal components from wild-isolate data.
Supplemental Data 17: Processed PCA GWA mappings.
Supplemental Data 18: Fine-mapping of PC1 GWA mapping confidence interval.
Supplemental Data 19: PCA phenotype data for DBT-1 allele swap experiment.
Supplemental Data 20: Raw HTA data from C15iso rescue experiment.
Supplemental Data 21: Processed HTA data from C15iso rescue experiment.
Supplemental Data 22: *C. elegans* metabolite measurements.
Supplemental Data 23: Fraction of human cells with DBT1 edits.
Supplemental Data 24: Human cell line metabolite measurements
Supplemental Data 25: VCF used for Tajima's D calculation.
Supplemental Data 25b: VCF index used for Tajima's D calculation.
Supplemental Data 26: Tajima's D of GWA mapping confidence interval.
Supplemental Data 27: Isolation locations of strains used in GWA mapping.

Supplemental Tables:

Supplemental Table 1: Strains constructed for this study
Supplemental Table 2: Oligonucleotides used in this study

1070 **Supplemental Figures:**

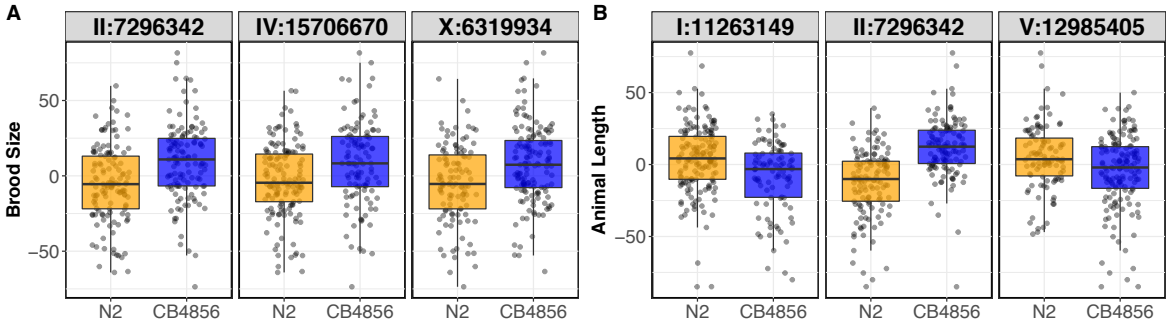
1071



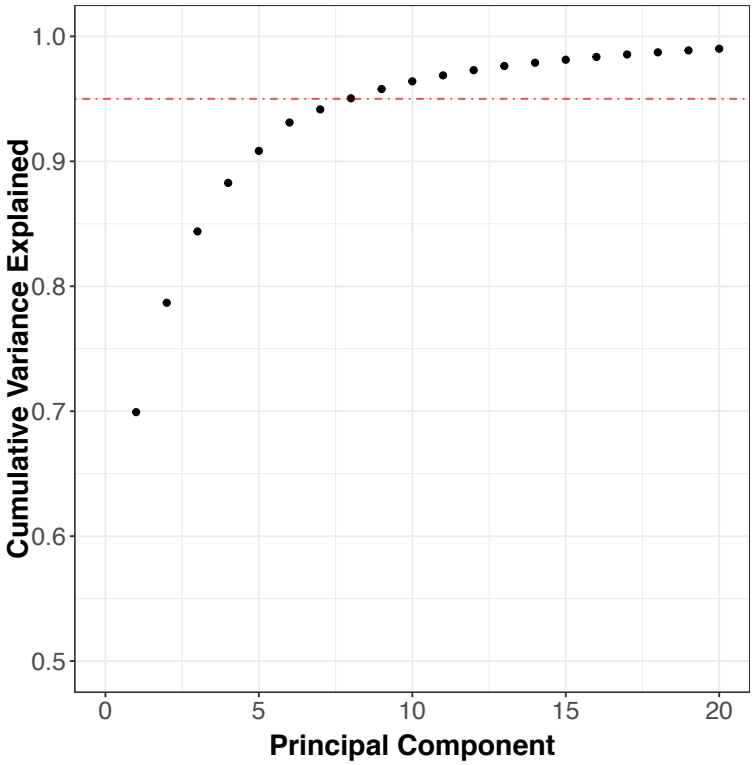
1072

1073 **Supplemental Figure 1: Arsenic trioxide dose response of four diverged *C. elegans***
1074 **strains**

