

NORTHWESTERN UNIVERSITY

Natural Genetic Variation Underlies Chemotherapeutic Drug Responses

*in Caenorhabditis elegans*

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biological Sciences

By

Shannon Colleen Brady

EVANSTON, ILLINOIS

September 2019

All contents of this dissertation  
not previously published and  
covered under separate licenses  
should be considered licensed under the  
creative commons CC0 (public domain) license



[creativecommons.org/publicdomain/zero/1.0/](https://creativecommons.org/publicdomain/zero/1.0/)

## Abstract

Individuals within a species vary in complex phenotypes, such as responses to toxins. This drug-response variation causes patients who are treated with the same medicine to experience a range of side effects, ultimately decreasing the efficacy of some drugs. Particular genetic variants among individuals might contribute to differential drug responses, and these biomarkers could be used to predict treatment outcomes. However, detecting these genetic variants is difficult in human populations because of statistical power limitations and confounding environmental variation. The model organism *Caenorhabditis elegans* can be used to understand how genetic variants underlie drug responses across divergent strains. In this dissertation, I describe how I used linkage mapping to detect quantitative trait loci (QTL) that contribute to toxin-response differences between two divergent strains, N2 and CB4856. First, I discuss a study in which I identified a nematode-specific gene, *scb-1*, that causes differences in responses to the chemotherapeutic drug bleomycin. Variation in expression of this gene likely underlies bleomycin-response differences across recombinant lines derived from the N2 and CB4856 strains. Additionally, I localized epistatic regions that contribute to bleomycin hypersensitivity. Next, I discuss the identification of regions of the genome that are enriched for toxin-response QTL. The detection of these hotspots suggests that pleiotropic loci might modulate drug responses in *C. elegans*. I also discuss the relative contributions of additive and interacting loci toward responses to each of these toxins. The results of these two projects suggest that toxin resistance might be selected in nature, especially in the case of bleomycin. My work on linkage mapping in *C. elegans* highlights the power of this system to detect additive and epistatic QTL and frames the importance of these findings in the light of evolution.

## Acknowledgments

This dissertation could not have been possible without the support I received from certain individuals. First and foremost, I would like to acknowledge my advisor, Dr. Erik Andersen, who was a constant force pushing the progress of this dissertation. Erik offered guidance and support during challenging times, and continually dissuaded me from “cutting bait” on the bleomycin project. Additionally, Erik greatly enhanced my understanding of genetics, *C. elegans*, experimental design, grammar, and writing. I cannot acknowledge Erik without also thanking Dr. Robyn Tanny, an integral member of the Andersen Lab. Robyn is the most patient teacher and the most organized strain “czar”, and she served as a relatable role model when I was coping with the frustrations of a fussy project (or a “science baby”, as she called it). Erik and Robyn are a dynamic duo who will continue to impress the *C. elegans* community with their significant contributions to the field.

The first graduate student in Erik’s lab was Stefan Zdraljevic, who deserves an entire paragraph of acknowledgment to himself. Stefan was assigned to be my mentor through my rotation in the lab, and he continued to mentor me through the entirety of my Ph.D. He taught me how to pick worms, how to run a PCR, how to design CRISPR guides, how to code in R, and essentially any skill I needed to complete this dissertation. But Stefan was more than a scientific mentor; he was my confidant, my life coach, my protector, and my sounding board. Stefan will continue to be a role model for me and for future mentees that have the fortune of getting to know him.

Just as Stefan was my mentor, I had the pleasure of mentoring Katie Evans during her rotation. Katie and I quickly bonded over Disney, country music, and crafting, and we successfully

completed large projects together. The days that I worked with Katie were my happiest memories of my time in the lab (even when we would spend long hours battling the worm sorter, casually wearing onesies). Katie is such a talented scientist - she is organized, extremely intelligent, and has a true knack for designing scripts that output beautiful reports. Although I was the mentor, I learned a great deal from Katie, and I am so proud to have had the chance to work side-by-side with her. In the summer of 2017, Katie, Briana Rodriguez, and I formed "Team Moana" for the Hawaii trip. Our adventures navigating Kauai, Maui, and Molokai (on very limited sleep) yielded hilarious stories, special memories, a couple new *C. elegans* strains, and a strong bond that will last a lifetime.

In the Summer of 2016, I met Dr. Daehan Lee at The Allied Genetics Conference. I immediately felt Daehan's passion for science and was thrilled when he decided to join the Andersen lab as a postdoctoral fellow. That passion continued throughout my time working with Daehan, and he was my biggest cheerleader when I would present my scientific results. Daehan encouraged me to stay excited about science, especially during moments when I was losing sight of the impact of my work. His energy and enthusiasm were critical for my ability to write this dissertation, and I look forward to seeing him continue to grow as a mentor, a scientist, and a father to "Baby Rae".

I would like to acknowledge other members of the Andersen Lab who supported the development of the projects in this dissertation. My undergraduate mentee, Karol Bisaga, played a crucial role in creating strains for the bleomycin manuscript, and he served as a reliable right-hand man throughout his time as my mentee. My other undergraduate mentee, Ellen Chao, was the most organized undergraduate I have met, and her energy and intelligence made mentoring her a treat. Dan Cook graduated one year before I did, but he was always willing to help me with coding,

sequencing, or résumé building, and he was a source of so many laughs. Clay Dilks joined the lab during my fifth year but has already impressed me with his work ethic and brings a calm, friendly energy to the lab. Nicole Roberto joined the lab as a wonderfully organized and efficient technician toward the end of my time in graduate school, and led the lab's move from Cook Hall to Silverman Hall. Tim Crombie acted as the fearless wrangler of all the Hawaii data and brought an optimistic, helpful perspective to each conversation we shared. I had the pleasure of working with Gaotian Zhang, Ye Wang, Anita Huang, Ryan Chung, Steffen Hahnel, Mostafa Zamanian, Kreena Patel, Kristen Laricchia, Josh Roberts, Tyler Shimko, Sarah Giuliani, and Sam Rosenberg during my time in graduate school, and they each contributed to the powerful science and friendships that come out of the Andersen Lab.

I am grateful to my committee members who offered sage words of advice during my qualifying exam and during each committee meeting. Each of them made helpful comments regarding my scientific and career choices that led me to achieving my Ph.D. Dr. Richard Morimoto acted as the committee chair, and Dr. Robert Holmgren, Dr. Xiaozhong (Alec) Wang, and Dr. Erik Andersen were the other members of my committee. I am also grateful for my funding sources, the Biotechnology Training Program and the Dr. John N. Nicholson fellowship.

At the beginning of my fourth year of graduate school I had the opportunity to complete an internship in the industry of biotechnology. I used this opportunity to explore the Pacific Northwest of the United States and worked with Dr. Alyssa Rosenbloom in the Research and Development team at Nanostring Technologies. Alyssa fostered my growth as a scientist by encouraging my independence while offering guidance when needed. She was able to recognize my strengths

and boosted my confidence at a much needed time. She and Dr. SuFey Ong welcomed me to Seattle with open arms, and for that I am truly grateful.

During the majority of my time in graduate school, I had the fortune of an incredibly supportive roommate, Emma Coughlin. Emma and I were partners in countless outreach initiatives including HerStory, the AP Biology Symposium, and Morning Mentors at Nichols Middle School, which we proudly co-founded in 2015. Emma and I acted as sounding boards for each other during experimental design and grant writing and had many nights in our living room drawing ideas on a whiteboard. She helped me brainstorm and execute ideas (especially for escape-room designs), navigate the challenges of taxes, funding, literal navigation of campus, and all things graduate school, and was the most reliable confidant during my lowest points of graduate school. I am forever grateful to Emma for the friendship she provided and likely would not have made it this far without her.

In addition to my fellow scientists, I was fortunate to have an expansive network of family and friends who supported me during this journey. My parents, Sue and Dan Brady, and my sister, Amanda Brady, were always a short call away when I needed some words of encouragement. My grandparents, Bill and Marilyn Helfers, and my aunt and uncle, Ellen and Bill Joens, were always excited to hear about my adventures in graduate school, and my grandpa even helped me collect samples during winter breaks. My close friends, Abbey Shipp, Amanda Myers, Rachael Lewin, Beth Ann Likes, Elizabeth Shultz, and Brittany Bernhagen, as well as my boyfriend, Paul McDonald, helped me celebrate my successes and take refreshing breaks. Last, but certainly not least, I'd like to acknowledge my dog, Kylie, who greets me with an optimistically wagging tail at the beginning and end of every day.

## List of Abbreviations

**ANOVA** - Analysis of variance statistical test

**BLMH** - Bleomycin hydrolase

**CRISPR** - Clustered regularly interspaced short palindromic repeats

**CSS** - Chromosome substitution strain

**eQTL** - Expression quantitative trait locus

**EXT** - Optical density, extinction

**FDR** - False discovery rate

**GFP** - Green fluorescence protein

**GWA** - Genome-wide association

**GWER** - Genome-wide error rate

**HDR** - Homology-driven repair

**HTA** - High-throughput assay

**Indel** - Insertion or deletion

**IQR** - Interquartile range

**LOD** - Log of the odds

**NIL** - Near-isogenic line

**NHEJ** - Non-homologous end joining

**PC** - Principal component

**PCA** - Principal component analysis

**PCR** - Polymerase chain reaction

**QTL** - Quantitative trait locus

**RIAIL** - Recombinant inbred advanced intercross line

**SNP** - Single-nucleotide polymorphism

**SNV** - Single-nucleotide variant

**STEM** - Science, technology, engineering, and math

**SV** - Structural variant

**TOF** - Time of flight

# Glossary

**BIOSORT** - A large particle flow cytometer developed by Union Biometrica

**CB4856** - A wild isolate of *C. elegans* from Hawaii

**CSV** - Comma separated value data format

**DL238** - A wild isolate of *C. elegans* from Hawaii

**FASTA** - A text-based format for representing genomic sequences

$H^2$  - Broad-sense heritability estimate

**JU258** - A wild isolate of *C. elegans* from Madeira

**K medium** - An enriched saline solution for *C. elegans* growth

**M9** - Minimal media for *C. elegans* growth

**N2** - The canonical laboratory strain of *C. elegans* from Bristol, England

**NGMA** - Normal growth media with 1% agar and 0.7% agarose to prevent burrowing

***scb-1*** - Previously *H19N07.3*, a gene underlying the sensitivity to the chemotherapeutic bleomycin

**TSV** - Tab-separated value data format

**VCF** - Variant caller format for representing sequence data

# Dedication

For Susan and Daniel Brady

# Table of Contents

<b>LIST OF TABLES</b> .....	<b>16</b>
<b>LIST OF FIGURES</b> .....	<b>17</b>
<b>CHAPTER 1 - INTRODUCTION</b> .....	<b>20</b>
CANCER AND CHEMOTHERAPEUTICS.....	20
VARIATION IN CHEMOTHERAPEUTIC RESPONSES ACROSS INDIVIDUALS .....	21
CHALLENGES OF BIOMARKER IDENTIFICATION IN HUMANS .....	23
<i>CAENORHABDITIS ELEGANS</i> AS A TRACTABLE MODEL ORGANISM.....	25
GENETICS AND GENOMICS OF <i>C. ELEGANS</i> .....	25
HIGH-THROUGHPUT ASSAYS OF <i>C. ELEGANS</i> .....	27
LINKAGE MAPPING IN <i>C. ELEGANS</i> .....	30
<b>CHAPTER 2 - VARIATION IN <i>SCB-1</i> FUNCTION UNDERLIES BLEOMYCIN RESPONSE</b>	
<b>DIFFERENCES</b> .....	<b>32</b>
PREFACE.....	32
ABSTRACT .....	32
INTRODUCTION .....	33
MATERIALS AND METHODS.....	35
RESULTS .....	47
DISCUSSION.....	86
FUTURE DIRECTIONS.....	89
CONTRIBUTIONS .....	92

	13
SUPPLEMENTAL TABLES .....	93
SUPPLEMENTAL FIGURES .....	99
<b>CHAPTER 3 - SHARED GENOMIC REGIONS UNDERLIE NATURAL VARIATION IN DIVERSE</b>	
<b>TOXIN RESPONSES.....</b>	<b>132</b>
PREFACE .....	132
ABSTRACT .....	132
INTRODUCTION .....	133
MATERIALS AND METHODS.....	135
RESULTS .....	145
DISCUSSION.....	168
FUTURE DIRECTIONS.....	173
CONTRIBUTIONS .....	174
SUPPLEMENTAL TABLES .....	175
SUPPLEMENTAL FIGURES.....	177
<b>CHAPTER 4 - ESCAPE-ROOM INSPIRED LEARNING ENRICHMENT.....</b>	<b>199</b>
PREFACE .....	199
ABSTRACT .....	200
INTRODUCTION .....	200
METHODOLOGY.....	202
WORKFLOW .....	203
IMPACT AND APPLICATIONS.....	220
CONTRIBUTIONS .....	223
SUPPLEMENT A – SETUP INSTRUCTIONS .....	223
SUPPLEMENT B – PROMPTS .....	228

	14
SUPPLEMENT C – WORKFLOW.....	233
SUPPLEMENT D – PRE-ACTIVITY SURVEY .....	241
SUPPLEMENT E – POST-ACTIVITY SURVEY .....	244
<b>CHAPTER 5 - DISCUSSION.....</b>	<b>248</b>
POWER OF LINKAGE MAPPING IN <i>C. ELEGANS</i> .....	248
LOOKING TOWARD FUTURE LINKAGE MAPPING STUDIES .....	250
EVOLUTIONARY IMPLICATIONS OF TOXIN-RESPONSE QTL .....	254
POTENTIAL FOR FUTURE EXPERIMENTS.....	256
<b>CHAPTER 6 - REFERENCES.....</b>	<b>260</b>
<b>APPENDIX A: CO-AUTHORED PUBLICATIONS .....</b>	<b>300</b>
SELECTION ON A SUBUNIT OF THE NURF CHROMATIN REMODELER MODIFIES LIFE HISTORY TRAITS IN A DOMESTICATED STRAIN OF <i>CAENORHABDITIS ELEGANS</i> .....	300
CORRELATIONS OF GENOTYPE WITH CLIMATE PARAMETERS SUGGEST <i>CAENORHABDITIS ELEGANS</i> NICHE ADAPTATIONS.....	302
THE GENETIC BASIS OF NATURAL VARIATION IN A PHORETIC BEHAVIOR .....	304
DISCOVERY OF GENOMIC INTERVALS THAT UNDERLIE NEMATODE RESPONSES TO BENZIMIDAZOLES.....	306
NATURAL VARIATION IN <i>C. ELEGANS</i> ARSENIC TOXICITY IS EXPLAINED BY DIFFERENCES IN BRANCHED CHAIN AMINO ACID METABOLISM.....	308
<b>APPENDIX B: COLLECTION OF WILD ISOLATES ON THE HAWAIIAN ISLANDS .....</b>	<b>310</b>
PREFACE.....	310
MOTIVATION FOR COLLECTIONS ON THE HAWAIIAN ISLANDS .....	310
RESULTS .....	310
CONTRIBUTIONS .....	312

**APPENDIX C: *EASYSORTER* PACKAGE ..... 314**

PREFACE .....314

EXPLANATION OF FUNCTIONALITY .....314

CONTRIBUTIONS .....318

ACKNOWLEDGEMENTS .....319

**APPENDIX D: *LINKAGEMAPPING* PACKAGE ..... 320**

PREFACE .....320

EXPLANATION OF FUNCTIONALITY .....320

CONTRIBUTIONS .....332

ACKNOWLEDGEMENTS .....334

## List of Tables

### MAIN TABLES

<i>Table 3-1 List of doses tested for each toxin.....</i>	<i>146</i>
<i>Table 3-2 Number of principal components that explain at least 90% of the variance in the RIALs for each toxin .....</i>	<i>147</i>
<i>Table 3-3 Toxin responses map to three QTL hotspots.....</i>	<i>154</i>
<i>Table 3-4 List of each toxin that maps to each QTL hotspot; bold denotes a trait tested in the NIL/CSS assays .</i>	<i>155</i>
<i>Table 3-5 All PCs assayed with NILs/CSSs and the traits that underlie each PC.....</i>	<i>156</i>
<i>Table 3-6 List of primary categorizations from NIL assay results.....</i>	<i>161</i>
<i>Table 3-7 Categorizations from combining the NIL and CSS assay results .....</i>	<i>165</i>

### SUPPLEMENTAL TABLES

<i>Table S2-1 Reagents used to generate NILs.....</i>	<i>93</i>
<i>Table S2-2 Reagents used to generate CRISPR/Cas9-mediated genome edits.....</i>	<i>93</i>
<i>Table S3-1 Reagents used to generate NILs and CSSs.....</i>	<i>175</i>

# List of Figures

## MAIN FIGURES

Figure 1-1 High-throughput assay (HTA).....	29
Figure 2-1 Responses to different doses of bleomycin vary across wild isolates .....	48
Figure 2-2 Bleomycin-response HTA traits are correlated .....	50
Figure 2-3 Bleomycin-response variation maps to four QTL .....	51
Figure 2-4 Linkage mapping QTL explain most of the broad-sense heritability for bleomycin responses .....	52
Figure 2-5 Broad NILs recapitulate the expected QTL effect on bleomycin responses.....	54
Figure 2-6 A two-factor genome scan for bleomycin responses does not detect epistatic loci .....	55
Figure 2-7 NILs localize the bleomycin-response QTL to a small region on chromosome V.....	57
Figure 2-8 Strains with <i>jmjd-5</i> deletion alleles are resistant to bleomycin .....	60
Figure 2-9 Strains with <i>jmjd-5</i> deletion alleles are sensitive in the control condition.....	61
Figure 2-10 The non-synonymous variant of <i>jmjd-5</i> does not affect bleomycin responses .....	63
Figure 2-11 Phenotypes in the modified HTA match previous bleomycin responses.....	65
Figure 2-12 Natural variation in <i>jmjd-5</i> does not underlie bleomycin-response differences .....	66
Figure 2-13 Strains with homozygous <i>jmjd-5</i> deletions are sensitive to the control condition, whereas hemizygous strains are not.....	67
Figure 2-14 <i>H19N07.3</i> expression differences map to chromosome V and correlate with bleomycin responses....	69
Figure 2-15 Linkage mapping of bleomycin-response differences among set 1 RIALs maps to the same location as set 2 RIALs .....	70
Figure 2-16 Strains with <i>H19N07.3</i> deletion alleles are sensitive to bleomycin.....	71
Figure 2-17 Strains with <i>H19N07.3</i> deletion alleles are not sensitive in the control condition .....	72
Figure 2-18 RNA-seq does not show <i>scb-1</i> expression differences in the parental strains.....	73
Figure 2-19 A two-factor genome scan for <i>scb-1</i> expression variation detects two significant genetic interactions .....	74

Figure 2-20 Natural variation in <i>scb-1</i> underlies bleomycin-response differences.....	75
Figure 2-21 Strains with <i>scb-1</i> deletion alleles are not sensitive in the control condition .....	76
Figure 2-22 A genome-wide association mapping for bleomycin-response variation detects six QTL .....	77
Figure 2-23 Rare and common variants are present near the <i>scb-1</i> gene .....	78
Figure 2-24 SCB-1 is conserved across nematodes .....	79
Figure 2-25 Near-isogenic lines suggest that three loci interact to cause bleomycin hypersensitivity.....	81
Figure 2-26 Near-isogenic lines suggest that three loci interact to cause bleomycin hypersensitivity.....	82
Figure 2-27 Near-isogenic lines suggest that three loci interact to cause bleomycin hypersensitivity.....	83
Figure 3-1 Toxin-response variation maps to many regions of the genome .....	148
Figure 3-2 Both additive and interacting loci underlie toxin response variation.....	149
Figure 3-3 The proportion of additive and interactive loci that underlie toxin responses differs among drugs ..	151
Figure 3-4 Three regions of the genome are enriched for toxin-response QTL.....	153
Figure 3-5 NIL and CSS results can be placed into distinct categories .....	159
Figure 3-6 NIL and CSS results represent several categories.....	160
Figure 3-7 HTA traits that comprise each principal component are placed into similar categories.....	163
Figure 3-8 NIL and CSS assay results can be compared to localize potential regions of epistasis.....	166
Figure 4-1 Mutagenesis screen answer .....	204
Figure 4-2 Complementation puzzle.....	205
Figure 4-3 Epistasis circuit.....	206
Figure 4-4 Expression plasmid.....	207
Figure 4-5 Gene expression poster .....	208
Figure 4-6 Mutant firefly box .....	209
Figure 4-7 Abdominal cell lineage .....	210
Figure 4-8 Time of gene action.....	212
Figure 4-9 Modes of inheritance pedigrees.....	215
Figure 4-10 Mysterious poster board .....	216
Figure 4-11 Triplet pedigrees.....	217

<i>Figure 4-12 Linkage pedigree</i> .....	218
<i>Figure 4-13 GWAS flight patterns</i> .....	219

### **SUPPLEMENTAL FIGURES**

<i>Figure S2-1 All linkage mapping results for bleomycin-response variation</i> .....	112
<i>Figure S2-2 Bleomycin responses for all wild isolates, colored by the genotype at each of the rare alleles around scb-1</i> .....	121
<i>Figure S2- 3 Chromosome V mappings after bleomycin QTL regression</i> .....	131
<i>Figure S3-1 Results of a dose-response assay for all toxins are shown</i> .....	182
<i>Figure S3-2 Traits measured by the HTA are correlated</i> .....	198
<i>Appendix D Figure 1 Example lodplot output</i> .....	324
<i>Appendix D Figure 2 Example maxlodplot output</i> .....	325
<i>Appendix D Figure 3 Example pxgplot output</i> .....	326
<i>Appendix D Figure 4 Example pxgplot_par output</i> .....	327
<i>Appendix D Figure 5 Example effplot output</i> .....	328
<i>Appendix D Figure 6 Example findN2fosmids output</i> .....	329
<i>Appendix D Figure 7 Example findRIAILSforNILs output</i> .....	331

# Chapter 1 - Introduction

## Cancer and chemotherapeutics

Cancer is the second leading cause of death worldwide, and approximately one in six deaths can be attributed to cancer [1]. Because this disease is such a prominent issue in the modern world, an enormous amount of scientific research focuses on understanding the biology of cancer and developing effective treatment strategies [2]. The field of cancer research has identified effective preventive and screening strategies as well as advances in cancer treatment regimens, and the number of cancer-related deaths is declining [3,4]. Many of the advances in treatment strategies are attributed to an immense arsenal of chemotherapeutic drugs at the fingertips of oncologists.

By definition, cancer cells abnormally divide without control, which can allow them to invade nearby tissues and spread to other parts of the patient's body [5]. Chemotherapeutic drugs aim to stop the growth of these cancerous cells, either by killing them or by preventing them from dividing [5,6]. Currently, more than 100 different FDA-approved chemotherapeutic drugs exist [7], and the mechanisms by which they affect cancerous cells can be grouped into several classes, some of which are detailed below.

The largest class of chemotherapeutic drugs comprises alkylating agents [6]. These drugs transfer alkyl groups to a broad range of molecules, including DNA bases, thereby disrupting molecular functions and leading to cytotoxicity [8]. Another common class of chemotherapeutics is topoisomerase poisons, which prevent the action of topoisomerase I and II. Topoisomerases

are necessary for DNA uncoiling during replication and transcription, and the disruption of their function by topoisomerase poisons leads to cell death [9]. Microtubule-targeting agents, such as vinca alkaloids [10] and taxanes [11,12] bind to and alter the dynamic polymerization of microtubules, thereby leading to apoptosis [13]. Finally, antitumor antibiotics cause breaks in DNA either directly or indirectly via superoxide [5].

Although each class of chemotherapeutic drug causes apoptosis in tumor cells, off-target effects can cause cell death in unintended tissues. In fact, the cytotoxic effects of mustard gas in WWII prompted its use as an early cancer treatment, eventually establishing alkylating agents as a promising class of chemotherapeutic drugs [14–16]. Because of the toxic nature of these tumor-cell killing drugs, a range of side effects can be expected from chemotherapy. Typically, the off-target effects of cancer treatment affect fast growing cells and lead to side effects such as chemotherapy-induced alopecia [17], nausea and vomiting [18,19], impaired immune system [20], and cardiovascular complications [21,22]. The presence and severity of these side effects can vary across individual patients, and oncologists aim to maximize the on-target while minimizing the off-target effects of chemotherapeutic regimens.

## Variation in chemotherapeutic responses across individuals

The balance between efficacy and toxicity of chemotherapy, or the therapeutic index, varies from patient to patient. Chemotherapeutic efficacy and treatment side effects can be influenced by environmental variables, such as diet [23], exercise [24], and smoking habits [25]. Some environmental factors can be used to predict patient responses to a broad range of chemotherapeutic regimens whereas others can be specifically predictive of a particular side

effect. For example, the general role of gut microbiota on efficacy and toxicity of many drugs, including chemotherapeutics, has been highlighted in several studies [26–28]. Social support is another environmental variable that broadly predicts patient quality of life after chemotherapy [29–31]. On the other hand, patient age, disease progression, and low BMI are specific predictors of low white blood cell counts after chemotherapy [32,33]. In addition to environmental variables that impact treatment outcomes, genetic variation among individuals can also account for differences in chemotherapeutic drug responses. Rapid advances in sequencing technology allow for genetic variants to be accurately measured, and particular genetic variants can inform patient treatment decisions.

The heritability of a trait is the amount of phenotypic variation across a population that can be explained by genetic differences among those individuals. Researchers can estimate the heritability of chemotherapeutic drug responses by treating cell lines [34–39], model organisms [40–44], or familial patient samples [45–48] with a drug of interest in a controlled environment. Although some of the observed variation in patient responses to chemotherapeutic drugs is often heritable, the challenge lies in identifying which genes underlie variation in these treatment outcomes.

The field of pharmacogenomics aims to identify genetic variants, or biomarkers, that underlie drug-response variation across patients. Variants that affect drug metabolism [49–55], drug transport [56–58], drug targets [59–62], or general immune responses [63–65] can lead to differences in patient drug responses. Approximately ten percent of all FDA approved drugs contain a biomarker on their label [66], and labels for 85 FDA-approved chemotherapeutic drugs contain the name of a genetic biomarker that can be used to predict treatment outcomes [67].

Some of these biomarkers predict adverse responses to chemotherapeutic agents. For example, toxicity of 5-fluorouracil (5FU) is associated with variants that impair the dihydropyrimidine dehydrogenase (DPD) enzyme [68,69]. Other biomarkers inform potential efficacy of chemotherapeutics. Patients with tumors that have mutated EGFR respond well to tyrosine kinase inhibitors, such as erlotinib [70], gefitinib [71], and afatinib [72]. Indeed, biomarkers are powerful tools used to inform treatment decisions, but many genetic variants that impact chemotherapeutic drug responses have yet to be identified [73].

## Challenges of biomarker identification in humans

Although human studies have identified many biomarkers that inform cancer treatment options, much of the heritability in drug responses remains to be explained. Many of the current biomarkers for drug-response variation were identified through genome-wide association (GWA) studies. Although the power of pharmacogenomic GWA studies continues to improve with advancing genome sequencing technology [74–76] and expanding panels of individuals [77–79], many limitations of human GWA studies remain [73,80–83].

As mentioned above, both genetic and environmental variables can affect drug responses. To maximize the heritability of drug-response variation, environmental variables must be tightly controlled among individuals. As one can imagine, removing variation in diet, exercise, tobacco use, microbiome, and all other environmental factors that might influence drug responses in human patients is nearly impossible. In the case of chemotherapeutic drug responses, tumor stage and heterogeneity varies across patients, even across patients that have the same cancer type, and this uncontrollable variation affects the efficacy of treatment [84]. Complicating matters

further, each cancer patient usually receives a combination of chemotherapeutic interventions [85–87]. Finding a cohort of patients who received identical treatments can be difficult, and drug-drug interactions can lead to additional variation in treatment responses across individuals in a GWA study.

Aside from the difficulty in minimizing environmental differences across individuals, human pharmacogenomic GWA studies are additionally limited by the complexity of drug-response traits. The precision of phenotypic measurements highly impacts the ability of a GWA study to identify a correlated genetic variant [88,89]. Tumor measurement strategies are subject to inaccuracies [90], and self-reported side effect information can be biased [91–94]. These imprecisions in phenotypic measurements make the identification of a biomarker more challenging.

Another obstacle of human GWA studies is the type of genetic variants that are typically detected. Many of the loci identified by human GWA studies are common variants with large effects, whereas rare and small-effect variants are less likely to be detected [95,96]. Larger, less geographically biased panels of individuals are necessary to identify rare variants, but creating such panels is a costly endeavor [81,83,95,97]. Finally, many genomic variants that underlie complex trait differences affect gene expression rather than protein changes [98–100]. Genomic variants that affect gene expression must be more accurately called to maximize the success of human GWA studies [101–103]. Alternatively, research with model organisms offers controllable environments, precise phenotypic measurements, and broad, well characterized genetic diversity. These model systems can be used as a proxy to understand how genetic variants impact drug-response differences across individuals.

## *Caenorhabditis elegans* as a tractable model organism

Eight years before his first publication on the species, Dr. Sydney Brenner sought a simple experimental organism that was suitable for genetic studies, and he settled on the roundworm *Caenorhabditis elegans* [104]. His original mutant screen established *C. elegans* as a “favorable organism for genetic analysis” and sparked an era of *C. elegans* research [104].

*C. elegans* is a microscopic, non-infectious nematode that can be found throughout the world [105]. Populations of *C. elegans* can be grown in the laboratory on bacteria-seeded agar petri dishes, and its transparent body can be easily viewed with a dissecting microscope [106]. These animals are primarily hermaphroditic, and their ability to self-propagate allows for easy maintenance of clonal populations, which can be cryopreserved for long-term storage with limited mutation accumulation [104,106]. Under standard laboratory conditions, the life cycle of *C. elegans* is predictable, with embryos hatching and developing through the first, second, third, and fourth larval stage before becoming sexually mature adults [104]. This whole life cycle lasts three to four days [104]. Thanks to a remarkably dedicated team of researchers, the entire cell lineage of *C. elegans*, from embryo to adult, is characterized and the anatomy of the organism is deeply understood [107]. These aspects of *C. elegans* make it an excellent model for many types of biological studies.

## Genetics and genomics of *C. elegans*

Although natural *C. elegans* strains exist primarily as hermaphrodites, males can be created using heat shock [106]. These males can be crossed to hermaphrodites to transfer genetic material.

Although hermaphrodites produce their own sperm, they preferentially use male sperm to fertilize their eggs, thereby increasing the efficiency of genetic crosses [108]. The ability to cross strains allows for the creation of powerful genetic tools, such as panels of recombinant strains and near-isogenic lines [109–112].

Genetic crosses in *C. elegans* are complemented by the facile use of gene editing in the species. Bacterial clustered regularly interspaced short palindromic repeat (CRISPR) proteins, such as Cas9, can be leveraged to precisely modify regions of the *C. elegans* genome [113,114]. The Cas9 protein can be directed to a particular genomic region by introducing a CRISPR RNA (crRNA) that contains the sequence of interest [115]. Once targeted to a genomic location, Cas9 cleaves the DNA and introduces a double-stranded break [114]. Broken DNA can be repaired through error-prone non-homologous end joining (NHEJ), which often leads to insertions and deletions in the repaired region and potential loss-of-function alleles [116–120]. When a more precise genome modification is desired, one can supply a repair construct with a genomic edit of interest, and homology-driven repair (HDR) can use this construct as a template to fix the broken DNA [119,121–125]. Cas9, crRNAs, and repair templates are easily injected into the gonad of adult *C. elegans* hermaphrodites, and a portion of their progeny will carry the mutation of interest [117,125].

*C. elegans* was the first multicellular animal to have its genome fully sequenced [126]. Its 100 MB genome contains six chromosomes (five autosomes and one sex chromosome) and 20,191 protein-coding genes (WormBase release WS269, December 2018), which is similar to the approximately 20,500 estimated human genes [127,128]. In fact, about 38% of the protein-coding genes in *C. elegans* have a human ortholog [129]. Many of the core components of the

RTK/Ras/MAPK, Notch, TGF- $\beta$ , and insulin pathways are conserved between *C. elegans* and humans [129]. These ancient pathways are commonly mutated in cancers, so a deeper understanding of cancer in humans can be facilitated by understanding the genetics of these pathways in *C. elegans* [130].

Notably, genomic variation across wild *C. elegans* strains has been well characterized, and whole-genome sequence data are available for 330 wild strains at [elegansvariation.org](http://elegansvariation.org) (CeNDR version 1.3.1) [131–133]. The incredible genetic diversity characterized across the worldwide population of *C. elegans* makes it an excellent model organism for understanding how natural genetic variants underlie phenotypic differences, such as variation in drug responses [40,133–135]. Effects of natural genetic variants can be directly assessed by generating precise allele replacements between different strains of the species [40,134,135].

