Natural variation in a glucuronosyltransferase modulates propionate sensitivity in a *C. elegans* propionic acidemia model

Huimin Na[1#], Stefan Zdraljevic[2#], Robyn E. Tanny[2], Albertha J.M. Walhout[1*] and Erik C. Andersen[2*]

1. Program in Systems Biology and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA

2. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

# These authors contribute equally to this work

* Corresponding authors

Email: erik.andersen@northwestern.edu; marian.walhout@umassmed.edu

ABSTRACT

Mutations in human metabolic genes can lead to rare diseases known as inborn errors of human metabolism. For instance, patients with loss-of-function mutations in either subunit of propionyl-CoA carboxylase suffer from propionic acidemia because they cannot catabolize propionate, leading to its harmful accumulation. Interestingly, both the penetrance and expressivity of metabolic disorders can be modulated by genetic background. However, modifiers of these diseases are difficult to identify because of the lack of statistical power for rare diseases in human genetics. Here, we use a model of propionic acidemia in the nematode *Caenorhabditis elegans* to identify genetic modifiers of propionate sensitivity. By genome-wide association mapping across wild strains exposed to excess propionate we identify several genomic regions correlated with reduced propionate sensitivity. We find that natural variation in the putative glucuronosyltransferase GLCT-3, a homolog of human B3GAT, partly explains differences in propionate sensitivity in one of these genomic intervals. Using genome-editing, we demonstrate that loss-of-function alleles in *glct-3* render the animals less sensitive to propionate. Additionally, we find that *C. elegans* has an expansion of the *glct* gene family, suggesting that the number of members of this family could influence sensitivity to excess propionate. Our findings demonstrate that natural variation in metabolic genes that are not directly associated with propionate breakdown can modulate propionate sensitivity. Our study provides a framework for using *C. elegans* to characterize the contributions of genetic background to inborn errors in human metabolism.

INTRODUCTION

Inborn errors of human metabolism are rare genetic diseases in which dietary nutrients or cellular metabolites cannot be broken down to generate energy, biomass, or remove toxic compounds. Most of these disorders are caused by loss-of-function mutations in genes encoding metabolic enzymes or metabolite transporters. Inborn errors of metabolism are often considered monogenic disorders. However, the penetrance and expressivity of these diseases can vary [1]. Therefore, it has been proposed that such diseases should be viewed as more complex traits in which not only environmental factors such as diet, but also genetic background, affect the age of onset and severity of the disease [1]. If true, modifier genes could harbor variation in different genetic backgrounds and affect the penetrance and expressivity of metabolic disorders. However, because such diseases are rare, often with incidences of 1:50,000 or fewer, identifying modifier genes in human populations has been exceedingly difficult [1, 2].

Propionic and methylmalonic acidemia are inborn errors of metabolism in which the short-chain fatty acid propionate cannot be broken down [3]. Patients with propionic acidemia carry loss-of-function mutations in both copies of either one of two genes, PCCA or PCCB, which encode the two proteins comprising propionyl-CoA carboxylase that converts propionyl-CoA to D-methylmalonyl-CoA. Methylmalonic acidemia is a bit more complicated because it can be caused by mutations in either methylmalonyl-CoA racemase, methylmalonyl-CoA mutase, or in enzymes involved in the processing of vitamin B12, which is an essential cofactor for methylmalonyl-CoA mutase [3, 4]. Propionyl-CoA is generated in the natural breakdown of the branched-chain amino acids isoleucine and valine, as well as the catabolism of methionine, threonine, and odd-chain

fatty acids. It can be inter-converted with propionate, which is generated by our gut microbiota in the digestion of plant fibers. Although propionate has been found to have beneficial functions [5, 6], it is toxic when it accumulates, as exemplified by patients with propionic acidemia [3]. Propionic acidemia is a rare disorder with a worldwide live birth incidence of 1:50,000 to 1:100,000. It is diagnosed in newborn screening by the detection of elevated levels of propionylcarnitine, 3-hydroxypropionate, and other aberrant metabolites [7].

The nematode *Caenorhabditis elegans* is a bacterivore found around the world [8-10]. In the laboratory, *C. elegans* can be fed different species and strains of bacteria [11, 12], but the vast majority of studies use the *Escherichia coli* strain OP50. However, *E. coli* OP50 cannot synthesize vitamin B12 and therefore cannot support the efficient breakdown of propionate by the canonical pathway [13, 14]. Previously, we found that *C. elegans* transcriptionally activates an alternative propionate breakdown pathway, or shunt, when flux through the canonical pathway is low due to genetic perturbations or low dietary vitamin B12 [15, 16]. This beta-oxidation pathway comprises five genes and generates acetyl-CoA [15](**Figure 1A**). *C. elegans* may have evolved a dedicated pathway for alternate propionate breakdown to be able to thrive eating bacteria that do not synthesize vitamin B12. It only activates the expression of propionate shunt genes when propionate accumulation is persistent, via a specific regulatory circuit known as a type 1 feed-forward loop with AND-logic gate using the nuclear hormone receptors *nhr-10* and *nhr-68* [16]. In propionic acidemia patients, the buildup of propionate shunt metabolites indicates the presence of the propionate shunt. However, its activity is not sufficient to mitigate

84 propionate toxicity likely because the enzymes functioning in other metabolic pathways
85 are repurposed [15].

86 The vast majority of *C. elegans* studies rely on the laboratory-adapted strain
87 named N2, from Bristol, England [17]. Over the last twenty years, hundreds of *C. elegans*
88 strains have been collected worldwide from natural habitats [10, 18-23]. *C. elegans* is a
89 self-fertilizing hermaphrodite and, therefore, different wild strains can be easily
90 maintained as fully isogenic strains. These different strains have been used to identify
91 quantitative trait loci that contribute to a variety of phenotypes, including anthelmintic and
92 cancer chemotherapeutic resistance, and in several cases the precise genotypic variation
93 that is causal to phenotypic variation has been determined [24-30]. Genomic information
94 about the different strains is organized in the *C. elegans* Natural Diversity Resource
95 (CeNDR), along with different tools for genome-wide association (GWA) mappings [31].

96 Here, we used wild *C. elegans* strains to identify natural variation in loci that modify
97 the resistance to exogenous propionate supplementation. To mimic propionic acidemia
98 metabolic conditions, we fed animals a diet of *Escherichia coli* OP50, which is low in
99 vitamin B12 and has low flux through the canonical propionate breakdown pathway [14,
100 15], and supplemented the animals with excess propionate. GWA mapping using 133 wild
101 strains identified several independent genomic regions or quantitative trait loci (QTL)
102 associated with propionate resistance. For one of these loci, we found the causal variant
103 in *glct-3,* which encodes a predicted beta-1,3-glucuronosyltransferase, and is an ortholog
104 of human B3GAT1, 2, and 3. This family of enzymes recognizes nonreducing terminal
105 sugars and their anomeric linkages. A human homolog, B3GAT3, catalyzes the formation
106 of the glycosaminoglycan-protein linkage by way of a glucuronyl transfer reaction in the

final step of the biosynthesis of proteoglycans [32]. Glucuronosyltransferases also catalyze reactions between metabolites, specifically the addition of glucuronic acid to toxic metabolites such as drugs [33]. Interestingly, we found that loss-of-function mutations in *glct-3* confer resistance to propionate. Our data show that quantitative toxicity phenotyping can be used to identify candidate modifier genes of traits associated with inborn errors in human metabolism.

RESULTS

C. elegans wild strains differ in sensitivities to exogenous propionate

Previously, we established *C. elegans* larval survival assays after exposure to exogenous propionate as a model of propionic acidemia [14-16]. In these assays, first larval stage (L1) animals are exposed to propionate and the proportion of animals that develop beyond that stage are quantified. Propionate dose-response curves (DRCs) showed that the laboratory-adapted strain N2 has an LD_{50} of approximately 80 mM [14-16]. Supplementation of vitamin B12, which supports breakdown of propionate by the canonical pathway, confers resistance to propionate (LD_{50} =120 mM) and loss of the propionyl-CoA carboxylase ortholog (*pcca-1*) or the first gene of the propionate shunt (*acdh-1*) render the animals sensitive (LD_{50} = 50 mM) [14, 15]. Because it is technically difficult to perform DRCs for all wild strains, we first asked whether 12 wild *C. elegans* strains, which represent high genetic diversity [19], exhibit differences in propionate sensitivity. To mimic metabolic conditions of human propionic acidemia, we fed the animals vitamin B12-deplete *E. coli* OP50 bacteria, which ensures that flux through the canonical propionate breakdown pathway was low [14, 15]. We performed three biological

replicate experiments, each consisting of three technical replicates, and found that the 12 strains exhibited varying degrees of propionate sensitivity (**Figures 1B, 1C, Table S1**). Nine of the strains had similar propionate sensitivities as the N2 strain with an LD_{50} of approximately 85 mM. The other three strains were more resistant to propionate with an LD_{50} of 100-110 mM. Next, we carefully titrated propionate at concentrations between 80 and 120 mM with 10 mM increments and confirmed that most strains exhibited sensitivity similar to the N2 strain, but that the DL238 and EG4725 strains were significantly more resistant (**Figure 1C, Table S2**). This result suggests that some wild strains have natural mechanisms to cope with high levels of propionate that are independent of vitamin B12 and the canonical propionate breakdown pathway.

To perform GWA mapping, we needed to test propionate sensitivity across a large set of wild *C. elegans* strains. Propionate sensitivity assays can be noisy, in part because of slight differences in experimental and environmental factors such as incubator and room temperature, propionate concentrations (which can change slightly due to evaporation and dilution), etc. To identify the dose with the highest reproducibility, we calculated broad-sense heritability (H^2) and found 100 mM propionate to be the best dose for the GWA mapping experiment (H^2= 0.79) (**Figures 1E, 1F**). Additionally, we performed power analysis to determine the number of replicate experiments that needed to be performed prior to testing a large number of wild strains. We found that five independent experiments, each with four technical replicates, would give us 80% power to detect a 20% difference in propionate sensitivity (**Figure S1**).

Five genomic loci modify sensitivity to propionate across the *C. elegans* population

153 To identify the genetic basis of propionate response variation in *C. elegans*, we exposed
154 133 wild strains to 100 mM propionate and measured L1 survival (**Table S3**). We tested
155 the strains in three batches and included six strains in every batch to control for potential
156 batch effects (**Figure S2**). We observed a broad range of propionate sensitivities (**Figure**
157 **2A**). Using whole-genome sequence data and the L1 survival phenotype, we performed
158 GWA mapping and identified five QTL that were above the Bonferroni-corrected
159 significance threshold (**Figure 2B**, two on chromosome II: (II:1880662-1993488,
160 II:13859466-13979658); two on chromosome V: (V:3213649-4284434, V:19229887-
161 19390858); and one on chromosome X: (X:9987812-10370303), coordinates are from
162 WS245). To test the independence of these QTL, we calculated the pairwise linkage
163 disequilibrium (LD) between each of the peak QTL markers (**Figure S3 - LD**). We
164 observed low levels of LD for the majority of QTL pairs, with the exception of the two QTL
165 on chromosome V (r^2 = 0.87, peak markers - V:3929669 and V:19356375), suggesting
166 that these two QTL might not be independent. Because multiple QTL were associated
167 with propionate sensitivity, it was difficult to decide which QTL to characterize in more
168 detail. Therefore, we used the sequence kernel association test (SKAT), which tests an
169 association between the phenotype of interest and the cumulative variation on a gene-
170 by-gene basis [34]. This approach identified two QTL, one that overlaps with the QTL on
171 left of chromosome V (V:3213649-4284434) identified using the single-marker mapping
172 approach (**Figure 2B**) and one QTL on chromosome I that only overlaps with the single-
173 marker mapping approach at a lower significance threshold (I:12374204-12388791)
174 (**Figure 2C**). This additional support for the QTL on the left arm of chromosome V
175 motivated us to investigate this genomic region further.

