**Copy-number variation in the glucuronyltransferase enzyme family modulates propionate sensitivity in a *C. elegans* model of propionic acidemia**

Huimin Na1#, Stefan Zdraljevic2#, Albertha J.M. Walhout1\* and Erik C. Andersen2\*

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](mailto:erik.andersen@northwestern.edu); [marian.walhout@umassmed.edu](mailto: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 have not been identified 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 five 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 glucuronosyltransferase gene family, suggesting that copy number of members of this family influences sensitivity to excess propionate. Our findings demonstrate that natural variation in genes that are not directly associated with propionate breakdown can contribute to 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 {Argmann, 2016 #3436}. 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 {Argmann, 2016 #3436}. 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 {Argmann, 2016 #3436}{Saudubray, 2018 #3598}.

Propionic and methylmalonic acidemia are inborn errors of metabolism in which the short-chain fatty acid propionate cannot be broken down {Deodato, 2006 #3135}. 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 {Banerjee, 2003 #2684}{Deodato, 2006 #3135}. 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 {Kasubuchi, 2015 #3142}{Hosseini, 2011 #2770}, it is toxic when it accumulates, as exemplified by patients with propionic acidemia {Deodato, 2006 #3135}. 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 {Matsumoto, 1996 #3068}.

The nematode *Caenorhabditis elegans* is a bacterivore found around the world {Frezal, 2015 #3306}{Felix, 2010 #3307}{Crombie, 2019 #3633}. In the laboratory, *C. elegans* can be fed different species and strains of bacteria {MacNeil, 2013 #2647}{Yilmaz, 2014 #2854}, but the vast majority of studies use the *Escherichia coli* strain OP50, a uracil auxotroph. However, *E. coli* OP50 cannot synthesize vitamin B12 and therefore cannot support the efficient breakdown of propionate by the canonical pathway {Watson, 2013 #2432;Watson, 2014 #2714}. 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 {Watson, 2016 #3248}{Bulcha, 2019 #3502}. This -oxidation pathway comprises five genes and generates acetyl-CoA {Watson, 2016 #3248}(**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* {Bulcha, 2019 #3502}. 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 propionate toxicity likely because the enzymes functioning in other metabolic pathways are repurposed {Watson, 2016 #3248}.

The vast majority of *C. elegans* studies rely on the laboratory-adapted strain named N2, from Bristol, England {Sterken, 2015 #3634}. Over the last twenty years, hundreds of *C. elegans* strains have been collected worldwide from natural habitats {Rockman, 2009 #3635}{Andersen, 2012 #3363}{Barriere, 2005 #3638}{Barriere, 2007 #3637}{Dolgin, 2008 #3636}{Petersen, 2015 #3639}{Crombie, 2019 #3633}. *C. elegans* is a self-fertilizing hermaphrodite and, therefore, different wild strains can be easily maintained as fully isogenic strains. These different strains have been used to identify quantitative trait loci that contribute to a variety of phenotypes, including anthelmintic and cancer chemotherapeutic resistance, and in several cases the precise genotypic variation that is causal to phenotypic variation has been determined {Reddy, 2009 #3369}{Ghosh, 2012 #3368}{Zdraljevic, 2017 #3594}{Brady, 2019 #3595}{Greene, 2016 #3377}{Burga, 2019 #3611}. Genomic information about the different strains is tabulated in the *C. elegans* Natural Diversity Resource (CeNDR), along with different tools for genome-wide association (GWA) mappings {Cook, 2017 #3365}.