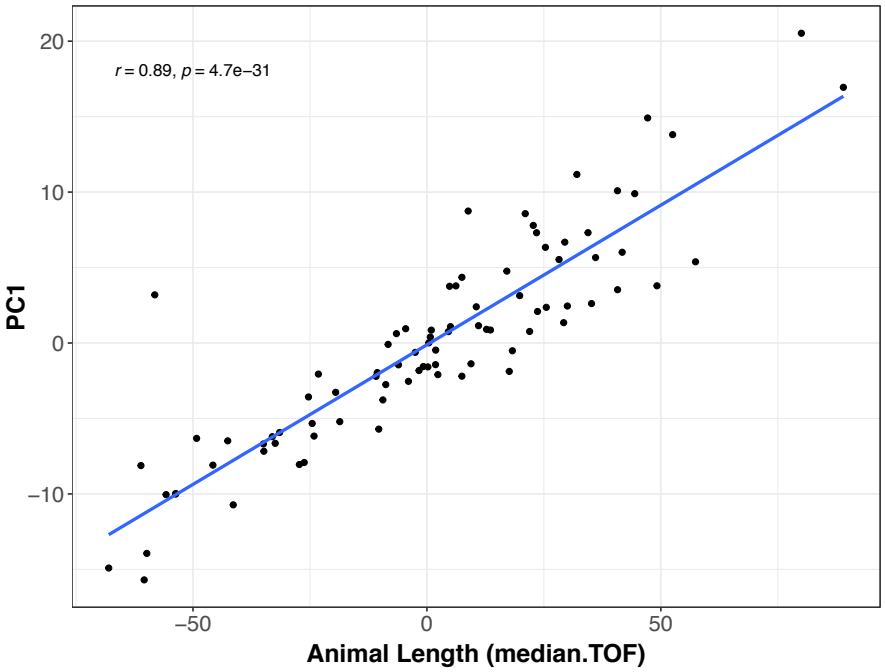
1075 Arsenic trioxide concentration in μM is plotted on the x-axis and the (A) normalized brood size or
1076 (B) median progeny length are plotted on the y-axis. Y-axis values represent individual
1077 phenotypic measurements subtracted from the mean value in 0 μM arsenic trioxide. At least 15
1078 replicates for each strain and condition are represented by Tukey box plots. Box plots are
1079 colored by strain (CB4856:blue, DL238:teal, JU775:pink, and N2:orange).



Supplemental Figure 2: RIAIL phenotypes from the linkage mapping experiment. Tukey box plots of regressed (A) brood sizes and (B) median animal lengths of the N2 and CB4856 RIAIL panel after exposure to arsenic trioxide. Each dot corresponds to the phenotype for a single RIAIL. The RIAILs are separated by the N2 (orange) or CB4856 (blue) genotype at the peak QTL positions identified by linkage mapping.