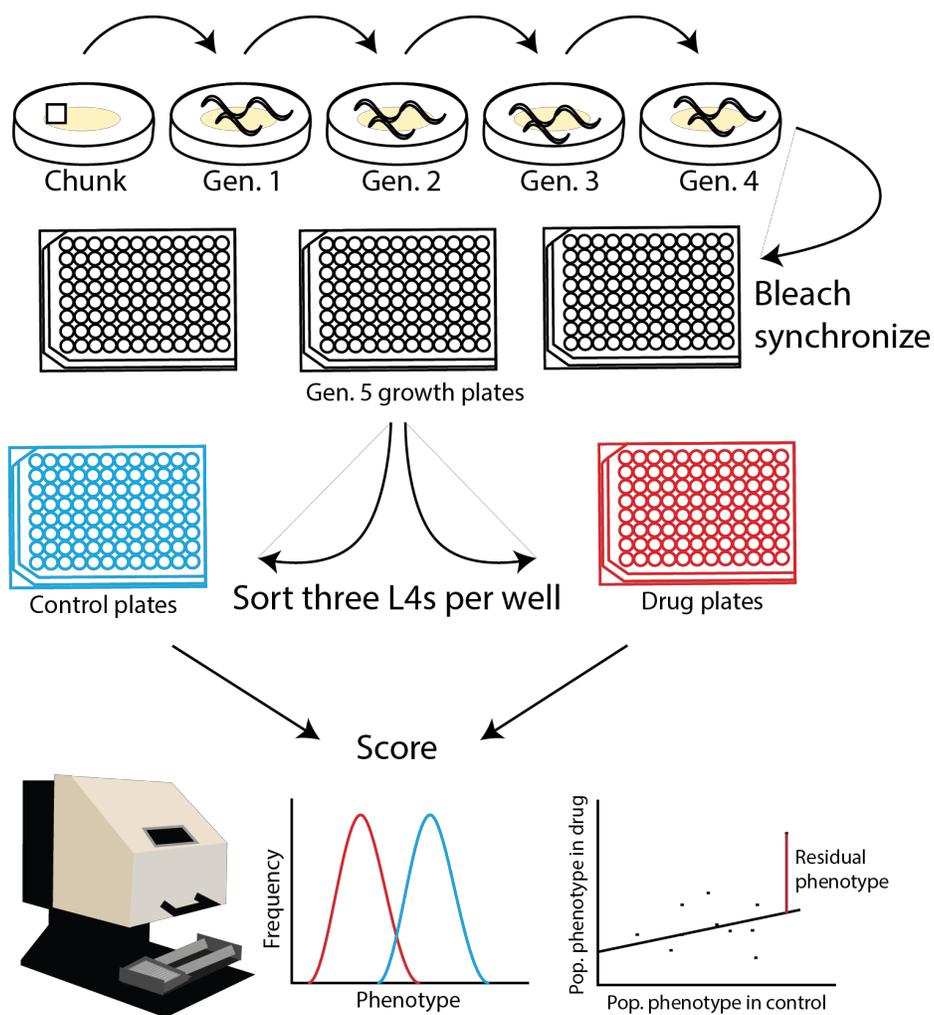
## High-throughput assays of *C. elegans*

Another advantage of research in *C. elegans* is the ability to measure phenotypes rapidly and with high precision while controlling for environmental variables. Several of these high-throughput assays have been developed, including automated microscopy, microplate readers, and flow cytometers [109,136–141]. Researchers can use these assays to measure drug responses of many *C. elegans* animals in a short amount of time.

In particular, the Andersen Lab has used the COPAS BIOSORT (Union Biometrica) to measure fecundity and body size for many strains of *C. elegans* in response to a multitude of drugs [40,41,109,135,142,143]. During this high-throughput assay (HTA), *C. elegans* strains of interest

are synchronized at the L1 larval stage and then grown for 48 hours under highly controlled conditions. When these animals reach the L4 larval stage, they are sorted into 96-well microtiter plates containing either control or drug conditions using the BIOSORT, with three animals in each well. These three animals grow and reproduce in the presence of either the control or drug condition for 96 hours. Finally, the BIOSORT is used to measure the length and optical density of each animal in each well. Animal size and fecundity measurements reflect how sensitive a given strain is to a particular drug. The BIOSORT can score all animals in each well of a 96-well plate in approximately 23 minutes, so adding biological replicates increases the precision of drug-response measurements without adding copious amounts of experiment time. A diagram of the high-throughput fitness assay is shown in **Figure 1-1**.

Each HTA generates an enormous amount of animal size and fecundity measurements. To efficiently and reproducibly analyze HTA results, members of the Andersen Lab have developed the *COPASutils* and *easysorter* pipelines [144–146]. These analysis tools read in raw data from the BIOSORT, remove data from contaminated wells, summarize measurements within each well, prune biological impossibilities and outliers, and control for day-to-day assay variation as well as phenotypic variation present in the control condition. After passing HTA data through these pipelines, drug responses can be compared easily across assayed strains.



**Figure 1-1 High-throughput assay (HTA)**

A diagram shows the high-throughput fitness assay protocol. From top to bottom, left to right: Each strain is chunked onto a fresh plate, and animals of the L4 stage are transferred for four generations. Gravid adults from the fourth generation are bleached to synchronize growth of all strains. Embryos are aliquoted to 96-well growth plates containing K medium and 5 mg/mL of bacterial lysate. After 48 hours, three L4 larvae are sorted from each well of the growth plate to the corresponding well of either a control plate (which contains K medium, 10 mg/mL bacterial lysate, and 1% distilled water) or a drug plate (which contains K medium, 10 mg/mL bacterial lysate, 1% distilled water, and bleomycin). Animals are grown in the control and drug plates for four days at 20°C with shaking. Then, animals are scored using the COPAS BIOSORT. For each well, all animals are measured for length (TOF) and optical density (EXT). The normal distributions represent hypothetical distributions of all animals in a given well of a control (blue) plate or a drug (red) plate. Summary statistics of these distributions are measured to obtain a population growth estimate for all animals within a well. For each trait measured, the average control phenotype for a given strain is plotted against the drug phenotype for each replicate of that strain. A linear model is fit to those data, and the residual phenotype is calculated and used for further analysis.

Similar to human variation in drug responses, differences in drug responses across strains of *C. elegans* is apparent [40,41,109,135,142,143,147]. Heritability calculations show that much of these drug-response differences can be attributed to genetic variants among strains [40,41,142]. Given the challenges of identifying genetic variants that underlie these drug-response differences in humans, *C. elegans* can be used as a proxy to locate genetic factors that impact animal growth and fecundity during drug exposure. In fact, variants that underlie drug-response differences in *C. elegans* have been identified by GWA studies across many wild strains and by linkage mapping of recombinant lines derived from two divergent strains [40,135,142].

## Linkage mapping in *C. elegans*

Two highly diverged strains of *C. elegans* can be studied to understand how genetic variation between them can cause phenotypic differences. N2 is the canonical laboratory strain; it is derived from an isolate found in Bristol, England in 1951 and is one of the earliest strains in the field [104,148]. Across the globe, in 1972, the CB4856 strain was isolated from a pineapple field on the Hawaiian island of Maui [149]. For decades, these two strains have been extensively studied and are genetically divergent [131,150–154]. Additionally, these two strains display variation in many phenotypes, including social behavior [155], temperature responsiveness [156], RNAi susceptibility [157], gene expression profiles [158], and notably, drug responses [40,41,141–143,147].

To isolate regions of the genome that might underlie drug-response variation between the N2 and CB4856 strains, panels of recombinant inbred advanced intercross lines (RIALs) have been generated by crossing the two genotypes for many generations [109,110]. Each strain in the RIAL

panel has a unique set of genetic markers derived from either the N2 or the CB4856 parental strain. Linkage mapping is used to correlate genetic markers with phenotypic variation across the RIALs, and mappings are facilitated by the *linkagemapping* pipeline developed by members of the Andersen Lab [41,159]. Recently, the set of genetic markers used for a linkage mapping has increased from 1,454 to 13,003, which allows for the identification of more precise regions of the genome that correlate with phenotypic variation [160].

Regions of the genome that are strongly correlated with phenotypic variation are called quantitative trait loci (QTL). Near-isogenic lines that isolate pieces of the QTL in a constant genetic background can be used to confirm the QTL effect and narrow the QTL to a smaller region of the genome. Ultimately, researchers are able to use CRISPR/Cas9-mediated genome editing to modify precise genomic loci and test the effect of those loci on the phenotype of interest. Several linkage-mapping studies in *C. elegans*, sometimes in combination with GWA studies, have identified a causal locus that underlies a quantitative trait [40,141,142,156,161–168].

*C. elegans* is an excellent model system that can be leveraged to understand how genetic variation causes phenotypic differences among individuals. In the case of drug-response variation in the clinic, it is sometimes difficult to attribute these life-threatening differences to particular genetic markers. Especially for drugs that target conserved signaling pathways, *C. elegans* can be a powerful tool for understanding how genetic differences underlie drug-response variation. Therefore, biomarkers that predict drug responses could be identified using *C. elegans*, and physicians might be able to use this information to screen patients before prescribing particular medicines, ultimately improving treatment outcomes.

# Chapter 2 - Variation in *scb-1* function underlies bleomycin response differences

## Preface

In the spring quarter of 2015, I chose to complete my final first-year rotation in the Andersen Lab. I had a strong interest in joining the lab for my thesis work, so I worked with Erik to select a project that could bridge into my time as a Ph.D. candidate. Early members of the Andersen Lab had mapped responses to various drugs using genome-wide association and linkage mapping. The massive mapping dataset offered seemingly endless quantitative trait loci that I could choose to follow. Given its impressively significant correlation with a region on chromosome V, I chose bleomycin response as my trait of interest for my thesis work. This “low hanging fruit” seemed to be a very straightforward project, but sometimes science is more complex than it seems. The following chapter highlights the exciting project that took the majority of my time in the Andersen Lab, identifying the genetic variant that causes bleomycin-response differences across *C. elegans* strains, and this work was published in *Genetics* in 2019.

## Abstract

Bleomycin is a powerful chemotherapeutic drug used to treat a variety of cancers. However, individual patients vary in their responses to bleomycin. The identification of genetic differences that underlie this response variation could improve treatment outcomes by tailoring bleomycin dosages to each patient. We used the model organism *Caenorhabditis elegans* to identify genetic

determinants of bleomycin-response differences by performing linkage mapping on recombinants derived from a cross between the laboratory strain (N2) and a wild strain (CB4856). This approach identified a small genomic region on chromosome V that underlies bleomycin-response variation. Using near-isogenic lines and strains with CRISPR/Cas9-mediated deletions and allele replacements, we discovered that a novel nematode-specific gene (*scb-1*) is required for bleomycin resistance. Although the mechanism by which this gene causes variation in bleomycin responses is unknown, we suggest that a rare variant present in the CB4856 strain might cause differences in the potential stress-response function of *scb-1* between the N2 and CB4856 strains, thereby leading to differences in bleomycin resistance.

## Introduction

Cancer is the second leading cause of death worldwide [1], which has led to extensive research for treatments, including the identification of over 100 effective chemotherapeutic drugs [7]. One of these drugs is bleomycin, an anti-tumor antibiotic that interacts with oxygen and transition metals to cause double-stranded DNA breaks [169]. Although the cytotoxicity of bleomycin can reliably induce cell death in tumor cells, off-target effects can lead to a range of harmful consequences from mild gastrointestinal irritation to severe bleomycin-induced pulmonary fibrosis [170]. The tradeoff between efficacy and toxicity varies across individuals, and understanding the genetic variants that affect bleomycin response might yield opportunities to broaden the therapeutic range [171,172].

Bleomycin sensitivity has been shown to be heritable, suggesting that genetic markers can be used to predict bleomycin responses [173]. Many studies have attempted to identify the genetic

variant(s) that underlie bleomycin-response differences across cancer patients, and some have identified potential connections between the metabolic enzyme bleomycin hydrolase (BLMH) and patient outcomes. However, none of these studies established a causal connection between genetic differences in BLMH and variation in bleomycin responses [171,174–177]. The inability to identify a genetic variant that causes differences in bleomycin responses in humans might be attributed to limited sample sizes [178], confounding environmental factors [179,180], variation in drug regimens across patients [181], or tumor complexity and progression [182,183]. However, the DNA-damage pathways that might be implicated in bleomycin responses are evolutionarily conserved across eukaryotes [184]. Therefore, studying bleomycin responses in a model organism with natural genetic variation can offer insights into how bleomycin response differs across individuals and can potentially be applied to the clinic [185].

*Caenorhabditis elegans* is a soil-associated microscopic roundworm that is an excellent model for basic cellular and organismal processes [105]. Not only does *C. elegans* have a well annotated reference genome ([126,186,187], [www.wormbase.org](http://www.wormbase.org) WS268), but this species also has broad genomic diversity across global populations [133]. Notably, the N2 strain and the CB4856 strain are well characterized and genetically divergent with approximately one single nucleotide variant per 850 bp [150–152]. These two strains were used to generate a panel of recombinant inbred advanced intercross lines (RIAILs) [109,110], which has been used to correlate genetic variants with differences in quantitative traits [40,134,141,156,161–168,188].

Here, we used a high-throughput fitness assay to measure bleomycin responses across a panel of 249 RIAILs [109] and then performed linkage mapping to identify quantitative trait loci (QTL) that underlie bleomycin-response variation. We used near-isogenic lines (NILs) to validate the

largest effect QTL on chromosome V. Our results from the NIL assays suggested that epistatic loci underlie bleomycin responses, but a two-factor genome scan was unable to detect significant genetic interactions. Next, we created and tested CRISPR-Cas9 mediated deletion alleles to investigate all six candidate genes in the QTL region. We identified a nematode-specific gene, *H19N07.3*, that underlies this QTL. Although this gene does not contain a protein-coding variant between the N2 and CB4856 strains, its gene expression varies across the RIAIL panel. Interestingly, a genome-wide association (GWA) approach identifies different QTL than the linkage mapping approach, suggesting that both common and rare variants underlie bleomycin response variation. Given the genetic complexity underlying the bleomycin response phenotype, this study highlights the power of the *C. elegans* model system to identify elusive causal genes.

## Materials and Methods

### Strains

Animals were grown at 20°C on 6 cm plates of modified nematode growth medium (NGMA), which contained 1% agar and 0.7% agarose, spotted with OP50 bacteria [147]. The two parental strains used in this study were N2 and CB4856. N2 is the canonical laboratory strain of *C. elegans* that has been extensively studied [104]. CB4856 is a well studied Hawaiian wild isolate that is genetically divergent from N2 and has a characterized genome [150–152]. The N2 and CB4856 strains were crossed for several generations to create a panel of recombinant inbred advanced intercross lines (RIAILs) that contain regions of the genome derived from each parental strain. These RIAILs were constructed previously [109,110] and have well characterized genotypes and allele frequencies, and we used this panel of RIAILs in our study to identify regions of the genome

correlated with drug response. The construction of near-isogenic lines (NILs), as well as CRISPR-Cas9 mediated deletion and allele-replacement strains, is detailed below. All strains and reagents used in strain constructions are listed in **Table S2-1**.

## High-throughput fitness assays

We used the high-throughput assay (HTA) described above. Populations of each strain were passaged on 6 cm plates for four generations to amplify animal numbers and reduce the effects of starvation [141]. Gravid adults were bleached for stage synchronization, and approximately 25 embryos from each strain were aliquoted into 96-well plates at a final volume of 50  $\mu$ L of K medium [189]. The following day, arrested L1 larvae were fed 5 mg/mL HB101 bacterial lysate in K medium (Pennsylvania State University Shared Fermentation Facility, State College, PA; [190]) and were grown for 48 hours at 20° with constant shaking. A large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston, MA) was used to sort three L4 larvae into each well of a 96-well plate that contained 50  $\mu$ L K medium plus HB101 lysate at 10 mg/mL, 50  $\mu$ M kanamycin, and either 1% distilled water (control) or 1% distilled water and bleomycin (drug). The sorted L4 larvae were grown and propagated for 96 hours at 20° with constant shaking. The population of parents and progeny were treated with sodium azide (50 mM in M9) and quantified by the BIOSORT for several fitness parameters. Because bleomycin exposure can affect animal proliferation (brood size), animal growth (length), and animal development (optical density), the fitness parameters we measured with the BIOSORT included brood size (n), animal length (time of flight, TOF), and optical density (extinction time, EXT).

## Bleomycin-response trait measurements and processing

Phenotypic measurements collected by the BIOSORT were processed using the R package *easysorter* [144]. Using this package, *read\_data* imported measurements from the BIOSORT and *remove\_contamination* was used to remove contaminated wells from analysis. The *sumplate* function then calculated normalized measurements (norm.n -- brood size normalized to number of animals sorted, norm.EXT -- EXT normalized by TOF measurements) and summary statistics (mean, median, 10<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup> percentile, interquartile range, covariance, and variance) of each trait for the population of animals. A total of 26 HTA traits were measured. When strains were phenotyped across multiple days, the *regress(assay=TRUE)* function was used to fit a linear model with the formula (*phenotype ~ assay*) to account for variation among assay days. Next, the *prune\_outliers()* function removed phenotypic values that were beyond two standard deviations of the mean (unless at least 5% of the strains were outside this range in the case of RIAIL assays). Finally, bleomycin-specific effects were calculated using the *regress(assay=FALSE)* function from *easysorter*, which fits a linear model with the formula (*phenotype ~ control phenotype*). The residual phenotypic values account for differences among strains that were present in control conditions.

## Bleomycin dose response

A dose-response high-throughput assay was performed using quadruplicates of four genetically divergent strains (N2, CB4856, JU258, and DL238) tested in various concentrations of bleomycin. The broad-sense heritability at each concentration was calculated using the *lmer* function within the *lme4* R package with the phenotype as the dependent variable and strain as a random effect (*phenotype ~ 1 + (1|strain)*). The concentration of bleomycin that provided the highest mean broad-

sense heritability across the 26 HTA traits was selected for linkage mapping experiments (50  $\mu$ M, mean  $H^2 = 0.58$ ). Bleomycin sulfate was purchased from Biotang Inc via Fisher Scientific (Catalog No. 50-148-546).

## Whole-genome sequence library prep and analysis

Whole-genome sequencing was performed on recombinant advanced intercross lines (RIAILs) and near-isogenic lines (NILs) using low-coverage sequencing. DNA was isolated from 100-300  $\mu$ L of packed worms using Omega BioTek's EZ 96 Tissue DNA Kit (catalog no. D1196-01). All samples were diluted to 0.2 ng/ $\mu$ L and incubated with diluted Illumina transposome (catalog no. FC-121-1031). Tagmented samples were amplified with barcoded primers. Unique libraries (192) were pooled by adding 8  $\mu$ L of each library. The pooled material was size-selected by separating the material on a 2% agarose gel and excising the fragments ranging from 400-600 bp. The sample was purified using Qiagen's Gel Extraction Kit (catalog no. 28706) and eluted in 30  $\mu$ L of buffer EB. The concentration of the purified sample was determined using the Qubit dsDNA HS Assay Kit (catalog no. Q32851). RIAILs and NILs were sequenced at low coverage (mean = 2.13x) using the Illumina HiSeq 2500 platform with a paired-end 100 bp reaction lane. RIAIL and NIL genotypes were imputed using VCF-kit [191]. To determine genotypes, a list of filtered, high-quality sites ( $n = 196,565$ ) where parental strains possess different genotypes was extracted from a previously established variant dataset [132]. All RIAIL genotypes can be accessed by downloading the *linkagemapping* R package at [github.com/AndersenLab/linkagemapping](https://github.com/AndersenLab/linkagemapping). NIL genotypes are described below.

## Linkage mapping

Linkage mapping was performed on each of the 26 HTA traits described above using the R package *linkagemapping* ([www.github.com/AndersenLab/linkagemapping](http://www.github.com/AndersenLab/linkagemapping)). The function *load\_cross\_obj*("N2xCB4856cross\_full") was executed to load a cross object containing 13,003 SNPs that describe locations of genetic recombination in the RIAIL panel (out of the 195,565 high-quality SNPs at which genotypes were called). The genotypic data and residual phenotypic data (after control-condition regression) were merged into a cross object using the *merge\_pheno* function with *set* = 2 to include strains with a reduced allele-frequency skew on chromosome I. The *fsearch* function was used to calculate logarithm of odds (LOD) scores for each marker and each trait as  $-n(\ln(1 - R^2)/2\ln(10))$ , where  $R$  is the Pearson correlation coefficient between RIAIL genotypes at the marker and trait values [40,41,162,192]. A 5% genome-wide error rate was calculated by permuting phenotype and genotype data 1,000 times. Mappings were repeated iteratively, each time using the marker with the highest LOD score as a cofactor until no significant QTL were detected. The *annotate\_lods* function was used to calculate the percent of variance in RIAIL phenotypes explained by each QTL and a 95% confidence interval around each peak marker, defined by any marker on the same chromosome within a 1.5-LOD drop from the peak marker.

## Generation of near-isogenic lines (NILs)

A near-isogenic line (NIL) is genetically identical to another strain aside from a small genomic region that is derived from an alternate strain. NILs are used to test the effect of modifications to particular genomic regions in a consistent genetic background. To make each NIL, males and hermaphrodites of the desired RIAIL (QX131 for ECA230 and QX450 for ECA232) and parental

background (CB4856 for ECA230 and N2 for ECA232) were crossed in bulk, then male progeny were crossed to the parental strain in bulk for another generation. For each NIL, eight single-parent crosses were performed followed by six generations of propagating isogenic lines to ensure homozygosity of the genome. For each cross, PCR was used to select non-recombinant progeny genotypes within the introgressed region by amplifying insertion-deletion (indel) variants between the N2 and CB4856 genotypes on the left and right side of the introgressed region. NIL strains ECA411 and ECA528 were generated by crossing ECA230 and CB4856 in bulk. Heterozygous F1 hermaphrodites were crossed to CB4856 males and the F2 L4 hermaphrodites were placed into individual wells of a 96-well plate with K medium and 5 mg/mL bacterial lysate and grown to starvation. After starvation, each well of the 96-well plate was genotyped to identify recombinants in the desired region. Recombinant strains were plated onto 6 cm plates and individual hermaphrodites were propagated for several generations to ensure homozygosity across the genome. NIL strains were whole-genome sequenced as described above to confirm their genotypes. Reagents used to generate all NIL strains and a summary of each introgressed region are detailed in the **Table S2-1**.

## Two-factor genome scan

We performed a two-factor genome scan to identify potentially epistatic loci that might contribute to traits of interest (either bleomycin responses or gene-expression levels). We used the *scantwo* function in the R *qtl* package to perform this analysis. Each pairwise combination of loci were tested for correlations with trait variation in the RIALs. The summary of each scantwo output includes the maximum interactive LOD score for each pair of chromosomes. To determine a threshold for significant interactions, we performed 1000 permutations of the scantwo analysis and selected the 95<sup>th</sup> percentile from these permutations. For the bleomycin-response variation

scantwo, the significant interaction threshold was a LOD score of 4.08. The significant interaction threshold for the *scb-1* expression variation scantwo was 4.05.

## Identification of genes with non-synonymous variants

The *get\_db* function within the *cegwas* R package was used to query WormBase build WS245 for genes in the QTL confidence interval (V:11,042,745-11,189,364). Our initial linkage mappings used the 1,454 marker set [110] and had a QTL confidence interval larger than the interval found using the whole-genome marker set (described above). Because this confidence interval was larger and more conservative, we kept it for subsequent testing of candidate genes. This expanded interval contained an additional 20 kb on either side of the whole-genome marker set confidence interval. The *snpeff* function within the *cegwas* R package was used to identify variants within the region of interest. We identified variants predicted to have MODERATE (coding variant, inframe insertion/deletion, missense variant, regulatory region ablation, or splice region variant) or HIGH (chromosome number variant, exon loss, frameshift variant, rare amino acid variant, splice donor/acceptor variant, start loss, stop loss/gain, or transcript ablation) phenotypic effects according to Sequence Ontology [193] and selected variants at which the CB4856 strain contains the alternate allele. This search found five candidate genes in the interval: *C45B11.8*, *C45B11.6*, *jmjd-5*, *srg-42*, and *cnc-10*, which each contain one non-synonymous variant between the N2 and CB4856 strains.

## Generation of deletion strains

Deletion alleles for genes of interest were generated to test the effects of loss-of-function variants on bleomycin responses. For each desired deletion, we designed a 5' and a 3' guide RNA with

the highest possible on-target and off-target scores, calculated using the Doench algorithm [194]. The N2 and CB4856 L4 hermaphrodite larvae were picked to 6 cm agar plates seeded with OP50 *E. coli* 24 hours before injection. The CRISPR injection mix was assembled by first incubating 0.88  $\mu\text{L}$  of 200  $\mu\text{M}$  AltR<sup>®</sup> CRISPR-Cas9 tracrRNA (IDT, catalog no. 1072532), 0.82  $\mu\text{L}$  of each of the 5' and 3' AltR<sup>®</sup> CRISPR-Cas9 crRNAs at a stock concentration of 100  $\mu\text{M}$  in Duplex Buffer (IDT), and 0.12  $\mu\text{L}$  of 100  $\mu\text{M}$  *dpy-10* co-injection crRNA at 95° for five minutes. 2.52  $\mu\text{L}$  of 69  $\mu\text{M}$  AltR<sup>®</sup> *S. pyogenes* Cas9 Nuclease (IDT, catalog no. 1081058) was added to the tracrRNA/crRNA complex mixture and incubated at room temperature for five minutes. Finally, 0.5  $\mu\text{L}$  of 10  $\mu\text{M}$  *dpy-10* repair construct and distilled water were added to a final volume of 10  $\mu\text{L}$ . The injection mixture was loaded into the injection capillary using a mouth pipet to avoid bubbles in the injection solution. Young adult animals were mounted onto agar injection pads, injected in either the anterior or posterior arm of the gonad, and allowed to recover on 6 cm plates in bulk. Twelve hours after injection, survivors were transferred to individual 6 cm plates. Broods with successful edits were easily observed because of the *dpy-10* co-injection marker, which created animals with an obvious locomotive defect, roller (Rol), or morphological phenotype, dumpy (Dpy). Three to four days after injections, plates were checked for the presence of Rol or Dpy F1 progeny. These Rol F1 progeny were transferred to individual 6 cm plates, allowed to lay offspring, and genotyped for the desired edit 24 hours later. Genotyping was performed with PCR amplicons designed around the desired site of the deletion. Plates with heterozygous or homozygous deletions were propagated and genotyped for at least two more generations to ensure homozygosity and to cross out the Rol mutation. Deletion amplicons were Sanger sequenced to identify the exact location of the deletion. Reagents used to generate deletion alleles are detailed in **Table S2-2**.

## Generation of CRISPR-mediated *jmjd-5* allele replacements

Allele replacement strains were created to test the effect of a particular amino acid substitution on bleomycin responses. A guide RNA was designed to cut two bp downstream of the the natural variant, with an on-target score of 31 and off-target score of 47 [194]. Two repair constructs were designed, one for the N2 to CB4856 replacement and *vice versa*. Repair oligonucleotides were homologous to the background strain except for the nucleotide variant, a silent mutation in the PAM site (A339A) to eliminate repair construct cleavage, and a silent mutation that introduces a *Bsa*I restriction enzyme cut site (T336T). Repair constructs contained a 35-bp homology arm on the PAM-proximal side of the edit and a 91-bp homology arm on the PAM-distal side of the edit. Injection mixes were made as above, with 0.6  $\mu$ L of 100  $\mu$ M *jmjd-5* repair construct added with the *dpy-10* repair construct in the last step of the protocol. Animals were injected as above and Rol F1 progeny were genotyped using PCR and restriction enzyme digestion. As with the deletion alleles, edited strains were homozygosed and their genotypes were confirmed with Sanger sequencing. Reagents used to generate these allele replacement strains are detailed in the **Table S2-2**.

## Linkage mapping of expression QTL

Microarray data for 13,107 probes were collected for synchronized young adult populations of 209 recombinant lines previously [158]. We performed linkage mapping as described above on these expression data and identified significant peaks for 3,298 probes.

## Hemizyosity high-throughput assay

Hemizyosity assays were used to test the effect of zero (two deletion alleles), one (one deletion allele and one wild-type allele), or two (two wild-type alleles) functional copies of a gene product on bleomycin responses. If the phenotype is affected by gene function, one would expect to see bleomycin responses scale with the number of functional alleles of the gene present in each strain. For each heterozygous/hemizygous genotype, 30 hermaphrodites and 60 males were placed onto each of four 6 cm plates and allowed to mate for 48 hours. The same process was completed for homozygous strains to remove biases introduced by the presence of male progeny in the assay. Mated hermaphrodites were transferred to a clean 6 cm plate and allowed to lay embryos for eight hours. After the egg lay period, adults were manually removed from egg lay plates, and embryos were washed into 15 mL conicals using M9 and a combination of pasteur pipette rinsing and scraping with plastic inoculation loops. Embryos were rinsed and centrifuged four times with M9 before being resuspended in K medium and 50  $\mu$ M kanamycin. Embryos hatched and arrested in the L1 larval stage in conicals overnight at 20° with shaking. The next morning, 50 L1 larvae were sorted into each well of a 96-well plate containing K medium, 10 mg/mL bacterial lysate, 50  $\mu$ M kanamycin, and either 1% distilled water or 1% distilled water plus bleomycin using the BIOSORT (dose-response assay) or by titering (hemizyosity assays). Animals were incubated in these plates for 48 hours at 20° with shaking and were paralyzed with 50  $\mu$ M sodium azide in M9 before being scored for phenotypic parameters using the BIOSORT.

## Statistical analysis

All statistical tests of phenotypic differences in the NIL, deletion strain, and allele-replacement strain assays were performed in R (version 3.3.1) using the *TukeyHSD* function [195] and an

ANOVA model with phenotype as the dependent variable and strain as the independent variable (*phenotype ~ strain*). The p-values of individual pairwise strain comparisons were reported, and p-values less than 0.05 were deemed significant. For each figure (with the exception of hemizyosity tests), phenotypes of NILs, deletion strains, and allele-replacement strains were compared to phenotypes of their respective background strains (either N2 or CB4856), and statistical significance is denoted by an asterisk above the boxplot for each strain. Correlation between RIAIL *H19N07.3* expression and median optical density in bleomycin was tested using a Spearman's correlation test. Statistical difference between N2 and CB4856 expression of *H19N07.3* measured by RNA-seq was tested using a likelihood-ratio and a Wald test with a threshold of  $p < 0.05$ .

## RNA-seq

Three independent replicates of RNA were sampled as follows. Bleach-synchronized embryos (~2,000) of the N2 and CB4856 strains were grown on 10 cm plates of NGMA for 66-69 hours. When F1 embryos appeared on the plate, synchronized young adults were collected by washing each plate twice with M9 buffer and incubating for 30 minutes in M9 to remove remaining bacteria. Then, samples were washed twice again with M9 buffer, and then washed with sterile water. Animals were pelleted and homogenized in Trizol (Ambion) by repeating freezing-thawing with liquid nitrogen five times. To extract RNA from each sample, chloroform, isopropanol, and ethanol were used for phase separation, precipitation and washing steps, respectively. RNA pellets were resuspended in TE buffer, and RNA quality was measured with a 2100 Bioanalyzer (Agilent). Library preparation and RNA sequencing (HiSeq4000, Illumina) were performed by the Genomics Facility at the University of Chicago. RNA-seq data were quantified with kallisto and then within-sample and between-sample normalization was performed using sleuth, which is based on

DESeq [196–198]. Significant differences between samples were determined by likelihood-ratio and Wald tests.

## Genome-wide association mapping

Bleomycin responses were measured for 83 *C. elegans* isotypes using the high-throughput fitness assay (**Figure 1-1**). Genome-wide association mapping was performed as described previously [134] using genotype data from CeNDR Release 20180527 [133]. In short, BCFtools was used to remove variants with missing genotype calls and variants with a minor allele frequency below 5% [199], and PLINK v1.9 was used to prune the genotypes at a linkage disequilibrium threshold of  $r^2 < 0.8$  [200,201], for a total of 59,241 pruned markers. A kinship matrix was generated using the *A.mat* function in the *rrBLUP* R package [202,203]. The *GWAS* function in the *rrBLUP* package was used to perform genome-wide association mapping with EMMA algorithm to correct for kinship [204,205]. The relatedness among these wild isolates was described previously [131,134,135].

## Identification of rare variants

VCF release 20180527 was downloaded from [elegansvariation.org](http://elegansvariation.org) [133]. The VCF was filtered to select all variants within the linkage mapping confidence interval (V:11042745-11189364) where CB4856 contains the alternate allele. Variants with a minor allele frequency less than 0.05 within the 83 wild isolates that have a bleomycin median optical density measurement were deemed “rare”.