176

177 **Chromosome V near-isogenic lines do not recapitulate propionate resistance**

178 To validate the effect of the QTL on the left arm of chromosome V, we constructed near-
179 isogenic lines (NILs) in which the region associated with propionate resistance
180 (V:3213649-4284434) was crossed from a resistant strain into the genome of a sensitive
181 strain. To identify candidate parental strains for NIL construction, we focused on the 12
182 strains that were phenotyped in the dose-response experiment (**Figure 1C**). Of these 12
183 strains, two were significantly resistant to propionate and ten were sensitive. Next, we
184 verified that the propionate-resistant strains had the alternative genotype at the peak QTL
185 marker identified using the single-marker mapping method and were compatible with
186 propionate sensitive strains at the *peel-1 zeel-1* [35] and *sup-35 pha-1* [36] incompatibility
187 loci. Using these criteria, we identified DL238 (propionate-resistant) and BRC20067
188 (propionate-sensitive) as suitable parental strains for NIL construction. We constructed
189 nine NILs that contained the DL238 genomic region surrounding the chromosome V QTL
190 introgressed into the BRC20067 genetic background (**Figure 3A, Table S4**). When we
191 exposed these NILs to propionate, we observed that the DL238 introgressed regions that
192 correspond to the chromosome V QTL confidence interval did not confer propionate
193 resistance (**Figure 3B, Table S5**). Because the genomic region spanned by these NILs
194 is larger than the QTL confidence interval, these results suggested that the chromosome
195 V QTL we identified might have been the result of a spurious association with the QTL on
196 the right of chromosome V. The LD between these two loci supports this hypothesis.
197 Because the chromosome V QTL might have a complex relationship, we focused on the

198 QTL identified on the right of chromosome I where the gene-based mapping overlapped
199 the marker-based GWA mapping at the lower significance threshold (**Figure 2B and 2C**).
200
201 **Variation in the *glct-3* gene confers propionate resistance**
202 Using a gene-based GWA mapping strategy, we found that variation in the gene *glct-3*
203 on the right arm of chromosome I was associated with propionate sensitivity among the
204 wild isolates (**Figure 2C**). Additionally, the most correlated marker from the marker-based
205 GWA mapping was in close proximity to *glct-3* (Figure 4A). Within the *glct-3* gene, we
206 observed eight distinct combinations of alleles (haplotypes) among the phenotyped wild
207 strains (**Figure S4**). Five of these distinct haplotypes all included the same stop-gained
208 variant at amino acid position 16 (Gly16*), along with other variants (Gly17Arg, Ser50Ala,
209 Ser111Thr, Tyr231Cys in the QX1793 strain; Gly17Arg Ser50Ala in the CX11276, DL238,
210 ED3046, ED3049, and NIC252 strains; Leu184Phe in the ECA36 strain; Ile46Thr in the
211 QX1792 strain, and only Gly16* in the MY23 and QX1791 strains). Strains with the Gly16*
212 variant were 20% more resistant to propionate treatment than strains with no variation in
213 *glct-3* (**Figure 4B**). This genomic region explains 13.1% of the total genetic variation in
214 response to exogenous propionate, indicating that, although other loci underlie this trait,
215 this gene is major contributor to natural differences in resistance to propionate.
216 To test whether variation in the *glct-3* is causal for the difference in propionate
217 sensitivity, we generated two independent *glct-3* alleles in the propionate sensitive
218 BRC20067 strain using CRISPR-Cas9 genome editing [37]. The *ww62* allele has a one
219 base pair deletion at position 57 in the first exon that causes a frameshift in the reading
220 frame leading to an early stop codon, and the *ww63* allele contains the same Gly16*

variant found in DL238. In line with previous experiments, we found that DL238 was more resistant to propionate treatment than BRC20067 (Cohen's F = 2.433). The strains harboring the *ww62* and *ww63 alleles* recapitulate 23.7% and 57.7% of the difference in propionate sensitivity between DL238 and BRC20067 as measured by Cohen's F, respectively (**Figure 4C, Table S6**), demonstrating that the loss of *glct-3* function confers resistance to propionate.

To understand how variation in *glct-3* might have arisen in the *C. elegans* species, we investigated the frequency of this allele across the population and the strains that harbor strongly deleterious variants. More than 330 wild *C. elegans* strains are currently available in CeNDR [31], 42 of these strains contain the Gly16* variant in *glct-3*. The majority of strains that contain the Gly16* variant (33/42) were isolated on the Hawaiian islands (**Figure S5**), which are known to harbor the most genetically divergent *C. elegans* individuals [10]. An additional three strains also have variants that are predicted to cause a loss of *glct-3* function (ECA733, JU1395, and ECA723). In agreement with the geographic distribution of the Gly16* allele, strains that harbor this allele are among the set of highly genetically divergent *C. elegans* strains (**Figure S6**). However, not all of the genetically divergent strains harbor variation in *glct-3*.

To further explore the evolutionary history of the *glct-3* gene, we examined the conservation of *glct-3* paralogs and their orthologs. The *glct-3* gene encodes a glucuronosyl transferase-like protein that has six paralogs in *C. elegans*, including five closely related genes *glct-1*, *glct-2*, *glct-4*, *glct-5*, and *glct-6*, and one distantly related paralog *sqv-8*, which we will not discuss further (**Figure 5A**). Five *glct* genes (1-5) are located on an 80 kb region on chromosome I, and *glct-6* is located on chromosome IV.

The close proximity of five of the six paralogs suggests that these genes are the products of gene duplication events, as observed for other gene families in *C. elegans* [38-40]. We observed elevated levels of variation in the genomic region that contains *glct-1* through *glct-5* (**Figure 5B, Table S7**), which supports the hypothesis that these sequences duplicated at some point in the *C. elegans* lineage and then diverged. Furthermore, the pattern of polymorphism in the six *C. elegans glct* paralogs suggests that after the initial duplication event, the function of the *glct-6* gene was retained, which is indicated by the absence of deleterious variants in this gene among wild isolates (**Figure 5B**) [31]. Similarly, *glct-4* has no variation that is predicted to be deleterious. By contrast, *glct-1*, *glct-2*, *glct-3*, and *glct-5* contain variants predicted to have large effects on gene function. Among the 330 *C. elegans* strains, 24 have variation that is predicted to remove the function of two or more of these four genes, with two strains that have predicted loss-of-function alleles in all four genes.

Copy number of the *glct* gene family varies across *Caenorhabditis* species

Next, we explored the conservation of the *glct* gene family across 20 species of *Caenorhabditis* nematodes, including ten for which the genome assembly was recently released [41]. We found that nine species contained only one *glct-3* ortholog, five contain two *glct-3* orthologs, three contain three *glct-3* orthologs, and one species each contain four, five, or six *glct-3* orthologs (**Figure S7**). The prevalence of low-copy numbers of *glct* genes among a majority of *Caenorhabditis* species suggests that the ancestral copy number is fewer than the six copies found in the *C. elegans* genome. This hypothesis is supported by the presence of one and two *glct-3* orthologs in the outgroup species

Heterorhabditis bacteriophora and *Oscheius tipulae*, respectively. Because the GLCT-6 protein and DNA sequences more closely resemble orthologous sequences among *Caenorhabditis* species than its paralogs in *C. elegans* (**Figure S8**), this gene is likely the ancestral state of this gene family. Taken together, these results suggest that the copy number of *glct* genes might affect fitness in the wild.

DISCUSSION

In this study, we identify mutations that naturally occur in *C. elegans glct-3* as modifiers of propionate sensitivity. This gene encodes a glucuronosyltransferase-like protein, which belongs to a family that includes the other GLCT proteins, as well as UGT enzymes. These enzymes generally catalyze the transfer of glucuronic acid to small molecules as part of the phase II detoxification system [42]. The addition of these adducts can makes the small molecules more easily secreted or less able to interact with targets, decreasing the toxicity of these compounds. The mechanism by which mutations in *glct-3* render the animal less sensitive to exogenous propionate supplementation remains unclear. Because natural variants in *glct-3* are predicted to cause loss of function, enzyme function is likely eliminated and small molecules would not be modified and detoxified. This result suggests that loss of *glct-3* would causes sensitivity to propionate. However, our data suggest the opposite conclusions, so GLCT-3 likely does not directly modify propionate as a detoxifying mechanism. Instead, our data suggest that modification of a small molecule, or perhaps protein, by GLCT-3 increases the toxicity of propionate. Future studies will determine which molecules are modified by GLCT-3 and what the functional consequences of such modifications are.

In the natural environment, *C. elegans* likely encounters a variety of bacteria and fungi that produce a plethora of small molecules, including short-chain fatty acids like propionate. These small molecules can accumulate and decrease fitness in the niche. When bacteria that produce vitamin B12 are also present in the niche, propionate toxicity can be reduced. For these reasons, natural strains of *C. elegans* might vary in their complements of *glct-3* paralogs. Strains that inhabit niches with high propionate but low levels of vitamin B12 might have a more active propionate shunt [15] or fewer members of the *glct-3* family to limit toxicity. Niches with less propionate and/or high levels of vitamin B12 could support strains that might have lost the propionate shunt or have more copies *glct-3* family members. Microevolution of similar metabolic regulators could act through differences in copy number and not through specific changes to enzymatic function or differences in gene expression. Clearly, natural changes in metabolic flux are important for how organisms deal with the complex milieu of their natural environment.

Interestingly, we could not validate the most significant QTL that we detected by GWA. This QTL on chromosome V was in strong LD with another QTL on chromosome V. Long-range LD is common in selfing organisms, especially *C. elegans* [19], but it can confound GWA mappings. Our results emphasize that all QTL need to be validated by independent strains. NILs, as we used here, offer an effective approach to rapidly test genomic intervals for correlations with observed phenotypic differences. It is likely that the chromosome V right QTL underlies the trait difference that we tested at the chromosome V left QTL. Additional near-isogenic lines need to be constructed to test this hypothesis and narrow this genomic region to a causal gene.

Our study demonstrates that natural variation can modify sensitivity to the cellular metabolite propionate. We used a *C. elegans* model that mimics metabolic conditions found in patients with propionic acidemia. These data indicate that *C. elegans* is a fruitful "simple" model to identify genetic modifiers of inborn errors in human metabolism, which is extremely difficult with human populations as these diseases are usually rare.

MATERIALS AND METHODS

Strains

All the wild strains were obtained from CeNDR (**Table S3**) and maintained at 20°C on nematode growth medium (NGM) plates on a diet of *E. coli* OP50. Near-isogenic lines (NILs) were generated using a procedure described previously {Evans, 2018 #3548] by crossing BRC20067 and DL238. Each NIL strain harbors recombination breakpoints at different locations on chromosome V generated by crossing two single recombinant strains, followed by six times backcrossing with BRC20067 to change the other five chromosomes into the BRC20067 background.