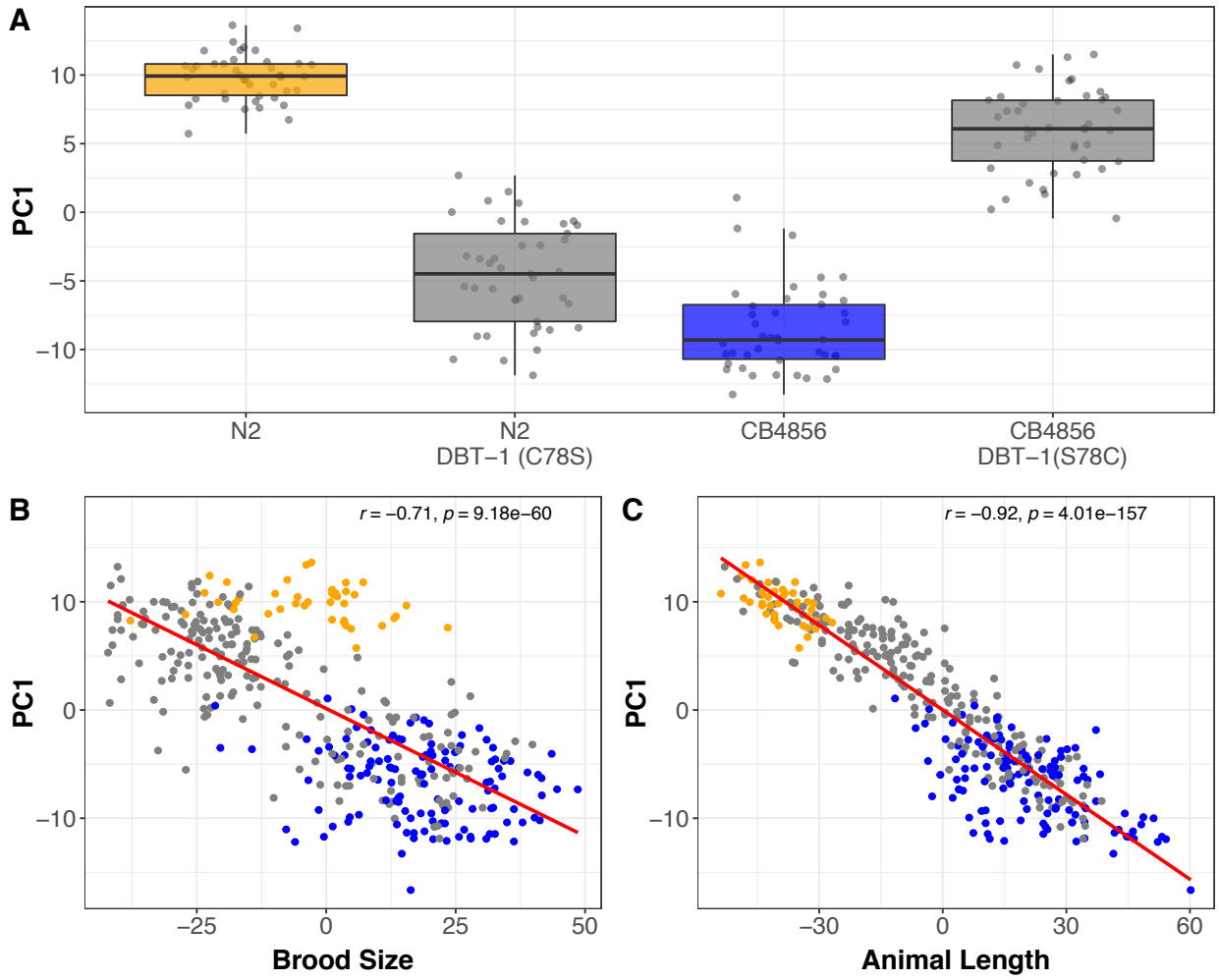


1087
1088 **Supplemental Figure 3: Variance explained by principal components from GWA mapping**
1089 **experiment.**
1090 Principal component analysis was performed on 86 wild-isolate phenotypes measured by the
1091 COPAS BIOSORT after arsenic trioxide exposure. The x-axis corresponds to individual PCs.
1092 The cumulative proportion of variance explained by the first 20 principal components is shown
1093 on the y-axis. 95% of the phenotypic variance in this data set (red line) is captured by the first
1094 eight PCs.



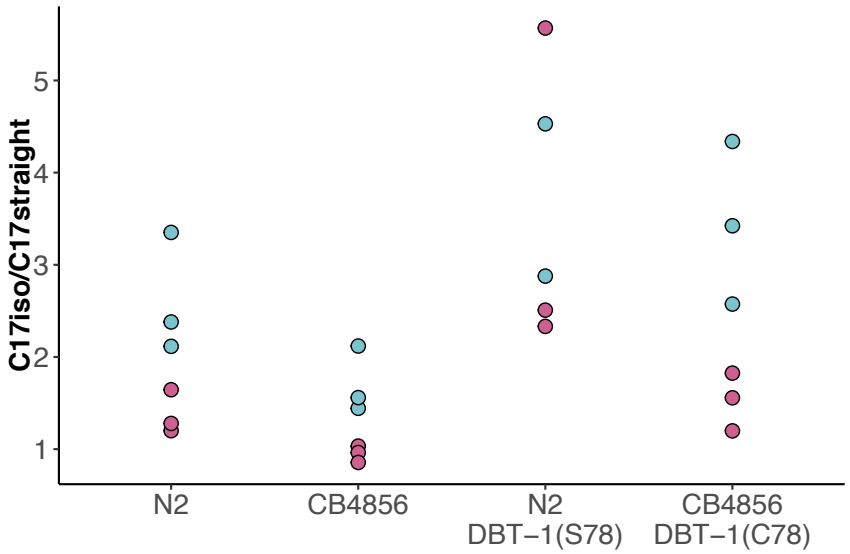
Supplemental Figure 4: Correlation between first principal component and animal length phenotypes

A scatter plot of the first principal component from PCA analysis of 67 phenotypes measured by the COPAS BIOSORT is plotted on the y-axis. The regressed median animal length is plotted on the x-axis. The Pearson correlation coefficient and the associated *p*-value of the correlation are inset in the upper left of the plot.