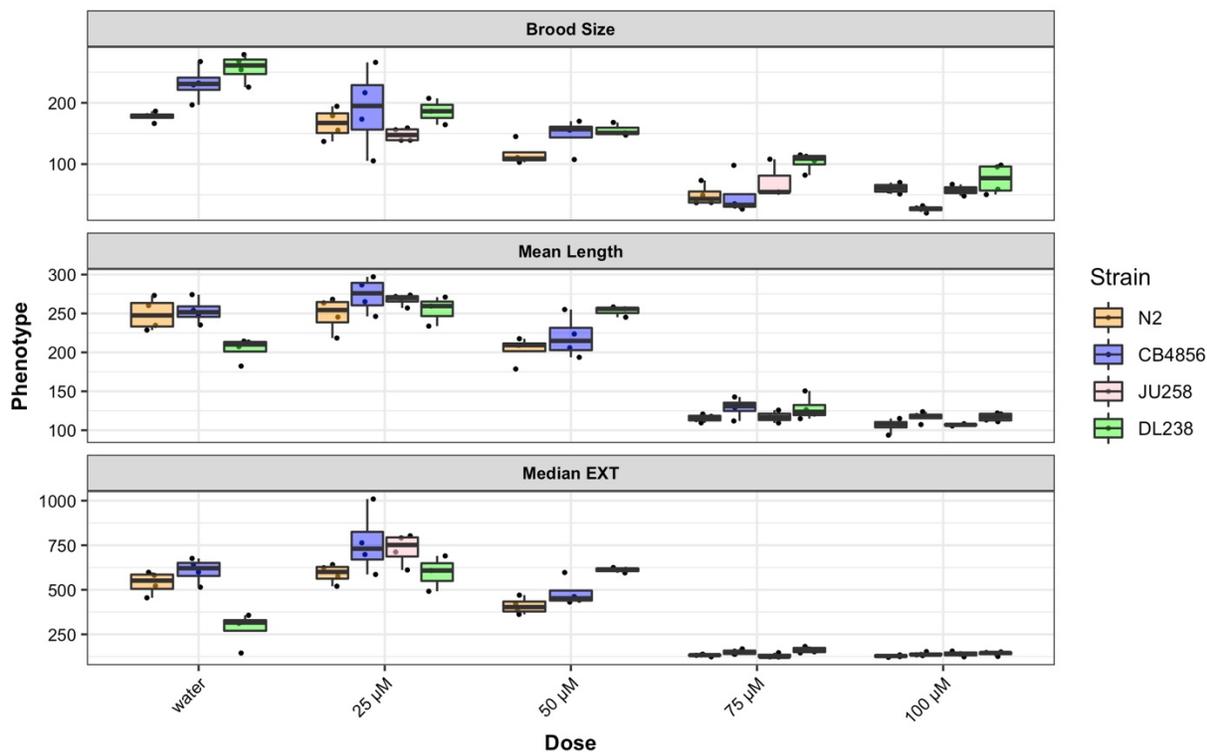
## Creation of neighbor-joining tree

Protein sequences for homologs of the *C. elegans* H19N07.3 protein were input to MUSCLE [206,207] to generate a multiple-sequence alignment. CLUSTALW was then used to generate a neighbor-joining tree and output as a Newick formatted file.

## Results

### Genetic differences underlie bleomycin-response variation

Bleomycin causes double-stranded DNA breaks, which ultimately lead to cytotoxicity of rapidly dividing cell populations. Therefore, exposure to bleomycin can affect the development of *C. elegans* larvae as well as germ-cell production of adult animals. We used a high-throughput assay (HTA) to measure the effects of bleomycin on development and brood size (**Figure 1-1**, Materials and Methods). To determine the concentration of bleomycin that would maximize among-strain while minimizing within-strain phenotypic variation, we used the HTA to perform a dose-response assay. We assessed bleomycin responses for four divergent strains (N2, CB4856, JU258, and DL238) across each of 26 HTA traits (**Figure 2-1**). For each concentration of bleomycin, we calculated the broad-sense heritability of the traits (Materials and Methods) and found that heritability was maximized at 50  $\mu\text{M}$  bleomycin (mean  $H^2$  across all traits = 0.58). Given these results, we exposed animals to 50  $\mu\text{M}$  bleomycin for all future HTA experiments.



**Figure 2-1 Responses to different doses of bleomycin vary across wild isolates**

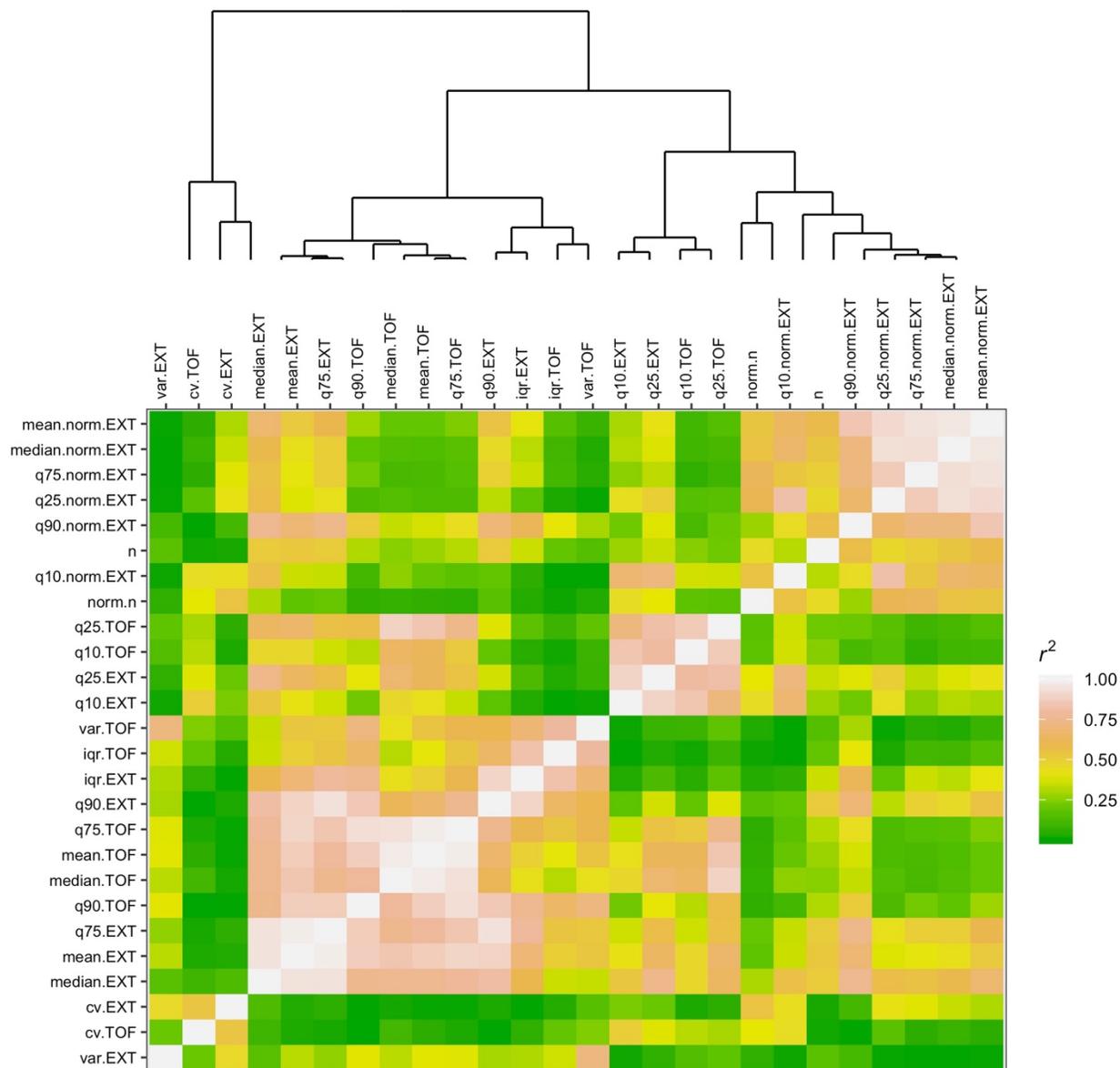
Dose-response phenotypes are shown for three high-throughput fitness traits: brood size (norm.n), mean length (mean.TOF), and median optical density (median.EXT). Phenotypes for each of the four wild isolates are shown as Tukey boxplots, colored by strain (N2 - orange, CB4856 - blue, JU258 - pink, DL238 - green). The x-axis shows the concentration of bleomycin (or water) to which the animals were exposed, and the y-axis shows the pruned phenotype. Each point is a biological replicate. The JU258 strain was pruned from the water and 50 µM bleomycin conditions because the wells were contaminated.

Two of the strains used in the dose response assay, N2 and CB4856, have been extensively characterized at the genome level [150–152] and displayed divergent bleomycin responses (**Figure 2-1**). Recombinant inbred advanced intercross lines (RIAILs) were previously constructed between these two strains [109,110], and these RIAILs have been leveraged to identify genetic variants that cause phenotypic differences between the N2 and CB4856 strains [40,141,156,161–168,188]. We used these RIAILs to identify genetic variants that contribute to differential bleomycin responses between the N2 and CB4856 strains. Using our HTA, we measured each

of the 26 fitness parameters for 249 RIALs. Correlations between each pairwise combination of the 26 HTA measurements revealed several clusters of highly correlated traits (**Figure 2-2**). Therefore, the summary statistics measured by the BIOSORT should not be considered independent traits for linkage mapping. We selected median optical density (median.EXT) for future analyses, which is related to both animal length and optical extinction, because this trait was highly correlated with many of the 26 HTA traits and was highly heritable ( $H^2 = 0.73$ ).

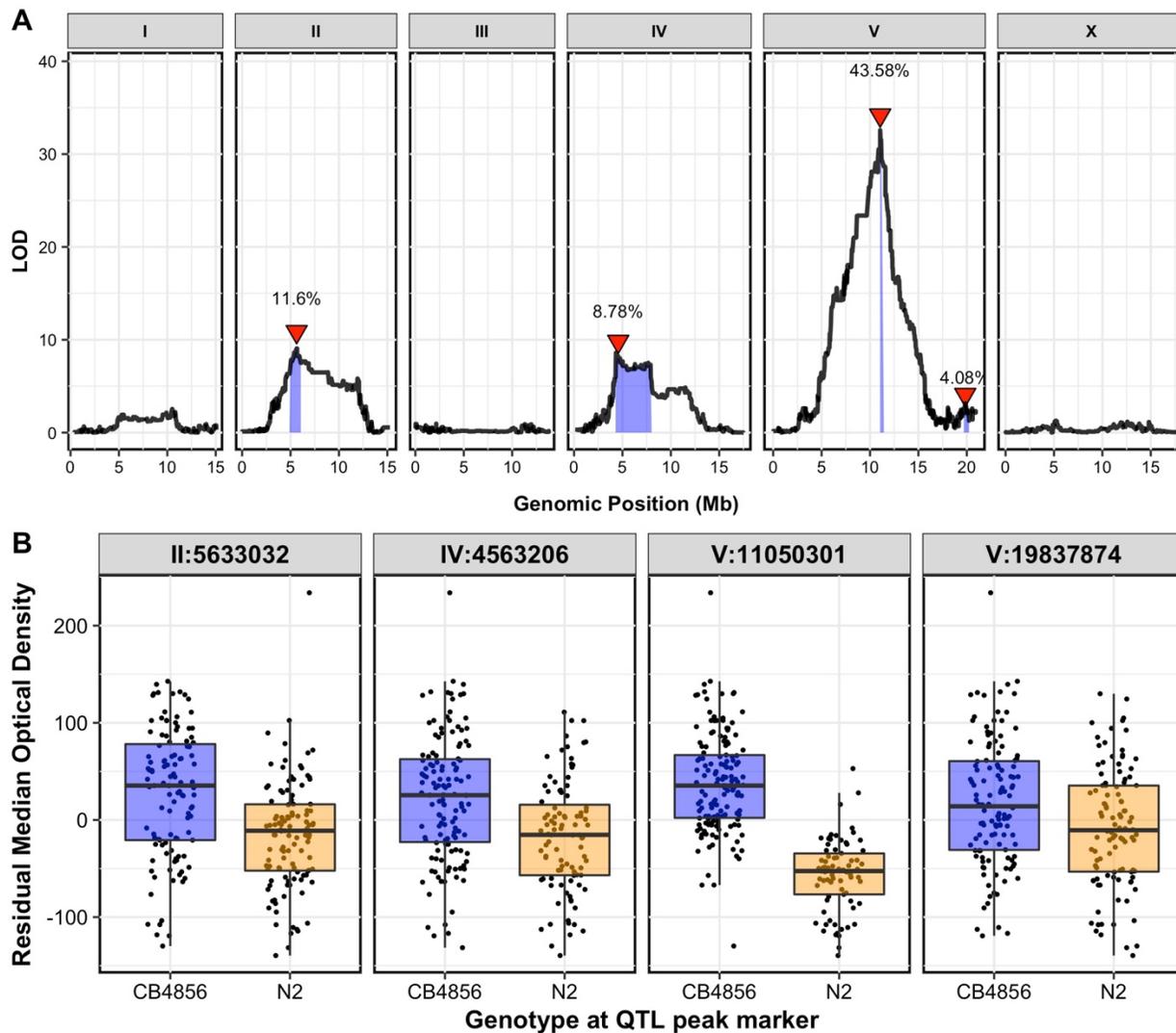
### The QTL on the center of chromosome V strongly impacts bleomycin response

We performed linkage mapping on the residual median optical density measurements in bleomycin and identified four significant quantitative trait loci (QTL, **Figure 2-3A**). The QTL on the center of chromosome V was highly significant (explained 43.58% of the total variation and 55.60% of the genetic variation) with a LOD score of 32.57, and it was detected for 25 of the 26 HTA traits (**Figure S2-1**). The QTL 95% confidence interval was approximately 147 kb. Strains with the N2 allele at the peak marker had a lower median optical density in bleomycin and were interpreted to be more sensitive than those RIALs with the CB4856 allele (**Figure 2-3B**). For each of the 26 HTA traits, we compared the broad-sense heritability (calculated from the dose responses) to the narrow-sense heritability (the sum of all variance-explained percentages of each significant QTL) and found that in many cases the QTL explained most of the broad-sense heritability (**Figure 2-4**).



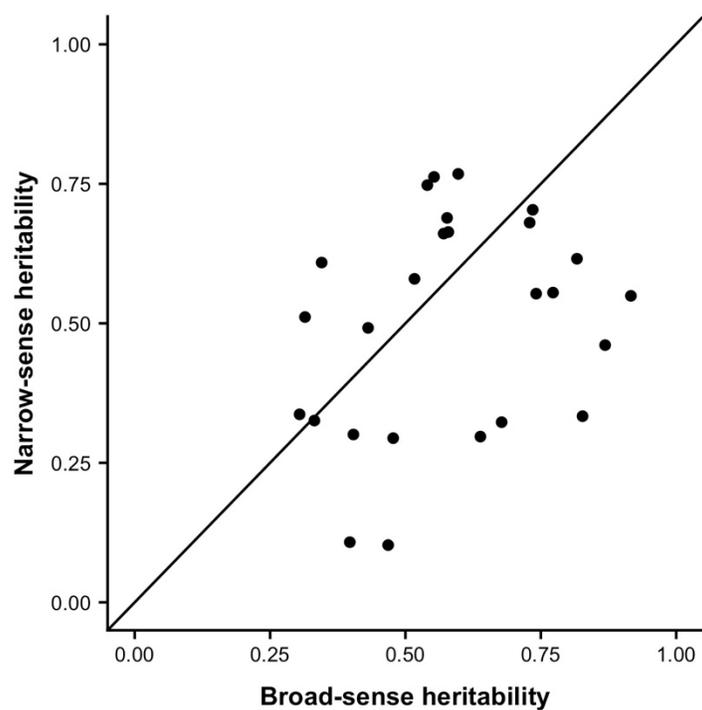
**Figure 2-2 Bleomycin-response HTA traits are correlated**

Pairwise correlations of high-throughput assay (HTA) traits are shown. Top: A dendrogram of HTA traits, calculated from pairwise correlations of RIAIL phenotypes for each trait, is shown. Bottom: The x- and y-axes list HTA traits for which the pairwise correlation coefficients ( $r^2$ ) are shown as colored tiles, ranging from green to yellow to pink to white.



**Figure 2-3 Bleomycin-response variation maps to four QTL**

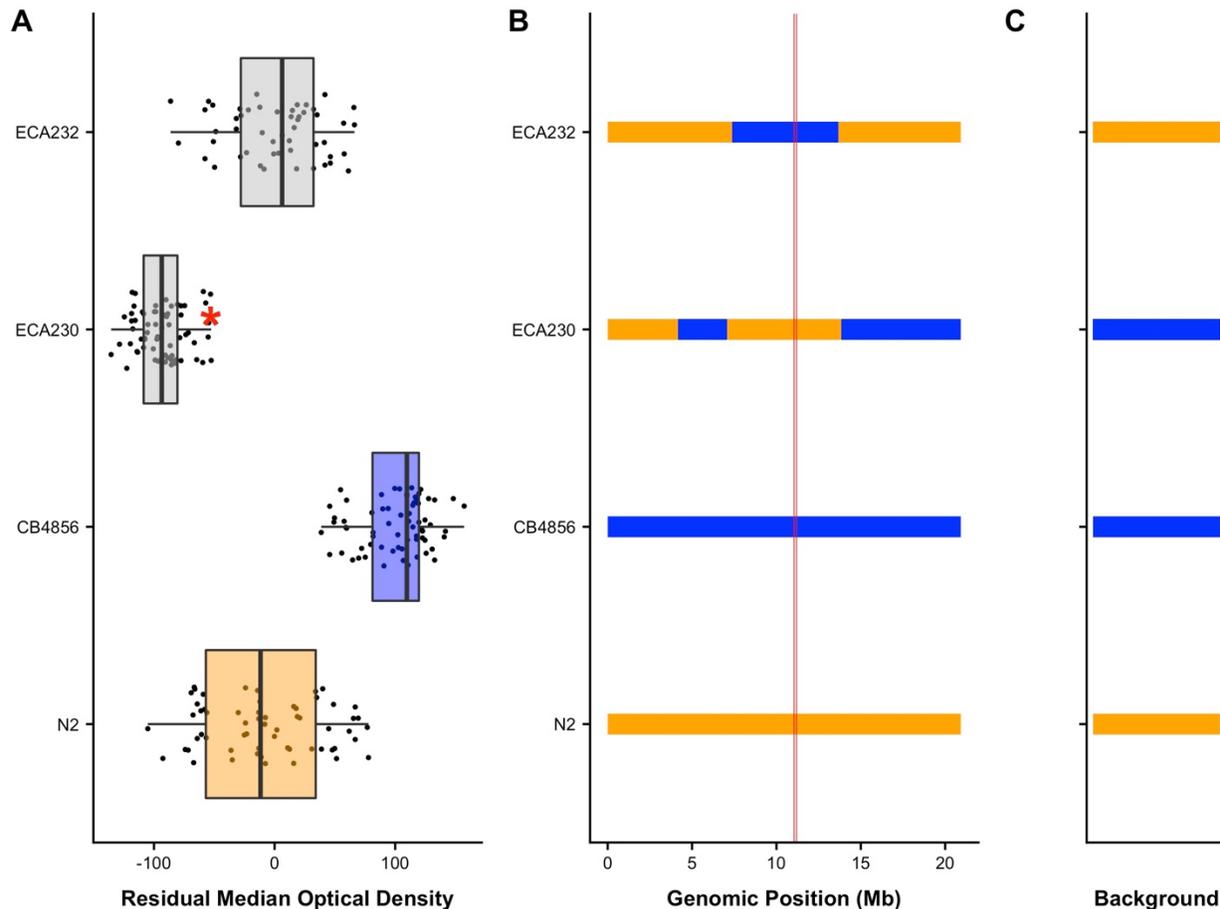
Linkage-mapping analysis of bleomycin-response variation is shown for residual median optical density in bleomycin. **A.** On the x-axis, each of 13,003 genomic markers, split by chromosome, were tested for correlation with phenotypic variation across the RIAIL panel. The log of the odds (LOD) score for each marker is reported on the y-axis. Each significant quantitative trait locus (QTL) is indicated by a red triangle at the peak marker, and a blue ribbon shows the 95% confidence interval around the peak marker. The total amount of phenotypic variance across the RIAIL panel explained by the genotype at each peak marker is shown as a percentage. **B.** Residual median optical density phenotypes (y-axis), split by allele at each QTL peak marker (x-axis), are shown. For each significant QTL, phenotypes of RIAILs that contain the N2 allele (orange) are compared to RIAILs that contain the CB4856 allele (blue). Phenotypes are shown as Tukey box plots with the phenotypes of each individual strain shown as points behind the box plots.



**Figure 2-4 Linkage mapping QTL explain most of the broad-sense heritability for bleomycin responses**

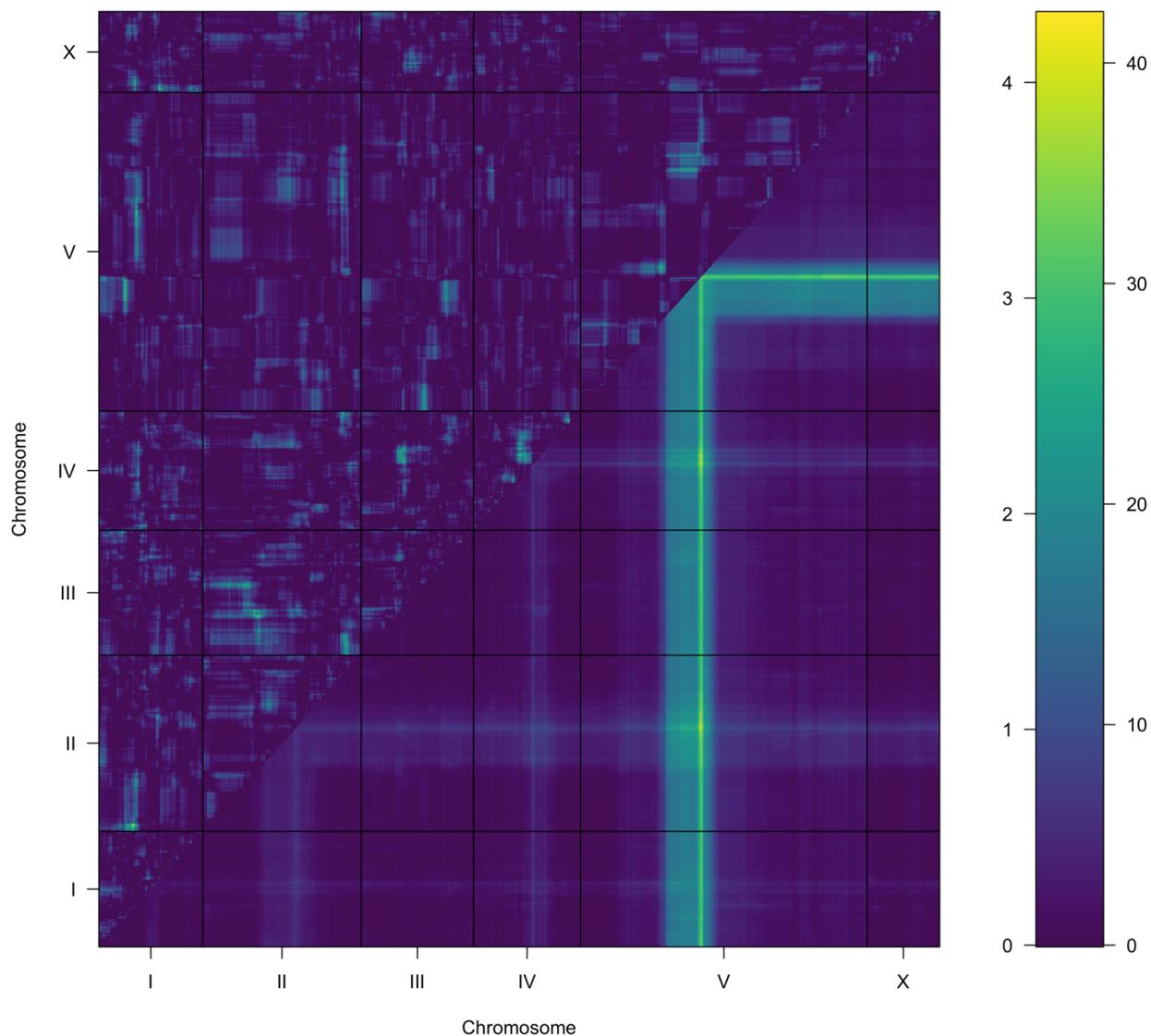
For each of the 26 HTA traits, the broad-sense heritability estimate, calculated from the dose-responses, is plotted on the x-axis, and the narrow-sense heritability estimate, calculated as the sum of all variance-explained percentages of significant QTL, is plotted on the y-axis. The line  $x=y$  is shown as a visual reference.

We isolated this QTL in a controlled genetic background by generating near-isogenic lines (NILs) that each contain a genetic background derived from either the N2 or CB4856 strain and a region of chromosome V from the opposite parental genotype (**Table S2-1**). We used the HTA to test these strains in response to bleomycin. The NIL with the N2 genotype across this QTL introgressed into the CB4856 background (ECA230) was statistically more sensitive to bleomycin than CB4856 (**Figure 2-5**,  $p = 1.3e-14$ , Tukey HSD). This phenotype indicated that the N2 genotype within the introgressed region (which includes the QTL confidence interval) confers sensitivity to bleomycin. However, the reciprocal NIL with the CB4856 locus introgressed into the N2 background (ECA232) had a bleomycin-response phenotype that was not significantly different from the N2 strain (**Figure 2-5**,  $p = 0.053$ , Tukey HSD), suggesting that interacting loci could underlie bleomycin responses in a background-dependent manner. We performed a two-factor genome scan to map potential epistatic loci but did not identify a significant interaction between the QTL on chromosome V and other loci (**Figure 2-6**). However, the failure to detect significant interacting QTL could be because we have too few recombinant strains or because too few replicates of each RIAIL were phenotyped. Alternatively, more than two loci might underlie the transgressive phenotype of ECA230 and a two-factor genome scan might not be able to capture this complexity.



**Figure 2-5 Broad NILs recapitulate the expected QTL effect on bleomycin responses**

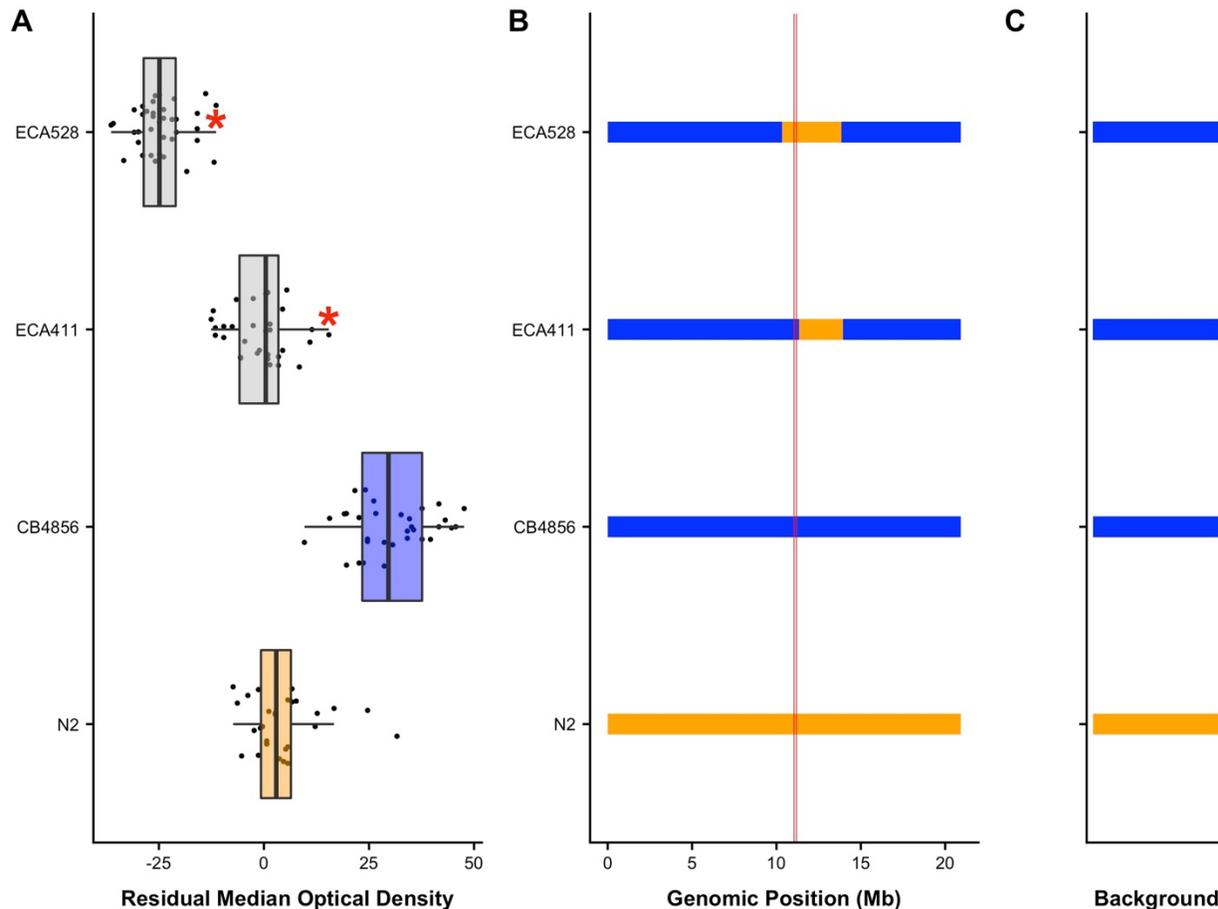
Phenotypes and genotypes of NIL strains are shown. **A.** Phenotypes for each strain are shown as Tukey box plots, with strain name on the y-axis and residual bleomycin median optical density on the x-axis. Each point is a biological replicate. Parental strain box plots are colored by their genetic background, with orange indicating an N2 background and blue indicating a CB4856 genetic background. NILs are shown as grey box plots. A red asterisk indicates a significant difference between the phenotype of a given strain and the phenotype of the corresponding parental strain ( $p < 0.05$ , Tukey HSD). **B.** Chromosomal representations of chromosome V are shown for each of the strains in **A.** Strain names are reported on the y-axis, and genomic position (Mb) is shown on the x-axis. Blocks of color indicate genotypes of genomic regions with orange indicating the N2 genotype and blue indicating the CB4856 genotype. Vertical red lines mark the confidence interval of the QTL from linkage mapping. **C.** Background genotypes are represented as rectangles with colors indicating N2 (orange) or CB4856 (blue) genetic backgrounds.



**Figure 2-6 A two-factor genome scan for bleomycin responses does not detect epistatic loci**

A two-factor genome scan for residual optical density in bleomycin is shown. Log of the odds (LOD) scores are shown for each pairwise combination of loci, split by chromosome. The upper-left triangle contains the epistasis LOD scores, and the lower-right triangle contains LOD scores for the full model. LOD scores are shown as colors, with lower scores in blue and higher scores in yellow, as shown in the color scale. The epistatic LOD score axis is on the left of the color scale and the full LOD score axis is on the right of the color scale.

Nonetheless, because ECA230 recapitulated the expected QTL phenotype, we generated two NILs (ECA411 and ECA528) that narrowed this introgressed region to more precisely locate the causal variant. In addition, the N2 region on the left of chromosome V was removed from both NIL strains to ensure that this region of introgression did not underlie the phenotypic difference between ECA230 and CB4856. The genotypes of ECA411 and ECA528 differ in a small region of chromosome V that includes the QTL confidence interval (**Figure 2-7**). Both of these strains were more sensitive to bleomycin than the background parental strain, CB4856. This result could suggest that the introgressed region shared between these strains, which does not include the QTL, conferred some bleomycin-response variation between the N2 and CB4856 strains (**Figure 2-7**). Alternatively, the hypersensitivity of these NILs could suggest the presence of Dobzhansky-Muller incompatibilities between the N2 and CB4856 genotypes [208] that might affect stress responses of the NILs. However, ECA528 was much more sensitive to bleomycin than ECA411 (**Figure 2-7**). Because ECA528 has the N2 genotype across the QTL region and ECA411 has the CB4856 genotype, these results suggest that the QTL genotype strongly affects bleomycin sensitivity (**Figure 2-7**, ECA528 vs. each other strain  $p < 1e-14$ , Tukey HSD). The empirically defined region lies between 10,339,727 and 11,345,443 bp on chromosome V and fully encompasses the linkage mapping confidence interval (from 11,042,745 to 11,189,364 bp on chromosome V).