Here, we used wild *C. elegans* strains to identify natural variation in loci that modify the resistance to exogenous propionate supplementation. To mimic propionic acidemia metabolic conditions, we fed animals a diet of *Escherichia coli* OP50, which is low in vitamin B12 and has low flux through the canonical propionate breakdown pathway {Watson, 2014 #2714;Watson, 2016 #3248}. GWA mapping using 133 strains identified five independent genomic regions or quantitative trait loci (QTL) associated with propionate resistance. For one of these loci, we found the causal variant in *glct-3,* which encodes a predicted beta-1,3-glucuronyltransferase, and is an ortholog of human B3GAT1, 2, and 3. This family of enzymes recognizes nonreducing terminal sugars and their anomeric linkages. A human homolog, B3GAT3, catalyzes the formation of the glycosaminoglycan-protein linkage by way of a glucuronyl transfer reaction in the final step of the biosynthesis of proteoglycans {Jones, 2015 #3649}. Glucuronyltransferases also catalyze reactions between metabolites, specifically the addition of glucuronic acid to toxic metabolites such as drugs {Rowland, 2013 #3612}. Interestingly, we determined 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 {Watson, 2014 #2714}{Watson, 2016 #3248}{Bulcha, 2019 #3502}. 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 LD50 of approximately 80 mM {Watson, 2014 #2714}{Watson, 2016 #3248}{Bulcha, 2019 #3502}. Supplementation of vitamin B12, which supports breakdown of propionate by the canonical pathway, confers resistance to propionate (LD50 =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 (LD50 = 50 mM) {Watson, 2014 #2714}{Watson, 2016 #3248}. 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 {Andersen, 2012 #3363}, 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 {Watson, 2014 #2714}{Watson, 2016 #3248}. 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**). Nine of the strains had similar propionate sensitivities as the N2 strain with an LD50 of approximately 85 mM. The other three strains were more resistant to propionate with an LD50 of approximately 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 1D**). This result suggests that some wild strains have natural mechanisms to cope with high levels of propionate independent of vitamin B12 and the canonical propionate breakdown pathway.

In order to perform GWA, we needed to test propionate sensitivity across a large set of wild *C. elegans* strains. We reasoned that a single dose of propionate should allow us to discriminate between sensitive and resistant strains. Propionate sensitivity assays can be noisy, in part because of slight differences in experimental factors such as incubator and room temperature, propionate concentrations (which can change slightly due to evaporation and dilution), etc. We needed to determine the dose of propionate to use in a larger set of strains and found that a concentration of 100 mM had the highest reproducibility (broad-sense heritability, *H2* = 0.79) (**Figures 1E, 1F**). Therefore, we selected this dose for further experiments. As mentioned above, propionate sensitivity assays are noisy beyond the propionate concentration as shown by the substantial variability across the dose-response experiments. Therefore, 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**

To identify the genetic basis of propionate response variation in *C. elegans*, we exposed 133 wild strains to 100 mM propionate and measured L1 survival (**Table S1** – GWA strains and phenotypes). We tested the strains in three batches and included six strains in every batch to control for potential batch effects (**Figure S2**). We observed a broad range of propionate sensitivities (**Figure 2A**). Next, we performed GWA mapping using these data and identified five QTL that were above the Bonferroni-corrected significance threshold (**Figure 2B**, two on chrII: (II:1880662-1993488, II:13859466-13979658); two on chrV: (V:3213649-4284434, V:19229887-19390858); and one on chrX: (X:9987812-10370303), coordinates are from WS245). To test the independence of these QTL, we calculated the pairwise linkage disequilibrium (LD) between each of the peak QTL markers (**Figure S3 - LD**). We observed low levels of LD for the majority of QTL pairs, with the exception of the two QTL on chromosome V (*r2* = 0.87, peak markers - V:3929669 and V:19356375), suggesting that these two QTL might not be independent. Because multiple QTL were associated with propionate sensitivity, it was difficult to decide which QTL to characterize in more detail. Therefore, we used the sequence kernel association test (SKAT), which tests an association between the phenotype of interest and the cumulative variation on a gene-by-gene basis {Wu, 2011 #3599}. This approach identified two QTL, one that overlaps with the QTL on left of chromosome V (V:3213649-4284434) identified using the single-marker mapping approach (**Figure 2B**), and a one QTL on chromosome I that only overlaps with the single-marker mapping approach at a lower significance threshold (I:12374204-12388791) (**Figure 2C**). This additional support for the QTL on the left arm of chromosome V motivated us to investigate this genomic region further.