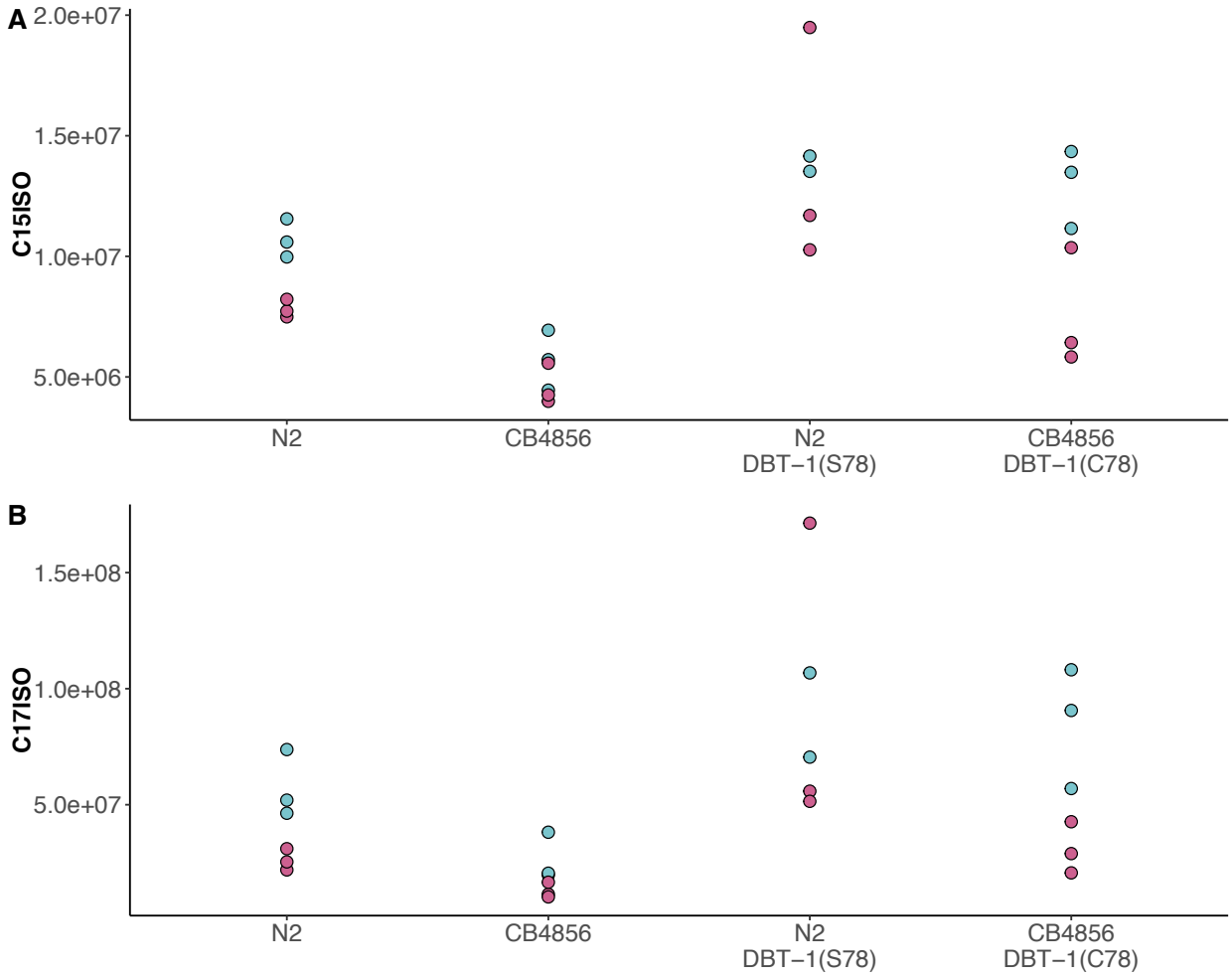


Supplemental Figure 5: Allele replacement strains recapitulate first principal component trait