**Figure 2-7 NILs localize the bleomycin-response QTL to a small region on chromosome V**

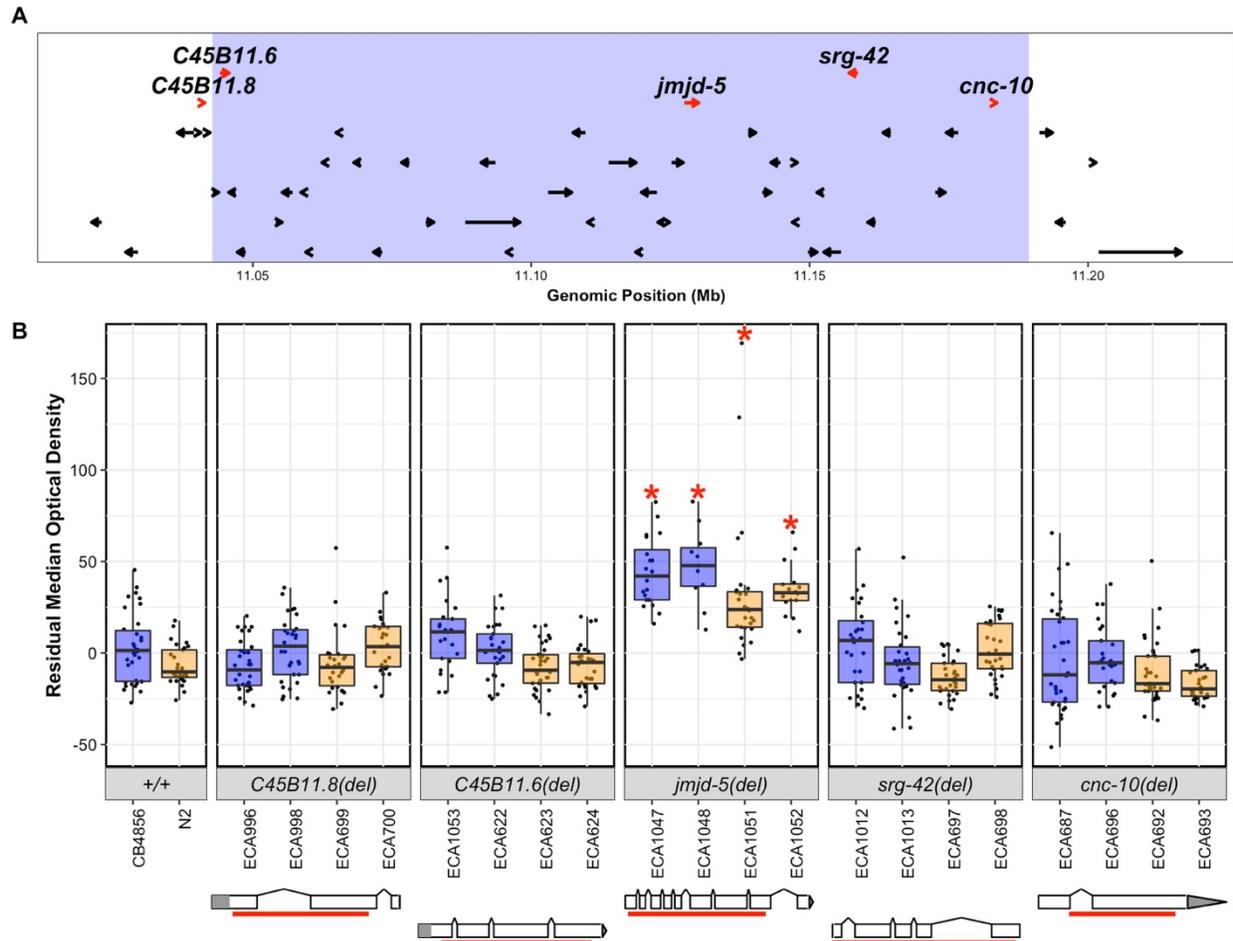
Phenotypes and genotypes of near-isogenic lines (NILs) are shown. **A.** Phenotypes for each strain are shown as Tukey box plots, with strain name on the y-axis and residual bleomycin median optical density on the x-axis. Each point is a biological replicate. Parental strain box plots are colored by their genetic background, with orange indicating an N2 background and blue indicating a CB4856 genetic background. NILs are shown as grey box plots. A red asterisk indicates a significant difference between the phenotype of a given strain and the phenotype of the corresponding parental strain ( $p < 0.05$ , Tukey HSD). **B.** Chromosomal representations of chromosome V are shown for each of the strains in **A.** Strain names are reported on the y-axis, and genomic position (Mb) is shown on the x-axis. Blocks of color indicate genotypes of genomic regions with orange indicating the N2 genotype and blue indicating the CB4856 genotype. Vertical red lines mark the confidence interval of the QTL from linkage mapping. **C.** Background genotypes are represented as rectangles with colors indicating N2 (orange) or CB4856 (blue) genetic backgrounds.

## Genes with non-synonymous variants in the QTL region do not impact bleomycin responses

Because the recombination rate in the centers of *C. elegans* chromosomes is lower than chromosome arms [110], it was difficult to generate additional NILs to narrow the QTL region further. Therefore, we took a targeted approach and created CRISPR-Cas9 directed modifications of candidate genes in the region. The 147 kb confidence interval on chromosome V contains 93 genes, including pseudogenes, piRNA, miRNA, ncRNA, and protein-coding genes. Given the narrow confidence interval, we expanded our search to include an additional 20 kb on each side of the 147 kb interval (Materials and Methods). Of the 118 genes included in the wider region, five genes, *C45B11.8*, *C45B11.6*, *jmjd-5*, *srg-42*, and *cnc-10*, contain predicted non-synonymous variants between the N2 and CB4856 strains (**Figure 2-8A**). These variants could cause differential bleomycin sensitivity between the N2 and CB4856 strains.

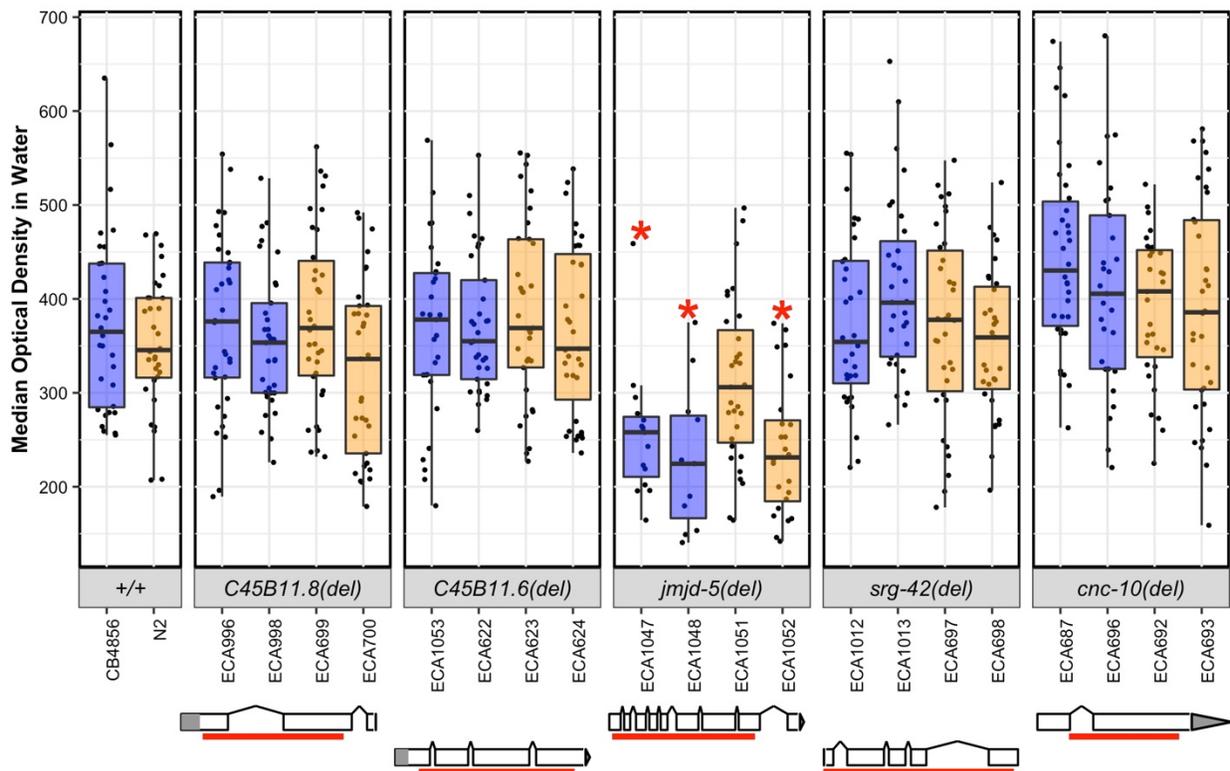
To test these genes in bleomycin-response variation, we systematically deleted each of the candidate genes in both the N2 and CB4856 backgrounds. We used CRISPR-Cas9 mediated genome editing to generate two independent deletion alleles of each gene in each genetic background to reduce the possibility that off-target mutations could cause phenotypic differences. We tested the bleomycin response of each deletion allele in comparison to the N2 and CB4856 parental strains (**Figure 2-8B**). The deletion alleles of *C45B11.8*, *C45B11.6*, *srg-42*, and *cnc-10* each had a bleomycin response similar to the respective parent genetic background, which suggested that the functions of each of these genes did not affect bleomycin responses (**Figure 2-8B**,  $p > 0.05$ , Tukey HSD). By contrast, the *jmjd-5* deletion alleles in the N2 and the CB4856 backgrounds were each more resistant to bleomycin than their respective parental strains (**Figure**

**2-8B**, ECA1047 vs. CB4856  $p = 3.8e-10$ , ECA1048 vs. CB4856  $p = 0.026$ , ECA1051 vs. N2  $p = 7.4e-4$ , ECA1052 vs. N2  $p = 2.9e-6$ , Tukey HSD). However, we also noted that these strains were more sensitive in the control condition than other deletion strains (**Figure 2-9**). Therefore, the relative increased bleomycin resistance observed in the *jmjd-5* deletion strains could be caused by their increased sensitivity in the control condition.



**Figure 2-8 Strains with *jmjD-5* deletion alleles are resistant to bleomycin**

Bleomycin responses of the deletion alleles for each candidate gene are shown. **A.** The linkage mapping QTL confidence interval (light blue) with 20 kb on the left and the right is displayed. Each protein-coding gene in the region is indicated by an arrow that points in the direction of transcription. Genes with non-synonymous variants between the N2 and CB4856 strains are shown as red arrows and are labeled with their gene name. **B.** Deletion alleles for each of these genes were tested in response to bleomycin. Bleomycin responses are shown as Tukey box plots, with the strain name on the x-axis, split by gene, and residual median optical density on the y-axis. Each point is a biological replicate. Strains are colored by their background genotype (orange indicates an N2 genetic background, and blue indicates a CB4856 genetic background). For each gene, two independent deletion alleles in each background were created and tested. Red asterisks indicate a significant difference ( $p < 0.05$ , Tukey HSD) between a strain with a deletion and the parental strain that has the same genetic background. Depictions of each deletion allele are shown below the phenotype for each candidate gene. White rectangles indicate exons and diagonal lines indicate introns. The 5' and 3' UTRs are shown by grey rectangles and triangles, respectively. The region of the gene that was deleted by CRISPR-Cas9 directed genome editing is shown as a red bar beneath each gene model.

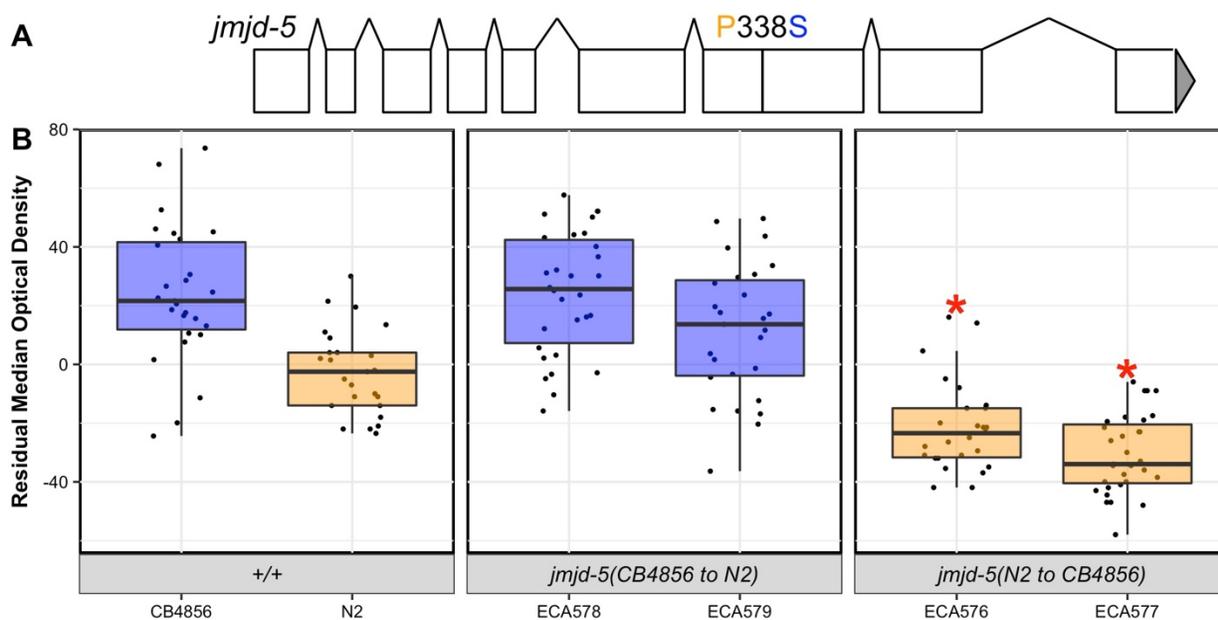


**Figure 2-9 Strains with *jmjD-5* deletion alleles are sensitive in the control condition**

Control-condition phenotypes of deletion alleles for each candidate gene are shown. Deletion alleles for each candidate gene were tested in response to the control condition (lysate in K medium with 1% distilled water). Control-condition responses are shown as Tukey box plots, with the strain name on the x-axis, split by gene, and median optical density on the y-axis. Each point is a biological replicate. Strains are colored by their background genotype (orange indicates an N2 genetic background, and blue indicates a CB4856 genetic background). For each gene, two independent deletion alleles in each background were created and tested. Red asterisks indicate a significant difference ( $p < 0.05$ , Tukey HSD) between a strain with a deletion and the parental strain that has the same genetic background. Depictions of each deletion allele are shown below the phenotype for each candidate gene. White rectangles indicate exons and diagonal lines indicate introns. The 5' and 3' UTRs are shown by grey rectangles and triangles, respectively. The region of the gene that was deleted by CRISPR-Cas9 directed genome editing is shown as a red bar beneath each gene model.

We tested if the non-synonymous variant in *jmjd-5* between the N2 and CB4856 strains caused bleomycin-response differences. At residue 338 of JMJD-5, the N2 strain has a proline, whereas the CB4856 strain has a serine (S338P, **Figure 2-10A**). We used CRISPR-Cas9 to generate reciprocal allele replacements of the *jmjd-5* single-nucleotide polymorphism that encodes the putative amino-acid change in the N2 background *jmjd-5(N2 to CB4856)* and in the CB4856 background *jmjd-5(CB4856 to N2)* (**Table S2-2**, Materials and Methods). We created two independent allele replacements in each genetic background and measured each strain for bleomycin-response differences as compared to the parental strains (**Figure 2-10B**). Although the allele-replacement strains with the CB4856 allele in the N2 genetic background *jmjd-5(N2 to CB4856)* were significantly different from the N2 parental strain, these strains were more sensitive to bleomycin than N2 (**Figure 2-10B**, ECA576 vs. N2  $p = 0.006$ , ECA577 vs. N2  $p = 1.6e-6$ , Tukey HSD). This increased sensitivity was unexpected, because the CB4856 allele at the *jmjd-5* locus should confer resistance. However, the NIL with the CB4856 genotype across the QTL was not different from the N2 parental strain (ECA232 in **Figure 2-5**), suggesting that the QTL might only confer increased sensitivity when the N2 allele is in the CB4856 background. Therefore, it remained unclear whether an allele replacement of *jmjd-5* in the N2 parental background could confer resistance. Neither of the two strains with the N2 allele in the CB4856 background, *jmjd-5(CB4856 to N2)*, conferred a significantly more sensitive phenotype than the CB4856 parental strain (**Figure 2-10B**). Given that the QTL explained 43.58% of phenotypic variation among the RIALs, the causal variant should have a clear impact on bleomycin response. Additionally, the NILs with the N2 allele at the QTL introgressed into the CB4856 background displayed a significant increase in bleomycin sensitivity compared to the parental CB4856 strain (**Figure 2-5**). Taken together, the phenotypes of the reciprocal allele-replacement strains showed that the amino-acid change in JMJD-5 likely does not underlie bleomycin-response variation between the

N2 and CB4856 strains, although deletion of this gene did cause resistance to bleomycin regardless of the genetic background.



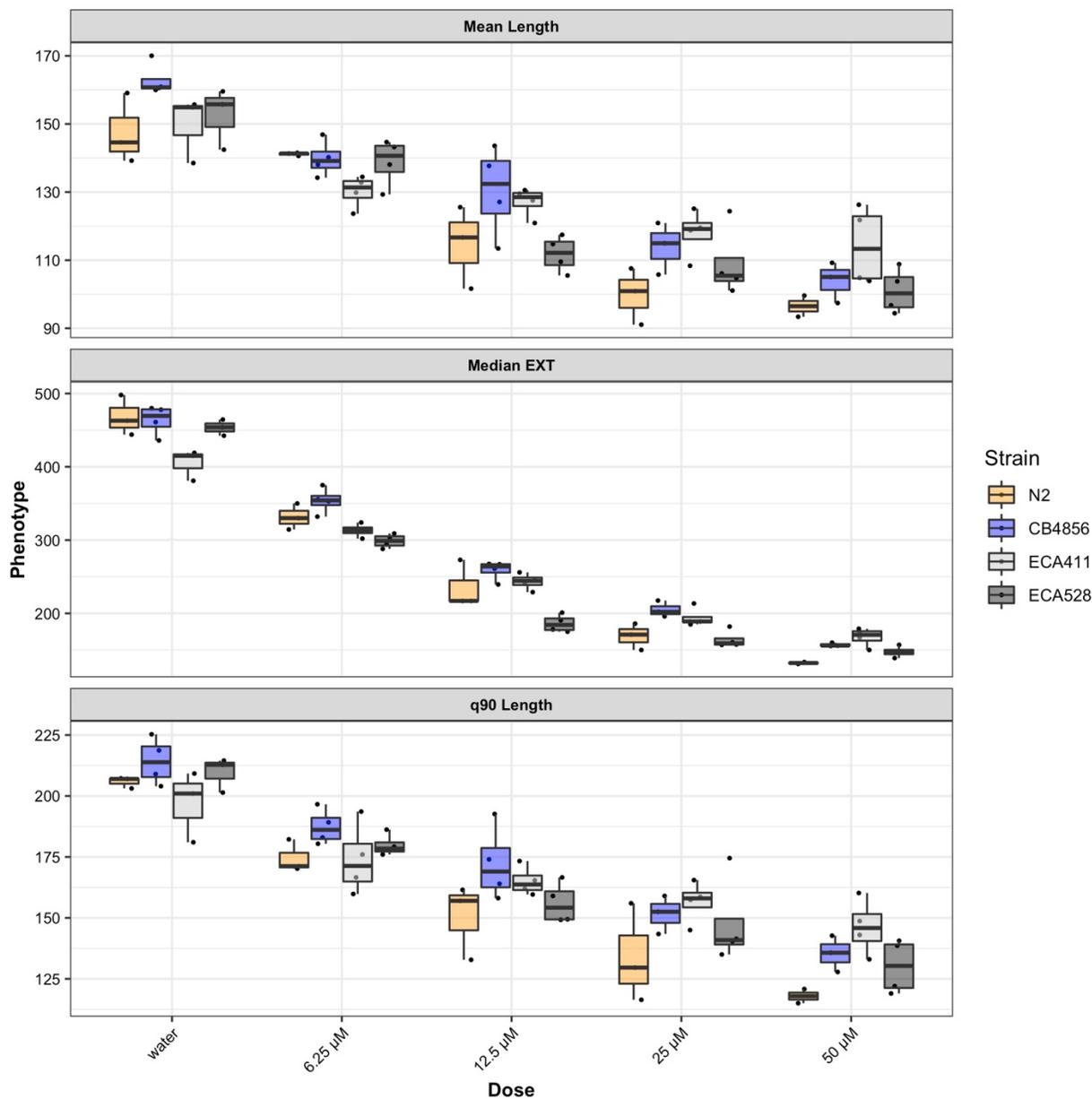
**Figure 2-10 The non-synonymous variant of *jmjD-5* does not affect bleomycin responses**

Reciprocal allele-replacement strains of *jmjD-5* were tested in response to bleomycin. **A.** A model of *jmjD-5* is shown with white rectangles indicating exons and black diagonal lines indicating introns. The 5' and 3' UTRs are shown as a grey rectangle and triangle, respectively. The location of the amino-acid variant between the N2 and CB4856 strains is shown in black text above the gene depiction with the N2 residue in orange and the CB4856 residue in blue. **B.** Bleomycin responses of allele-replacement strains are shown as Tukey box plots, split and colored by genetic background (N2 in orange and CB4856 in blue) with the strain name on the x-axis and residual median optical density on the y-axis. Each point is a biological replicate. Red asterisks indicate a significant difference between an allele-replacement strain and the parental strain sharing its genetic background ( $p < 0.05$ , Tukey HSD).

To test whether other variation in *jmjD-5* function between N2 and CB4856 impacts bleomycin responses, we designed a modified version of the HTA. This modified version allowed us to test bleomycin responses of heterozygous (two different functional alleles of *jmjD-5*) and hemizygous (one functional allele and one deletion allele of *jmjD-5*) animals. Thereby, this approach allowed us to determine the effect of adding or removing single N2 or CB4856 *jmjD-5* alleles on bleomycin sensitivity. During the modified HTA, strains were crossed and approximately 50 heterozygous or

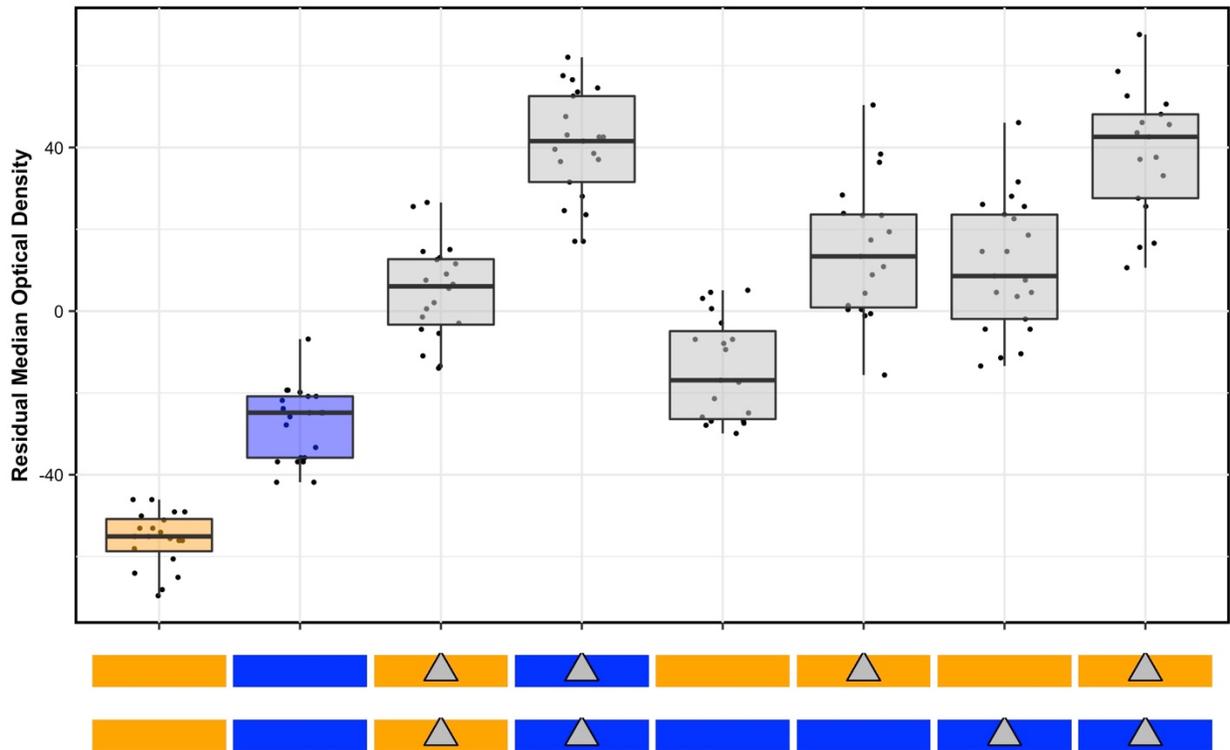
hemizygous F1 progeny were placed directly into either control or drug conditions. Whereas the original HTA allows animals to propagate for a generation in the presence of control or drug conditions, the modified HTA scores the F1 individuals that are sorted into the 96-well plates before they reproduce. Therefore, the scored F1 population remains heterozygous or hemizygous and a marker that differentiates heterozygous versus homozygous animals is not needed. First, we performed a dose response with this modified HTA to confirm that bleomycin responses differed among N2, CB4856, ECA411, and ECA528 when the modified HTA was used and to identify the dose of bleomycin with the highest heritability in the new approach. The most heritable concentration of bleomycin was 12.5  $\mu\text{M}$ , and the same ranking of bleomycin resistance was recapitulated among the strains tested with the modified HTA as with the original HTA (**Figure 2-11**).

We performed a reciprocal hemizyosity assay to test if natural variation in *jmjd-5* function affected bleomycin responses. The results of this assay supported the previously identified increase in bleomycin resistance of homozygous *jmjd-5* deletions in both parental backgrounds (**Figure 2-12**,  $P < 0.05$ , Tukey HSD), which again might be caused by an increased sensitivity in the control condition (**Figure 2-13**). However, the increases in bleomycin resistance between each *jmjd-5* deletion strain and the strain with the same genetic background were similar, and the reciprocal hemizygous strains show equivalent bleomycin responses (**Figure 2-12**). Taken together, these results suggest that natural variation in *jmjd-5* function does not underlie this QTL.



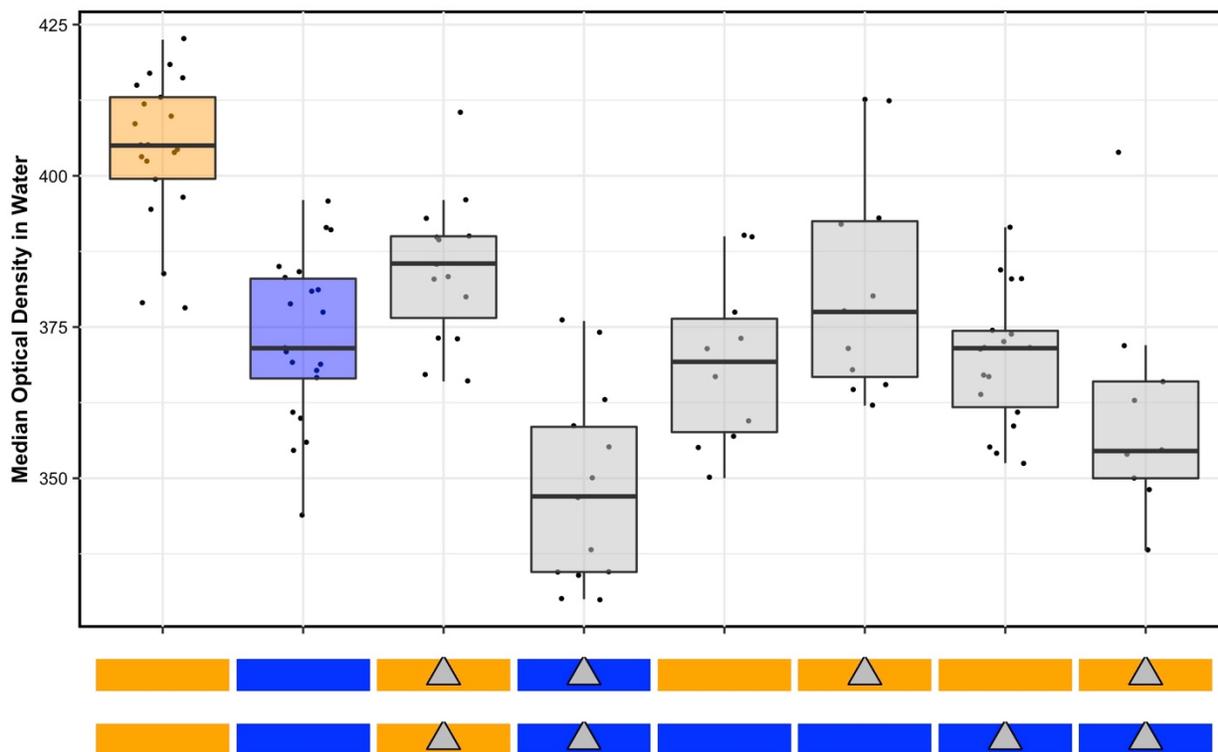
**Figure 2-11 Phenotypes in the modified HTA match previous bleomycin responses**

Dose-response phenotypes from the modified HTA are shown for three high-throughput fitness traits: mean length (mean.TOF), median optical density (median.EXT), and the 90<sup>th</sup> quantile of optical density (q90 EXT). Phenotypes for each of the four tested strains are shown as Tukey boxplots, colored by strain (N2 - orange, CB4856 - blue, ECA411 - light grey, ECA528 - dark grey). The x-axis shows the concentration of bleomycin (or water) to which the animals were exposed, and the y-axis shows the pruned phenotype. Each point is a biological replicate.



**Figure 2-12 Natural variation in *jmjd-5* does not underlie bleomycin-response differences**

Results of the *jmjd-5* reciprocal hemizygosity assay are shown. The y-axis shows the residual median optical density for each strain reported along the x-axis. Bleomycin responses are reported as Tukey box plots where each point is a biological replicate. The genotypes of each strain are shown as colored rectangles beneath each box plot, where each rectangle represents a homolog (orange rectangles are an N2 genotype, and blue rectangles are a CB4856 genotype). The maternal homolog is shown on top and the paternal homolog is shown on bottom. Grey triangles indicate a deletion of *jmjd-5*, placed on the rectangle showing the background into which the deletion was introduced. The box plots for the parental strains (N2 and CB4856, on the left) are colored according to genotype.



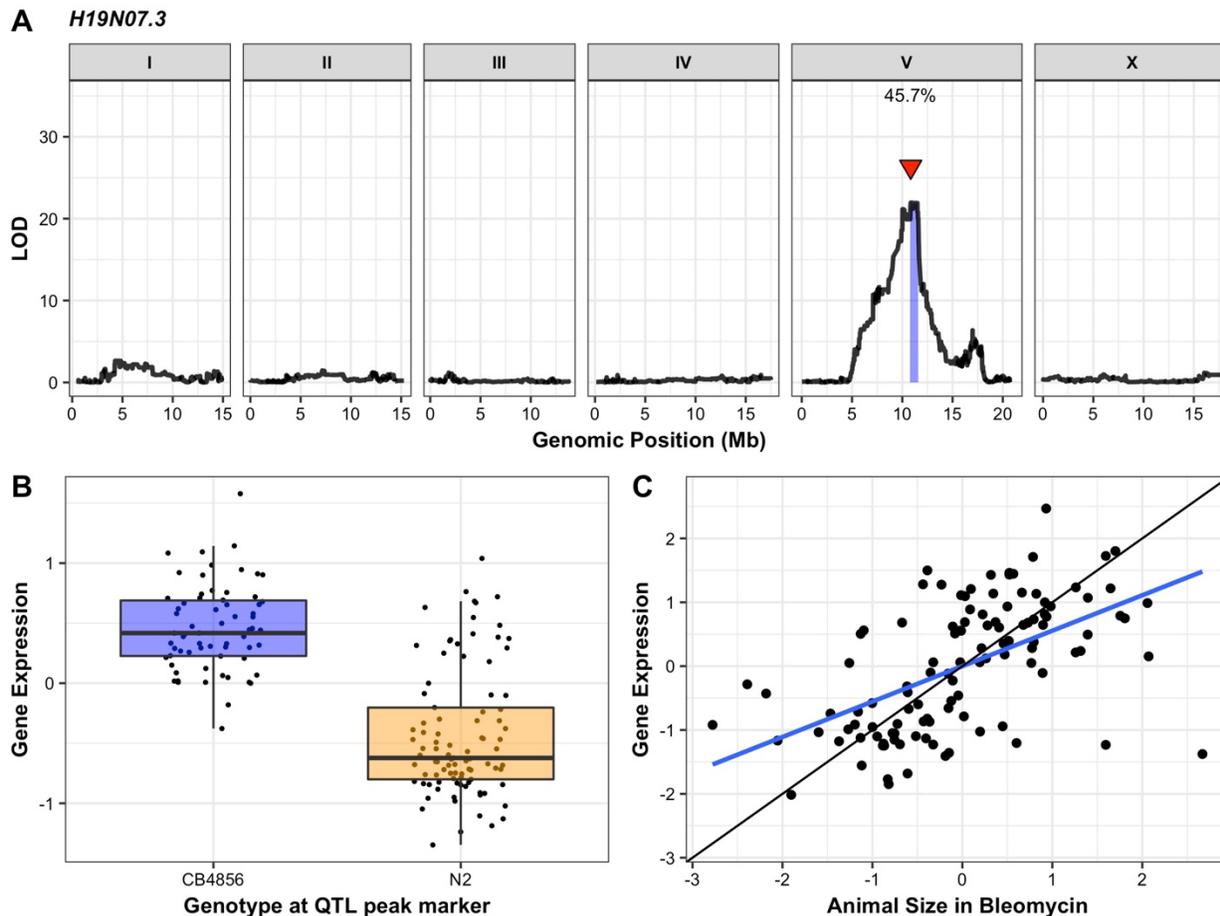
**Figure 2-13 Strains with homozygous *jmjD-5* deletions are sensitive to the control condition, whereas hemizygous strains are not**

Control-condition phenotypes of the strains tested in the reciprocal hemizygosity experiment of *jmjD-5* are shown. The y-axis shows the pruned median optical density upon exposure to the control condition (lysate in K medium plus 1% distilled water) for each strain reported along the x-axis. Control-condition responses are reported as Tukey box plots where each point is a biological replicate. The genotypes of each strain are shown as colored rectangles beneath each box plot, where each rectangle represents a homolog (orange rectangles are an N2 genotype, and blue rectangles are a CB4856 genotype). The maternal homolog is shown on top and the paternal homolog is shown on bottom. Grey triangles indicate a deletion of *jmjD-5*, placed on the rectangle showing the background into which the deletion was introduced. The box plots for the parental strains (N2 and CB4856, on the left) are colored according to genotype.

