Figures and figure supplements

Natural variation in C. elegans arsenic toxicity is explained by differences in branched chain amino acid metabolism

Stefan Zdraljevic et al
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 the first principal component trait in the presence of 1000 μM arsenic trioxide is 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 blue boxes. The phenotypic variance explained by each QTL is shown above the peak QTL marker, which is marked by red triangles. (B) The correlation between brood size (blue; $r^2 = 0.38$, p-value=1.65E-27) or animal length (pink; $r^2 = 0.74$, p-value=3.16E-74) with the first principal component trait. Each dot represents an individual RIAIL’s phenotype, with the animal length and brood size phenotype values on the x-axis and the first principal component phenotype on the y-axis. (C) Tukey box plots of near-isogenic line (NIL) phenotype values for the first principal component trait in the presence of 1000 μM arsenic trioxide is shown. NIL genotypes are indicated below the plot as genomic ranges. The N2 trait is significantly different than the CB4856 and NIL traits (Tukey HSD p-value<1E-5).

DOI: https://doi.org/10.7554/eLife.40260.002
Figure 1—figure supplement 1. Arsenic trioxide dose response of four diverged C. elegans strains. Arsenic trioxide concentration in μM is plotted on the x-axis and the (A) normalized brood size, (B) mean progeny length, (C) mean normalized optical density, (D) mean normalized yellow fluorescence, (E) PC 1.

Zdraljevic et al. eLife 2019;8:e40260. DOI: https://doi.org/10.7554/eLife.40260
or (E) the first principal component are plotted on the y-axes. For panels A-D, the y-axis values represent individual phenotypic measurements subtracted from the mean value in 0 μM arsenic trioxide. At least 15 replicates for each strain and condition are represented by Tukey box plots. Box plots are colored by strain (CB4856: blue, DL238: teal, JU775: pink, and N2: orange).

DOI: https://doi.org/10.7554/eLife.40260.003
**Figure 1—figure supplement 2.** Trait correlations and principal component loadings of arsenic trioxide dose response. (A) The trait Pearson’s correlation coefficient of the assay- and control-regressed measured traits from the dose response experiment with 1000 µM arsenic trioxide is shown. (B) The contribution of each measured trait to the principal components that explain 90% of the total variance is shown. For each plot, the tile color corresponds to the value, where yellow colors represent higher values.

DOI: https://doi.org/10.7554/eLife.40260.007
Figure 1—figure supplement 3. Effect size and broad-sense heritability estimates for the arsenic trioxide dose response. Each panel corresponds to a concentration of arsenic trioxide concentration, which is indicated above the plot panel. The x-axis represents the partial omega squared ($\omega^2_p$) effect-size estimate, and the y-axis represent the broad-sense heritability estimate ($H^2$). Each dot represents a different BIOSORT-measured or principal component trait.

DOI: https://doi.org/10.7554/eLife.40260.009
Figure 1—figure supplement 4. RIAIL phenotypes from the linkage mapping experiment. Tukey box plots of the first principal component are shown: (A), assay- and control-regressed mean animal length (B), brood sizes (C), optical density (D), and mean normalized yellow fluorescence (E) 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 each QTL detected by linkage mapping.

DOI: https://doi.org/10.7554/eLife.40260.012
Figure 1—figure supplement 5. Linkage mapping results for brood size, animal length, and the first principal component. Linkage mapping plots for regressed brood size (orange), mean normalized yellow fluorescence (pink), mean animal length (blue) and mean normalized optical density (red) in the presence of 1000 μM 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 colored boxes below significant QTL.

DOI: https://doi.org/10.7554/eLife.40260.013
Figure 1—figure supplement 6. Genomic-heritability estimates of linkage mapping traits. The genomic broad ($H^2$)- and narrow ($h^2$)-sense heritability estimates calculated using the expectation ($E(A)$) of the realized relatedness matrix or the realized relatedness matrix ($A$) are shown. Each dot represents a measured or principal component trait. Dots are colored black if that trait mapped to the center of chromosome II and red if no QTL was detected on the center of chromosome II.