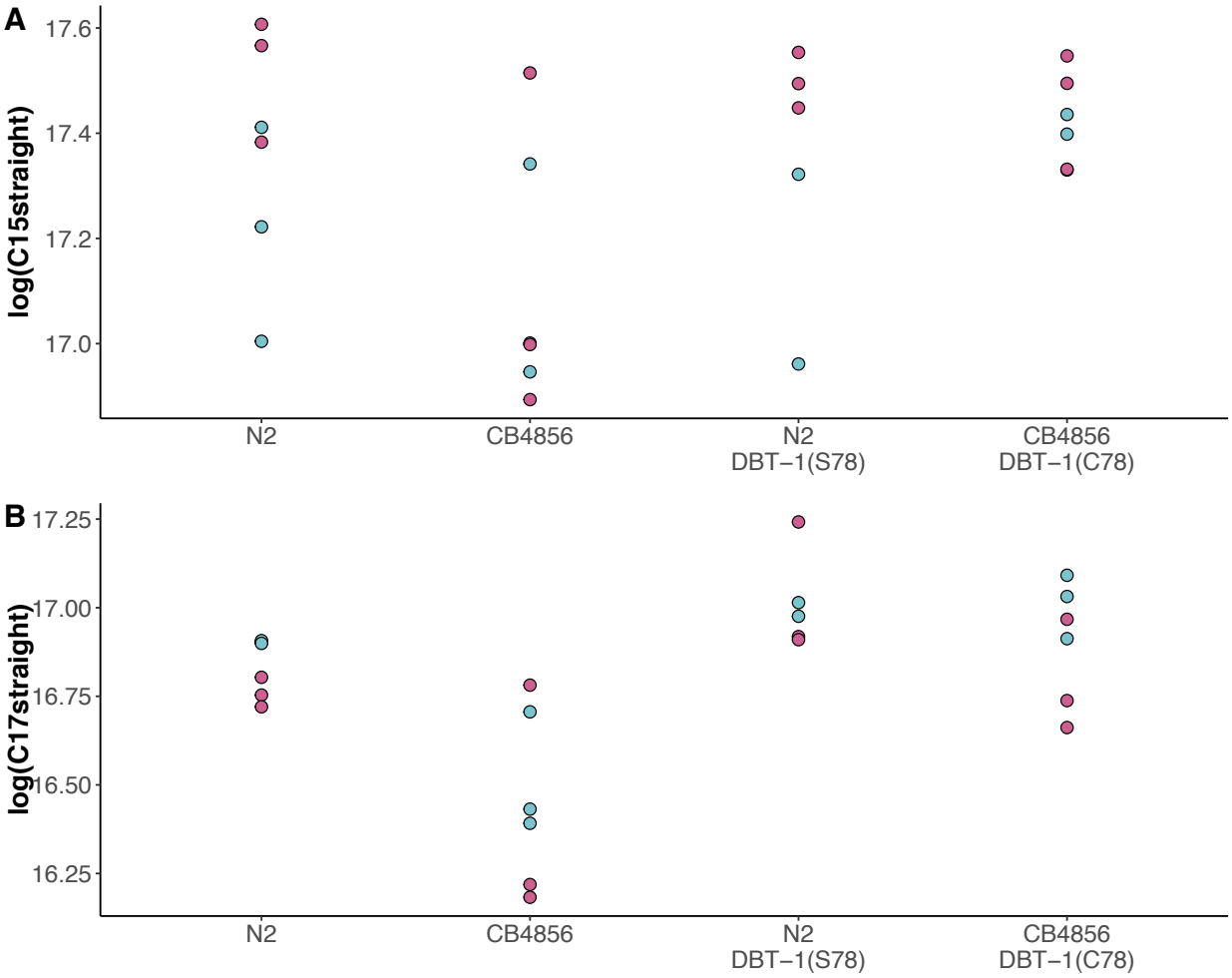
(A) Tukey box plots of the first principal component generated by PCA on allele-replacement strain phenotypes measured by the COPAS BIOSORT after arsenic trioxide exposure are shown (N2, orange; CB4856, blue; allele replacement strains, gray). Labels correspond to the genetic background and the corresponding residue at position 78 of DBT-1 (C for cysteine, S for serine). Scatter plots showing a strong negative correlation between the first principal component and (B) brood size or (C) median animal length phenotypes from of the allele-replacement strains. Each dot corresponds to a strain replicate and are colored as in panel B. The Pearson correlation coefficient and the associated *p*-value of the correlation are inset in the upper right of the plot.