## The nematode-specific gene *H19N07.3* impacts bleomycin variation

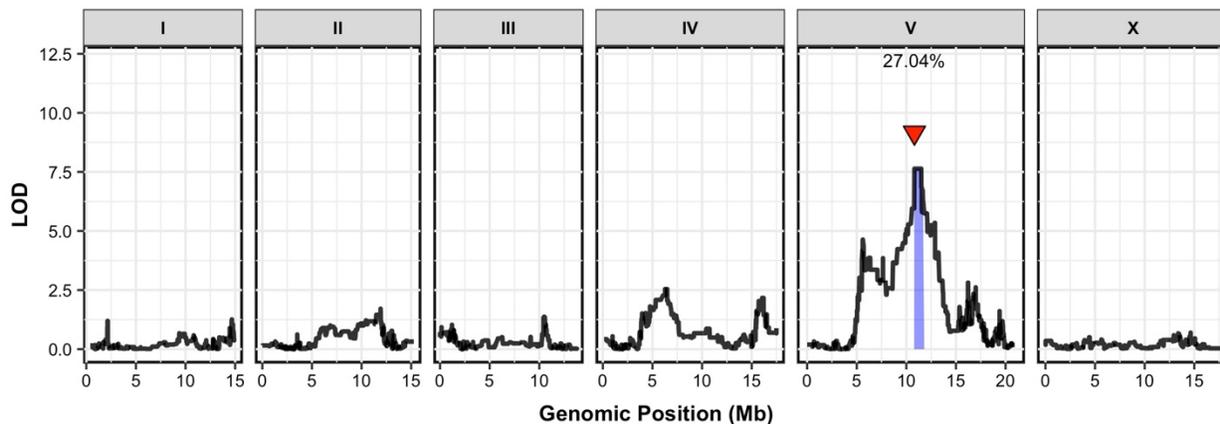
Because none of the genes with a non-synonymous variant between the N2 and CB4856 strains explained the QTL, we explored other ways in which natural variation could impact bleomycin responses. We used the 13,003 SNPs to perform linkage mapping on the gene expression data of the RIALs and identified 4,326 expression QTL (eQTL) across the genome ([158], Materials and Methods). Of the 118 genes in the 187 kb surrounding the bleomycin-response QTL,

expression for 50 genes were measured in the previous microarray study. We identified a significant eQTL for eight of these 50 genes, four of which mapped to chromosome V. eQTL for two of those four genes, *H19N07.3* and *cnc-10*, mapped to the center of chromosome V and overlapped with the bleomycin-response QTL. Because *cnc-10* did not underlie bleomycin response variation (**Figure 2-8B**), we hypothesized that *H19N07.3* might underlie the bleomycin-response QTL. The *H19N07.3* eQTL explains 45.70% of the variation in *H19N07.3* expression among the RIALs (**Figure 2-14A, Figure 2-14B**). The length of animals and expression of *scb-1* was correlated in the RIAL strains (**Figure 2-14C**,  $r^2 = 0.61$ ,  $p < 9.5e-13$  Spearman's correlation). Although this gene does not have a non-synonymous variant between the N2 and CB4856 strains, natural variation in gene expression could impact bleomycin responses. We note that expression variation was measured for a different panel of RIALs (set 1 RIALs in the *linkagemapping* package [110]) than the linkage mapping shown in **Figure 2-3** (mapped with set 2 RIALs [109]). However, linkage mapping of bleomycin-response variation maps to the same location using the set 1 panel of RIALs [110] as the mapping using set 2 RIALs [109] (**Figure 2-15**), suggesting that the same variant underlies bleomycin-response differences in both strain sets.



**Figure 2-14** *H19N07.3* expression differences map to chromosome V and correlate with bleomycin responses

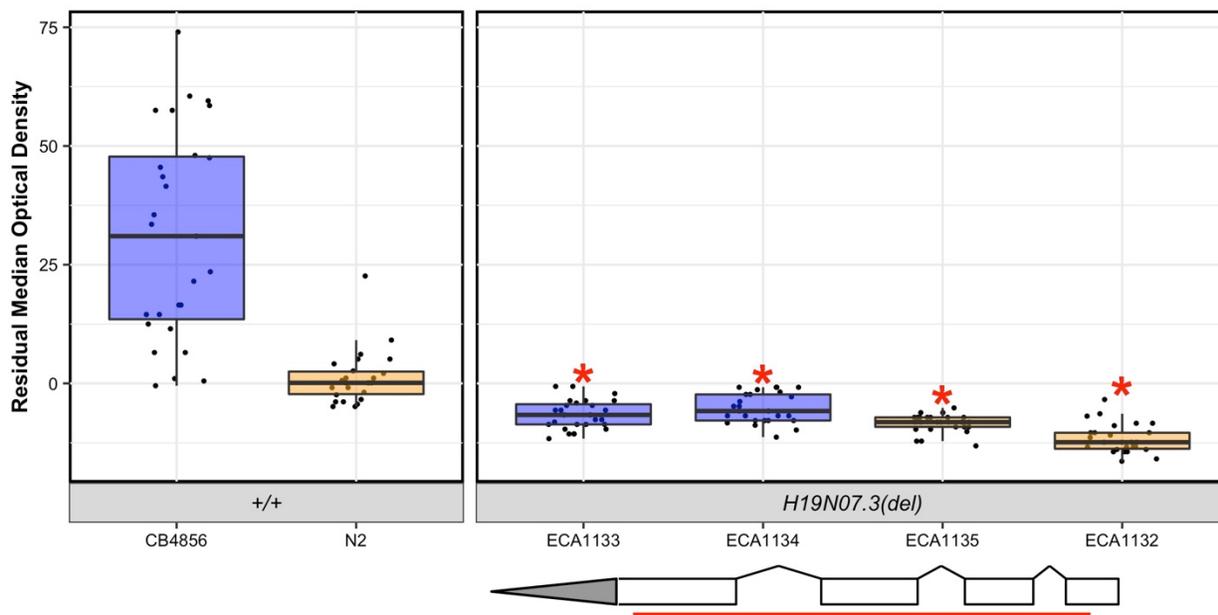
Linkage mapping of the *H19N07.3* expression difference among RIALs is shown. **A.** On the x-axis, each of 13,003 genomic markers, split by chromosome, were tested for correlation with *H19N07.3* expression variation across the RIAL panel. The log of the odds (LOD) score for each marker is reported on the y-axis. The significant quantitative trait locus (QTL) is indicated by a red triangle at the peak marker, and a blue ribbon shows the 95% confidence interval around the peak marker. The total amount of expression variance across the RIAL panel explained by the genotype at the peak marker is printed as a percentage. **B.** RIAL gene expression (y-axis), split by allele at the QTL peak marker (x-axis) is shown. Phenotypes of RIALs containing the N2 allele (orange) are compared to RIALs containing the CB4856 allele (blue). Phenotypes are shown as Tukey box plots, and each point is the *H19N07.3* expression of an individual strain. **C.** The correlation between animal size in bleomycin and *H19N07.3* expression is shown as a scatterplot, with each RIAL shown as a point. Each axis was scaled to have a mean of zero and a standard deviation of one. The line of best fit is shown in blue. The identity line is shown in black for reference.



**Figure 2-15 Linkage mapping of bleomycin-response differences among set 1 RIALs maps to the same location as set 2 RIALs**

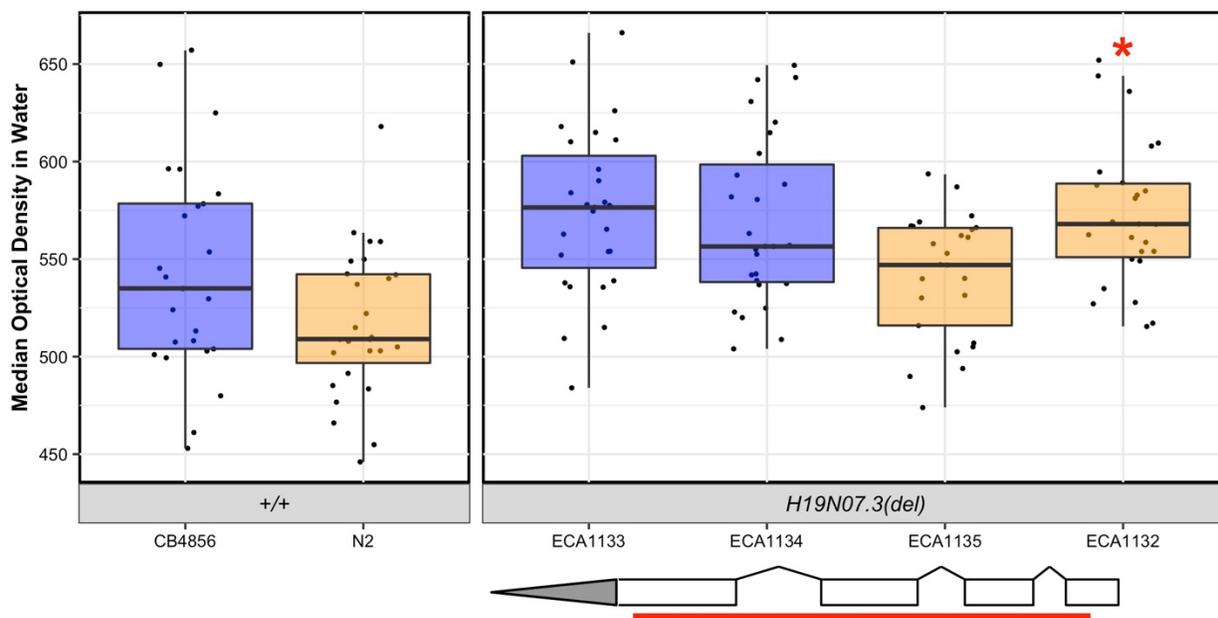
Bleomycin responses were measured for set 1 RIALs [110] and linkage mapping was performed to identify QTL that underlie drug-response differences. On the x-axis, each of 13,003 genomic markers, split by chromosome, were tested for correlation with phenotypic variation across the RIAL panel. The log of the odds (LOD) score for each marker is reported on the y-axis. Each significant quantitative trait locus (QTL) is indicated by a red triangle at the peak marker, and a blue ribbon shows the 95% confidence interval around the peak marker. The total amount of phenotypic variance across the RIAL panel explained by the genotype at each peak marker is shown as a percentage.

We created two independent CRISPR-Cas9 mediated deletion alleles of *H19N07.3* in the N2 and the CB4856 backgrounds and measured the bleomycin responses of these strains compared to the parental strains (**Figure 2-16, Table S2-2, Materials and Methods**). Each *H19N07.3* deletion strain was more sensitive to bleomycin than the respective parental strain (**Figure 2-16, ECA1133 vs. CB4856**  $p < 1.4e-14$ , **ECA1134 vs. CB4856**,  $p < 1.4e-14$ , **ECA1132 vs. N2**,  $p = 6.9e-5$ , **ECA1135 vs. N2**,  $p = 0.006$ , Tukey HSD). These results suggest that *H19N07.3* function is required for resistance to bleomycin. Therefore, we renamed this gene *scb-1* for sensitivity to the chemotherapeutic bleomycin. Unlike with the *jmjd-5* deletion strains, the *scb-1* deletion strains had no significant differences in the control condition (**Figure 2-17**). Therefore, the bleomycin sensitivity of the *scb-1* deletion strains were not caused by control-condition phenotypes.



**Figure 2-16 Strains with *H19N07.3* deletion alleles are sensitive to bleomycin**

Bleomycin responses of *H19N07.3* deletion alleles are shown as Tukey box plots, with the strain name on the x-axis, split by genotype, and residual median optical density on the y-axis. Each point is a biological replicate. Strains are colored by their background genotype (orange indicates an N2 genetic background, and blue indicates a CB4856 genetic background). Two independent deletion alleles in each genetic background were created and tested. Red asterisks indicate a significant difference ( $p < 0.05$ , Tukey HSD) between a strain with a deletion and the parental strain that has the same genetic background. A depiction of the deletion allele is shown below the box plots. White rectangles indicate exons, and diagonal lines indicate introns. The 5' and 3' UTRs are shown by grey rectangles and triangles, respectively. The region of the gene that was deleted by CRISPR-Cas9 directed genome editing is shown as a red bar beneath the gene model.

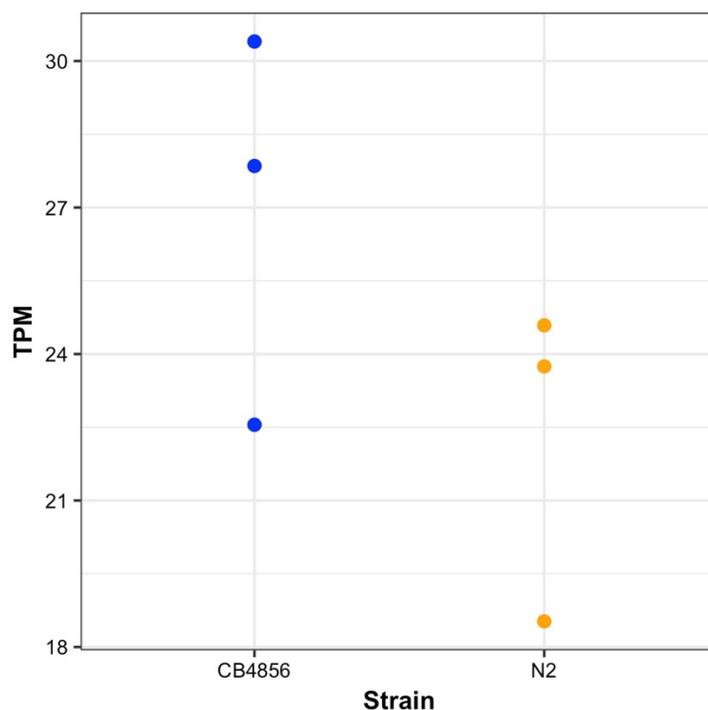


**Figure 2-17 Strains with *H19N07.3* deletion alleles are not sensitive in the control condition**

Control-condition phenotypes of deletion alleles *H19N07.3* are shown. Deletion alleles for *H19N07.3* were tested in response to the control condition (lysate in K medium with 1% distilled water). Control-condition responses are shown as Tukey box plots, with the strain name on the x-axis and median optical density on the y-axis. Each point is a biological replicate. Strains are colored by their background genotype (orange indicates an N2 genetic background, and blue indicates a CB4856 genetic background). Two independent deletion alleles in each background were created and tested. Red asterisks indicate a significant difference ( $p < 0.05$ , Tukey HSD) between a strain with a deletion and the parental strain that has the same genetic background. Depictions of the deletion allele is shown below the phenotype for each candidate gene. White rectangles indicate exons and diagonal lines indicate introns. The 5' and 3' UTRs are shown by grey rectangles and triangles, respectively. The region of the gene that was deleted by CRISPR-Cas9 directed genome editing is shown as a red bar beneath the gene model.

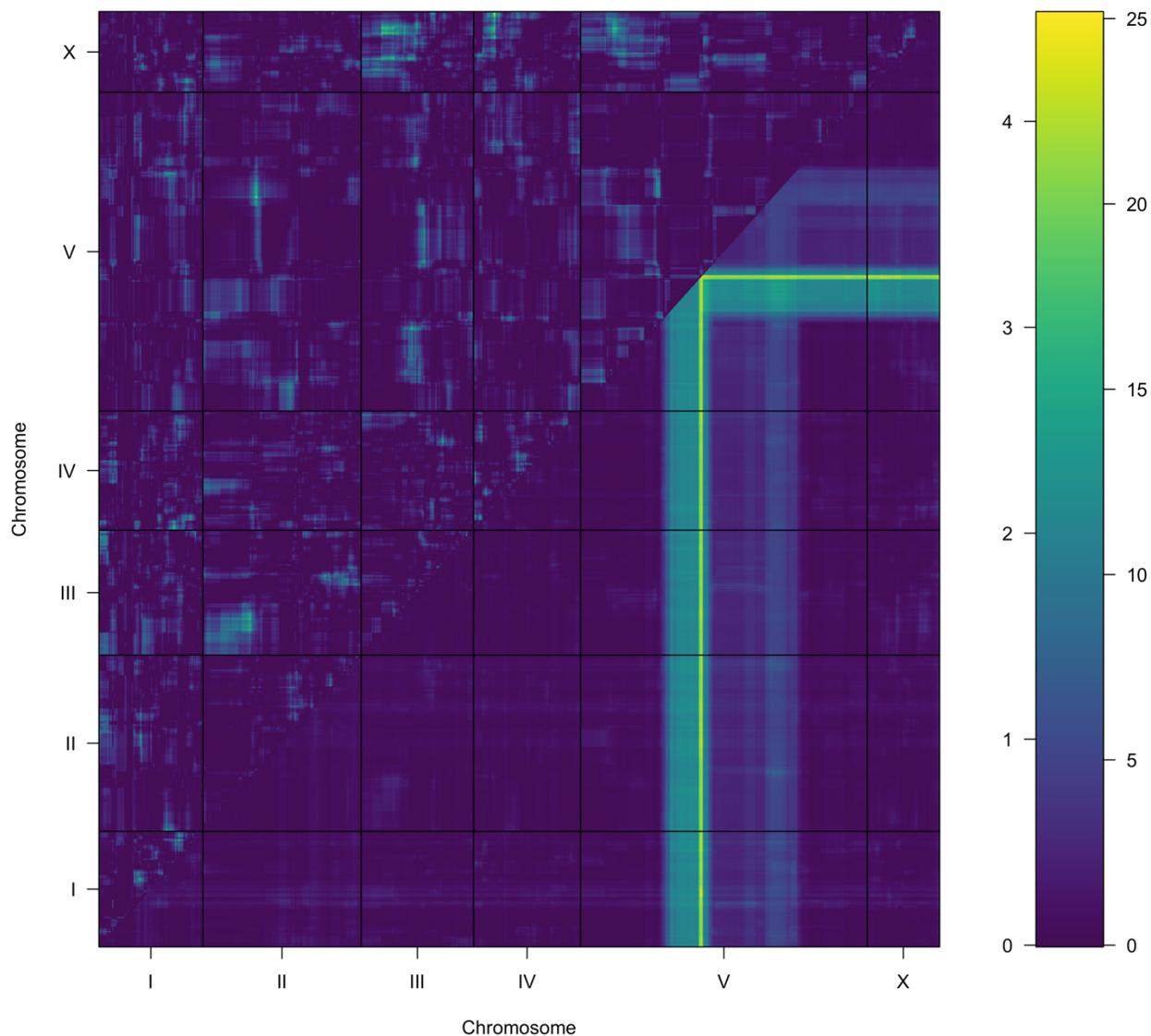
Because an *scb-1* non-synonymous variant has not been identified between the N2 and CB4856 strains, changes to protein function likely do not cause bleomycin response differences. RIALs with the CB4856 allele at the QTL peak marker have increased expression of *scb-1* and increased bleomycin resistance compared to RIALs with the N2 allele (**Figure 2-4, Figure 2-14**). Therefore, *scb-1* expression differences might cause the bleomycin-response variation between the parental strains. We performed RNA-seq of the N2 and CB4856 strains to assess *scb-1* expression differences between the parental strains and did not identify a significant increase in expression

in the CB4856 strain (**Figure 2-18**,  $p = 0.20$ , Wald test;  $p = 0.17$ , likelihood ratio test). This result could be caused by the low sample size ( $n = 3$ ) in the RNA-seq experiment, or the RIAIL strains could have a novel variant that arose during strain construction that causes *scb-1* expression variation. Alternatively, the expression difference observed in the RIAIL strains could be attributed to epistatic loci. We performed a two-factor genome scan to identify epistatic loci that underlie *scb-1* expression variation in the RIAILs and identified two significant interactions: one between loci on chromosomes IV and X and another between loci on chromosomes II and V (**Figure 2-19**). This result might suggest that epistatic loci underlie *scb-1* expression variation in the RIAILs and could explain why *scb-1* expression is not variable in the parental strains.



**Figure 2-18 RNA-seq does not show *scb-1* expression differences in the parental strains**

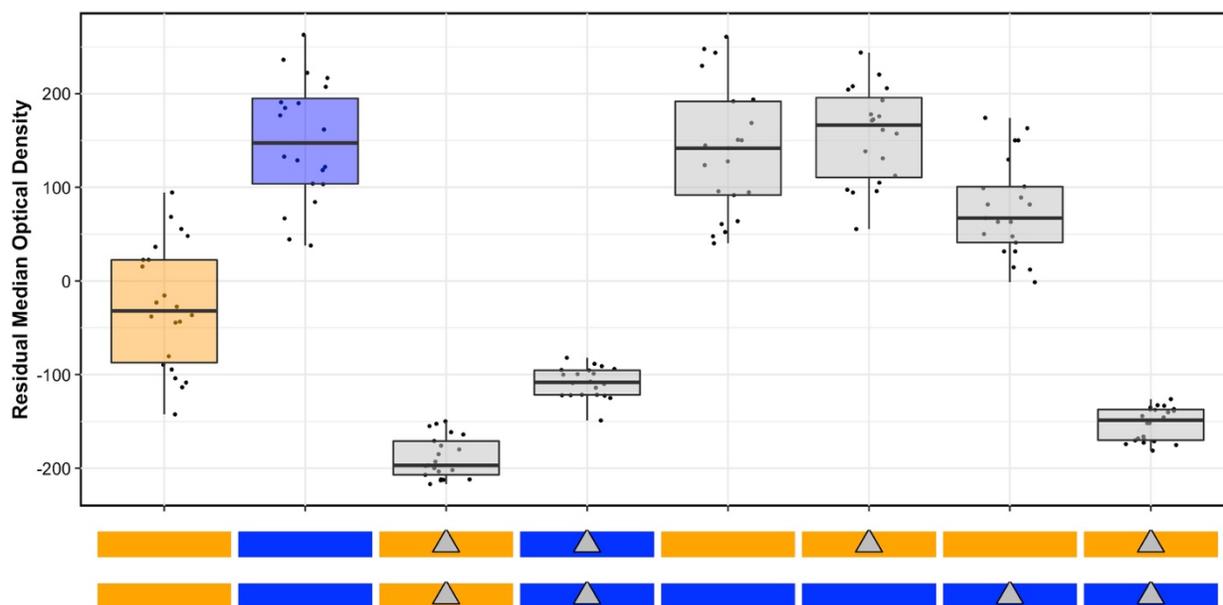
RNA-seq measurements of *scb-1* expression for young adult populations of N2 and CB4856 are reported. The x-axis indicates the sample (N2 or CB4856), and the y-axis shows the transcripts per million (TPM) estimate for each replicate. Replicates are plotted as dots, colored by strain (orange for N2 and blue for CB4856).



**Figure 2-19 A two-factor genome scan for *scb-1* expression variation detects two significant genetic interactions**

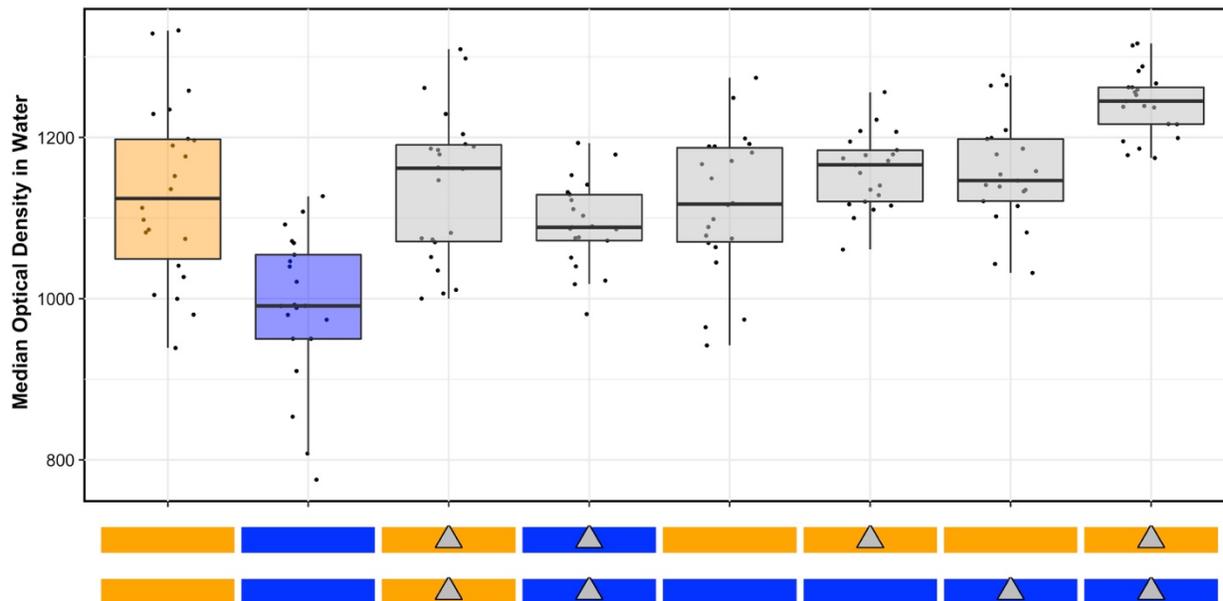
A two-factor genome scan for *scb-1* expression variation is shown. Log of the odds (LOD) scores are shown for each pairwise combination of loci, split by chromosome. The upper-left triangle contains the epistasis LOD scores, and the lower-right triangle contains LOD scores for the full model. LOD scores are shown as colors, with lower scores in blue and higher scores in yellow, as shown in the color scale. The epistatic LOD score axis is on the left of the color scale and the full LOD score axis is on the right of the color scale. The interactions between loci on chromosome II and V as well as between chromosomes IV and X are significant.

To test the role of natural variation in *scb-1* function, we performed a reciprocal hemizyosity test in bleomycin and control conditions (**Figure 2-20, Figure 2-21**). These results matched the increase in sensitivity of homozygous deletions in both parental backgrounds observed previously. The hemizygous strain with the CB4856 allele of *scb-1* had a bleomycin phenotype similar to the heterozygous strain, whereas the hemizygous strain with the N2 allele of *scb-1* was more sensitive to bleomycin than the heterozygous strain. Taken together, these results suggest that natural variation in *scb-1* function underlies the bleomycin-response difference between the N2 and CB4856 strains.



**Figure 2-20 Natural variation in *scb-1* underlies bleomycin-response differences**

Results of the *scb-1* reciprocal hemizyosity assay are shown. The y-axis shows the residual median optical density for each strain reported along the x-axis. Bleomycin responses are reported as Tukey box plots where each point is a biological replicate. The genotypes of each strain are shown as colored rectangles beneath each box plot, where each rectangle represents a homolog (orange rectangles are an N2 genotype, and blue rectangles are a CB4856 genotype). Maternal homologs are shown on top and paternal homologs are shown on bottom. Grey triangles indicate a deletion of *scb-1*, placed on the rectangle showing the background into which the deletion was introduced. The box plots for the parental strains (N2 and CB4856, on the left) are colored according to genotype.



**Figure 2-21 Strains with *scb-1* deletion alleles are not sensitive in the control condition**

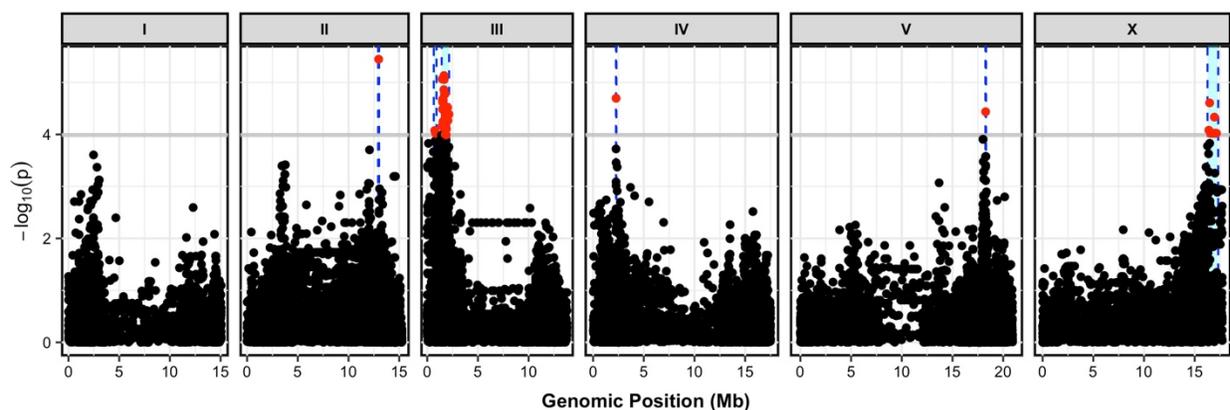
Control-condition phenotypes of the *scb-1* reciprocal hemizygosity assay are shown. The y-axis shows the median optical density in the control condition (K media and 1% water) for each strain reported along the x-axis. Control-condition responses are reported as Tukey box plots where each point is a biological replicate. The genotypes of each strain are shown as colored rectangles beneath each box plot, where each rectangle represents a homolog (orange rectangles are an N2 genotype, and blue rectangles are a CB4856 genotype). Maternal homologs are shown on top and paternal homologs are shown on bottom. Grey triangles indicate a deletion of *scb-1*, placed on the rectangle showing the background into which the deletion was introduced. The box plots for the parental strains (N2 and CB4856, on the left) are colored according to genotype.

## Differences in *scb-1* function might be regulated by a rare variant

The *scb-1* natural variant that underlies the bleomycin-response differences remains unknown.

Because this gene does not have a predicted non-synonymous variant between the N2 and CB4856 strains, *scb-1* gene expression might underlie bleomycin response differences. Potential candidate variants that could cause this expression difference include one variant two kilobases upstream of the gene and one variant in the third intron of *scb-1*. However, gene expression can be regulated by distant loci, so the identification of the specific variant is difficult. To understand whether natural variation of *scb-1* underlies bleomycin-response differences in other strains, we

compared the bleomycin-response linkage mapping to a genome-wide association mapping (GWA). We used the HTA to measure median optical density in bleomycin for 83 divergent wild isolates and performed GWA mapping (**Figure 2-22**). Six QTL were identified from the GWA, but none of these QTL regions overlapped the QTL from linkage mapping (**Figure 2-22**). Therefore, the CB4856 strain might have a rare variant that underlies its increase in bleomycin resistance compared to the N2 strain.

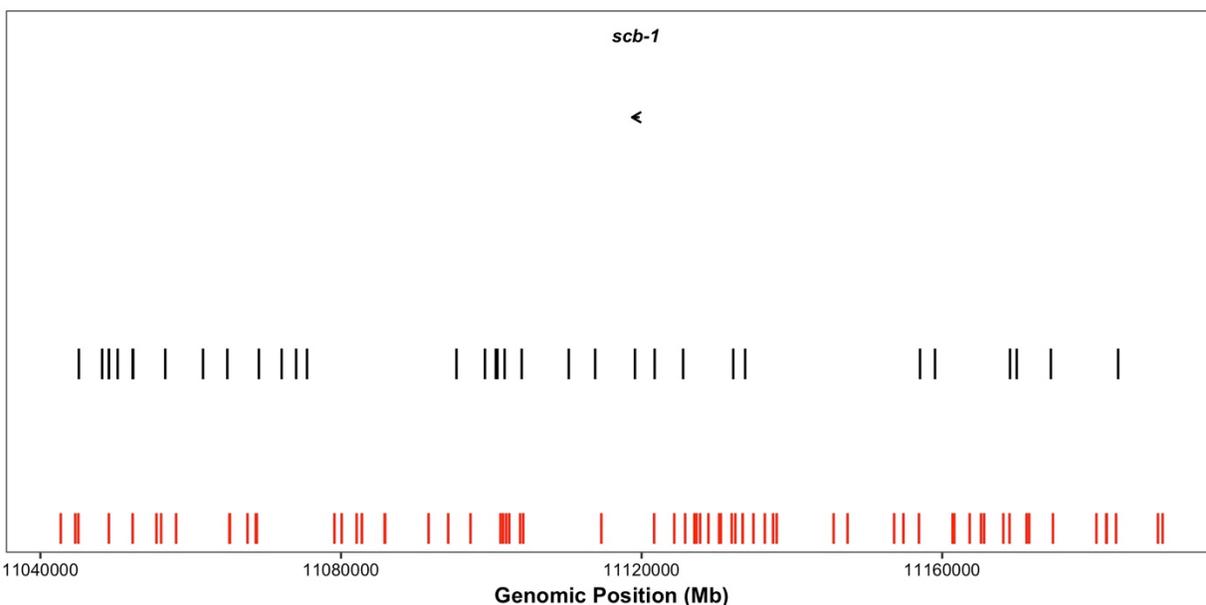


**Figure 2-22 A genome-wide association mapping for bleomycin-response variation detects six QTL**

A genome-wide association study for residual median optical density in bleomycin for 83 wild isolates is shown. On the x-axis, each genomic marker, split by chromosome, was tested for correlation with phenotypic variation across the wild isolates. The  $\log_{10}(p)$  value of these correlations are reported on the y-axis. Each marker that reached a significance threshold determined by eigenvalue decomposition of the SNP correlation matrix is colored in red. QTL regions of interest are indicated by blue regions surrounding the significant markers.