DOI: https://doi.org/10.7554/eLife.40260.014
Figure 1—figure supplement 7. Linkage mapping QTL summary. All QTL identified by linkage mapping are shown. Traits are labeled on the y-axis and the genomic position in Mb is plotted on the x-axis. Only PCs that explain greater than 90% of the variance are shown. Triangles represent the peak QTL position and bars represent the associated 1.5-LOD drop QTL confidence interval. Triangles and bars are colored based on the LOD score, where red colors correspond to higher LOD values.

DOI: https://doi.org/10.7554/eLife.40260.016
Figure 1—figure supplement 8. NIL recapitulation of chromosome II QTL. Tukey box plots of near-isogenic line (NIL) phenotype values for, from top to bottom: the brood size, mean animal length, mean normalized optical density, and mean normalized yellow fluorescence traits in the presence of 1000 μM arsenic trioxide are shown. NIL genotypes are indicated below the plot as genomic ranges. The N2 phenotypes are significantly different from all other strains (Brood size: Tukey HSD p-value=2E-7; Animal Length: Tukey HSD p-value<1E-7; Optical Density: Tukey HSD p-value<1E-7; Fluorescence: Tukey HSD p-value<1E-7).

DOI: https://doi.org/10.7554/eLife.40260.017
Figure 1—figure supplement 9. Trait correlations and principal component loadings of NIL and allele-replacement recapitulation experiment. (A) The Pearson’s correlation coefficients of the assay- and control-regressed measured traits are shown. (B) The contribution of each measured trait to the principal components that explain 90% of the total variance in the NIL and allele replacement-recapitulation experiment, which was performed at 1000 μM, is shown. For each plot, the tile color represents the value, where yellow colors represent higher values.

DOI: https://doi.org/10.7554/eLife.40260.018
Figure 1—figure supplement 10. Brood size and animal length are correlated with the first principal component for the NIL recapitulation experiment. The correlation between brood size (blue) or animal length (pink) (A), mean normalized optical density (B) or mean normalized yellow fluorescence (C) traits with the first principal component (PC1) trait for the NIL recapitulation experiment are shown. Each dot represents an individual NIL or parental strain replicate phenotype with the animal length and brood size phenotype values on the x-axis and the first principal component (PC1) phenotype on the y-axis.

DOI: https://doi.org/10.7554/eLife.40260.021
Figure 1—figure supplement 11. Trait correlations and principal component loadings of linkage mapping experiment. (A) The trait Pearson’s correlation coefficient of the assay- and control-regressed measured traits is shown. (B) The contribution of each measured trait to the principal components that explain 90% of the total variance in the linkage mapping experiment, which was performed at 1000 μM, is shown. For each plot, the tile color represents the value, where yellow colors represent higher values.

DOI: https://doi.org/10.7554/eLife.40260.022
Figure 2. Variation in *C. elegans* wild isolates responses to arsenic trioxide maps to the center of chromosome II. (A) A manhattan plot for the first principal component in the presence of 1000 μM arsenic trioxide is shown. Each dot represents an SNV that is present in at least 5% of the assayed wild population. The genomic position in Mb, separated by chromosome, is plotted on the x-axis and the \(-\log_{10}(p)\) for each SNV is plotted on the y-axis. SNVs are colored red if they pass the genome-wide Bonferroni-corrected significance (BF) threshold, which is denoted by the gray horizontal line. SNVs are colored pink if they pass the genome-wide eigen-decomposition significance (ED) threshold, which is denoted by the dotted gray horizontal line. The genomic region of interests surrounding the QTL that pass the BF and ED thresholds are represented by cyan and pink rectangles, respectively. (B) Tukey box plots of phenotypes used for association mapping in (A) are shown. Each dot corresponds to the phenotype of an individual strain, which is plotted on the y-axis. Strains are grouped by their genotype at the peak QTL position (red SNV from panel A, ChrII:7,931,252), where REF corresponds to the allele from the reference N2 strain. The N2 (orange) and CB4856 (blue) strains are highlighted. (C) Fine mapping of the chromosome II region of interest (cyan region from panel A, 7.60–8.21 Mb) is shown. Each dot represents an SNV present in the CB4856 strain. The association between the SNV and first principal component is shown on the y-axis and the genomic position of the SNV is shown on the x-axis. Dots are colored by their SnpEff predicted effect.