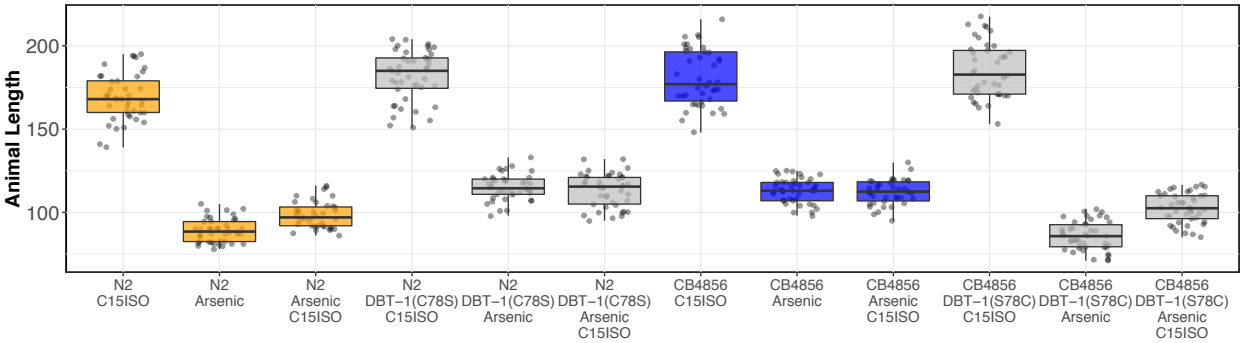
1114
1115 **Supplemental Figure 6: C17ISO/C17straight ratios.**
1116 The ratio of C17ISO/C17straight-chain is plotted on the y-axis for three independent replicates
1117 of the N2, CB4856, and allele swap strains exposed to control (teal) or 100 μM arsenic trioxide
1118 (pink) conditions. The difference between N2 control and N2 arsenic trioxide conditions is
1119 significant (Tukey HSD *p*-value = 0.036), and the difference between CB4856 mock and arsenic
1120 conditions is (Tukey HSD *p*-value = 0.024). In The difference between N2 swap control and
1121 arsenic trioxide conditions is not statistically significant (Tukey HSD *p*-value = 0.89) and the
1122 difference between CB4856 swap mock and arsenic conditions is significant (Tukey HSD *p*-
1123 value = 0.024).