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

To validate the effect of the QTL on the left arm of chromosome V, we constructed near-isogenic lines (NILs) in which the region associated with propionate resistance (V:3213649-4284434) was crossed from a resistant strain into the genome of a sensitive strain. To identify candidate parental strains for NIL construction, we focused on the 12 strains that were phenotyped in the dose-response experiment (**Figure 1C**). Of these 12 strains, two were significantly resistant to propionate and ten were sensitive. Next, we verified that the propionate-resistant strains had the alternative genotype at the peak QTL marker identified using the single-marker mapping method and were compatible with propionate sensitive strains at the *peel-1-zeel-1* {Seidel, 2008 #3607} and *sup-35-pha-1* {Ben-David, 2017 #3608} incompatibility loci. Using these criteria, we identified DL238 (propionate-resistant) and BRC20067 (propionate-sensitive) as suitable parental strains for NIL construction. We constructed nine NILs that contained the DL238 genomic region surrounding the chromosome V QTL introgressed into the BRC20067 genetic background (**Figure 3A**). When we exposed these NILs to propionate, we observed that the DL238 introgressed regions that correspond to the chromosome V QTL confidence interval did not confer propionate resistance (**Figure 3B**). Because the genomic region spanned by these NILs is larger than the QTL confidence interval, these results suggested that the chromosome V QTL we identified might have been the result of a spurious association with genetic variation in this region. Alternatively, the other QTL on the right arm of chromosome V identified by the single-marker mapping approach or complex epistatic interactions might underlie differences in propionate sensitivity between these two strains. The observed LD between the two QTL on chromosome V supports that hypothesis. Because the marker-based GWA mapping might have detected a spurious association, we focused on the QTL identified on the right of chromosome I where the gene-based mapping overlapped the marker-based GWA mapping at the lower significance threshold (**Figure 2B and 2C**).

**Variation in the *glct-3* gene confers propionate resistance**

To narrow and validate the genetic differences that underlie the chromosome I QTL, we closely inspected the mapped genomic interval. The QTL region on chromosome I contains only four genes [I suggested a cartoon of the four genes as Figure 4A – also panels A and B are crazy large and don’t need to be. Importantly, gene-based GWA mapping revealed that variation in the gene *glct-3* is most correlated with propionate sensitivity across the *C. elegans* population (**Figure S4**). Within the *glct-3* gene, we observed eight distinct combinations of alleles (haplotypes) among the phenotyped wild strains (**Figure S4**). Five of these distinct haplotypes all included the same stop-gained variant at amino acid position 16 (Gly16\*), along with other variants (Gly17Arg, Ser50Ala, Ser111Thr, Tyr231Cys in the QX1793 strain; Gly17Arg Ser50Ala in the CX11276, DL238, ED3046, ED3049, and NIC252 strains; Leu184Phe in the ECA36 strain; Ile46Thr in the QX1792 strain, and only Gly16\* in the MY23 and QX1791 strains). Strains with the Gly16\* variant were 20% more resistant to propionate treatment than strains with no variation in *glct-3* (**Figure 4A**). This genomic regions explains on 13.1% of the variation in response to exogenous propionate, indicating that, although other loci underlie this trait, this gene is major contributor to natural differences in resistance to propionate.

To test whether variation in the *glct-3* is causal for the difference in propionate sensitivity, we generated two independent *glct-3* alleles in the propionate sensitive BRC20067 strain using CRISPR-Cas9 genome editing {Kim, 2014 #3034}. The *ww62* allele has a one base pair deletion at position 57 in the first exon that causes a frameshift in the reading 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 effect size difference between DL238 and BRC20067 as measured by Cohen’s F, respectively (<https://doi.org/10.4324/9780203771587>) (**Figure 4B**), demonstrating that the loss of *glct-3* function confers resistance to propionate. [for discussion: how much of the overall heritability is explained by *glct-3* variation?].

**(calculated with both the additive and epistatic kinship matrices)**

Marker-based heritability (*H2*) from GWA is 0.46 and additive (*h2*) is 0.15. *glct-3* stop-gained accounts for 40.9% of the *h2* and 13.1% of *H2*.

***glct* gene family copy number varies across *C. elegans* species**

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 and 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* {Lee, 2019 #3640, PMID: 16825662,18837995}. We observed elevated levels of variation in the genomic region that contains *glct-1* through *glct-5* (**Figure 5B-C**), which supports the hypothesis that these sequences duplicated at some point in the *C. elegans* lineage and then diverged.

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 (PMID: 31007946). 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 SX**). The prevalence of low-copy numbers of *glct* genes among a majority of *Caenorhabditis* species suggests that the ancestral copy number state 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 SX**), this gene is likely the ancestral state of this gene family.