DOI: https://doi.org/10.7554/eLife.40260.030
Figure 2—figure supplement 1. GWA mapping QTL summary. All QTL identified by GWA mapping above the eigen significance threshold are shown. Traits are labeled on the y-axis and the genomic position in Mb is plotted on the x-axis. Only PCs that explain greater than 90% of the variance are shown. Triangles represent the peak QTL position, and bars represent the associated QTL region of interest. Triangles and bars are colored based on the significance value where red colors correspond to higher significance values.

DOI: https://doi.org/10.7554/eLife.40260.031
Figure 2—figure supplement 2. Fine-mapping of the chromosome II QTL identified by GWA mapping. Fine mapping of the chromosome II region of interest (cyan region from Figure 1A, 7.71–8.18 Mb) is shown. Each dot represents an SNV present in the phenotyped population. SNVs present in the CB4856 strain are shown in the left panel and SNVs present in other phenotyped strains, but REF in CB4856, are shown in the right panel. The association between the SNV and first principal component (PC1) is shown on the y-axis, and the genomic position of the SNV is shown on the x-axis. Dots are colored by their SnpEff predicted effect.

DOI: https://doi.org/10.7554/eLife.40260.033
Figure 2—figure supplement 3. Tajima’s D across the arsenic trioxide QTL confidence interval. Divergence, as measured by Tajima’s D, is shown across the arsenic trioxide QTL confidence interval (II:7,430,000–8,330,000). The whole-genome SNV data set (Cook et al., 2016; Cook et al., 2017) was used for Tajima’s D calculations. Window size for the calculations was 500 SNVs with a 10 SNV sliding window size. The vertical red line marks the position of the *dbt-1* locus.
DOI: https://doi.org/10.7554/eLife.40260.034
Figure 2—figure supplement 4. 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.
DOI: https://doi.org/10.7554/eLife.40260.036
Figure 2—figure supplement 5. Genomic-heritability estimates of GWA mapping traits. The genomic broad ($H^2$)- and narrow ($h^2$)-sense heritability estimates calculated using the realized relatedness matrix are shown. Each dot represents a measured or principal component trait. The five traits discussed throughout the manuscript are highlighted (purple: first principal component, pink: mean animal length, blue: brood size, orange: mean normalized optical density, and red: mean normalized yellow fluorescence). The four input traits for PCA in the left panel are mean normalized optical density, brood size, and mean animal length. For the right panel, mean normalized yellow fluorescence was also added.

DOI: https://doi.org/10.7554/eLife.40260.037
Figure 2—figure supplement 6. Trait correlations and principal component loadings of GWA mapping experiment. (A) The Pearson’s correlation coefficients of the assay- and control-regressed for measured traits are shown. (B) The contribution of each measured traits to the principal components that explain 90% of the total variance in the GWA mapping experiment, which was performed at 1000 μM, is shown. For each plot, the tile color represents the value, where yellow colors represent higher values.

DOI: https://doi.org/10.7554/eLife.40260.039
Figure 3. The DBT-1(C78S) variant contributes to arsenic trioxide responses. Tukey box plots of the first principal component generated by PCA on allele-replacement strain phenotypes measured by the COPAS BIOSORT 1000 μM 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). All pair-wise comparisons are significantly different (Tukey HSD, p-value < 1E-7).

DOI: https://doi.org/10.7554/eLife.40260.046
Figure 3—figure supplement 1. The DBT-1 C78S variant underlies arsenic trioxide sensitivity in *C. elegans*. From top to bottom: Tukey box plots of residual mean animal length, brood size, mean normalized yellow fluorescence, and mean normalized optical density after exposure to 1000 μM arsenic trioxide 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 was significant (Brood size: Tukey’s HSD p-value<1.55E-5; Animal length: p-value<2E-73; Fluorescence: p-value<1E-7; Optical Density: p-value<1E-7). DOI: https://doi.org/10.7554/eLife.40260.047
**Figure 3—figure supplement 2.** Brood size and animal length are correlated with the first principal component for the allele-replacement recapitulation experiment. The correlations between brood size (blue) or animal length (pink) with the first principal component trait for the allele-replacement recapitulation experiment are shown. Each dot represents an individual allele-replacement or parental strain replicate phenotype, with the animal length and brood size phenotype values on the x-axis and the first principal component (PC1) phenotype on the y-axis.