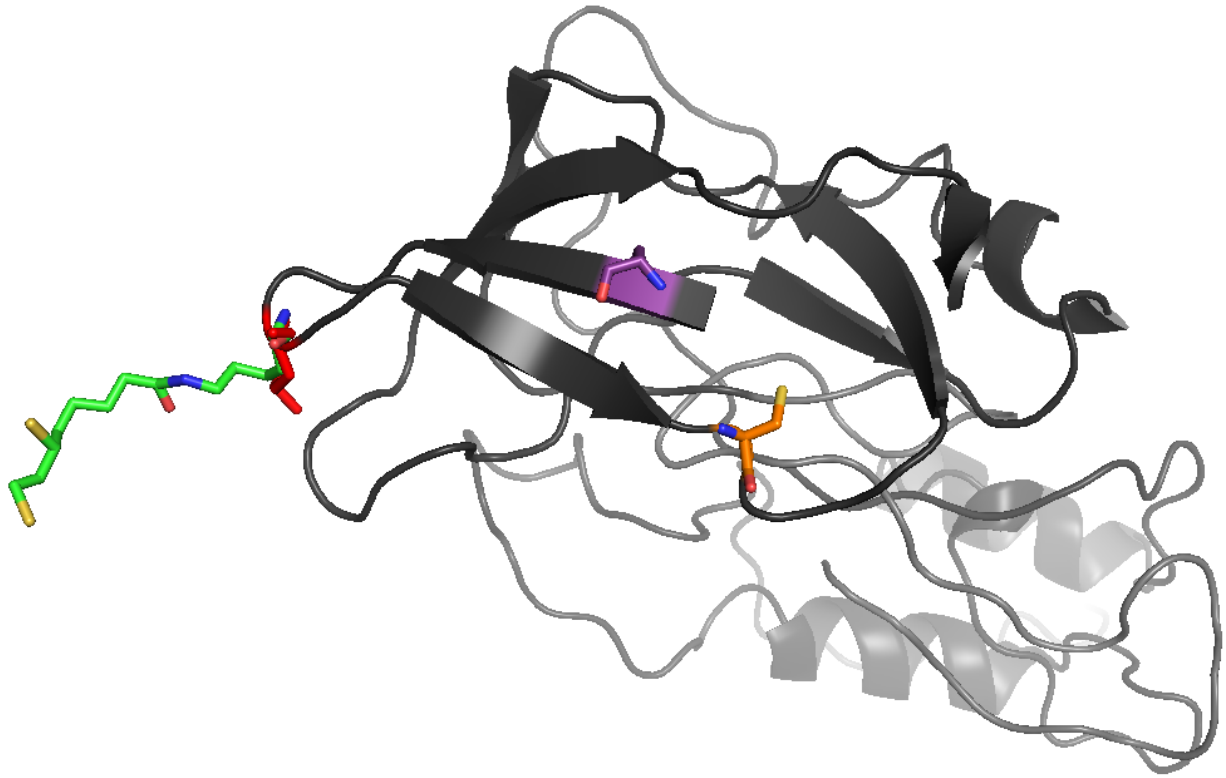
Supplemental Figure 7: Raw abundance of C15ISO and C17ISO.
The raw A) C15ISO and B) C17ISO is plotted on the y-axis for three independent replicates of the N2, CB4856, and allele swap strains exposed to control (teal) or 100 μM arsenic trioxide (pink) conditions. A) The difference between N2 control and N2 arsenic trioxide conditions is not significant (Tukey HSD *p*-value = 0.0046), but the difference between CB4856 mock and arsenic conditions is not (Tukey HSD *p*-value = 0.275). In The difference between N2 swap control and arsenic trioxide conditions is not statistically significant (Tukey HSD *p*-value = 0.99) and the difference between CB4856 swap mock and arsenic conditions is significant (Tukey HSD *p*-value = 0.033). B) The difference between N2 control and N2 arsenic trioxide conditions is not significant (Tukey HSD *p*-value = 0.024), but the difference between CB4856 mock and arsenic conditions is not (Tukey HSD *p*-value = 0.10). In The difference between N2 swap control and arsenic trioxide conditions is not statistically significant (Tukey HSD *p*-value = 0.94) and the difference between CB4856 swap mock and arsenic conditions is significant (Tukey HSD *p*-value = 0.029)



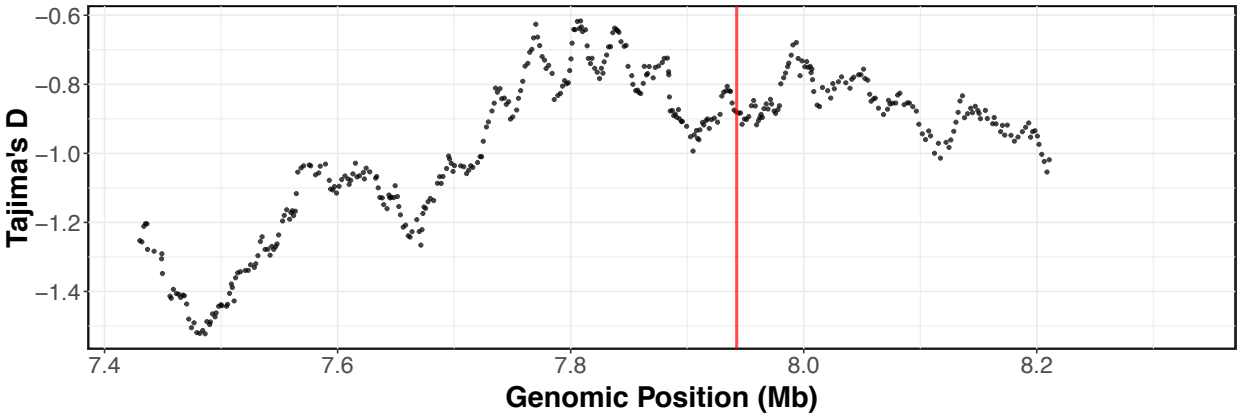
Supplemental Figure 8: C15 and C17 straight-chain metabolite levels.
The log-transformed raw abundance of A) C15 straight-chain and B) C17 straight-chain is plotted on the y-axis for three independent replicates of the N2, CB4856, and allele swap strains exposed to control (teal) or 100 μM arsenic trioxide (pink) conditions.