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 {Cook, 2017 #3365}, 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 4C**), which are known to harbor the most genetically divergent *C. elegans* individuals {Crombie, 2019 #3633}. 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 4D**). However, not all of the genetically divergent strains harbor variation in *glct-3*. This result suggests that the ancestral *C. elegans* state had an expanded *glct* gene family but the number of *glct* genes varies across this species. 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 {Cook, 2017 #3365}. 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 a 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. Taken together, these results suggest that the copy number of *glct* genes might affect fitness in the wild.

|  |  |
| --- | --- |
| **Species** | ***glct-3* homologs** |
| *C. elegans* | 6 |
| *C. castelli* | 5 |
| *CSP26* | 4 |
| *C. briggsae* | 3 |
| *CSP40* | 3 |
| *C. sinica* | 3 |
| *C. nouraguensis* | 2 |
| *CSP28* | 2 |
| *CSP39* | 2 |
| *C. Brenneri* | 2 |
| *C. latens* | 2 |
| *CSP31* | 1 |
| *CSP38* | 1 |
| *C. virilis* | 1 |
| *C. japonica* | 1 |
| *C. kamaaina* | 1 |
| *C. remanei* | 1 |
| *C. tropicalis* | 1 |
| *C. doughertyi* | 1 |
| *Oscheius tipulae* | 2 |
| *Heterorhabditis bacteriophora* | 1 |

**DISCUSSION**

Points for Discussion

* Our findings
* The strongest QTL was not the right one – statistics does not always rule, but LD told us it might not be the right gene. Need to pay attention to LD and validate all QTL
* How much does glct-3 explain heritability of propionate sensitivity
* GLCTs and their function and anything related to human disease, copy number changes underlie mechanism to deal with propionate poisoning, varies in nature because local microbiome might vary in composition (B12-producing bugs)
* Evolution of metabolic regulators through copy number changes not through evolution of novel functions, suggests that flux is important and that metabolism is most easily altered by changing flux than by making new shunts (make sense?)
* End with how awesome *C. elegans* is as model to study IEM and their modifiers

**MATERIALS AND METHODS**

**Strains**

All the wild strains were obtained from CeNDR (**Table S1**){Cook, 2017 #3365 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 of ChrV 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 (**Supp Data XX - input mapping phenotype**). The phenotype data were used for association mappings.

**Heritability calculations**

For dose-response experiments, broad-sense heritability () estimates were calculated using the *lmer* function in the lme4 package with the following linear mixed-model (phenotype ~ 1 + (1|strain)). 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 per dose. For the fine-scale dose response experiment, we subsampled three replicates twelve independent times for calculations.

**Power analysis**

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

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

GWA mapping was performed using phenotype data from 133 *C. elegans* wild strains. We performed the same mapping procedure as described previously {Zdraljevic, 2019 #3600}. Briefly, genotype data were acquired from the latest variant call format (VCF) release (Release 20180527) from CeNDR that was imputed using IBDseq, with the following parameters: minalleles = 5%, r2window = 1500, ibdtrim = 0, r2max = 0.8 {Browning, 2013 #3601}. We used BCFtools to filter variants that had any missing genotype calls and variants that were below 5% minor allele frequency {Li, 2011 #3602}. We used PLINK v1.9 to LD-prune the genotypes at a threshold of 0.8, using *--indep-pairwise 50 10 0.8* {Chang, 2015 #3603}{Purcell, 2007 #3604})*.* 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 (doi:10.3835/plantgenome2011.08.0024). 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 {Li, 2005 #3606}( 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>.

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

In parallel to marker-based GWA mappings, *cegwas2-nf* performs gene-based GWA mappings using SKAT, which is implemented in the RVtests software {Wu, 2011 #3599}{Zhan, 2016 #3605}). We set the maximum allele frequency for SKAT to 5% using the --freqUpper from flag *cegwas2-nf*, and the minimum number of strains to share a variant to two using the --minburden flag.

**Linkage Disequilibrium**

We used the *LD* function from the *genetics* package in R to calculate linkage disequilibrium and report the *r2* correlation coefficient between the markers (<https://cran.r-project.org/package=genetics>).

**Phylogenetic analysis**

**DNA and protein FASTA files for each species were downloaded from <http://download.caenorhabditis.org/v1/sequence/> (**PMID: 31007946)**. DNA and protein for each species FASTA files were combined and custom DNA and protein BLAST databases were built using *makeblastdb* (PMID: 20003500). The glct-3 CDS was used to query the DNA BLAST database using the *blastn* command with the *-evalue* threshold set to 1. Homologous sequences were extracted from the databse using the *blastdbcmd* command. Next, a multiple sequence alignment of the homolgous sequences was generated using MUSCLE (PMID: 15318951) with default settings and output in the phylip format.**

**For DNA sequences, the *raxmlHPC-AVX* command from RAxML 8 (v** **8.2.12) with the GTRGAMMA substitution model was used to generate initial phylogenies (PMID: 24451623). 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 (PMID: 11382360) 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 (PMID: 30931475) or the *ggtree* R package (PMID: 30351396).**

**REFERENCES**

**FIGURES**

**A close up of a map

Description automatically generated**

**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. LD50 values for the L1 survival trait in response to propionate supplementation for the 12 wild *C. elegans* strains. Three biological replicates each with three technical replicates were performed. (\*indicates p<0.05).