DOI: https://doi.org/10.7554/eLife.40260.048
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 these breakdown can then serve as building blocks for the mmBCFA that are required for developmental progression. (B) The difference in the C15ISO/C15SC (left panel) or C17ISO/C17SC (right panel) ratios between 100 μM arsenic trioxide and control conditions is plotted on the y-axis for three independent replicates of the CB4856 and CB4856 allele replacement strains and six independent replicates of the N2 and N2 allele replacement strains. Figure 4 continued on next page
strains. The difference between the C15 ratio for the CB4856-CB4856 allele replacement comparison is significant (Tukey HSD p-value = 0.0427733), but the difference between the C17 ratios for these two strains is not (Tukey HSD p-value = 0.164721). The difference between the C15 and C17 ratios for the N2-N2 allele replacement comparisons are both significant (C15: Tukey HSD p-value = 0.0358; C17: Tukey HSD p-value = 0.003747). (C) Tukey box plots median animal length after 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 pair-wise strain comparison is significant except for the N2 DBT-1(S78) - CB4856 comparisons (Tukey’s HSD p-value < 1.43E-6).

DOI: https://doi.org/10.7554/eLife.40260.049
Figure 4—figure supplement 1. Raw abundance of C17ISO for CB4856 and CB4856 allele replacement. The raw abundance of C17ISO is plotted on the y-axis for three independent replicates of the CB4856 and CB4856 allele replacement strains exposed to control (teal) or 100 μM arsenic trioxide (pink) conditions. The difference between CB4856 replacement mock and arsenic conditions was significant (Tukey HSD p-value=0.029), but the difference between CB4856 mock and arsenic conditions was not significant (Tukey HSD p-value=0.10).

DOI: https://doi.org/10.7554/eLife.40260.050
Figure 4—figure supplement 2. Straight-chain fatty acids are not affected by arsenic trioxide. The difference in raw C15SC (left panel) or C17SC (right panel) abundances between 100 μM arsenic trioxide and control conditions is plotted on the y-axis for three independent replicates of the CB4856 and CB4856 allele replacement strains and six independent replicates of the N2 and N2 allele replacement strains. We found no significant differences when comparing the abundances between parental and allele-replacement strains.

DOI: https://doi.org/10.7554/eLife.40260.051
Figure 4—figure supplement 3. C15ISO and C17ISO to strait-chain ratios in control conditions. The C15ISO/C15SC (left panel) or C17ISO/C17SC (right panel) ratios in control conditions are plotted on the y-axis for three independent replicates of the CB4856 and CB4856 allele replacement strains and six independent replicates of the N2 and N2 allele replacement strains. The C15 and C17 ratios for the CB4856-CB4856 allele replacement comparison were significant (C15: Tukey HSD p-value=0.0168749; C17: Tukey HSD p-value=0.0342525). The difference between the C17 ratio for the N2-N2 allele replacement comparison was significant (Tukey HSD p-value=0.0044667), but the difference in the C15 ratio was not significant (Tukey HSD p-value=0.1239674).