Propionate sensitivity assays

A 2 M propionic acid stock solution was prepared in a chemical hood. For 40 ml solution, 6 ml propionic acid (sigma, #402907), 13.5 ml 5 M sodium hydroxide, and 20.5 ml water were mixed together, and the pH was adjusted to 6.0 with sodium hydroxide. The solution was filter sterilized and stored at 4°C. On day 0, arrested L1 animals were placed on seeded plates with propionate and after incubation for two days, animals that developed

beyond the L1 stage were evaluated as survivors. Propionic acid survival rate was calculated as the proportion of animals that have developed beyond the L1 stage over the total number of L1 animals at day 0. Biological triplicate experiments with three technical replicates were performed. For the panel of 133 wild isolates, 100 mM propionate was used in five biological replicates, each with four technical replicates. The scale of this experiment required us to split the strains into three sets. To process the data, we first took the mean of the four technical replicates and removed biological replicate outliers, which were defined by mean + 1.5 x standard deviation. Next, we corrected the strain phenotype data for biological replicate and strain set using a linear model with the formula (phenotype ~ biological replicate + strain set). After this regression analysis, we removed outlier replicates as above and took the mean of the remaining replicate residuals per strain (**Table S3**). The phenotype data were used for association mappings.

Heritability calculations

For dose-response experiments, broad-sense heritability (H^2) estimates were calculated using the *lmer* function in the lme4 package with the following linear mixed-model (phenotype ~ 1 + (1|strain)). $H^2$was then calculated as the fraction of the total variance that can be explained by the random component (strain) of the mixed model. For the complete dose-response experiment, we calculated H^2 per dose. For the fine-scale dose response experiment, we subsampled three replicates twelve independent times for H^2calculations.

358 **Power analysis**

359 To determine the number of replicate measures we needed to collect for association
360 mapping, we measured L1 survival of the DL238 strain after exposure to 100 mM
361 propionic acid in 40 replicates. The 40 replicates consisted of eight technical replicates
362 across five independent preparations of agar plates with propionate. For a range of mean
363 differences (0.01 to 1, in increments of 0.01), we subsampled two to eight replicates for
364 each of the five plate preparations 100 times and calculated the standard deviation of L1
365 survival for the subsamples. To calculate the power to detect a difference across a range
366 of replicates and mean differences, we used the *power.t.test* function in the pwr R
367 package with the following parameters - *n* = number of subsampled replicates, *delta* =
368 (0.01 to 1, in increments of 0.01), *sd* = mean of the standard deviation subsamples,
369 *sig.level* = 0.00001, alternative = “two.sided”, type = “two.sample”. With four technical
370 replicates across five independent plate preparations we were able to detect a 20%
371 difference in means 80% of the time.

372

373 **Marker-based genome-wide association mappings**

374 GWA mapping was performed using phenotype data from 133 *C. elegans* wild strains.
375 We performed the same mapping procedure as described previously [43]. Briefly,
376 genotype data were acquired from the latest variant call format (VCF) release (Release
377 20180527) from CeNDR that was imputed using IBDseq, with the following parameters:
378 minalleles = 5%, r2window = 1500, ibdtrim = 0, r2max = 0.8 [44]. We used BCFtools to
379 filter variants that had any missing genotype calls and variants that were below 5% minor
380 allele frequency [45]. We used PLINK v1.9 to LD-prune the genotypes at a threshold of

$r^2 < 0.8$, using *--indep-pairwise 50 10 0.8* [46, 47]). This genotype data set consisted of 59,241 markers that were used to generate the realized additive kinship matrix using the *A.mat* function in the *rrBLUP* R package [48]. These markers were also used for genome-wide mappings. However, because these markers still have substantial LD within this genotype set, we performed eigen decomposition of the correlation matrix of the genotype matrix using *eigs_sym* function in Rspectra package [49](https://github.com/yixuan/RSpectra). The correlation matrix was generated using the *cor* function in the correlateR R package (https://github.com/AEBilgrau/correlateR). We set any eigenvalue greater than one from this analysis to one and summed all of the resulting eigenvalues to obtain 772 independent tests within this genotype matrix. We used the *GWAS* function in the rrBLUP package to perform genome-wide mapping with the following command: *rrBLUP::GWAS(pheno = PC1, geno = Pruned_Markers, K = KINSHIP, min.MAF = 0.05, n.core = 1, P3D = FALSE, plot = FALSE)*. To perform fine-mapping, we defined confidence intervals from the genome-wide mapping as +/- 100 single-nucleotide variants (SNVs) from the rightmost and leftmost markers above the Bonferroni significance threshold. We then generated a QTL region of interest genotype matrix that was filtered as described above, with the one exception that we did not perform LD pruning. We used PLINK v1.9 to extract the LD between the markers used for fine mapping and the peak QTL marker identified from the genome-wide scan. We used the same command as above to perform fine mapping but used the reduced variant set. The workflow for performing GWA mapping can be found at https://github.com/AndersenLab/cegwas2-nf.

404 **Sequence Kernel Association Test (SKAT) Mapping**

405 In parallel to marker-based GWA mappings, *cegwas2-nf* performs gene-based GWA
406 mappings using SKAT, which is implemented in the RVtests software [34, 50]). We set
407 the maximum allele frequency for SKAT to 50% using the --freqUpper from flag *cegwas2-*
408 *nf*, and the minimum number of strains to share a variant to two using the --minburden
409 flag.

410

411 **Linkage Disequilibrium**

412 We used the *LD* function from the *genetics* package in R to calculate linkage
413 disequilibrium and report the r^2 correlation coefficient between the markers (https://cran.r-
414 project.org/package=genetics).

415

416 **Phylogenetic analysis**

417 DNA and protein FASTA files for each species were downloaded from
418 http://download.caenorhabditis.org/v1/sequence/ [41]. DNA and protein for each species
419 FASTA files were combined and custom DNA and protein BLAST databases were built
420 using *makeblastdb* [51]. The *glct-3* coding sequence (CDS) was used to query the DNA
421 BLAST database using the *blastn* command with the *-evalue* threshold set to 1.
422 Homologous sequences were extracted from the database using the *blastdbcmd*
423 command. Next, a multiple sequence alignment of the homolgous sequences was
424 generated using MUSCLE [52] with default settings and output in the phylip format.

425 For DNA sequences, the *raxmlHPC-AVX* command from RAxML 8 (v 8.2.12) with
426 the GTRGAMMA substitution model was used to generate initial phylogenies [53]. Next,

we preformed bootstrapping with the following command *raxmlHPC-AVX* -p 12345 -x 12345 -# autoFC -m GTRGAMMA and extracted the best tree with bootstrap support.

For protein sequences, we used the bayesian information criterion model selection feature of RAxML 8 to identify VT [54] as the best substitution model with the following command: *raxmlHPC-AVX -p 12345 -m PROTGAMMAAUTO --auto-prot=bic*. Next, we performed bootstrapping of the phylogentic tree using the following command: *raxmlHPC-AVX -p 12345 -x 12345 -# autoFC -m PROTGAMMAAUTO --auto-prot=bic*. All phylogenies were visualized using the interactive tree of life software [55] or the *ggtree* R package [56].

ACKNOWLEDGEMENTS

We thank members of the Andersen and Walhout laboratories for critical comments on the manuscript. The research was funded by R01-DK115690 to A.J.M.W. and E.C.A.

REFERENCES

1. Argmann CA, Houten SM, Zhu J, Schadt EE. A Next Generation Multiscale View of Inborn Errors of Metabolism. Cell Metab. 2016;23(1):13-26. doi: 10.1016/j.cmet.2015.11.012. PubMed PMID: 26712461; PubMed Central PMCID: PMCPMC4715559.
2. Saudubray JM, Garcia-Cazorla A. Inborn Errors of Metabolism Overview: Pathophysiology, Manifestations, Evaluation, and Management. Pediatr Clin North Am. 2018;65(2):179-208. Epub 2018/03/06. doi: 10.1016/j.pcl.2017.11.002. PubMed PMID: 29502909.
3. Deodato F, Boenzi S, Santorelli FM, Dionisi-Vici C. Methylmalonic and propionic aciduria. Am J Med Genet C Semin Med Genet. 2006;142C(2):104-12. doi: 10.1002/ajmg.c.30090. PubMed PMID: 16602092.
4. Banerjee R, Ragsdale SW. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu Rev Biochem. 2003;72:209-47. Epub 2003/10/07. doi: 10.1146/annurev.biochem.72.121801.161828. PubMed PMID: 14527323.
5. Kasubuchi M, Hasegawa S, Hiramatsu T, Ichimura A, Kimura I. Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. Nutrients.

2015;7(4):2839-49. doi: 10.3390/nu7042839. PubMed PMID: 25875123; PubMed Central PMCID: PMCPMC4425176.

6. Hosseini E, Grootaert C, Verstraete W, Van de Wiele T. Propionate as a health-promoting microbial metabolite in the human gut. Nutrition reviews. 2011;69(5):245-58. Epub 2011/04/28. doi: 10.1111/j.1753-4887.2011.00388.x. PubMed PMID: 21521227.

7. Matsumoto I, Kuhara T. A new chemical diagnostic method for inborn errors of metabolism by mass spectrometry - rapid, practical, and simultaneous urinary metabolites analysis. Mass Spectrometry Reviews. 1996;15:43-57.

8. Frezal L, Felix MA. *C. elegans* outside the Petri dish. Elife. 2015;4. doi: 10.7554/eLife.05849. PubMed PMID: 25822066; PubMed Central PMCID: PMCPMC4373675.

9. Felix MA, Braendle C. The natural history of *Caenorhabditis elegans*. Curr Biol. 2010;20(22):R965-9. doi: 10.1016/j.cub.2010.09.050. PubMed PMID: 21093785.

10. Crombie TA, Zdraljevic S, Cook DE, Tanny RE, Brady SC, Wang Y, et al. Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. Elife. 2019;8. Epub 2019/12/04. doi: 10.7554/eLife.50465. PubMed PMID: 31793880; PubMed Central PMCID: PMCPMC6927746.

11. MacNeil LT, Walhout AJM. Food, pathogen, signal: The multifaceted nature of a bacterial diet. Worm. 2013;2:e26454.

12. Yilmaz LS, Walhout AJM. Worms, bacteria and micronutrients: an elegant model of our diet. Trends Genet. 2014;30:496-503.

13. Watson E, MacNeil LT, Arda HE, Zhu LJ, Walhout AJM. Integration of metabolic and gene regulatory networks modulates the *C. elegans* dietary response. Cell. 2013;153:253-66.

14. Watson E, MacNeil LT, Ritter AD, Yilmaz LS, Rosebrock AP, Caudy AA, et al. Interspecies systems biology uncovers metabolites affecting *C. elegans* gene expression and life history traits. Cell. 2014;156:759-70.

15. Watson E, Olin-Sandoval V, Hoy MJ, Li C-H, Louisse T, Yao V, et al. Metabolic network rewiring of propionate flux compensates vitamin B12 deficiency in *C. elegans.* Elife. 2016;5:pii: e17670.

16. Bulcha JT, Giese GE, Ali MZ, Lee Y-U, Walker M, Holdorf AD, et al. A persistence detector for metabolic network rewiring in an animal. Cell Rep. 2019;26:460-8.

17. Sterken MG, Snoek LB, Kammenga JE, Andersen EC. The laboratory domestication of *Caenorhabditis elegans*. Trends Genet. 2015;31(5):224-31. Epub 2015/03/26. doi: 10.1016/j.tig.2015.02.009. PubMed PMID: 25804345; PubMed Central PMCID: PMCPMC4417040.

18. Rockman MV, Kruglyak L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. PLoS Genet. 2009;5(3):e1000419. Epub 2009/03/14. doi: 10.1371/journal.pgen.1000419. PubMed PMID: 19283065; PubMed Central PMCID: PMCPMC2652117.

19. Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, et al. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. Nat Genet. 2012;44(3):285-90. doi: 10.1038/ng.1050. PubMed PMID: 22286215; PubMed Central PMCID: PMCPMC3365839.

20. Barriere A, Felix MA. High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. Curr Biol. 2005;15(13):1176-84. Epub 2005/07/12. doi: 10.1016/j.cub.2005.06.022. PubMed PMID: 16005289.
21. Barriere A, Felix MA. Temporal dynamics and linkage disequilibrium in natural Caenorhabditis elegans populations. Genetics. 2007;176(2):999-1011. Epub 2007/04/06. doi: 10.1534/genetics.106.067223. PubMed PMID: 17409084; PubMed Central PMCID: PMCPMC1894625.
22. Dolgin ES, Felix MA, Cutter AD. Hakuna Nematoda: genetic and phenotypic diversity in African isolates of *Caenorhabditis elegans* and *C. briggsae*. Heredity (Edinb). 2008;100(3):304-15. Epub 2007/12/13. doi: 10.1038/sj.hdy.6801079. PubMed PMID: 18073782.
23. Petersen C, Saebelfeld M, Barbosa C, Pees B, Hermann RJ, Schalkowski R, et al. Ten years of life in compost: temporal and spatial variation of North German *Caenorhabditis elegans* populations. Ecol Evol. 2015;5(16):3250-63. Epub 2015/09/19. doi: 10.1002/ece3.1605. PubMed PMID: 26380661; PubMed Central PMCID: PMCPMC4569023.
24. Reddy KC, Andersen EC, Kruglyak L, Kim DH. A polymorphism in *npr-1 i*s a behavioral determinant of pathogen susceptibility in *C. elegans*. Science. 2009;323(5912):382-4. doi: 10.1126/science.1166527. PubMed PMID: 19150845; PubMed Central PMCID: PMCPMC2748219.
25. Ghosh R, Andersen EC, Shapiro JA, Gerke JP, Kruglyak L. Natural variation in a chloride channel subunit confers avermectin resistance in *C. elegans*. Science. 2012;335(6068):574-8. doi: 10.1126/science.1214318. PubMed PMID: 22301316; PubMed Central PMCID: PMCPMC3273849.
26. Zdraljevic S, Strand C, Seidel HS, Cook DE, Doench JG, Andersen EC. Natural variation in a single amino acid substitution underlies physiological responses to topoisomerase II poisons. PLoS Genet. 2017;13(7):e1006891. Epub 2017/07/13. doi: 10.1371/journal.pgen.1006891. PubMed PMID: 28700616; PubMed Central PMCID: PMCPMC5529024.
27. Brady SC, Zdraljevic S, Bisaga KW, Tanny RE, Cook DE, Lee D, et al. A Novel Gene Underlies Bleomycin-Response Variation in *Caenorhabditis elegans*. Genetics. 2019;212(4):1453-68. Epub 2019/06/07. doi: 10.1534/genetics.119.302286. PubMed PMID: 31171655; PubMed Central PMCID: PMCPMC6707474.
28. Greene JS, Brown M, Dobosiewicz M, Ishida IG, Macosko EZ, Zhang X, et al. Balancing selection shapes density-dependent foraging behaviour. Nature. 2016;539(7628):254-8. doi: 10.1038/nature19848. PubMed PMID: 27799655; PubMed Central PMCID: PMCPMC5161598.
29. Burga A, Ben-David E, Lemus Vergara T, Boocock J, Kruglyak L. Fast genetic mapping of complex traits in *C. elegans* using millions of individuals in bulk. Nat Commun. 2019;10(1):2680. Epub 2019/06/20. doi: 10.1038/s41467-019-10636-9. PubMed PMID: 31213597; PubMed Central PMCID: PMCPMC6582151.
30. Gao AW, Sterken MG, Uit de Bos J, van Creij J, Kamble R, Snoek BL, et al. Natural genetic variation in *C. elegans* identified genomic loci controlling metabolite levels. Genome Res. 2018;28(9):1296-308. Epub 2018/08/16. doi: 10.1101/gr.232322.117. PubMed PMID: 30108180; PubMed Central PMCID: PMCPMC6120624.

31. Cook DE, Zdraljevic S, Roberts JP, Andersen EC. CeNDR, the *Caenorhabditis elegans* natural diversity resource. Nucleic Acids Res. 2017;45(D1):D650-D7. doi: 10.1093/nar/gkw893. PubMed PMID: 27701074; PubMed Central PMCID: PMCPMC5210618.
32. Jones KL, Schwarze U, Adam MP, Byers PH, Mefford HC. A homozygous B3GAT3 mutation causes a severe syndrome with multiple fractures, expanding the phenotype of linkeropathy syndromes. Am J Med Genet A. 2015;167A(11):2691-6. Epub 2015/06/19. doi: 10.1002/ajmg.a.37209. PubMed PMID: 26086840; PubMed Central PMCID: PMCPMC4654953.
33. Rowland A, Miners JO, Mackenzie PI. The UDP-glucuronosyltransferases: their role in drug metabolism and detoxification. Int J Biochem Cell Biol. 2013;45(6):1121-32. Epub 2013/03/19. doi: 10.1016/j.biocel.2013.02.019. PubMed PMID: 23500526.
34. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82-93. Epub 2011/07/09. doi: 10.1016/j.ajhg.2011.05.029. PubMed PMID: 21737059; PubMed Central PMCID: PMCPMC3135811.
35. Seidel HS, Rockman MV, Kruglyak L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science. 2008;319(5863):589-94. Epub 2008/01/12. doi: 10.1126/science.1151107. PubMed PMID: 18187622; PubMed Central PMCID: PMCPMC2421010.
36. Ben-David E, Burga A, Kruglyak L. A maternal-effect selfish genetic element in *Caenorhabditis elegans*. Science. 2017;356(6342):1051-5. Epub 2017/05/13. doi: 10.1126/science.aan0621. PubMed PMID: 28495877; PubMed Central PMCID: PMCPMC6251971.
37. Kim H, Ishidate T, Ghanta KS, Seth M, Conte D, Jr., Shirayama M, et al. A co-CRISPR strategy for efficient genome editing in *Caenorhabditis elegans*. Genetics. 2014;197(4):1069-80. doi: 10.1534/genetics.114.166389. PubMed PMID: 24879462; PubMed Central PMCID: PMC4125384.
38. Lee D, Zdraljevic S, Cook DE, Frezal L, Hsu JC, Sterken MG, et al. Selection and gene flow shape niche-associated variation in pheromone response. Nat Ecol Evol. 2019;3(10):1455-63. Epub 2019/09/25. doi: 10.1038/s41559-019-0982-3. PubMed PMID: 31548647; PubMed Central PMCID: PMCPMC6764921.
39. Thomas JH. Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. Genome Res. 2006;16(8):1017-30. Epub 2006/07/11. doi: 10.1101/gr.5089806. PubMed PMID: 16825662; PubMed Central PMCID: PMCPMC1524861.
40. Thomas JH, Robertson HM. The *Caenorhabditis* chemoreceptor gene families. BMC Biol. 2008;6:42. Epub 2008/10/08. doi: 10.1186/1741-7007-6-42. PubMed PMID: 18837995; PubMed Central PMCID: PMCPMC2576165.
41. Stevens L, Felix MA, Beltran T, Braendle C, Caurcel C, Fausett S, et al. Comparative genomics of 10 new *Caenorhabditis* species. Evol Lett. 2019;3(2):217-36. Epub 2019/04/23. doi: 10.1002/evl3.110. PubMed PMID: 31007946; PubMed Central PMCID: PMCPMC6457397.
42. Jancova P, Anzenbacher P, Anzenbacherova E. Phase II drug metabolizing enzymes. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2010;154(2):103-16. Epub 2010/07/30. doi: 10.5507/bp.2010.017. PubMed PMID: 20668491.

43. Zdraljevic S, Fox BW, Strand C, Panda O, Tenjo FJ, Brady SC, et al. Natural variation in *C. elegans* arsenic toxicity is explained by differences in branched chain amino acid metabolism. Elife. 2019;8. Epub 2019/04/09. doi: 10.7554/eLife.40260. PubMed PMID: 30958264; PubMed Central PMCID: PMCPMC6453569.
44. Browning BL, Browning SR. Detecting identity by descent and estimating genotype error rates in sequence data. Am J Hum Genet. 2013;93(5):840-51. Epub 2013/11/12. doi: 10.1016/j.ajhg.2013.09.014. PubMed PMID: 24207118; PubMed Central PMCID: PMCPMC3824133.
45. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987-93. Epub 2011/09/10. doi: 10.1093/bioinformatics/btr509. PubMed PMID: 21903627; PubMed Central PMCID: PMCPMC3198575.
46. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. Epub 2015/02/28. doi: 10.1186/s13742-015-0047-8. PubMed PMID: 25722852; PubMed Central PMCID: PMCPMC4342193.
47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75. Epub 2007/08/19. doi: 10.1086/519795. PubMed PMID: 17701901; PubMed Central PMCID: PMCPMC1950838.
48. Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. The plant genome. 2011;4:250-5.
49. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). 2005;95(3):221-7. Epub 2005/08/04. doi: 10.1038/sj.hdy.6800717. PubMed PMID: 16077740.
50. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016;32(9):1423-6. Epub 2016/05/08. doi: 10.1093/bioinformatics/btw079. PubMed PMID: 27153000; PubMed Central PMCID: PMCPMC4848408.
51. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. Epub 2009/12/17. doi: 10.1186/1471-2105-10-421. PubMed PMID: 20003500; PubMed Central PMCID: PMCPMC2803857.
52. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113. Epub 2004/08/21. doi: 10.1186/1471-2105-5-113. PubMed PMID: 15318951; PubMed Central PMCID: PMCPMC517706.
53. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. Epub 2014/01/24. doi: 10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID: PMCPMC3998144.
54. Muller T, Vingron M. Modeling amino acid replacement. J Comput Biol. 2000;7(6):761-76. Epub 2001/05/31. doi: 10.1089/10665270050514918. PubMed PMID: 11382360.