C. Dose response curves 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. Propionate concentrations were tested from 0 to 140 mM in 20 mM increments. For reference, the N2 (orange, sensitive) and DL238 (blue, resistant) strains are colored. The dashed red line indicates 100 mM propionate.

D. Dose response curves of 12 genetically distinct wild *C. elegans* strains. Propionate concentrations were tested from 80 to 120 mM in 10 mM increments. The dashed red line indicates 100 mM propionate.

E. Broad-sense heritability (*H2*) estimates for different concentrations of propionate.

F. Tukey boxplots of broad-sense heritability (*H2*) estimates. Each dot corresponds to an *H2* estimate after subsampling three replicate measures.

**A picture containing sky, tree

Description automatically generated**

**Figure 2. Multiple QTL are associated with variable propionate sensitivities among *C. elegans* strains**

(A) Normalized L1 survival in the presence of 100 mM propionate for 133 wild *C. elegans* strains. L1 survival percentages were normalized by dividing each strain measurement by the maximum L1 survival percentage of all strains. Error bars show the standard deviation of replicate strain measurements.

(B) Manhattan plot from marker-based GWA mapping for the 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 (ED) 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.

**A screenshot of a cell phone

Description automatically generated**

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

(A) Chromosome V genotypes of near-isogenic lines (NILs) generated between BRC20067 (pink) and DL238 (blue). The dotted lines denote the QTL region of interest from the marker-based GWA mapping.

(B) Tukey box plots of the L1 survival phenotypes of each NIL and parental strains. Each dot represents a replicate L1 survival measurement.

**A close up of a map

Description automatically generated**

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

(A) Box plots showing normalized L1 survival in response to 100 mM propionate supplementation for strains carrying the two different *glct-3* alleles. Each dot represents a strain.

(B) Tukey box plots of the L1 survival of each CRISPR-edited and parental strain upon supplementation of 100 mM propionate. Each dot represents a replicate L1 survival measurement.

(C) Map indicating where propionate resistant and sensitive strains were found.

(D) Phylogenetic tree indicating evolutionary history of different *C. elegans* strains.

**A close up of a map

Description automatically generated**

**Figure 5.** ***glct* Family Expansion [EA: please provide legend]**

(A) An unrooted maximum-likelihood phylogeny of the six glucuronosyl transferase-like gene coding sequences present in the *C. elegans* genome. Branch colors correspond to the bootstrap support for the split, with pink indicating higher support.