DOI: https://doi.org/10.7554/eLife.40260.053
Figure 4—figure supplement 4. Strains with the DBT-1(C78) allele produce more branched chain fatty acids in the L1 larval stage in control conditions. Branched chain (left panel) and straight chain (right panel) fatty acid measurements in L1 animals are represented on the y-axis. We found significant...
differences in abundances when comparing all parental and allele-replacement strains for C15ISO and C17ISO chain fatty acids (CB4856-C15ISO DBT-1 (C78): Tukey HSD p-value = 0.0036201, n=3; N2-C15ISO DBT-1(C78): Tukey HSD p-value = 0.0265059, n=6; CB4856-C17ISO DBT-1(C78): Tukey HSD p-value = 0.0086572, n=3; N2-C17ISO DBT-1(C78): Tukey HSD p-value = 0.0022501, n=6). Conversely, we observed no significant differences in straight chain fatty production except for C17n production in the CB4856 background (CB4856-C15n DBT-1(C78): Tukey HSD p-value = 0.0787388, n=3; N2-C15n DBT-1(C78): Tukey HSD p-value = 0.5817993, n=6; CB4856-C17n DBT-1(C78): Tukey HSD p-value = 0.0086572, n=3; N2-C17n DBT-1(C78): Tukey HSD p-value = 0.35827, n=6).

DOI: https://doi.org/10.7554/eLife.40260.054
Figure 4—figure supplement 5. Young adult C15ISO and C17ISO to strait-chain ratios in control conditions. The C15ISO/C18SC (top panel) or C17ISO/C18SC (bottom panel) ratios of young adult animals in control conditions are plotted on the y-axis for six independent replicates for all strains. We found no significant differences when comparing the abundances between parental and allele-replacement strains.

DOI: https://doi.org/10.7554/eLife.40260.056
Figure 4—figure supplement 6. Complete results from C15iso rescue experiment. Tukey box plots for the PC1 trait after 0.64 μM 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 was significant except for the N2 DBT-1(S78) - CB4856 comparisons (Tukey’s HSD p-value<1.43E-6).

DOI: https://doi.org/10.7554/eLife.40260.058
Figure 4—figure supplement 7. BIOSORT-quantified traits for C15ISO rescue experiment Tukey box plots for the mean normalized optical density (A), mean animal length (B), brood size (C), and mean normalized yellow fluorescence (D) traits after 0.64 μM 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). For all traits, except brood size, the comparison between arsenic and arsenic with C15ISO for strains with the DBT-1(C78) allele was significant (Tukey’s HSD p-value<2E-3).

DOI: https://doi.org/10.7554/eLife.40260.059
Figure 4—figure supplement 8. Trait correlations and principal component loadings of C15ISO rescue experiment. (A) The Pearson's correlation coefficients of the assay- and control-regressed measured traits are shown. (B) The contribution of each measured trait to the principal components that explain 90% of the total variance in the GWA mapping experiment, which was performed at 1000 µM, is shown. For each plot, the tile colors represent the value, where yellow colors represent higher values.

DOI: https://doi.org/10.7554/eLife.40260.060
**Figure 4—figure supplement 9.** Brood size and animal length are correlated with the first principal component for the C15ISO rescue experiment. The correlations between brood size (blue) or animal length (pink) with the first principal component trait for the C15ISO rescue experiment are shown. Each dot represents an individual allele-replacement or parental strain replicate phenotype, with the animal length and brood size phenotype values on the x-axis and the first principal component phenotype on the y-axis.

DOI: https://doi.org/10.7554/eLife.40260.061
**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 percent increase of edited human cells that contain the W84C, S112C, or R113C amino acid change in DBT1 in the presence 5 μM arsenic trioxide relative to control conditions are shown. The number of reads in 5 μM arsenic trioxide for all replicates are significantly different from control conditions (Fisher’s exact test, p-value<0.011).

DOI: https://doi.org/10.7554/eLife.40260.066
Figure 5—figure supplement 1. Three-dimensional homology model of *C. elegans* DBT-1. A three-dimensional homology model of *C. elegans* DBT-1 (black) aligned to human pyruvate dehydrogenase lipoamide domain (PDB: 1Y8N) is shown. The C78 residue that conferred resistance to arsenic trioxide is shown.

*Zdraljevic et al.* eLife 2019;8:e40260. DOI: https://doi.org/10.7554/eLife.40260

*Figure 5—figure supplement 1 continued on next page*
highlighted in orange, and the C65 residue is highlighted in purple. The C. elegans K71 residue is highlighted in red, and the human lipoylated lysine is highlighted in green.

DOI: https://doi.org/10.7554/eLife.40260.067