639 55. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
640 developments. Nucleic Acids Res. 2019;47(W1):W256-W9. Epub 2019/04/02. doi:
641 10.1093/nar/gkz239. PubMed PMID: 30931475; PubMed Central PMCID:
642 PMCPMC6602468.
643 56. Yu G, Lam TT, Zhu H, Guan Y. Two Methods for Mapping and Visualizing
644 Associated Data on Phylogeny Using Ggtree. Mol Biol Evol. 2018;35(12):3041-3. Epub
645 2018/10/24. doi: 10.1093/molbev/msy194. PubMed PMID: 30351396; PubMed Central
646 PMCID: PMCPMC6278858.
647

FIGURES

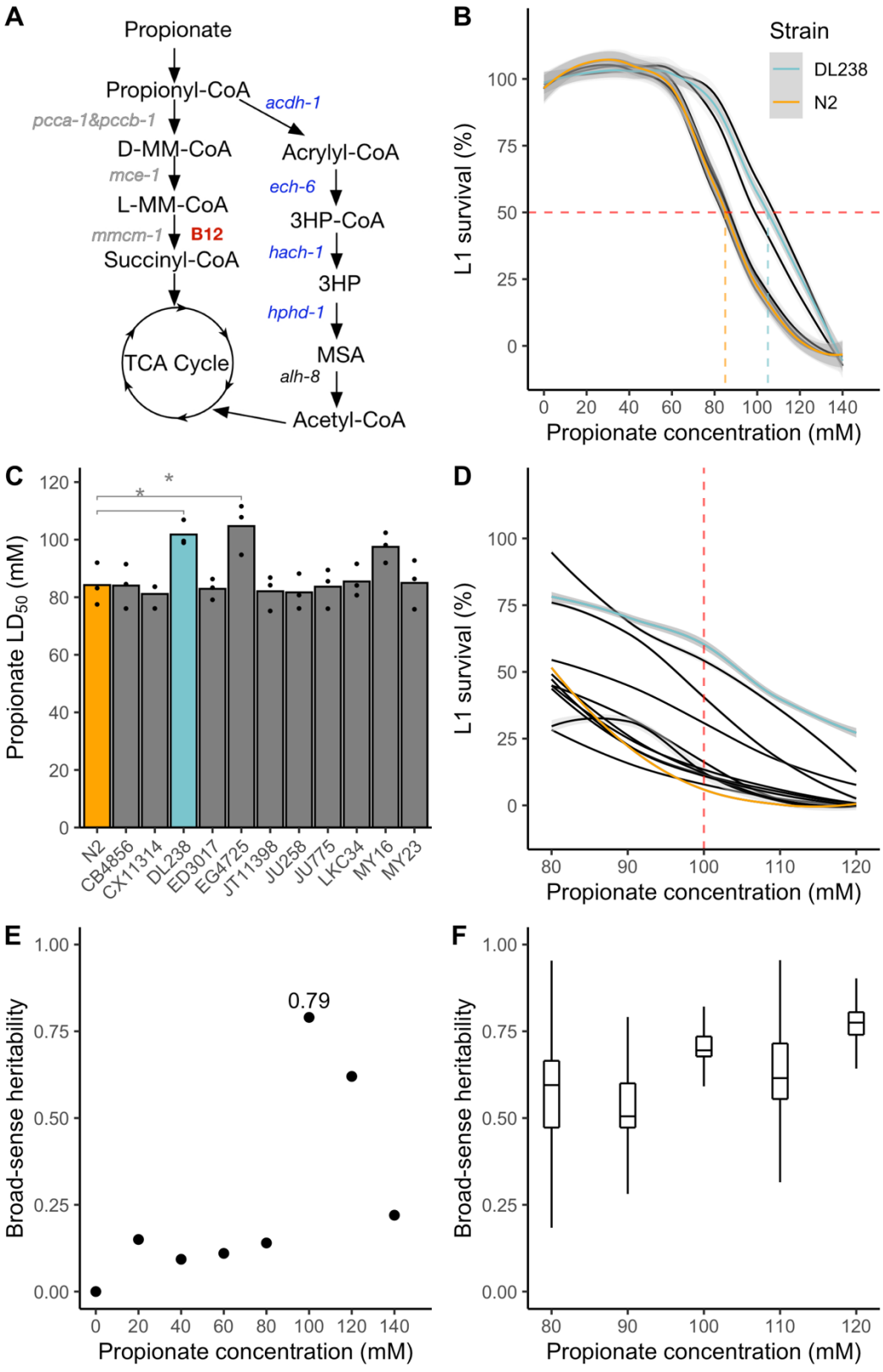


Figure 1. Natural variation in propionate sensitivity in 12 genetically diverse *C. elegans* strains

A. Propionate breakdown pathways in *C. elegans*. MM – methylmalonyl; TCA – tricarboxylic acid; MSA – malonic semialdehyde; HP – hydroxypropionate.

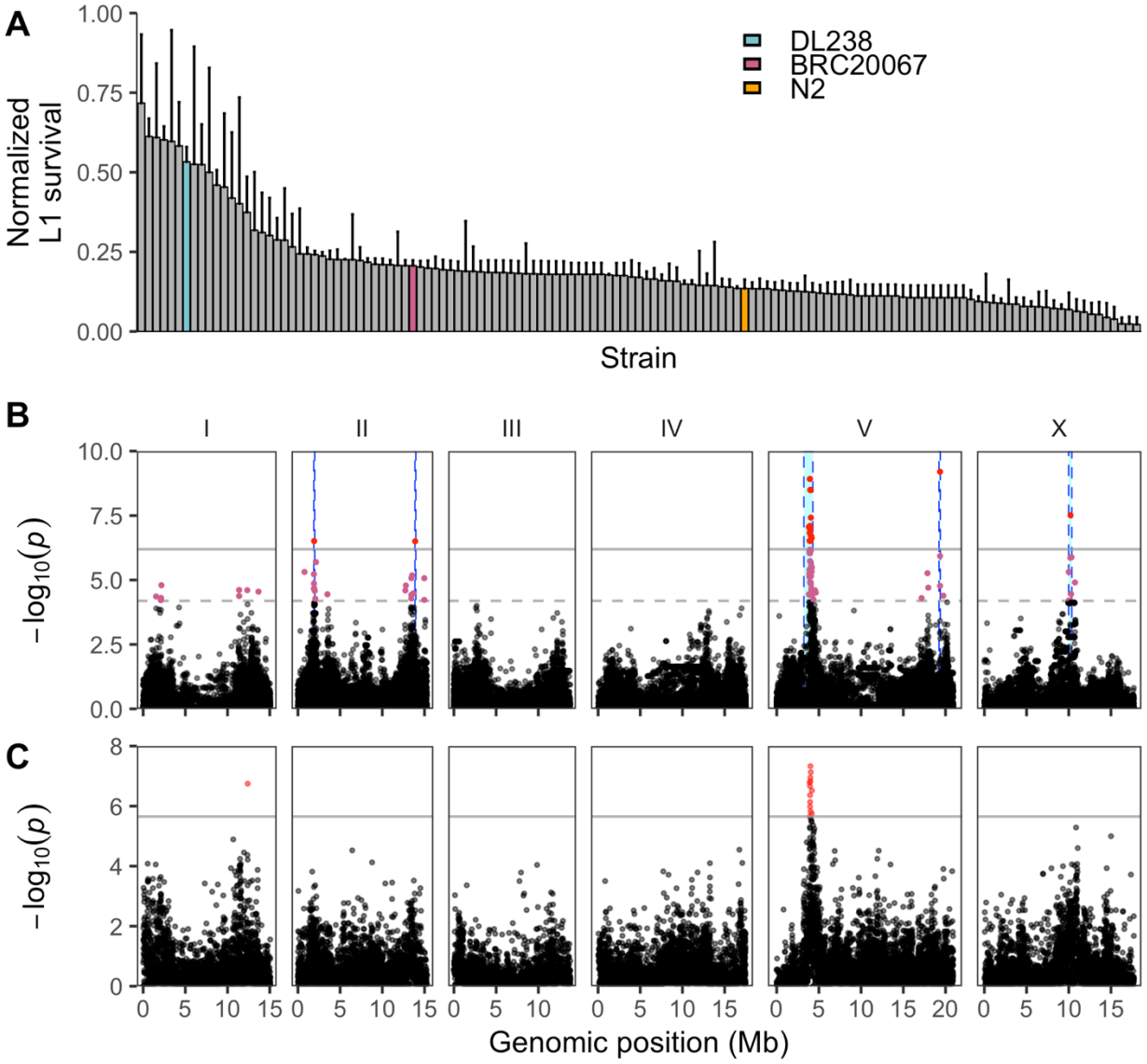
B. Propionate dose-response curves (DRCs) for 12 genetically distinct wild *C. elegans* strains. The Loess-smoothed fits of three biological replicates, each comprising three technical replicates, is shown by solid colored lines, and the standard error of the fit is shown in gray. For reference, the DRCs are colored for the N2 (orange, propionate-sensitive) and DL238 (blue, propionate-resistant) strains. The horizontal dashed red line indicates 50% L1 survival and the vertical colored lines represent the LD_{50} concentration for N2 and DL238.

C. LD_{50} values of L1 survival after propionate exposure for the 12 wild *C. elegans* strains. Three biological replicates each with three technical replicates were performed. (* indicates Student's t-Test $p < 0.05$).

D. Propionate DRCs of 12 wild *C. elegans* strains exposed to concentrations between 80 and 120 mM. The dashed red line indicates 100 mM propionate. For reference, the DRCs are colored for the N2 (orange, propionate-sensitive) and DL238 (blue, propionate-resistant) strains.

E. Broad-sense heritability (H^2) estimates from the dose response curves shown in A.

F. Tukey boxplots of broad-sense heritability (H^2) estimates for the dose response in panel D. Each boxplot represents 12 H^2 estimates after subsampling three replicate measures.



672

673 **Figure 2. Multiple QTL are associated with variable propionate sensitivities**

674 **among *C. elegans* strains**

675 (A) Normalized L1 survival in the presence of 100 mM propionate for 133 wild *C. elegans*

676 strains. L1 survival percentages were normalized by dividing each strain measurement

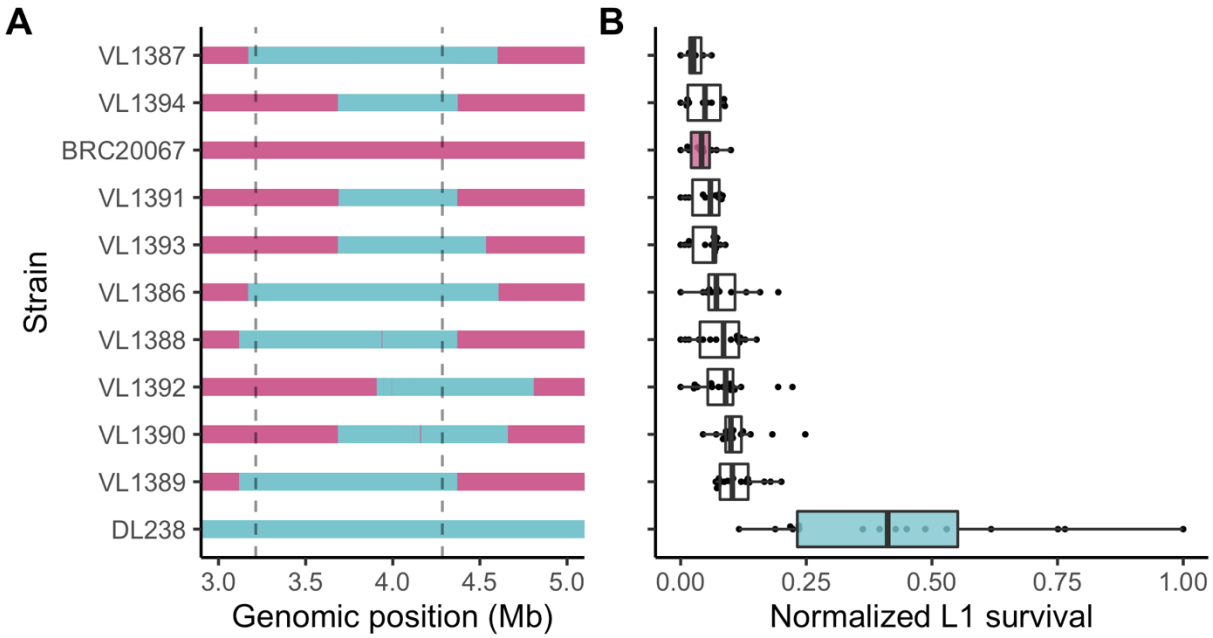
677 by the maximum L1 survival percentage of all strains. Error bars show the standard

678 deviation of replicate strain measurements. The reference strain N2 (orange) and the two

679 strains discussed throughout this work DL238 (blue) and BRC20067 (pink) are colored.

(B) Manhattan plot from marker-based GWA mapping for the normalized L1 survival percentage after propionate exposure. Each point 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 *-log10(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 threshold, which is denoted by the dotted gray horizontal line. The genomic regions of interest surrounding the QTL that pass the BF threshold are indicated in cyan.