Supplemental Figure 9: Complete results from C15iso rescue experiment
Tukey box plots median animal length after C15ISO, arsenic trioxide, or arsenic trioxide and 0.64 µM C15ISO exposure are shown (N2, orange; CB4856, blue; allele replacement strains, gray). Labels correspond to the genetic background and the corresponding residue at position 78 of DBT-1 (C for cysteine, S for serine). Every pairwise strain comparison is significant except for the N2 DBT-1(S78) - CB4856 comparisons (Tukey's HSD *p*-value < 1.43E-6)**.**

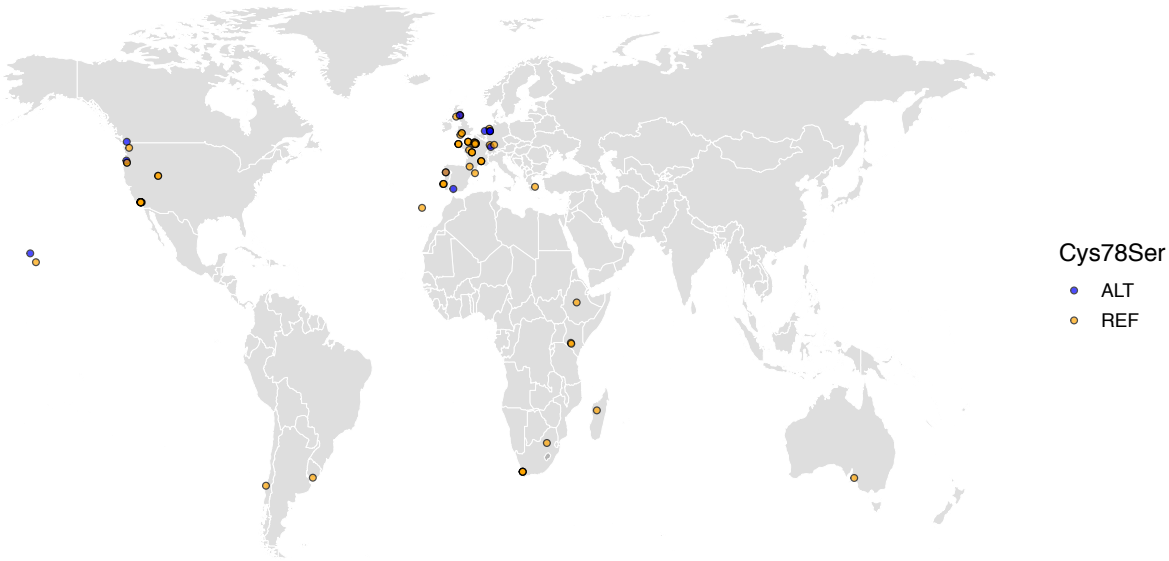


1151
1152 **Supplemental Figure 10: Three-dimensional homology model of *C. elegans* DBT-1**
1153 A three-dimensional homology model of *C. elegans* DBT-1 (black) aligned to human pyruvate
1154 dehydrogenase lipoyl domain (PDB:1Y8N) is shown. The C78 residue that confers resistance to
1155 arsenic trioxide is highlighted in orange, and the C65 residue is highlighted in purple. The
1156 *C. elegans* K71 residue is highlighted in red, and the human lipoylated lysine is highlighted in
1157 green.



1158

1159 **Supplemental Figure 11: Tajima's *D* across the arsenic trioxide QTL confidence interval**
1160 Divergence, as measured by Tajima's D, is shown across the arsenic trioxide QTL confidence
1161 interval (II:7,430,000-8,330,000). The whole-genome SNV data set [24,49] was used for
1162 Tajima's D calculations. Window size for the calculations was 500 SNVs with a 10 SNV sliding
1163 window size. The vertical red line marks the position of the *dbt-1* locus.



Supplemental Figure 12: The worldwide distribution of the DBT-1(C78S) allele
Cysteine (REF) is shown in orange and serine (ALT) is shown in blue. Latitude and longitude coordinates of sampling locations were used to plot individual strains on the map.