We identified all single nucleotide variants (SNVs), small insertion/deletions (indels), and structural variants (SVs) present in these 83 strains for which the CB4856 strain contains the alternate allele compared to the N2 reference strain. We found 105 variants within the QTL confidence interval (79 SNVs, 26 indels, 0 SVs) for which the CB4856 strain contains the alternate allele (**Figure 2-23**). We then identified SNVs and indels with a minor-allele frequency less than 5% within the 83 strains, because these low-frequency variants are likely to have insufficient

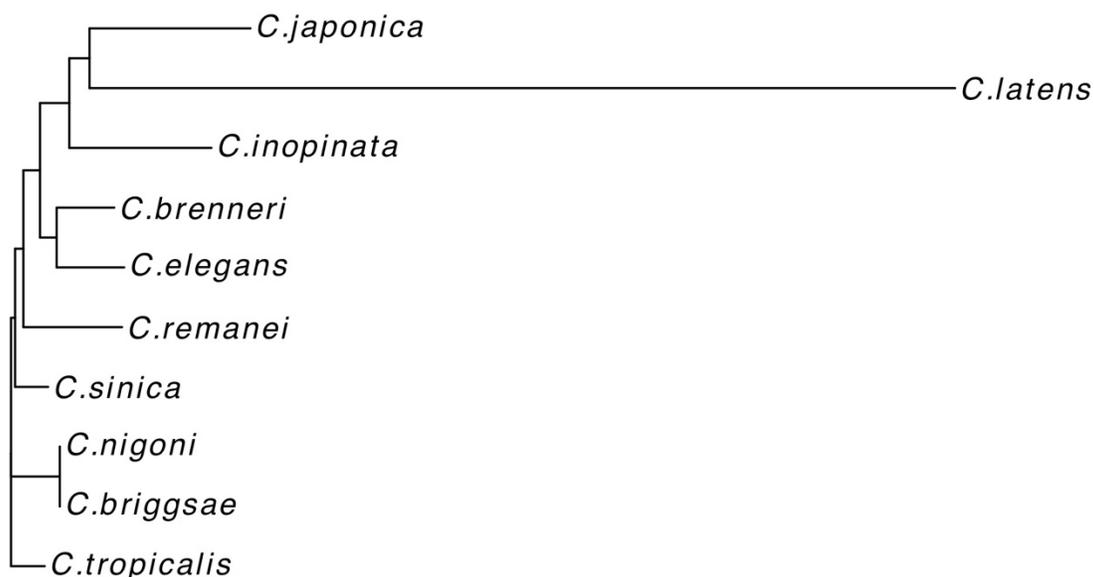
power to map by GWA. Seventy-two of the 105 variants in the region were identified as rare variants that might underlie the bleomycin-response difference between the N2 and CB4856 strains (**Figure 2-23, Figure S2-2**). Twenty-eight of these rare variants were not unique to CB4856, and other strains in the wild isolate panel shared these variants. However, none of these variants showed phenotypic trends consistent with an alternate allele conferring resistance to bleomycin (**Figure S2-2**). Forty-four of the 72 rare variants were unique to CB4856 within this set of 83 strains. One or more of these 44 variants could underlie the bleomycin-response QTL, but further work must be performed to identify which, if any, of these variants underlies the *scb-1* bleomycin-response difference between N2 and CB4856.



**Figure 2-23 Rare and common variants are present near the *scb-1* gene**

All variants in the QTL confidence interval for which CB4856 contains the alternate allele are plotted as vertical lines. On the x-axis, genomic position of the variant is shown. Variants are colored by minor allele frequency (less than 0.05 = red, greater than 0.05 = black). The location of *scb-1* is shown as an arrow, labeled with the gene name.

We searched for homologs of *scb-1* in other species using a BLASTp search ([www.wormbase.org](http://www.wormbase.org), Release WS268) and identified homologs in nine other *Caenorhabditis* species but none outside of the Nematoda phylum (**Figure 2-24**) [206,207]. None of the homologs of SCB-1 have previously identified functions. We used Phyre2 to predict protein domains within the SCB-1 protein and were unable to detect any functional domains by sequence homology. Twenty-three percent of the SCB-1 protein sequence matched a hydrolase of a Middle East respiratory syndrome-related coronavirus [209]. However, the low confidence of the model (21.5%) should be considered before making conclusions about the function of *scb-1* based on these results.



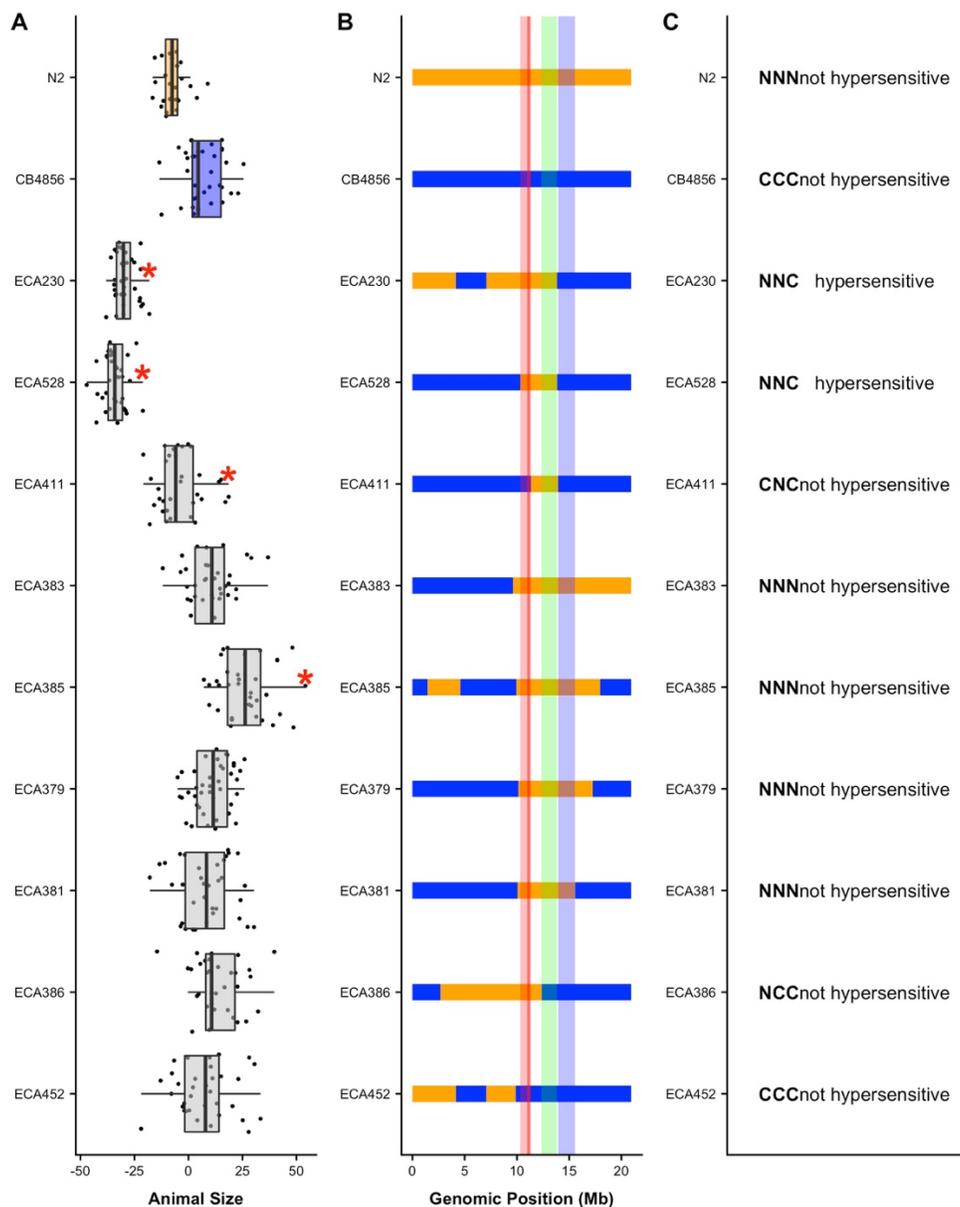
**Figure 2-24 SCB-1 is conserved across nematodes**

A neighbor-joining tree from a multiple-sequence alignment of SCB-1 homologs is shown.

## Evidence of unidirectional additive or epistatic effects on bleomycin response variation

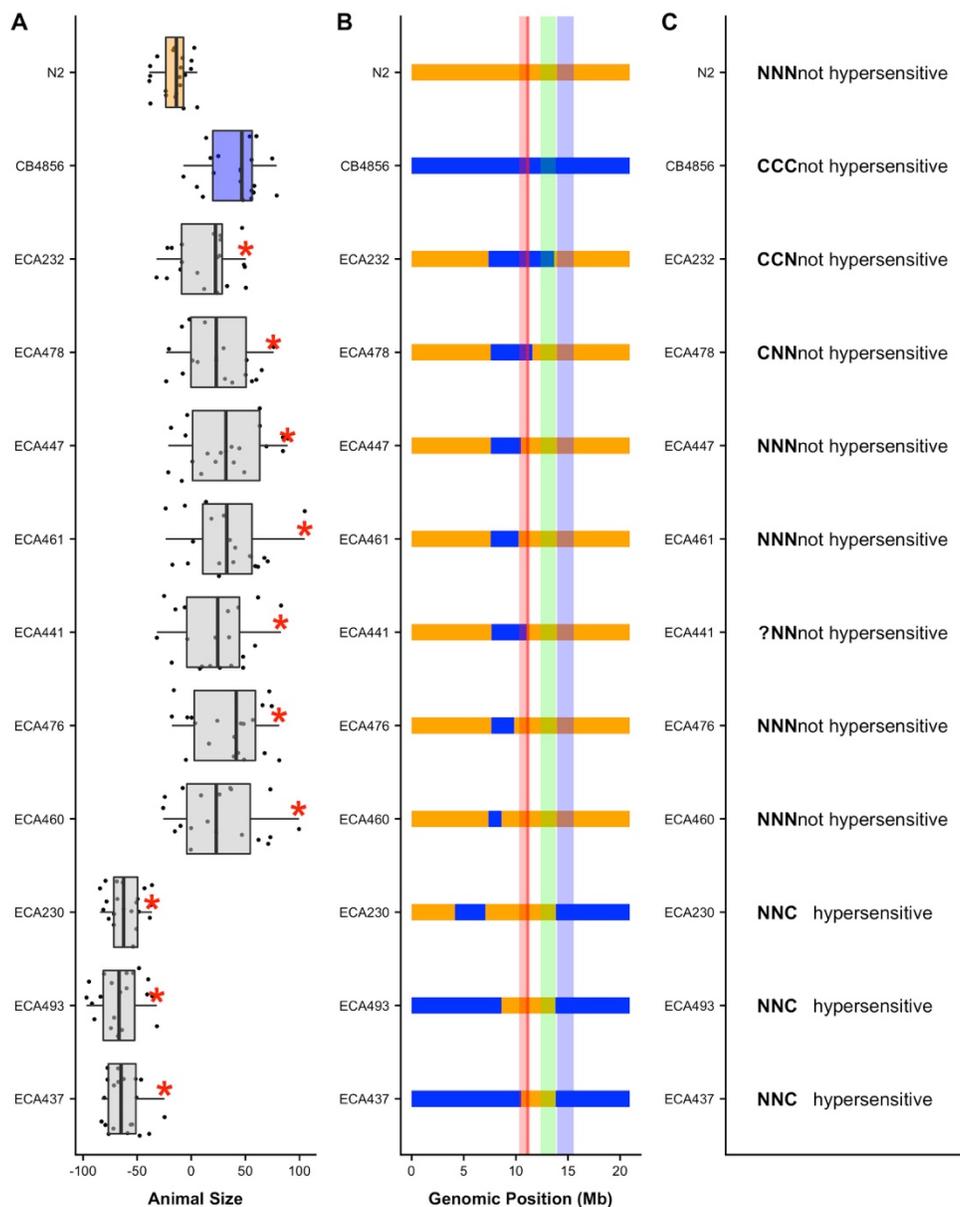
Although *scb-1* was identified as the causal gene that underlies the QTL on chromosome V, the effect on this chromosome likely is not completely explained by a single *scb-1* variant. The chromosome V NILs, ECA230 and ECA232, showed transgressive phenotypes in bleomycin. In particular, ECA230, which has the N2 QTL region introgressed into a CB4856 genetic background, is much more sensitive to bleomycin than the N2 parental strain. Given that this strain has the CB4856, resistant allele at each of the other three QTL identified with linkage mapping, we predicted that ECA230 would be more resistant to bleomycin than the N2 strain. One possible explanation for this phenomenon is that two or more loci act additively or interact in this NIL to cause this hypersensitivity phenotype. Because the transgressive phenotype is only seen in one direction (ECA230 is hypersensitive, but ECA232 is not hyper-resistant), epistasis is likely causing this phenomenon. We hypothesized that the N2 allele at one locus is incompatible with a CB4856 allele at another locus and sought to identify these interaction partners.

We generated a panel of NILs to test this hypothesis and to identify regions of the genome that might be candidates for such interacting loci. **Figures 2-25, 2-26, and 2-27** show the phenotypes of the strains in these experiments. We identified a minimum of three loci that interact to cause bleomycin hypersensitivity in the NILs. The locations of these regions, as well as the strains that define them, are described below.



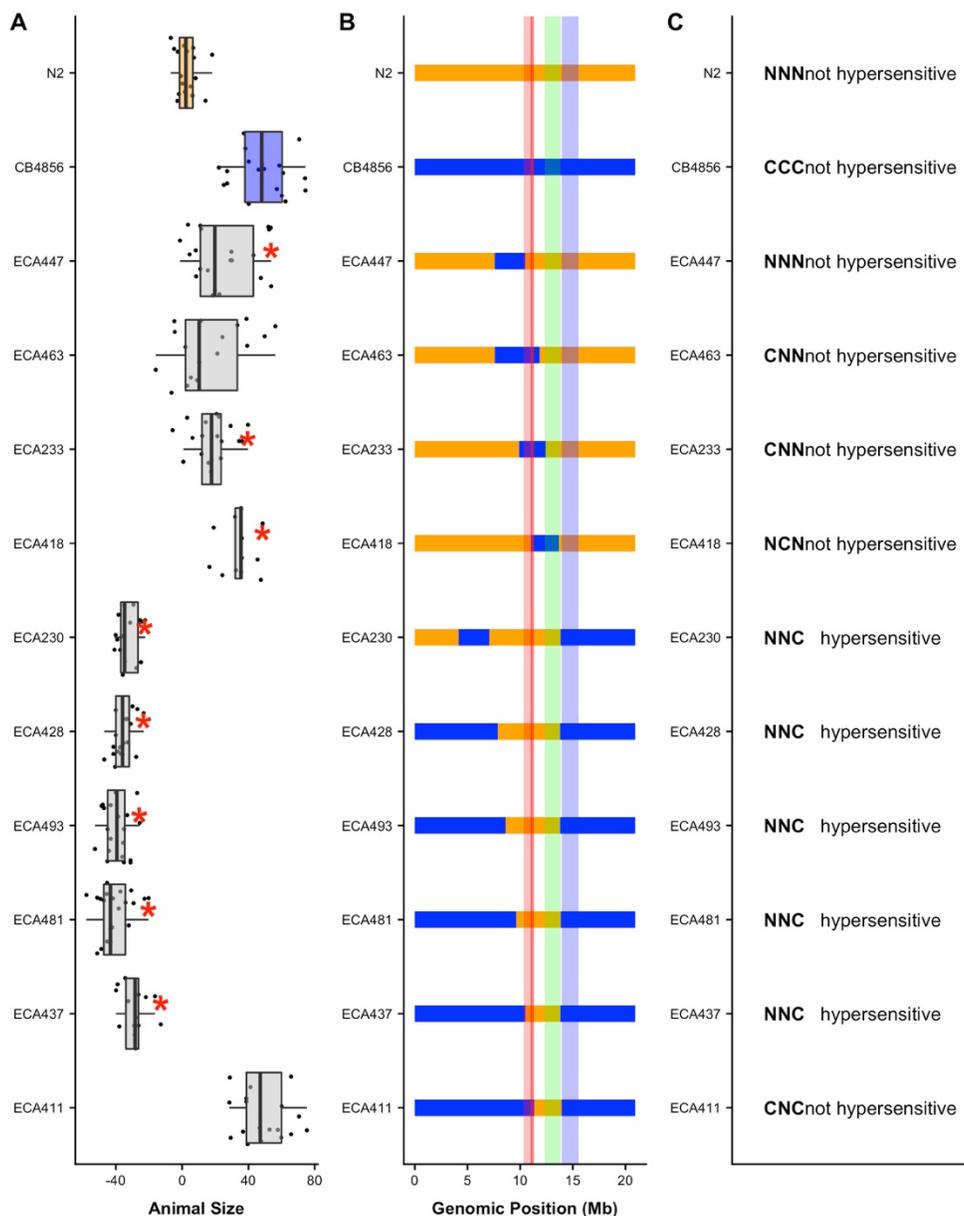
**Figure 2-25 Near-isogenic lines suggest that three loci interact to cause bleomycin hypersensitivity**

Phenotypes and genotypes of near-isogenic lines (NILs) are shown. **A**. Phenotypes for each strain are shown as Tukey box plots, with strain name on the y-axis and residual bleomycin median optical density on the x-axis. Each point is a biological replicate. Parental strain box plots are colored by their genetic background, with orange indicating an N2 background and blue indicating a CB4856 genetic background. NILs are shown as grey box plots. A red asterisk indicates a significant difference between the phenotype of a given strain and the phenotype of the corresponding parental strain ( $p < 0.05$ , Tukey HSD). **B**. Chromosomal representations of chromosome V are shown for each of the strains in **A**. Strain names are reported on the y-axis, and genomic position (Mb) is shown on the x-axis. Blocks of color indicate genotypes of genomic regions with orange indicating the N2 genotype and blue indicating the CB4856 genotype. Vertical red lines mark the confidence interval of the QTL from linkage mapping. Red, green, and blue vertical regions show potential epistatic loci. **C**. The allele combination of each strain is shown for the red, green, and blue regions. Phenotypic predictions based on this combination of alleles are written beside each genotype.



**Figure 2-26 Near-isogenic lines suggest that three loci interact to cause bleomycin hypersensitivity**

Phenotypes and genotypes of near-isogenic lines (NILs) are shown. **A**. Phenotypes for each strain are shown as Tukey box plots, with strain name on the y-axis and residual bleomycin median optical density on the x-axis. Each point is a biological replicate. Parental strain box plots are colored by their genetic background, with orange indicating an N2 background and blue indicating a CB4856 genetic background. NILs are shown as grey box plots. A red asterisk indicates a significant difference between the phenotype of a given strain and the phenotype of the corresponding parental strain ( $p < 0.05$ , Tukey HSD). **B**. Chromosomal representations of chromosome V are shown for each of the strains in **A**. Strain names are reported on the y-axis, and genomic position (Mb) is shown on the x-axis. Blocks of color indicate genotypes of genomic regions with orange indicating the N2 genotype and blue indicating the CB4856 genotype. Vertical red lines mark the confidence interval of the QTL from linkage mapping. Red, green, and blue vertical regions show potential epistatic loci. **C**. The allele combination of each strain is shown for the red, green, and blue regions. Phenotypic predictions based on this combination of alleles are written beside each genotype.



**Figure 2-27 Near-isogenic lines suggest that three loci interact to cause bleomycin hypersensitivity**

Phenotypes and genotypes of near-isogenic lines (NILs) are shown. **A**. Phenotypes for each strain are shown as Tukey box plots, with strain name on the y-axis and residual bleomycin median optical density on the x-axis. Each point is a biological replicate. Parental strain box plots are colored by their genetic background, with orange indicating an N2 background and blue indicating a CB4856 genetic background. NILs are shown as grey box plots. A red asterisk indicates a significant difference between the phenotype of a given strain and the phenotype of the corresponding parental strain ( $p < 0.05$ , Tukey HSD). **B**. Chromosomal representations of chromosome V are shown for each of the strains in **A**. Strain names are reported on the y-axis, and genomic position (Mb) is shown on the x-axis. Blocks of color indicate genotypes of genomic regions with orange indicating the N2 genotype and blue indicating the CB4856 genotype. Vertical red lines mark the confidence interval of the QTL from linkage mapping. Red, green, and blue vertical regions show potential epistatic loci. **C**. The allele combination of each strain is shown for the red, green, and blue regions. Phenotypic predictions based on this combination of alleles are written beside each genotype.

In **Figure 2-25**, ECA230 and ECA528 are both hypersensitive to bleomycin. These strains share particular genomic regions that are candidates for epistatic loci. First, these strains share the N2 allele at the QTL peak marker as well as the region surrounding the QTL from V:10,339,728 to V:13,851,058. Additionally they share the CB4856 genotype from V:4,165,561 to V:7,082,838, from V:13851058 to the end of chromosome V, and on all other chromosomes. We hypothesized that a locus in their shared N2 region interacts with a locus in one of their shared CB4856 regions to cause their hypersensitivity to bleomycin.

The ECA411 NIL shares much of its genome with ECA528, but ECA411 is not hypersensitive. Therefore, a genomic difference between ECA411 and ECA528 can localize one of the epistatic regions that causes ECA528 and ECA230 to be hypersensitive. ECA411 and ECA528 differ at the QTL peak marker and a region surrounding it from V:10,339,728 to V:11,345,444. At this region, ECA528 and ECA230 share the N2 genotype, whereas ECA411 contains the genotype across this region. Following this logic, we hypothesized that an N2 allele in this region (highlighted in red in **Figure 2-25**, **Figure 2-26**, **Figure 2-27**) interacts with a CB4856 allele at another region. Because ECA411 does not have the N2 allele in the red region, the interaction does not occur, and this strain is not hypersensitive.

With the N2 interaction partner localized to the red area around the QTL marker, we sought to localize the CB4856 interaction partner that ECA230 and ECA528 share. ECA381 has the N2 genotype across the red region and the CB4856 genotype across the majority of its genome, but ECA381 does not show the hypersensitivity to bleomycin that ECA230 and ECA528 share. Although ECA381 shares the N2 interaction partner with the hypersensitive strains, we hypothesized that ECA381 must not have the CB4856 interaction partner that ECA230 and

ECA528 share. With this logic, we were able to eliminate any CB4856 regions that are shared between ECA381 and ECA230/ECA528. Therefore, the CB4856 interacting partner could be localized to between V:13,851,058 and V:15,562,989 (highlighted in blue in **Figure 2-25, Figure 2-26, Figure 2-27**), where ECA381 contains the N2 genotype whereas ECA230 and ECA528 share the CB4856 genotype. In summary, this result determined that an N2 allele in the red region could interact with a CB4856 allele in the blue region to cause hypersensitivity to bleomycin.

The N2-in-red/CB4856-in-blue interaction model holds true for several strains (ECA230, ECA528, ECA411, ECA383, ECA385, ECA379, and ECA381), but ECA386 did not fit this model. ECA386 has the N2 genotype across the red region and the CB4856 genotype across the blue region and would therefore be predicted to have a hypersensitive phenotype. The fact that this strain is not hypersensitive indicates that the model must be more complex. We hypothesized that a third interaction partner exists in a region that ECA230 and ECA528 share but ECA386 does not. This third region could be between V:12,366,450 and V:13,943,824 (highlighted in green in **Figure 2-25, Figure 2-26, Figure 2-27**) where ECA386 has the CB4856 genotype whereas ECA230 and ECA528 share the N2 genotype. This model would predict that the combination of an N2 allele in the red region, an N2 allele in the green region, and a CB4856 allele in the blue region interact to cause hypersensitivity to bleomycin. Alternatively, the third interaction partner could lie between V:4,165,561 and V:7,082,838, where ECA230 and ECA528 share the CB4856 genotype whereas ECA386 has the N2 genotype. With this alternative model, a combination of a CB4856 allele in the V:4,165,561-7,082,838 interval, an N2 allele in the red region, and a CB4856 allele in the blue region would predict a hypersensitive phenotype. Currently, we are unable to differentiate between these two models.

**Figures 2-25, 2-26, and 2-27** show the phenotypes and genotypes of NILs. Potential locations of the three interaction regions are shown as colored rectangles on the NIL genotypes. To the right, the allele that each strain has across each of the three regions is written along with the phenotype that would be predicted by the model. The phenotypic predictions for this model match with the observed phenotypes for all three of these assays. The alternative model, where the green region is moved to V:4,165,561-7,082,838 also correctly predicts the observed phenotypes.

## Discussion

Here, we performed linkage mapping of bleomycin-response variation and identified a highly significant QTL on chromosome V. We tested all six candidate genes in the QTL region to identify a causal gene that underlies bleomycin-response variation between two divergent strains. Deletions of four of these genes, *C45B11.8*, *C45B11.6*, *srg-42*, and *cnc-10* did not impact bleomycin responses. Deletions in one gene, *jmjd-5*, showed increased bleomycin resistance in both parental backgrounds. However, we concluded that the QTL cannot be explained by differences in *jmjd-5* after further analysis of allele-replacement strains and hemizyosity tests. Deletions in a gene with an expression difference, *scb-1* (previously named *H19N07.3*), caused an increase in bleomycin sensitivity in both the N2 and the CB4856 strains. Results from a reciprocal hemizyosity assay indicated that natural variation in *scb-1* function caused differences in bleomycin responses between the N2 and CB4856 strains. Because loss-of-function alleles in *scb-1* caused increased bleomycin sensitivity (**Figure 2-16**) and the RIALs with lower *scb-1* expression levels show increased bleomycin sensitivity (**Figure 2-14**), natural differences in *scb-1* expression might cause the bleomycin-response variation between the N2 and CB4856 strains.

The function of *scb-1*, and particularly how it regulates bleomycin response, remains unknown. A previous study found that RNAi of *scb-1* impaired the DAF-16/FOXO-induced lifespan extension of *daf-2(e1370ts)* mutants, which suggests that *scb-1* might play a role in stress response [210]. Because bleomycin causes double-stranded DNA breaks and introduces oxidative stress to cells [211], reduction of *scb-1* function might inhibit the ability of an animal to respond to bleomycin. This model is in agreement with our observation that *scb-1* deletions and RIAILs with lower *scb-1* expression are sensitive to bleomycin (**Figure 2-16, Figure 2-14**). We used the amino acid sequence of SCB-1 to query the Phyre2 database and found weak homology to a viral hydrolase [209,212]. This result could suggest that SCB-1 might function as a hydrolase, which could be the mechanism by which *scb-1* regulates cellular stress. This finding would be similar to clinical studies that have suggested a role of bleomycin hydrolase (BLMH) in bleomycin-response variation [171,174–177]. Because *scb-1* is expressed in the nucleus of all somatic cells, this gene might impact the ability of bleomycin to cause DNA damage within the cell nucleus [213]. Alternatively, *scb-1* could impact bleomycin import, export, or another mechanism. If the mechanism of *scb-1* is conserved in humans, this discovery could offer insights into the clinical applications of bleomycin. Our results also suggested the presence of genes that interact with *scb-1* to cause bleomycin-response differences. These interacting genes could be conserved in humans and therefore inform the use of bleomycin in the clinic.

Despite its lack of conservation in humans, the SCB-1 protein is homologous to other proteins in other nematode species. Bleomycin is produced by the soil bacterium, *Streptomyces verticillus* [214–216], which might be found in association with nematodes such as *C. elegans* in the wild [217]. A shared niche between *C. elegans* and *S. verticillus* could cause bleomycin resistance to be selected. Additionally, the CB4856 wild isolate is more resistant to bleomycin than the

laboratory-adapted strain, N2. In fact, the N2 strain is the most sensitive to bleomycin across all strains tested in our HTA (**Figure S2-2**), which could indicate that bleomycin resistance is beneficial for wild isolates. Given its potential role in the highly conserved insulin-like pathway, *scb-1* could be beneficial in responses to multiple toxins. Interestingly, the *scb-1* gene lies within a toxin-response QTL hotspot on chromosome V [41]. Understanding the mechanism of the role of *scb-1* in toxin responses might offer insights into evolutionary processes that shaped the genomic diversity of *C. elegans* and other nematode species.

Previous studies have leveraged both linkage mapping and GWA in *C. elegans* to identify genetic variants that underlie drug-response differences [40,134]. In each of these studies, drug-response QTL overlap between linkage mapping and GWA, and variants in common between both mapping strain sets have been shown to underlie drug-response QTL. In the case of the bleomycin response, the linkage-mapping QTL did not overlap with the QTL identified through GWA. Therefore, the variant that underlies the QTL likely is not present at an allele frequency above 5% in the panel of wild isolates used for the bleomycin GWA. The difference between linkage mapping and GWA results indicates that both rare and common natural variants underlie bleomycin-response variation.

This study emphasizes the power of the *C. elegans* model system to dissect complex traits. Although linkage mapping detected a highly significant QTL, the manner in which genetic components affect bleomycin responses is not simple. Certain near-isogenic lines showed transgressive phenotypes (**Figure 2-5, Figure 2-7, Figure 2-25, Figure 2-26, Figure 2-27**), which indicates that multiple loci interact to create extreme bleomycin sensitivity in particular strains with mixed genetic backgrounds. Our attempts to identify epistatic loci in the RIALs by performing a

two-factor genome scan were unsuccessful (**Figure 2-6**), potentially because of the complexity of these epistatic interactions. We were able to identify particular regions of epistasis empirically (**Figure 2-25, Figure 2-26, Figure 2-27**), but the precise interacting loci remain elusive. Despite this complexity, *scb-1* deletions showed increased bleomycin sensitivity in both parental backgrounds and expression variation among the RIAL panel mapped to the same locus as the bleomycin response QTL. Interestingly, the parental strains do not seem to vary in *scb-1* expression, as measured by RNA-seq (**Figure 2-18**). Additional complexities of this trait include the lack of overlap between GWA and linkage mapping QTL and the potential effect of *jmjd-5* loss-of-function on bleomycin responses. Despite the complicated manner in which genetic variants seem to affect bleomycin responses, we leveraged the powerful model of *C. elegans* to identify a single gene that underlies this complex trait.

## Future directions

Because RNA-seq did not confirm *scb-1* expression level variation between N2 and CB4856, this expression variation still needs to be validated. Although RNA-seq is effective for measuring expression of a broad range of transcripts, qPCR offers a more targeted strategy to quantify levels of specific transcripts. Given its lower cost compared to RNA-seq, we could use qPCR to measure transcript abundance differences between N2 and CB4856 for a high number of replicates. However, *scb-1* expression differences were observed in RIALs, not in the parental strains. Therefore, novel mutations in the RIALs could be driving the QTL at *scb-1*. We plan to deeply sequence ECA411 and ECA528, the two NILs that show a stark phenotypic difference in bleomycin responses (**Figure 2-7**). These NILs are both derived from RIALs, so any novel mutation that underlies the *scb-1* QTL in linkage mapping should be present in these strains.

Two-factor genome scans were performed on bleomycin-response differences and on expression differences of *scb-1*. Although the bleomycin-response analysis did not identify evidence of epistasis, interactions that underlie *scb-1* expression differences were detected. These results suggest that interactions underlie *scb-1* expression variation but do not underlie bleomycin-response differences. However, two distinct panels of RIALs were used in each of the two-factor scans (set 1 for the gene-expression scan [110] and set 2 for the bleomycin-response scan [109]). The linkage mapping of bleomycin-response variation in the set 1 RIALs was less significant than that of the set 2 RIALs, potentially because the set 1 strains were propagated at higher temperatures and for longer amounts of time before freezing, which can cause accumulation of *de novo* variants and transposon insertions. Ideally, the two-factor scan for bleomycin-response differences and for *scb-1* expression differences would be performed on set 2 strains. However, expression data are not available for the set 2 RIALs. Future work should collect expression data for this panel of RIALs, given its increased heritability in the HTA compared to set 1 RIALs.

To experimentally validate the effect of *scb-1* expression differences on bleomycin responses, we can assay strains with over-expression of *scb-1*. Plasmids that contain GFP and functional *scb-1* under a heat-sensitive promoter can be injected into N2 and CB4856 animals to over-express *scb-1*. GFP-positive animals indicate that the overexpression plasmid is present, and *scb-1* over-expression can be induced by exposing animals to high temperatures. However, over-expression might have deleterious effects on the animals and therefore might not reflect the effects of natural variation in *scb-1* expression levels on bleomycin responses.

Another goal of this project is to locate the precise variant that causes bleomycin-response variation. The *scb-1* deletion assay and the *scb-1* hemizygosity assay confirmed that natural variation in *scb-1* function affects bleomycin responses (**Figure 2-16, Figure 2-20**). However, the precise variant that modulates the *scb-1* function differences between N2 and CB4856 remains unknown. We have identified 72 candidate SNPs that might affect *scb-1* function, thereby modulating bleomycin responses. Future experiments can use CRISPR/Cas9-mediated genome editing to modify each of these candidate variants in both genetic backgrounds, determining which variant underlies the *scb-1* QTL.

Although we were able to identify locations of potential epistasis that cause bleomycin hypersensitivity, more NILs are needed to narrow these three regions. First, assays with NILs that can differentiate between the model shown in the figures versus the alternative placement of the green region should be performed. Next, each region can be narrowed by constructing NILs with smaller regions of introgression and testing them. Although we proved that variation in *scb-1* underlies the bleomycin-response QTL, we cannot determine whether it is the red-region interacting partner or not. To test the impact of *scb-1* on the epistasis that underlies bleomycin hypersensitivity, the gene must be deleted in multiple NILs.

The bleomycin-response QTL maps to the toxin-response QTL hotspot in the center of chromosome V. Many QTL map to this same region, which might suggest a pleiotropic variant that affects responses to multiple toxins. We were able to identify *scb-1* as the gene that underlies the bleomycin-response QTL, and *scb-1* could underlie variation to multiple toxins. RNAi of *scb-1* has been shown to impair the DAF-26/FOXO-induced lifespan expansion of *daf-2(e1370ts)* mutants, which might implicate *scb-1* function in stress responses. We can computationally

remove the effect of *scb-1* on linkage mappings by setting the genotype at the bleomycin QTL peak marker as a cofactor and then see if chromosome V mappings are dependent on *scb-1* variation. With this analysis, many QTL in the center of chromosome V indeed no longer reach significance (**Figure S2-3**). However, loci in the center of chromosomes are tightly linked, so whether the disappearance of the chromosome V hotspot is dependent on *scb-1* variation or another gene in the region remains unclear. Future experiments should test the responses of the *scb-1* deletion strains to toxins that map to this hotspot (carmustine, chlorothalonil, cisplatin, irinotecan, mechlorethamine, paraquat, and silver). If *scb-1* does underlie the response to multiple drugs, this finding could help define the mechanism of *scb-1*, and it would be an exciting example of a single gene that impacts multiple quantitative traits.