(C) Manhattan plot from gene-based GWA mapping for L1 survival after propionate exposure. Each point represents a gene and is colored red if it passes the genome-wide BF threshold (gray line). The genomic position in Mb, separated by chromosome, is plotted on the x-axis and the *-log10(p)* for each gene is plotted on the y-axis.



693

694 **Figure 3. Chromosome V near-isogenic lines do not recapitulate the chrV right QTL**

695 **effect**

696 (A) Chromosome V genotypes of near-isogenic lines (NILs) generated between

697 BRC20067 (pink) and DL238 (blue). The dotted lines denote the QTL region of interest

698 from the marker-based GWA mapping.

699 (B) Tukey box plots of normalized L1 survival after exposure to 100 mM propionate

700 phenotypes of each NIL and parental strain. Each dot represents a replicate L1 survival

701 measurement.

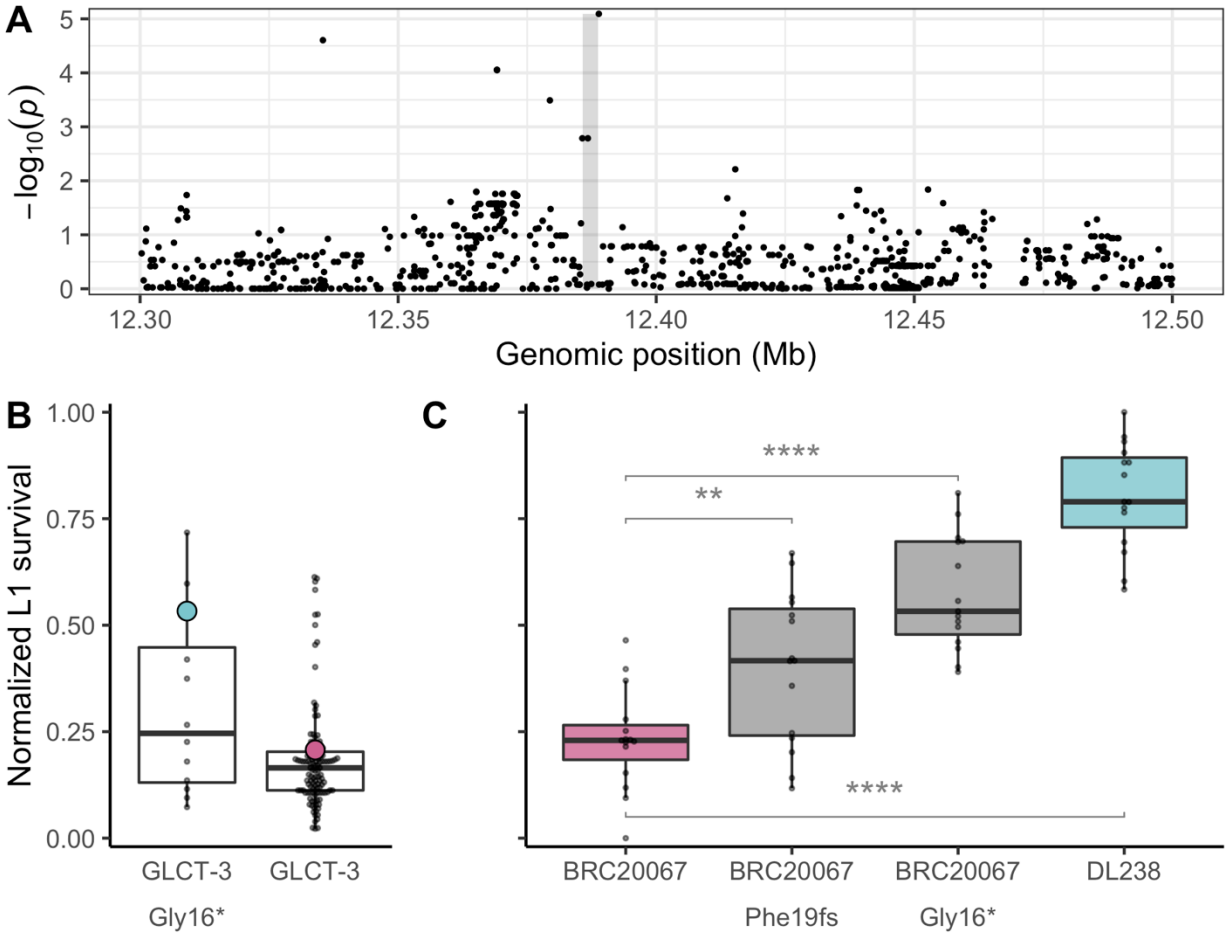


Figure 4. Variation in *glct-3* underlies differential propionate sensitivity in *C. elegans*

(A) Manhattan plot showing the strength of correlation between variants surrounding the *glct-3* gene identified by gene-based GWA mapping of the normalized L1 survival after propionate exposure phenotype. The gray shaded rectangle represents the *glct-3* gene (chrI:12385765-12388791).

(B) Tukey box plots of *C. elegans* wild isolate's normalized L1 survival after exposure to 100 mM propionate. Each dot represents the mean of 20 replicate measures for each strain. Strains are separated by the presence of a stop-gained variant at amino acid position 16 of GLCT-3. DL238 (blue) and BRC20067 (pink) are highlighted for reference.

713 (C) Tukey box plots of normalized L1 survival of each CRISPR-edited and parental strain
714 after exposure to 100 mM propionate are shown. Each dot represents a replicate L1
715 survival measurement. (** indicates Student's t-Test $p < 0.001$ and **** indicates $p <$
716 0.0001).

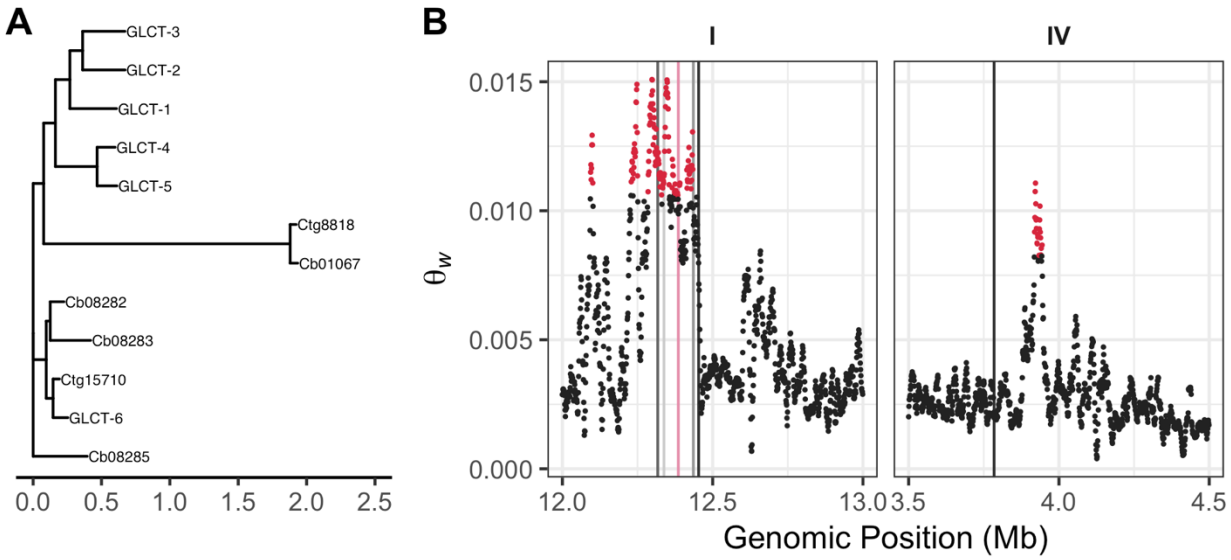


Figure 5. Expansion of the *glct* family

(A) An unrooted maximum-likelihood phylogeny of the six glucuronosyl transferase-like protein sequences encoded by the *C. elegans* genome and homologs from *C. briggsae* (Cb) and *C. tropicalis* (Ct).

(B) Watterson's theta (Θ_w) for the genomic regions that contain the six glucuronosyl transferase-like genes in *C. elegans*. Each dot represents a 10,000 bp genomic region and is colored red if the Θ_w value is greater than the 99th quantile of values across the chromosome.