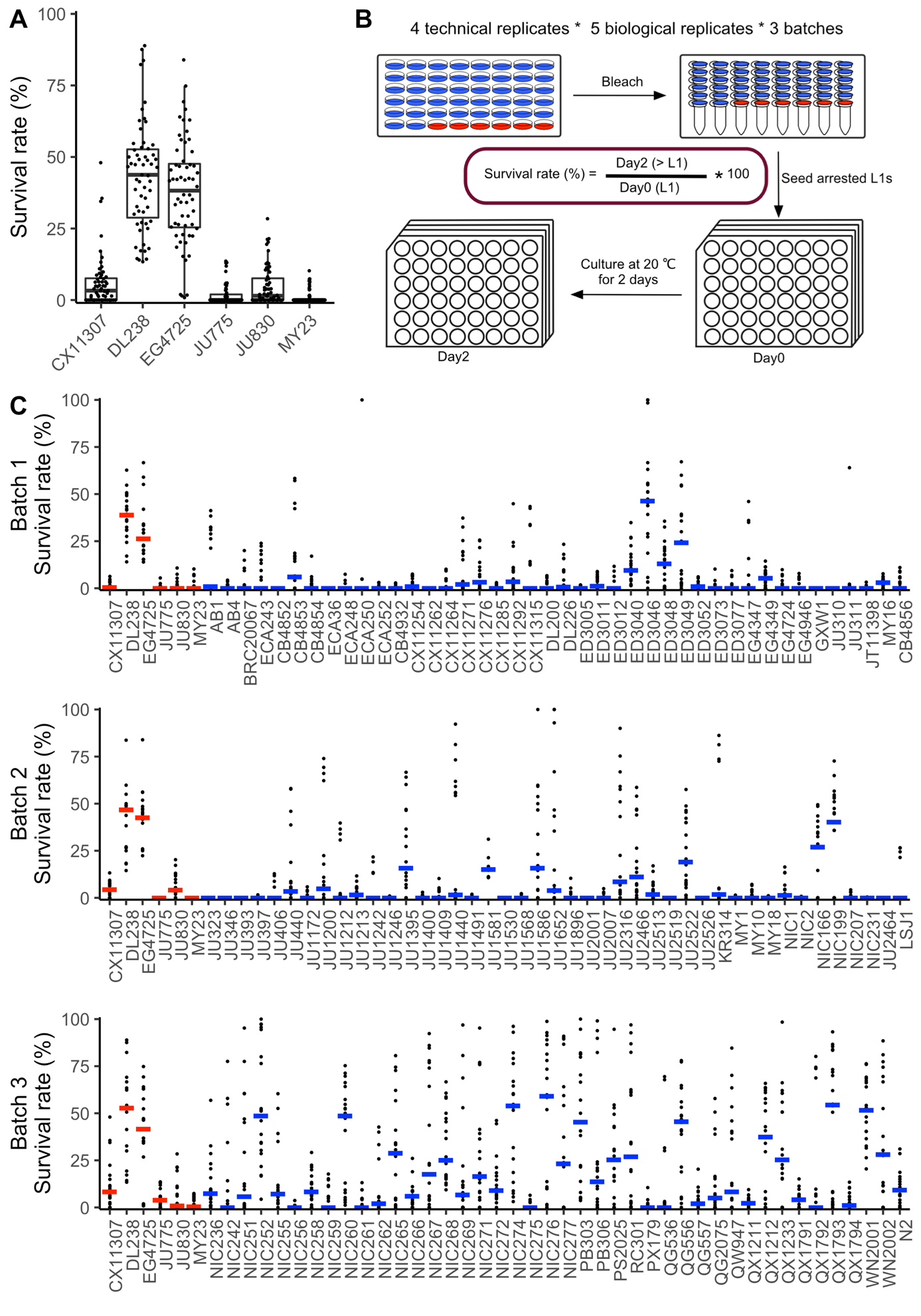
(B)

(C)

**A close up of a logo

Description automatically generatedFigure S1. Power Calculation**

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.

****

**Figure S2. Experimental Procedure and Individual Data Sets**

(A) The L1 survival in the presence of 100 mM propionate for six *C. elegans* strains with four technical plus five biological replicates is shown on the y-axis.

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

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

**A screenshot of a cell phone

Description automatically generated**

**Figure S3. Linkage Disequilibrium of Genomic Loci Significantly Associated with Propionate sensitivity**

The linkage disequilibrium (*r2*) of peak QTL markers identified by genome-wide association mapping is shown. The tile color represents the correlation between marker pairs.

A screenshot of a cell phone

Description automatically generated

**Figure S4. Variation in chromosome I genes associated with propionate sensitivity [I don’t have this figure so I stopped here]**

(A-D) The normalized L1 survival in the presence of propionate for each phenotyped *C. elegans* strain is shown on the x axis. The y axis represents unique haplotypes constructed from variants with moderate-to-severe predicted effects on gene function for each gene found to be significantly associated with propionate sensitivity. Haplotypes are numbered from 1:n for the genes significantly associated with *C. elegans* propionate responses: (A) *bgnt-1.7*, (B) *gly-17*, (C) *nhr-77*, and (D) *glct-3*. If a variant with a high predicted effect on gene function was identified, we plotted it separately. The red diamonds represent the median phenotype value for each unique haplotype. The blue and pink diamonds represent the DL238 and BRC20067 strains, respectively. E) The pairwise linkage disequilibrium (*r2*) between the allele that encodes the Gly16\* (red diamond) in GLCT-3 and all variants (black diamonds) in the genes significantly associated with *C. elegans* propionate responses is shown on the y axis. The x axis represents the genomic position (Mb) of each variant.

**Figure SXX.**

A screenshot of a cell phone

Description automatically generated

**Phylogenetic relationship of *glct-3* homologous cDNA sequences**

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

**Figure SXX.**

A close up of a map

Description automatically generated

**Phylogenetic relationship of GLCT-3 homologous protein sequences**

The maximum likelihood phylogenetic relationship of *glct-3* homologs is shown. Branch colors correspond to the bootstrap support for the split, with pink indicating higher support. The *C. elegans* GLCT protein sequences are bolded.