## Contributions

During this work, I was funded by the Biotechnology Training Program and the John N. Nicholson fellowship. Stefan Zdraljevic generated the ECA411 and ECA528 NILs that were used to confirm the QTL interval. He also generated many of the NILs used to localize the epistatic loci and helped run the experiments in **Figures 2-25, 2-26, and 2-27**. Karol Bisaga was my undergraduate mentee during this project and played a crucial role in running the lysates and PCRs of all of the deletion and allele-replacement strains in the project. Robyn Tanny and Dan Cook created libraries for sequencing and analyzed sequence data for all RIALs, NILs, and wild isolates. Daehan Lee and Ye Wang performed RNA-seq and analyzed those data. Erik Andersen oversaw the entirety of this work and offered critical advice that helped guide each step of the project. Finally, Stefan Zdraljevic, Daehan Lee, and Katie Evans acted as my sounding boards, motivators, and advisors during the successes and failures of the “low hanging fruit” project.

## Supplemental Tables

Table S2-1 Reagents used to generate NILs

<b>ECA230</b> <i>eanIR150(V</i> <i>N2&gt;CB4856)</i>	Constructed from: QX131 x CB4856	N2 into CB4856
<b>ECA232</b> <i>eanIR152(V</i> <i>CB4856&gt;N2)</i>	Constructed from: QX450 x N2	CB4856 into N2
<b>ECA411</b> <i>eanIR185(V</i> <i>N2&gt;CB4856)</i>	Constructed from: ECA230 x CB4856	N2 into CB4856
<b>ECA528</b> <i>eanIR302(V</i> <i>N2&gt;CB4856)</i>	Constructed from: ECA230 x CB4856	N2 into CB4856
Left indel - V: 7,862,556		
oECA799	TTCTCGCTACTGGAACACGC	
oECA800	TCAAGAAGCGTTGGGAAGTCT	
Right indel - V: 13,110,045		
oECA745	TGCAGAGGTGGAGTAACCCT	
oECA746	CTCGGTCTCTCCCCCACTAA	

Table S2-2 Reagents used to generate CRISPR/Cas9-mediated genome edits

crECA36 *dpy-10* guide RNA: GCUACCAUAGGCACCACGAG

crECA37 *dpy-10* repair construct: CACTTGAACTTCAATACGGCAAGATGAGAATGACTGGAAACC  
GTACCGCATGCGGTGCCTAGGTAGCGGAGCTTCACAT GGCTTCAGACCAACAGCCTAT

<b><i>jmjd-5(deletion)</i></b>	
ECA1047 <i>jmjd-5(ean136)</i>	CB4856 background
ECA1048 <i>jmjd-5(ean137)</i>	CB4856 background
ECA1051 <i>jmjd-5(ean141)</i>	N2 background

ECA1052 <i>jmjd-5(ean142)</i>	N2 background
guide RNAs	
crECA58	TGGTACAAAACACTATTTTCGGA
crECA59	AAAATTGACGAGTGTTCGCGA
confirmation primers - external	
oECA1153	TCCTCGTATTACAATCCGTTGTCCA
oECA1193	TGTCGTCTGGAAACATATGGCT
confirmation primers - internal	
oECA1194	CCGATAAAGGGCTGTGTATGGG
oECA1195	TCGAAAAGGCGATGTTGTGCAA

<b><i>C45B11.8(deletion)</i></b>	
ECA996 <i>C45B11.8(ean103)</i>	CB4856 background
ECA998 <i>C45B11.8(ean105)</i>	CB4856 background
ECA699 <i>C45B11.8(ean60)</i>	N2 background
ECA700 <i>C45B11.8(ean61)</i>	N2 background
guide RNAs	
crECA22	GCTGCAGTAGAGGTGACATT
crECA23	GAAGAAGTGAAAGAAGTGGG
confirmation primers - external	
oECA1269	TCTCCGTGACTCAAATTTTCGACA
oECA1270	AGATGAAGATCACACTCTTGCGA
confirmation primers - internal	
oECA1267	TCGCCTAATGTCACCTCTACTGC
oECA1268	TCCTCCCACTTCTTTCACTTCT

<b><i>C45B11.6(deletion)</i></b>	
ECA1053 <i>C45B11.6(ean142)</i>	CB4856 background
ECA622 <i>C45B11.6(ean46)</i>	CB4856 background
ECA623 <i>C45B11.6(ean47)</i>	N2 background
ECA624 <i>C45B11.6(ean48)</i>	N2 background
guide RNAs	
crECA9	TCATCAGGATCAATTTCAAG
crECA10	AGAATATCTGAATTGCCGAA
confirmation primers - external	
oECA1226	TCCTGGTTTTTCTTTTCAGTGGTTGT
oECA1227	TGTCTTCGGCAATTTTGTGCC
confirmation primers - internal	
oECA1224	GCTGGATTGCATTTGTCAAACCC
oECA1225	AGTTAAGAAAAGCAGCACCTGGA

<b><i>srg-42(deletion)</i></b>	
ECA1012 <i>srg-42(ean119)</i>	CB4856 background
ECA1013 <i>srg-42(ean120)</i>	CB4856 background
ECA697 <i>srg-42(ean58)</i>	N2 background
ECA698 <i>srg-42(ean59)</i>	N2 background
guide RNAs	
crECA19	TCAATTACAACTAGCGATT
crECA20	AGATGGTAAACCATAAATAG
confirmation primers - external	
oECA1262	TCACGCGTCACAATTATTGCTGA
oECA1263	AGCCATTGTTCAATTTCCCAGGT

confirmation primers - internal	
oECA1260	AGGGACAGTTATGATCACCAGT
oECA1261	GCCTGGCCCTTTTCAGAGACAA

<b><i>cnc-10(deletion)</i></b>	
ECA687 <i>cnc-10(ean53)</i>	CB4856 background
ECA696 <i>cnc-10(ean57)</i>	CB4856 background
ECA692 <i>cnc-10(ean55)</i>	N2 background
ECA693 <i>cnc-10(ean56)</i>	N2 background
guide RNAs	
oECA1186*	ACAACGTCTGCTCAATTTTA
crECA21*	ACGTCTGCTCAATTTTATGG
oECA1187	GCACTAATGGGAGCTGCAAT
confirmation primers	
oECA1170**	GTCCTTACTGAGGCGTGTCCAT
oECA1266**	TCCAGGATCTACGCAAAAATGAACT
oECA1171	CAGGTTCAAATCCTGCGGACAG

\*oECA1186 and oECA1187 were used to generate the CB4856 deletions. crECA21 and oECA1187 were used to generate the N2 deletions.

\*\*oECA1170 and oECA1171 were used to confirm CB4856 deletions. oECA1266 and oECA1171 were used to confirm N2 deletions.

<b><i>H19N07.3(deletion)</i></b>	
ECA1133 <i>H19N07.3(ean179)</i>	CB4856 background
ECA1134 <i>H19N07.3(ean180)</i>	CB4856 background
ECA1131 <i>H19N07.3(ean177)</i>	N2 background
ECA1132 <i>H19N07.3(ean178)</i>	N2 background
guide RNAs	

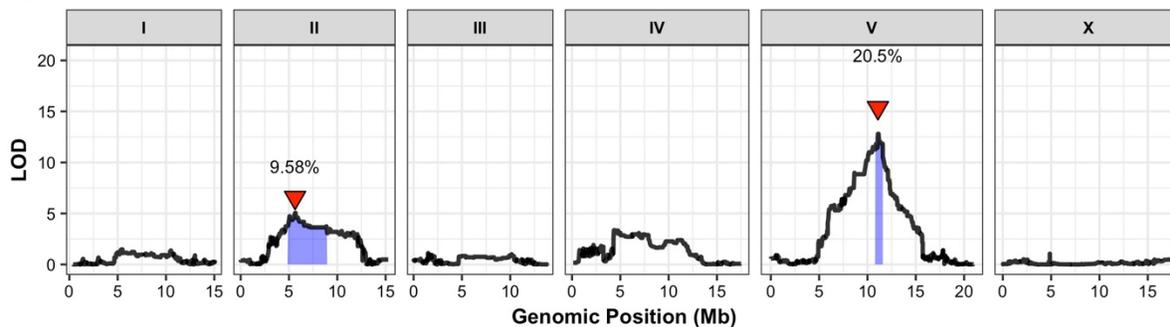
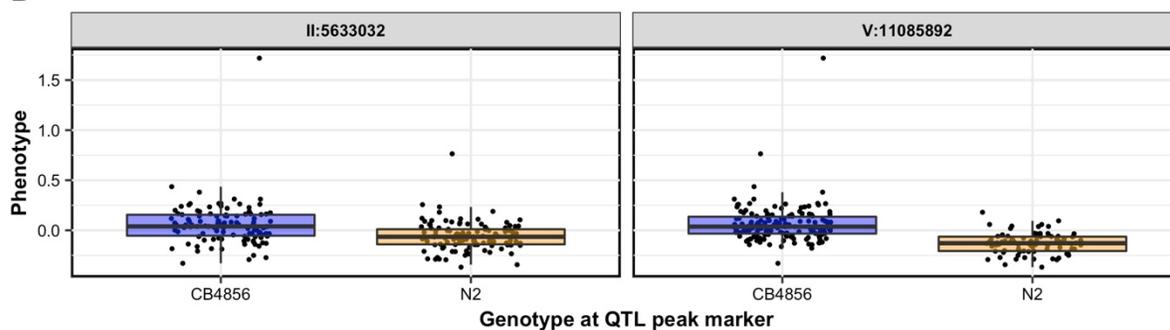
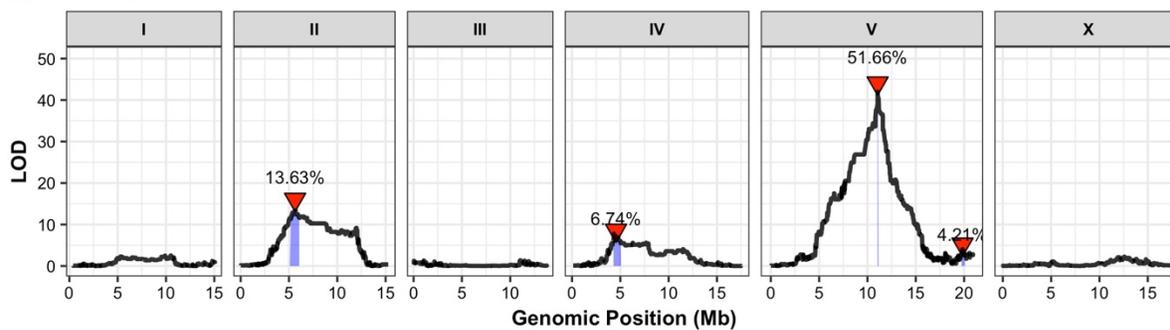
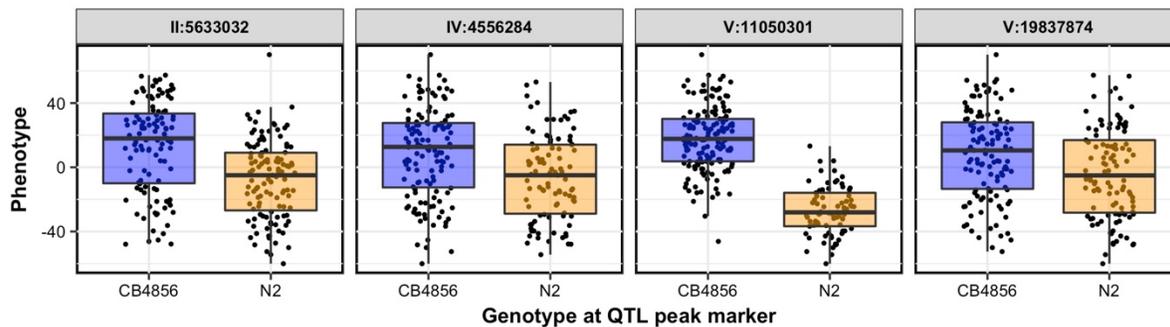
crECA84	GCGAGCACAACCTTCAAGAAA
crECA85	CGTATGGCTGCCAAGGCCAG
confirmation primers	
oECA1173	TCTTGCAGACACATGGGTCC
oECA1174	ATCGGTGGGCACAATGTGAT

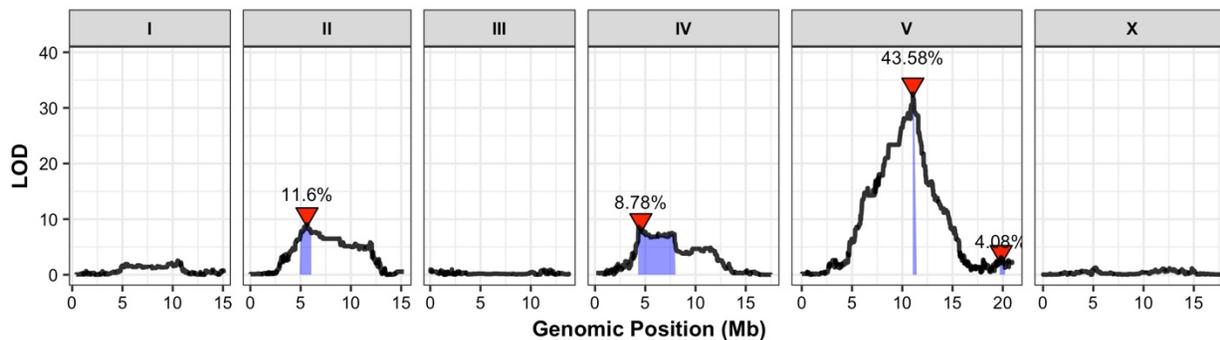
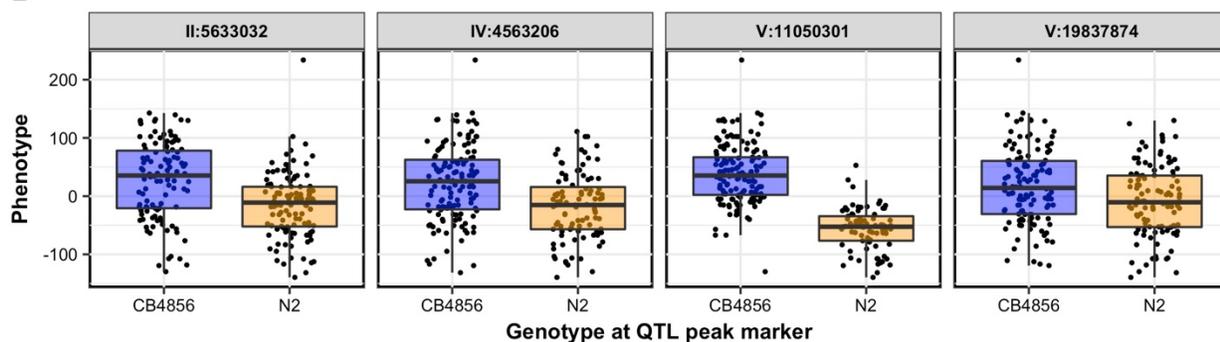
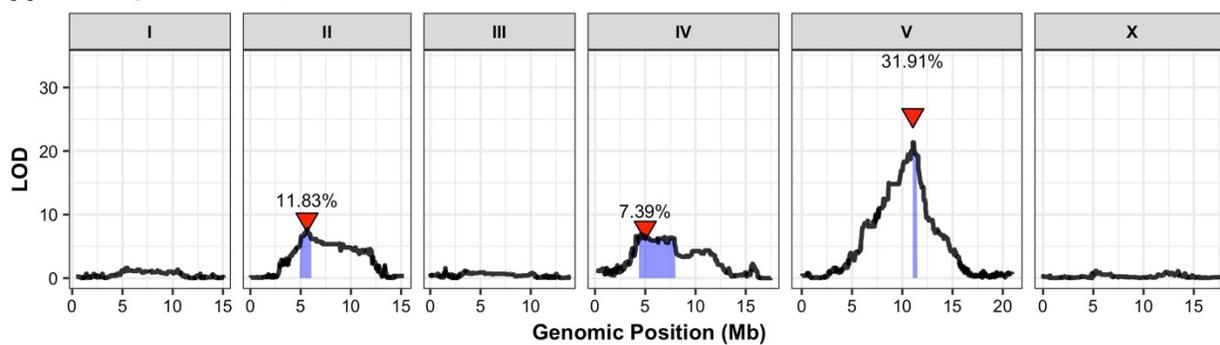
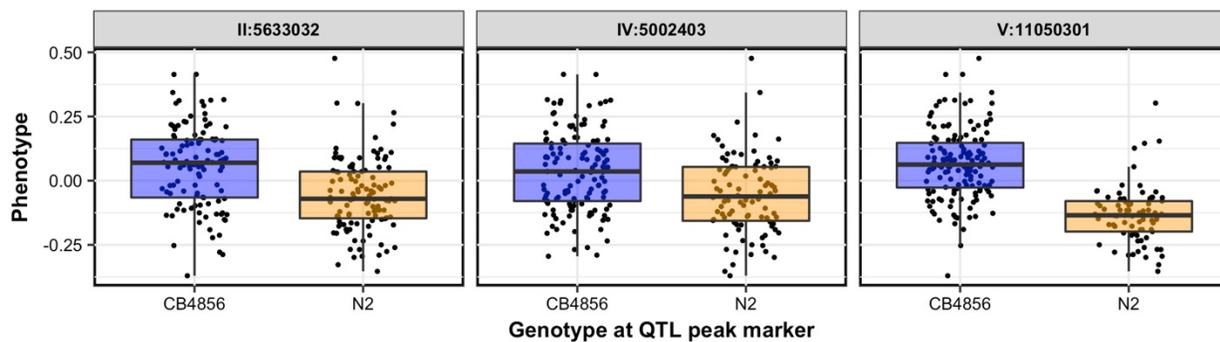
<b><i>jmjd-5(CB4856 to N2)</i></b>	
ECA578 <i>jmjd-5(ean12[S338P])</i>	CB4856 background
ECA579 <i>jmjd-5(ean13[S338P])</i>	CB4856 background
guide RNA	
oECA1196	GGAATTTGAAAGTGGAATTA
repair template	
oECA1199	ACTAGCATGGTTAATTCATGAAAATTTACCTGGTGTGTGCATCTG ATGATTGGATTCATTTCGAGTTTTTCAGTTCAATACAACATAACG TATCCTGCGTTAATTCCACTTTCAAATTCAAATCTATCGATGA ATGTGATGAAGATGA
confirmation primers - to check for <i>Bsa</i> I site introduction	
oECA1194	CCGATAAAGGGCTGTGTATGGG
oECA1195	TCGAAAAGGCGATGTTGTGCAA

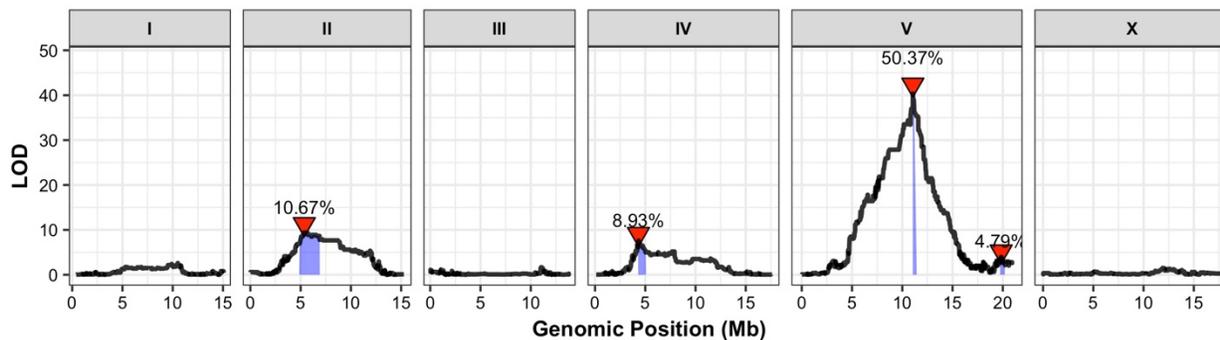
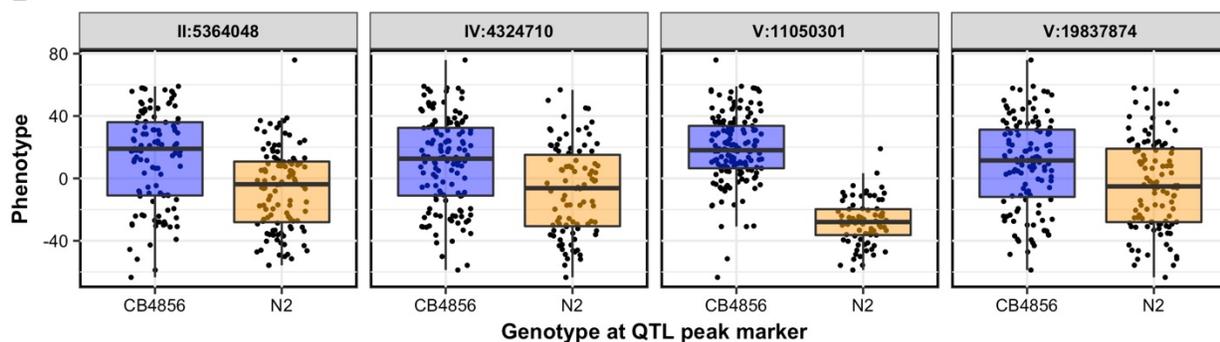
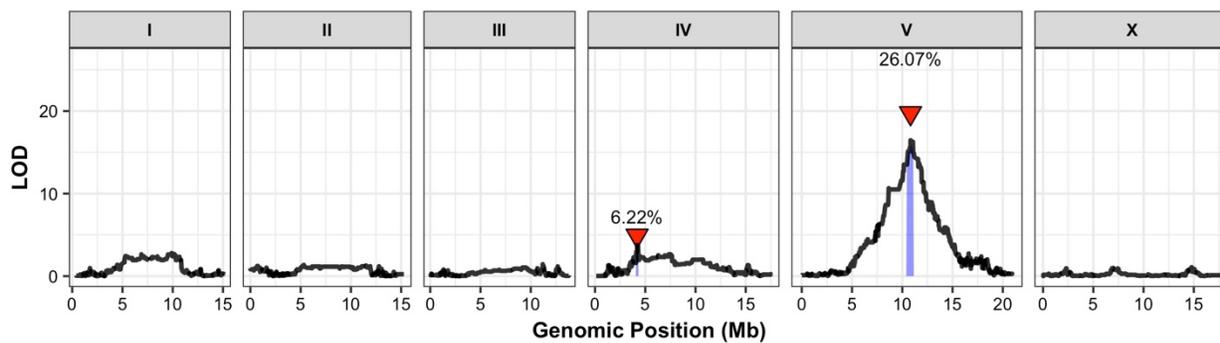
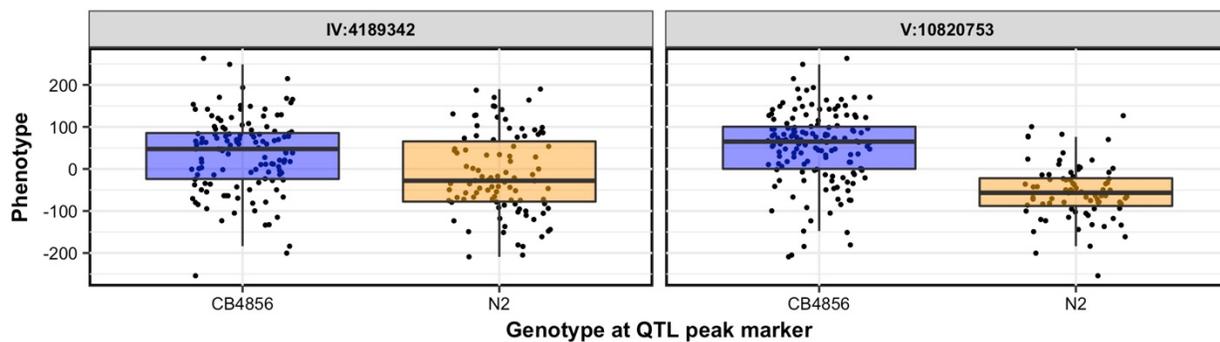
<b><i>jmjd-5(N2 to CB4856)</i></b>	
ECA576 <i>jmjd-5(ean10[P338S])</i>	N2 background
ECA577 <i>jmjd-5(ean11[P338S])</i>	N2 background
guide RNA	
oECA1196	GGAATTTGAAAGTGGAATTA
repair template	
oECA1198	ACTAGCATGGTTAATTCATGAAAATTTACCTGG TGTGTGCATCTGATGATTGGATTCATTTCGAGTTT TCAGTTCAATACAACATAACGTATTCTGCGTT

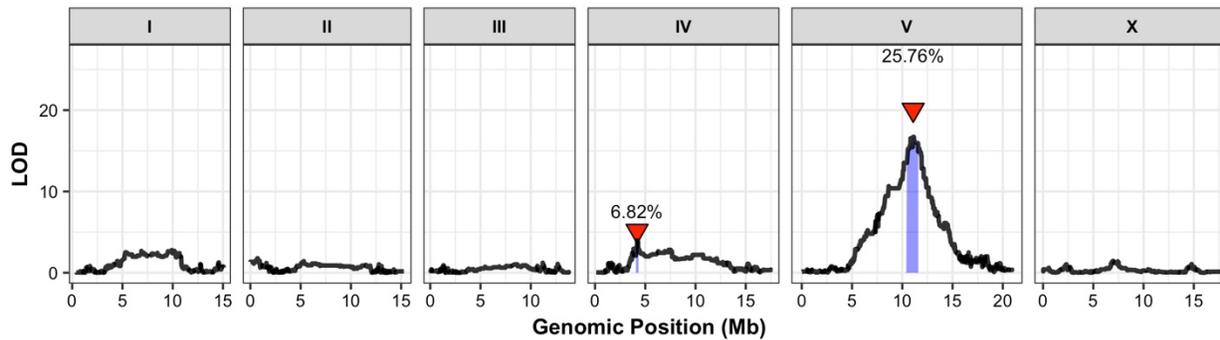
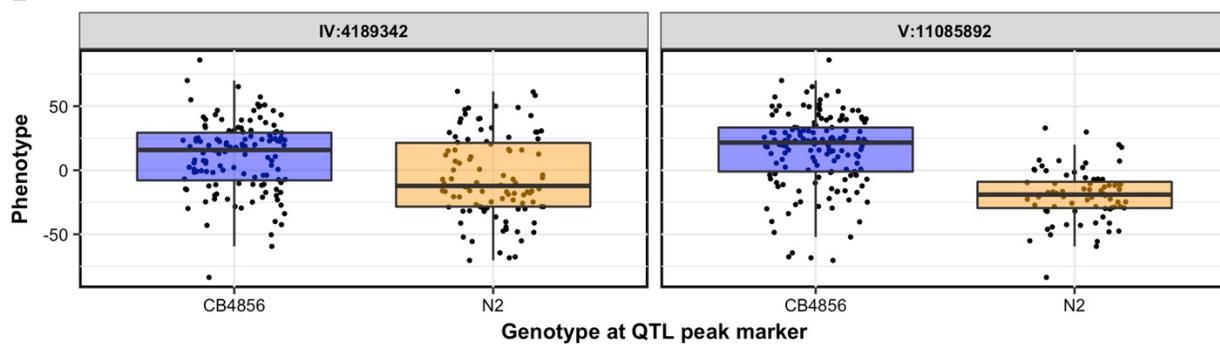
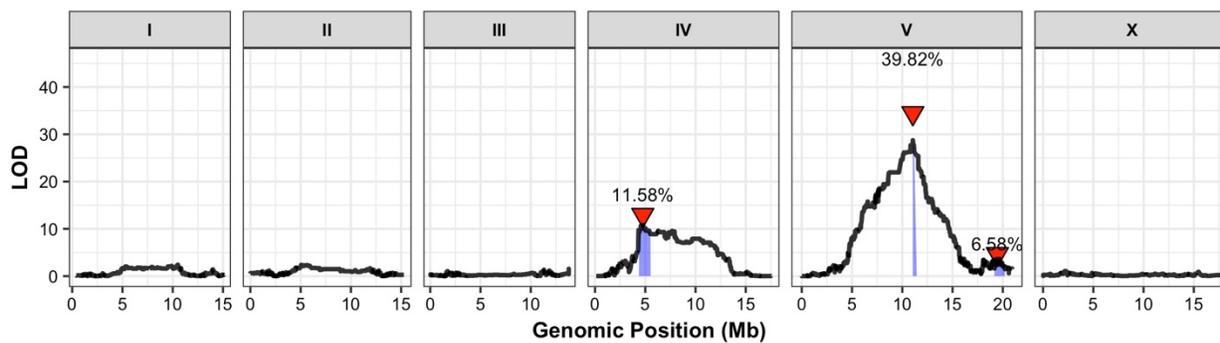
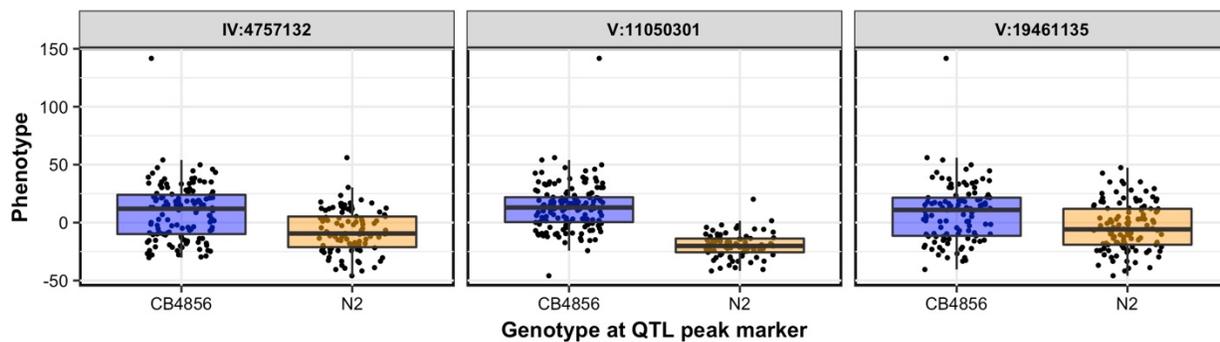
	AATCCACTTTCAAATTCCAAATCTATCGATGA ATGTGATGAAGATGA
confirmation primers - to check for <i>Bsa</i> I site introduction	
oECA1194	CCGATAAAGGGCTGTGTATGGG
oECA1195	TCGAAAAGGCGATGTTGTGCAA

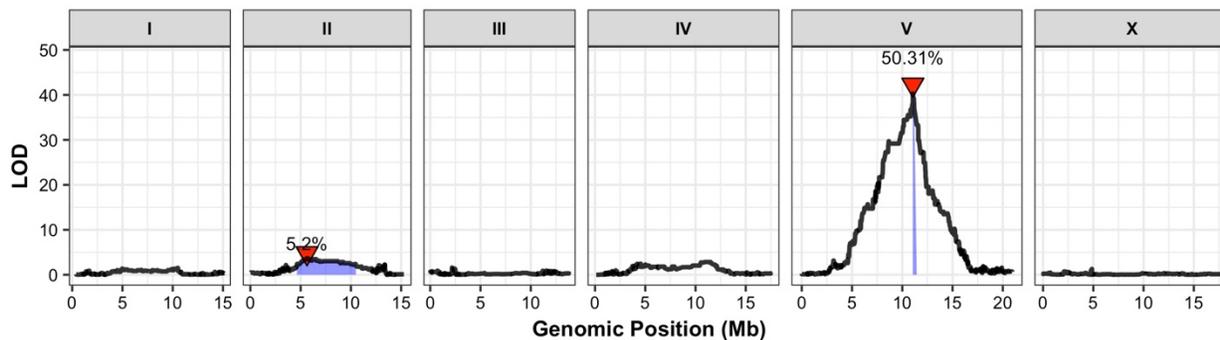
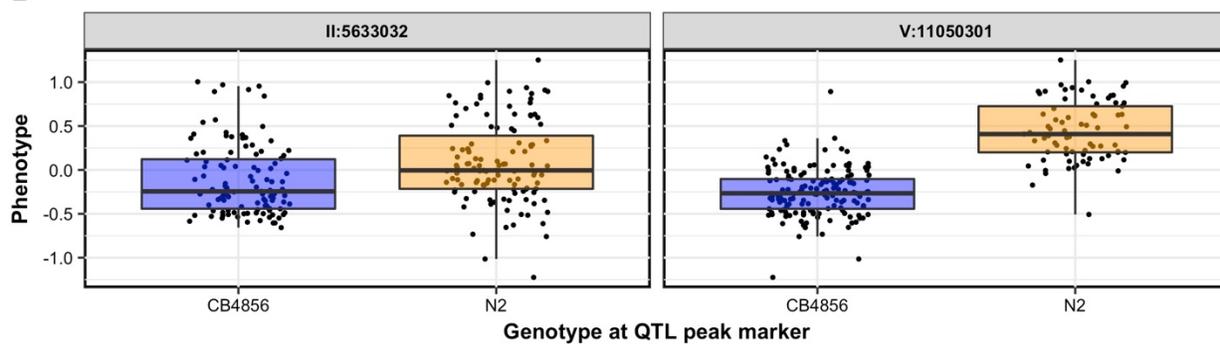
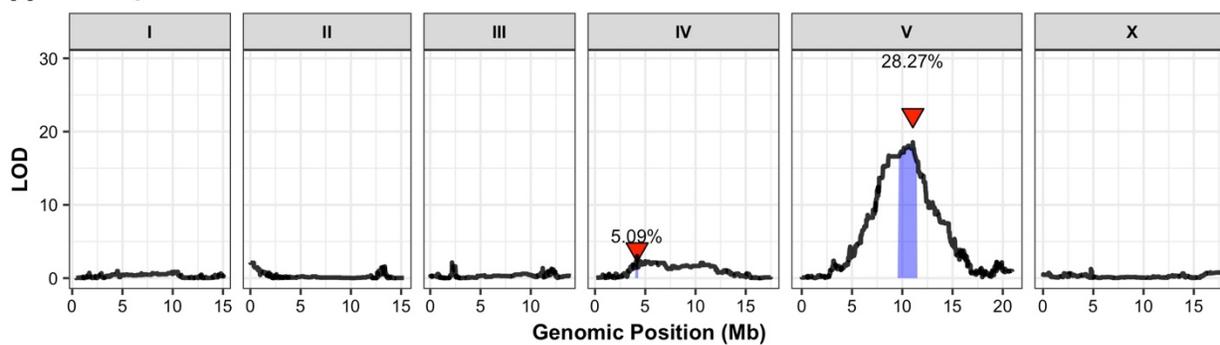
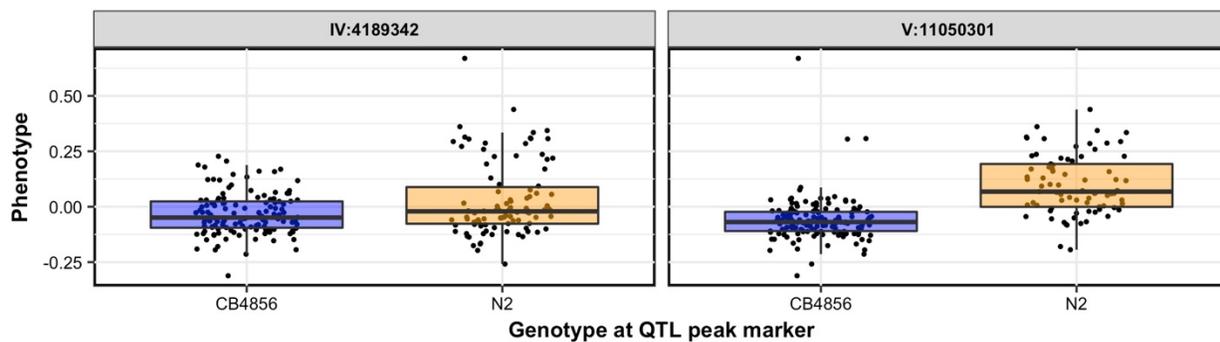
## Supplemental Figures