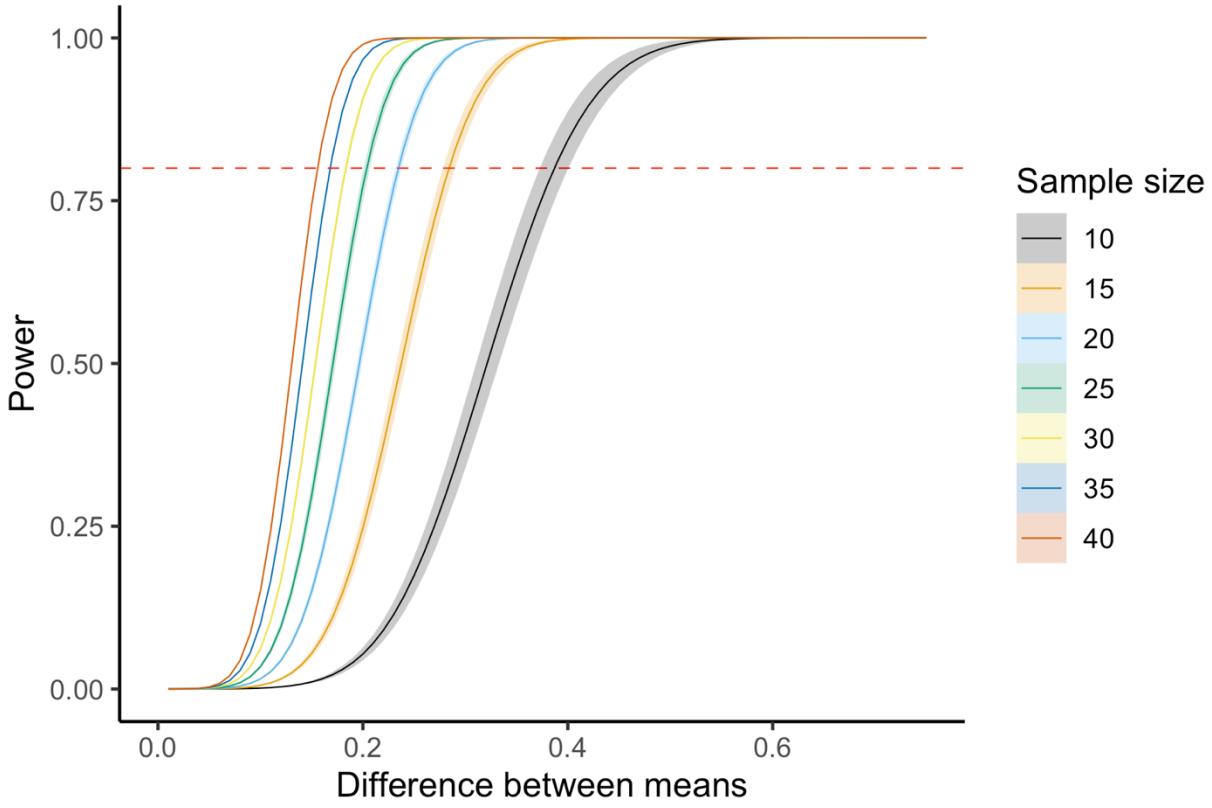
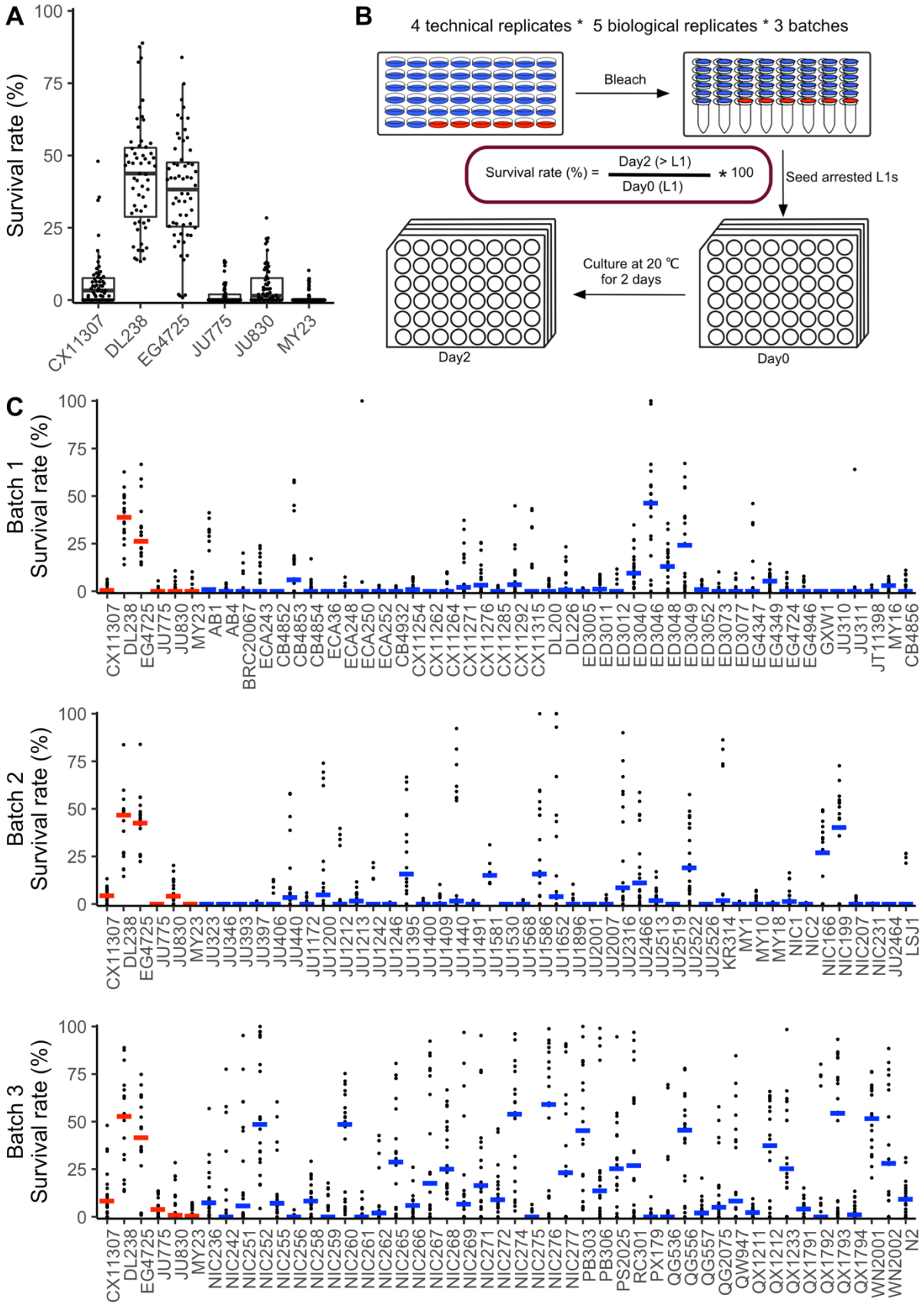


Figure S1. Power calculations

Power analysis of L1 survival after propionate exposure is shown. We calculated power for a range of mean differences from 0 to 1, using the average standard deviation of 100 subsamples from a large-scale experiment that measured DL238 propionate survival. The solid line represents the mean of 10 replicate power calculations and the shaded area around the solid lines represent the standard deviation of the replicates. The line colors represent the sample size. The dashed red line indicates 0.8 power.



734

735

736 **Figure S2. Experimental procedure and individual data sets**

737 (A) L1 survival in the presence of 100 mM propionate for six *C. elegans* strains with four
738 technical plus five biological replicates.

739 (B) Experimental setup to phenotype wild isolates for GWA mapping. 133 wild *C. elegans*
740 strains were divided into three batches to test their survival after exposure to 100 mM
741 propionate. Each batch contains 48 strains, including six control strains that control for
742 batch effects.

743 (C) L1 survival rate in the presence of 100 mM propionate for each 48 strain batch
744 described in B. The colored bars represent the median L1 survival for each *C. elegans*
745 strain. Batch-control strains are indicated by red median bars.

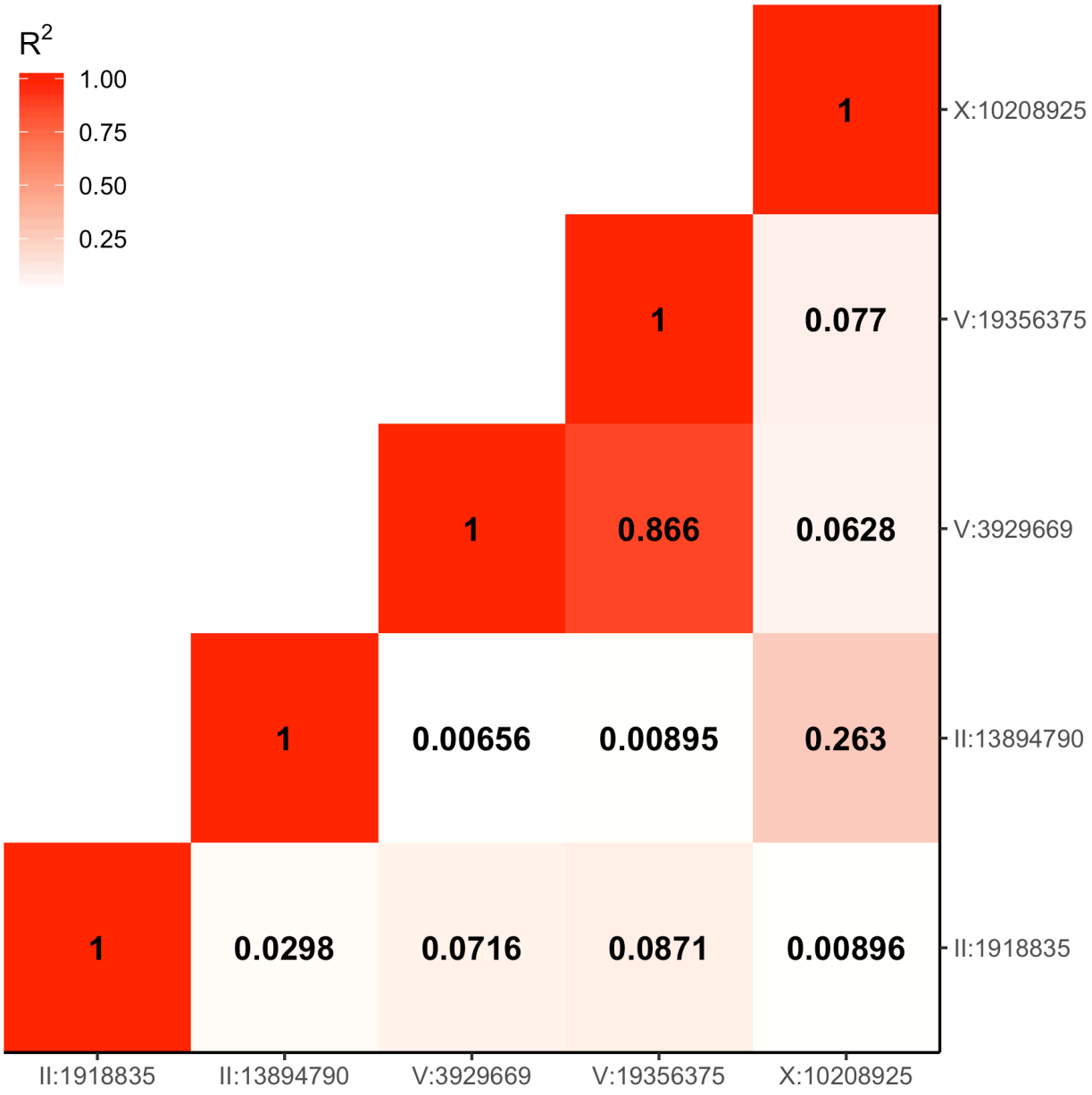


Figure S3. Linkage disequilibrium of genomic loci significantly associated with propionate sensitivity

Linkage disequilibrium (r^2) of peak QTL markers identified by genome-wide association mapping is shown. The tile color represents the correlation between marker pairs.

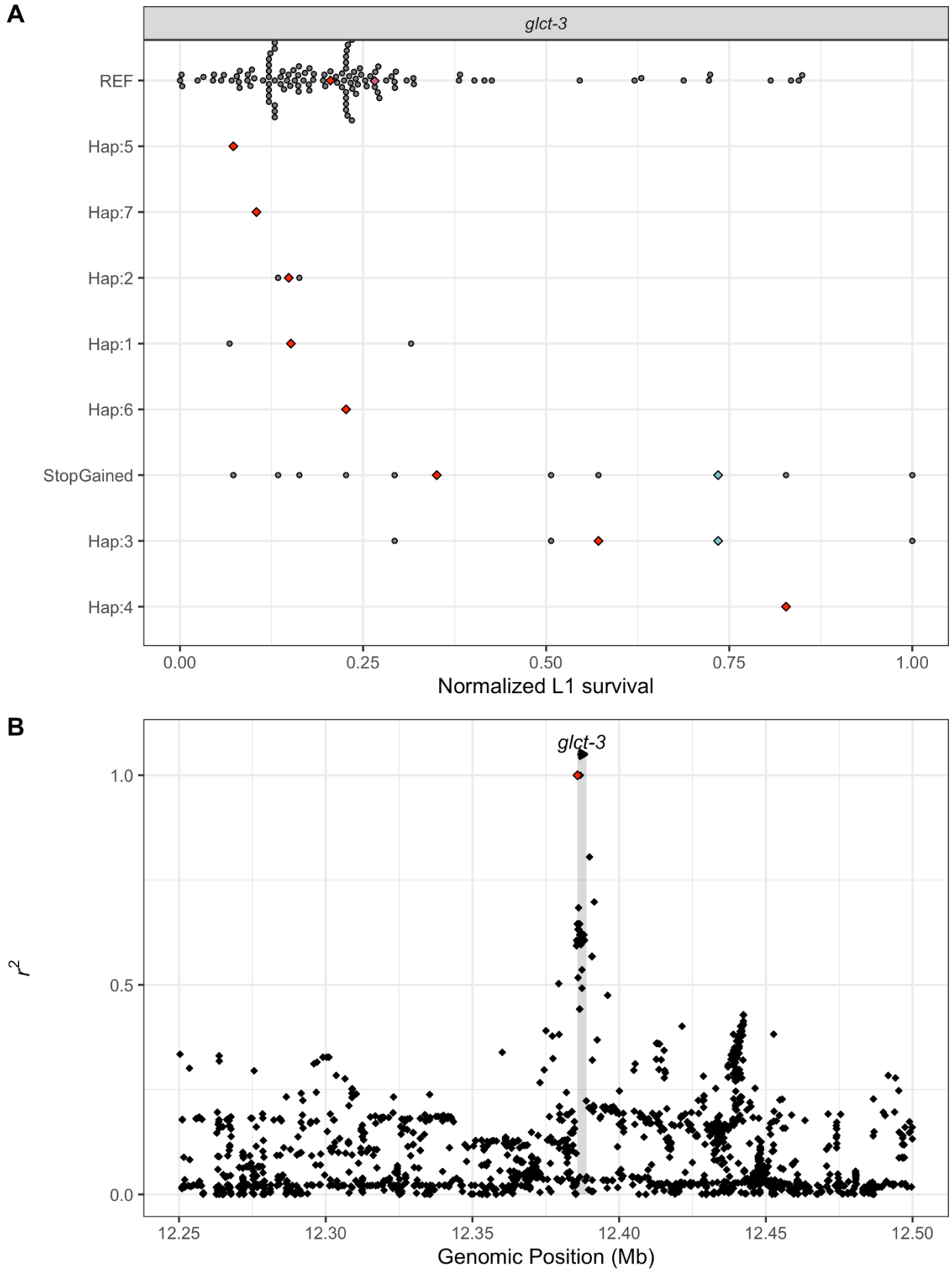


Figure S4. Variation in chromosome I genes associated with propionate sensitivity

753 (A) The normalized L1 survival in the presence of propionate for each phenotyped *C.*
754 *elegans* strain is shown on the x-axis. The y-axis represents unique haplotypes
755 (numbered from 1:n) constructed from variants with moderate-to-severe predicted effects
756 on *glct-3* found to be significantly associated with propionate sensitivity. If a variant with
757 a high predicted effect on gene function was identified, we plotted it separately. The red
758 diamonds represent the median phenotype value for each unique haplotype. The blue
759 and pink diamonds represent the DL238 and BRC20067 strains, respectively. (B) The
760 pairwise linkage disequilibrium (r^2) between the allele that encodes the Gly16* (red
761 diamond) in GLCT-3 and all variants (black diamonds) in the surrounding genomic region
762 is shown on the y-axis. The x-axis represents the genomic position (Mb) of each variant.

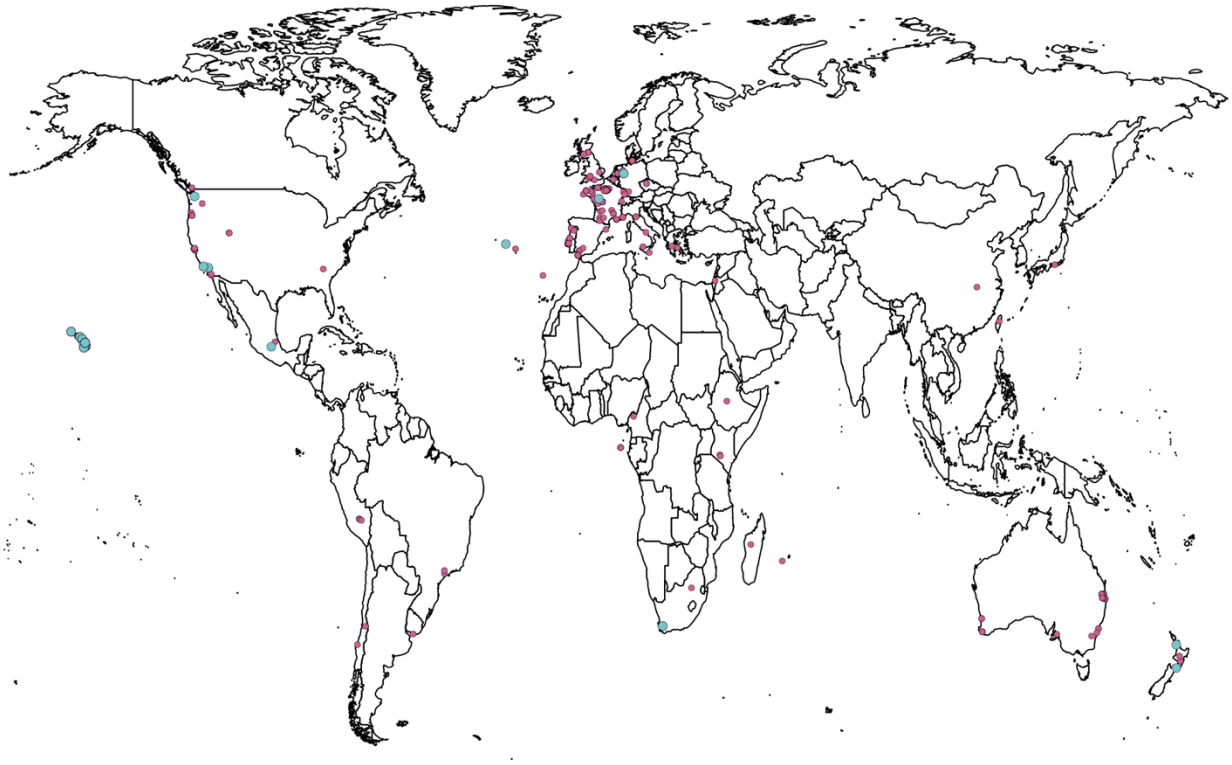
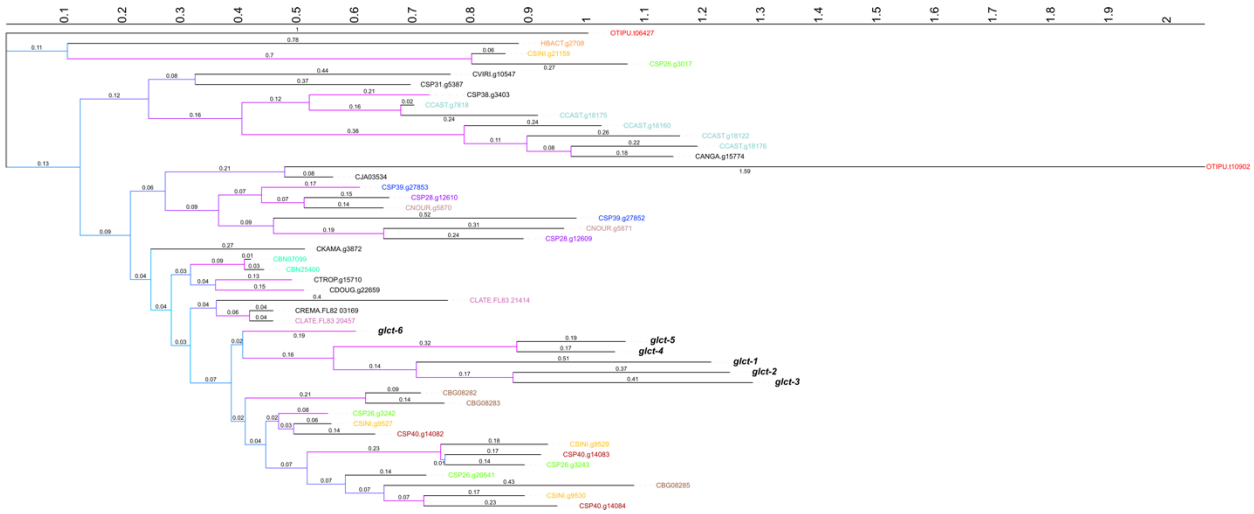


Figure S5. The global distribution of the GLCT-3 Gly16* allele

Sampling locations of wild *C. elegans* strains. Each dot represents the location where an individual strain was sampled. Pink dots represent strains carrying the REF allele at GLCT-3, and blue dots represent strains carrying the Gly16* allele.

Figure S6. Phylogenetic tree of the *C. elegans* population

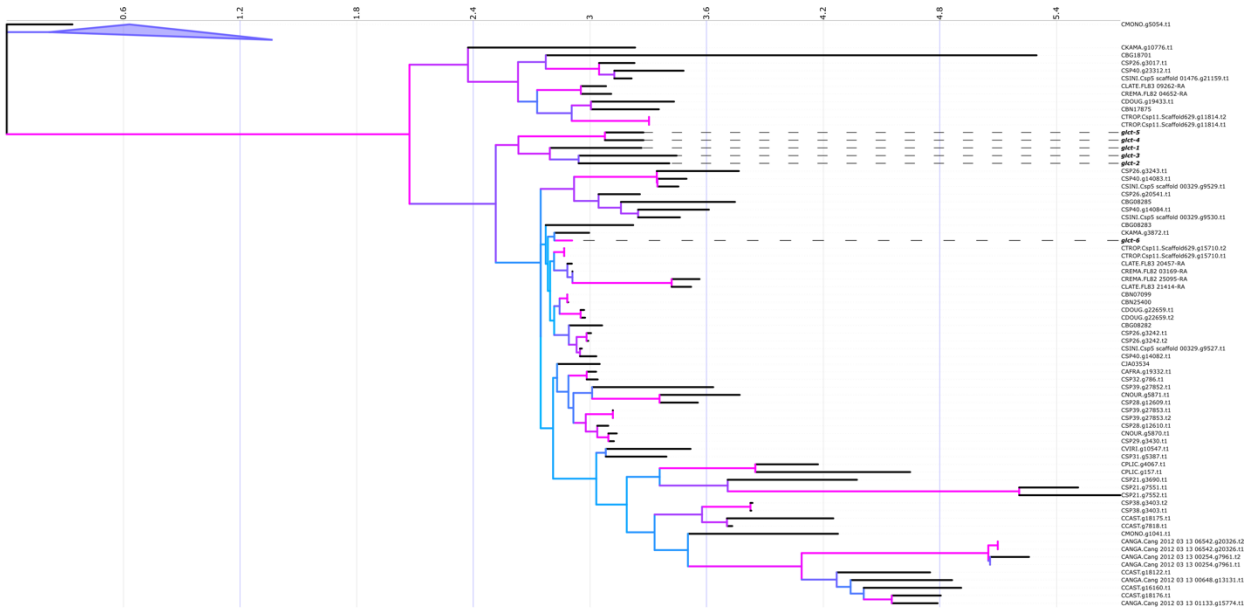
A maximum likelihood phylogenetic tree of the *C. elegans* population. Branches are colored based on the GLCT-3 allele the individual strain carries, blue represents strains with the GLCT-3 Gly16* allele, and pink represents strains with the reference allele.



774

775 **Figure S7. Phylogenetic relationship of *glct-3* homologous cDNA sequences**

776 The maximum likelihood phylogenetic relationship of *glct-3* homologs is shown. Branch
777 lengths are shown above each branch. Branch colors correspond to the bootstrap support
778 for the split, with pink indicating higher support. If a species contains more than homolog,
779 all homologs for that species are colored the same color. Species with only one homolog
780 are colored black. The *C. elegans glct* genes are colored in black and bolded.



781

782 **Figure S8. Phylogenetic relationship of GLCT-3 homologous protein sequences**

783 The maximum likelihood phylogenetic relationship of *glct-3* homologs is shown. Branch

784 colors correspond to the bootstrap support for the split, with pink indicating higher support.

785 The *C. elegans* GLCT protein sequences are bolded.

Supplemental Tables

Supplemental Table 1: Raw propionate dose response (0-140 mM) phenotypes

Supplemental Table 2: Raw propionate dose response (80-110 mM) phenotypes

Supplemental Table 3: Processed L1 survival phenotypes used for GWA mapping

Supplemental Table 4: Chromosome V NIL genotypes

Supplemental Table 5: Raw NIL L1 phenotypes

Supplemental Table 6: Raw CRISPR edit L1 phenotypes

Supplemental Table 7: Watterson's theta for chromosomes I and IV