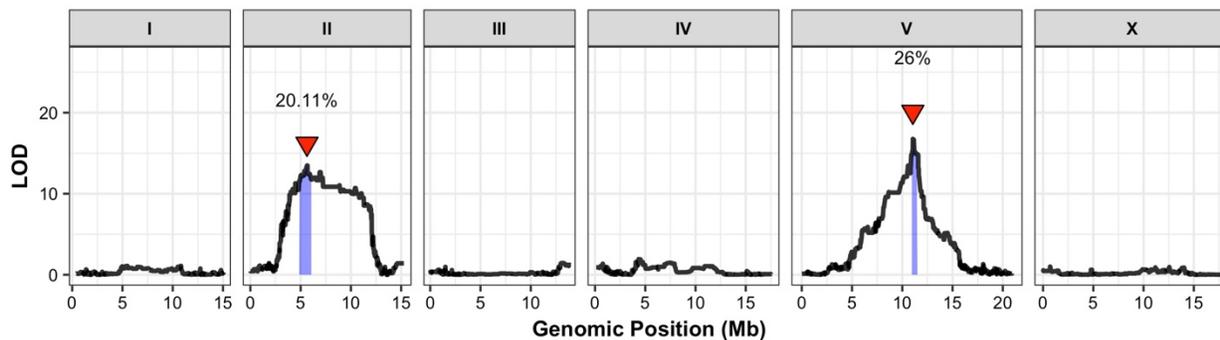
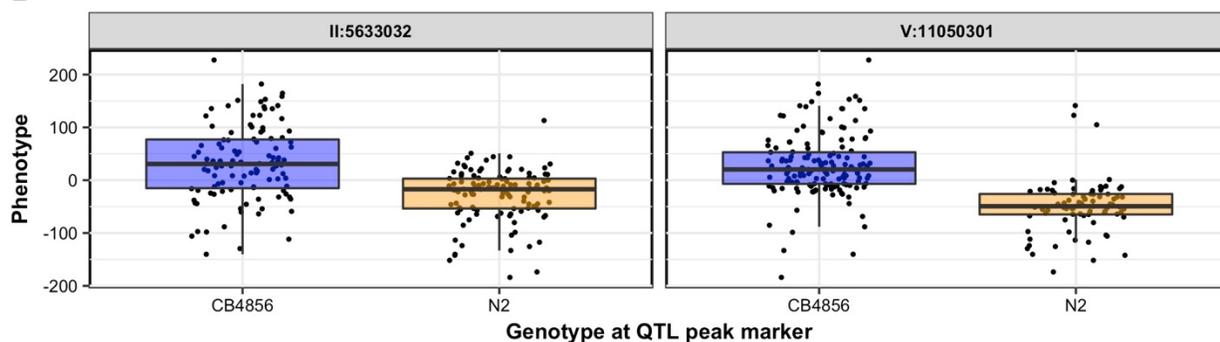
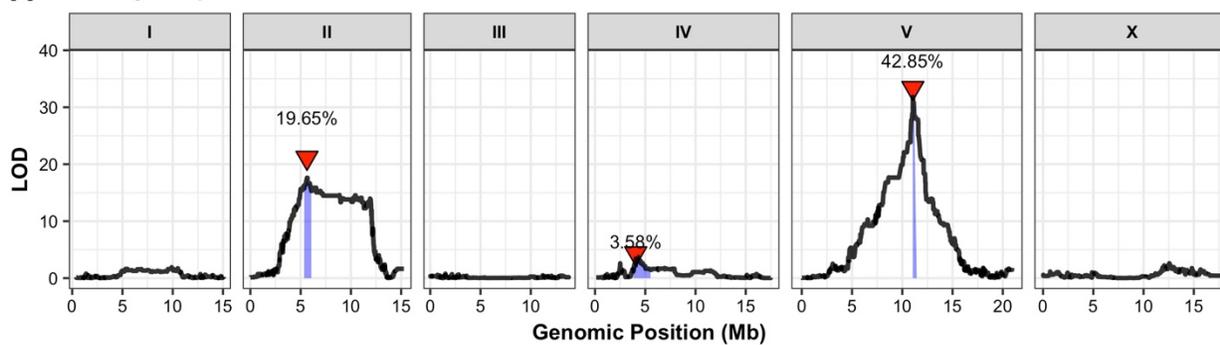
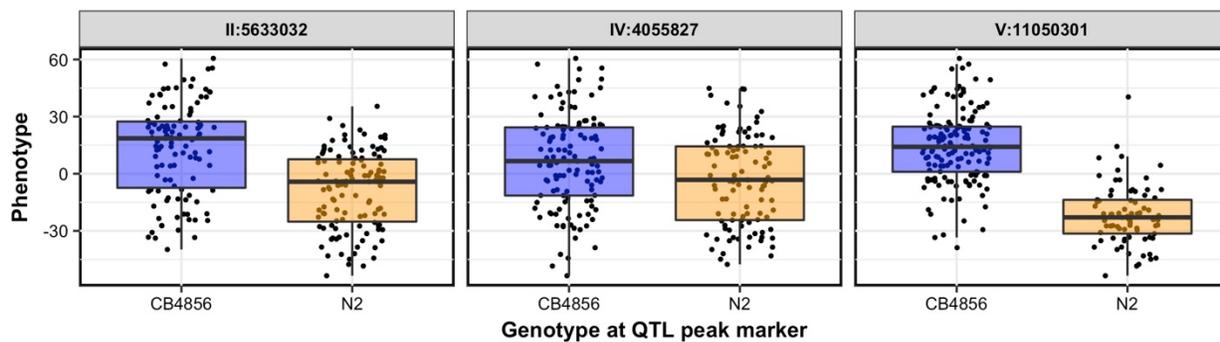
**A** Bleomycin mean.norm.EXT**B****A** Bleomycin mean.TOF**B**

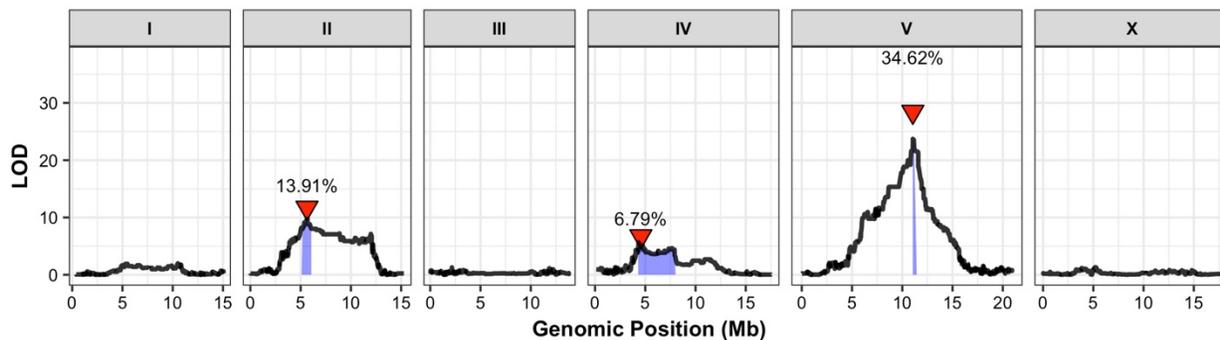
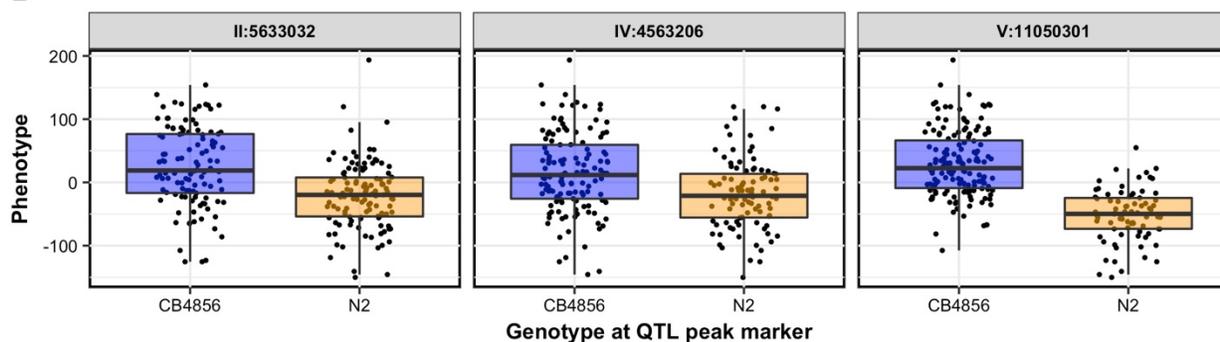
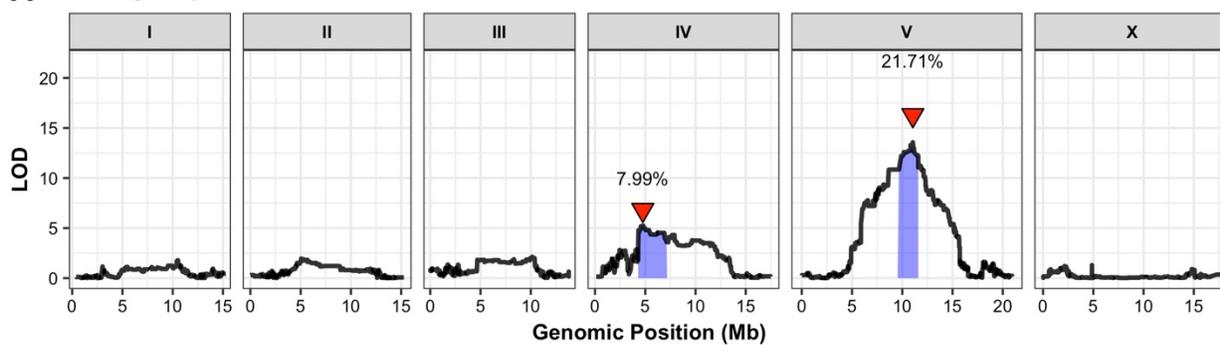
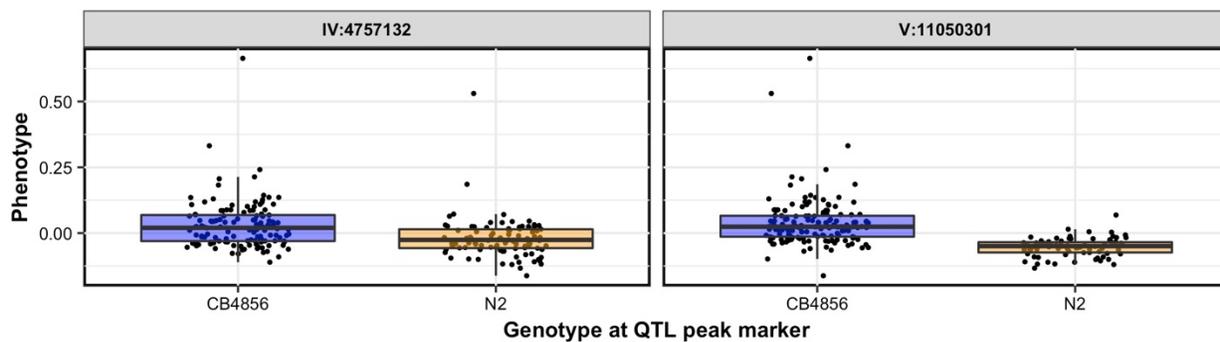
**A** Bleomycin median.EXT**B****A** Bleomycin median.norm.EXT**B**

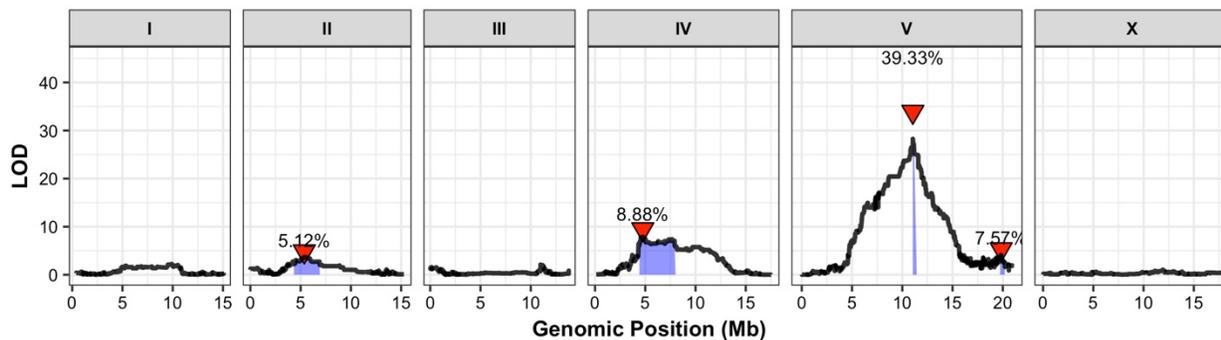
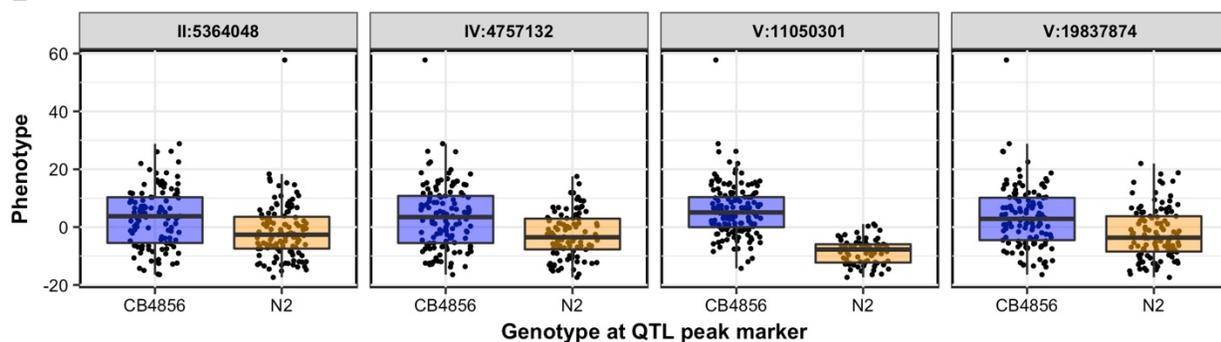
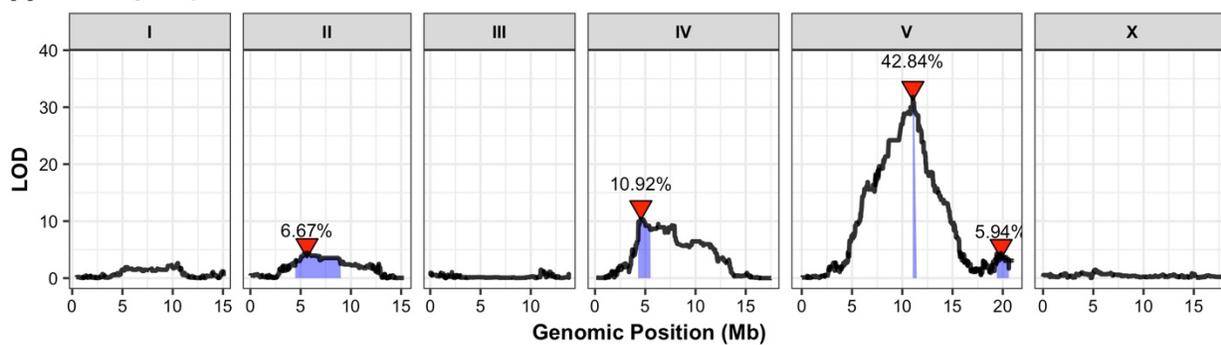
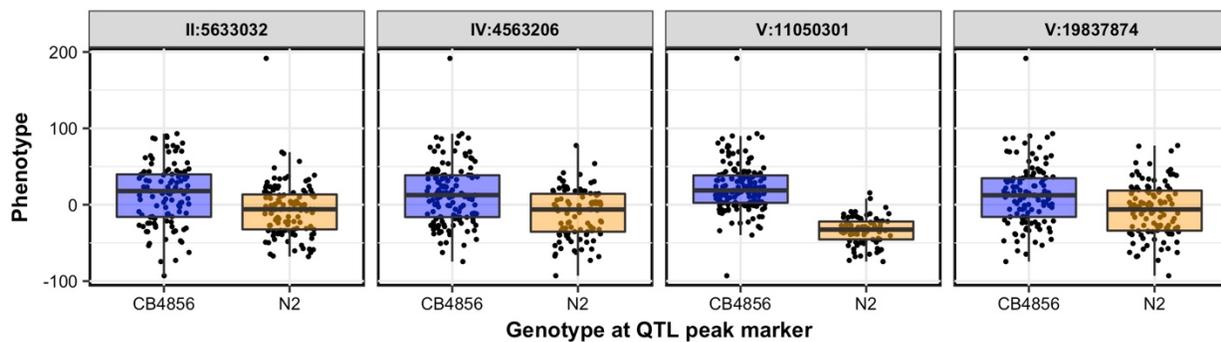
**A** Bleomycin median.TOF**B****A** Bleomycin n**B**

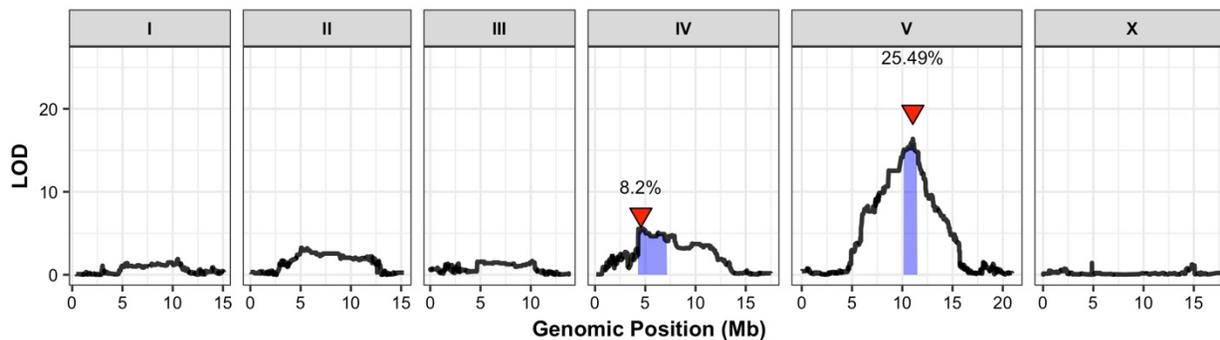
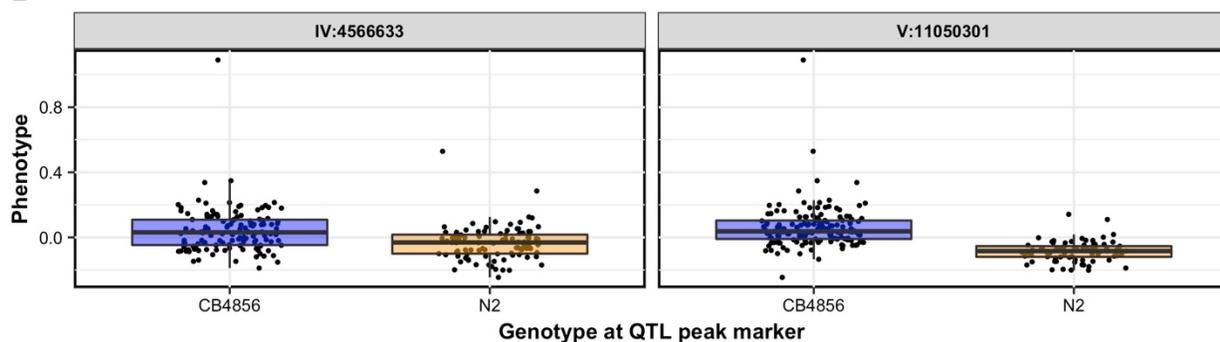
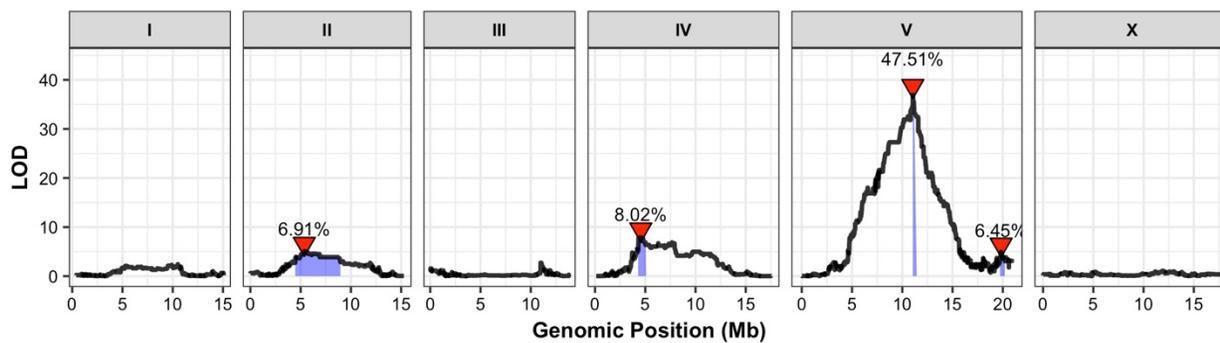
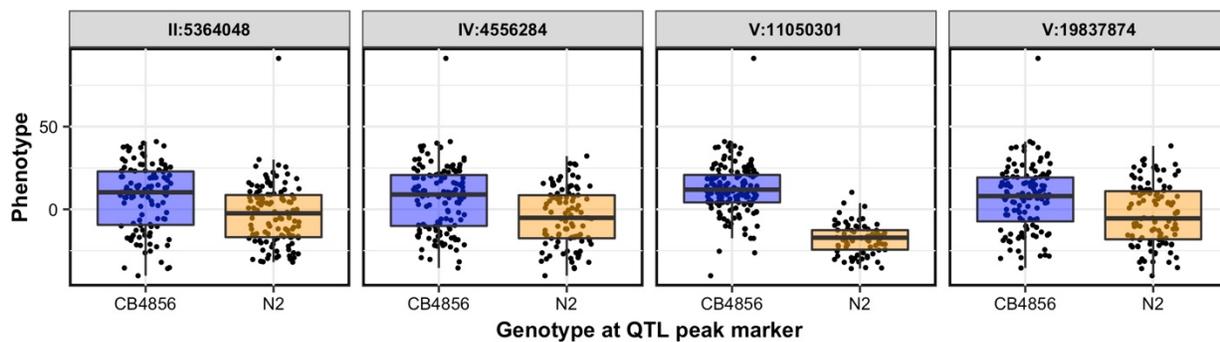
**A** Bleomycin norm.n**B****A** Bleomycin q10.EXT**B**

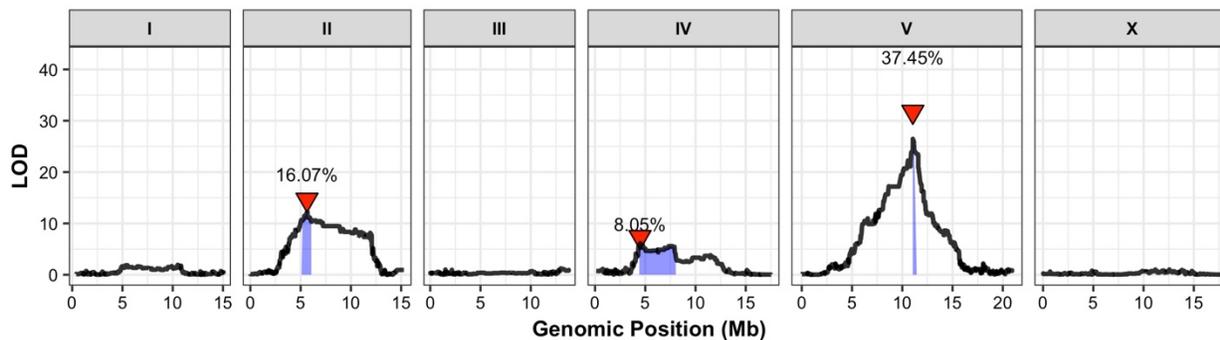
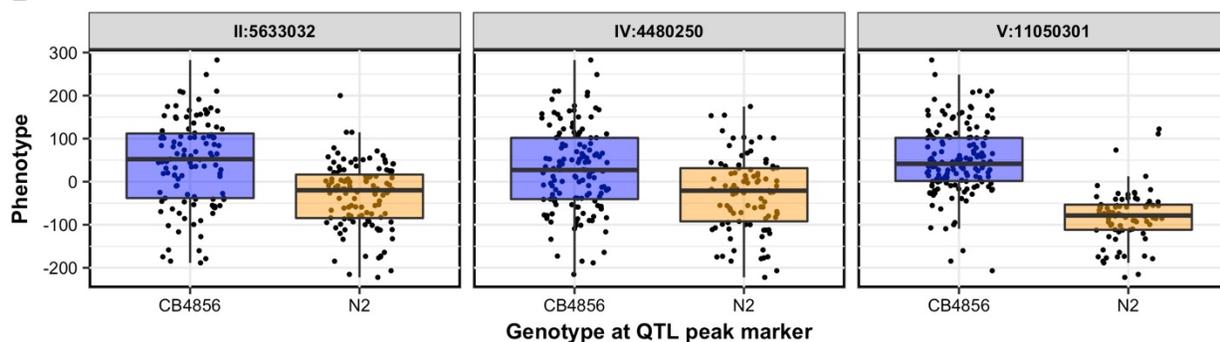
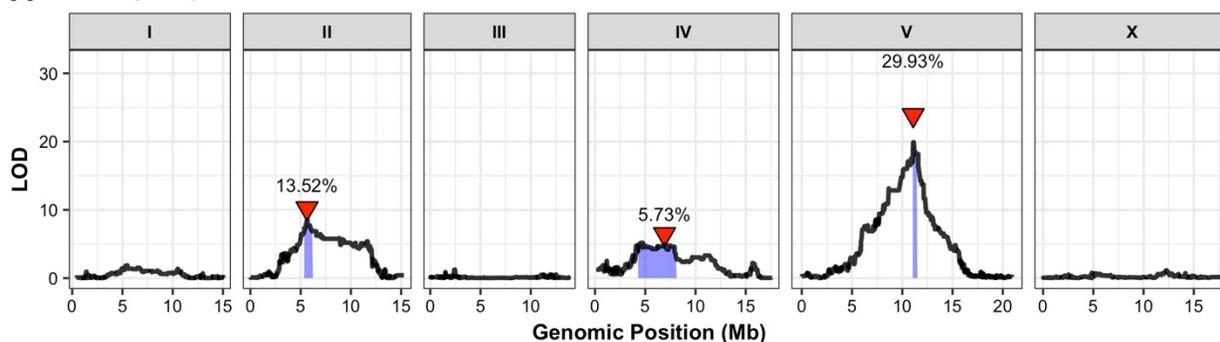
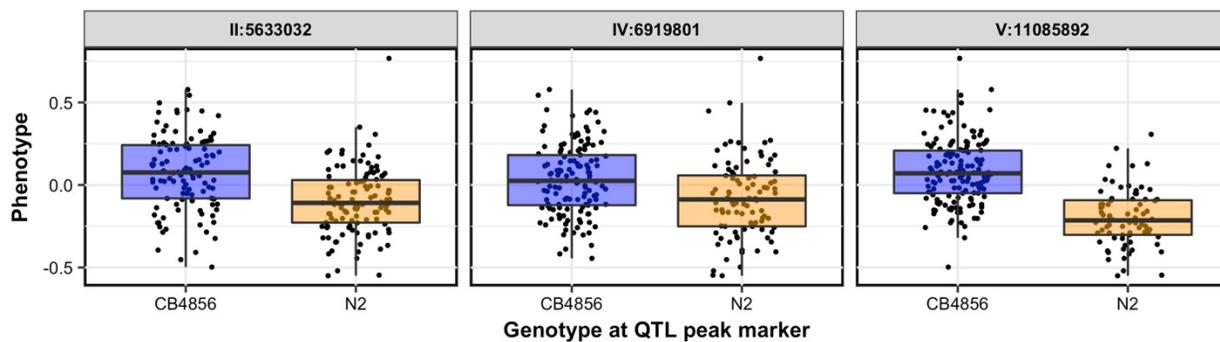
**A** Bleomycin cv.EXT**B****A** Bleomycin cv.TOF**B**

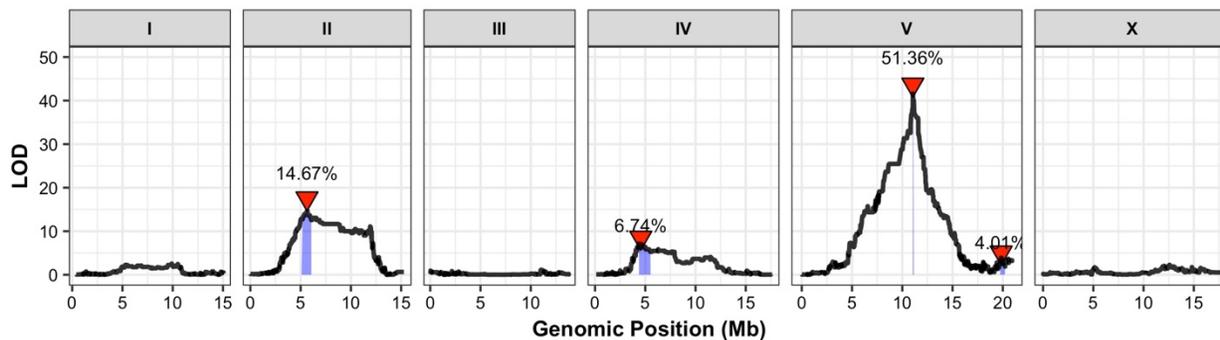
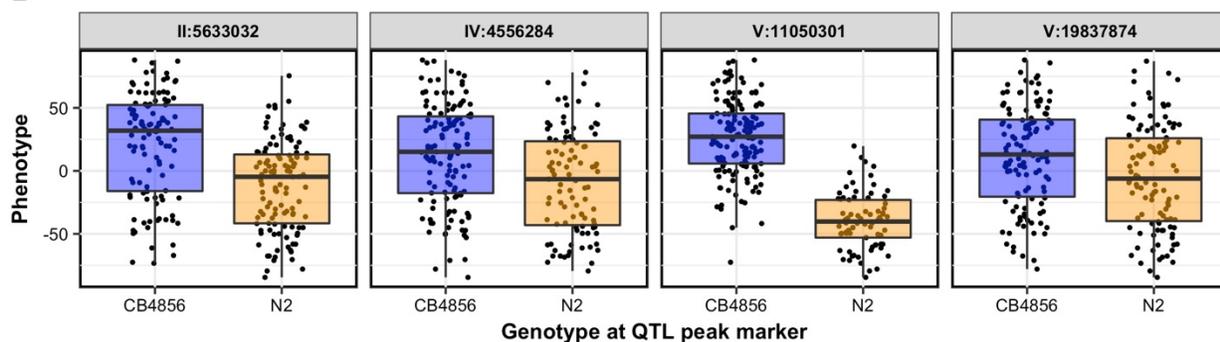
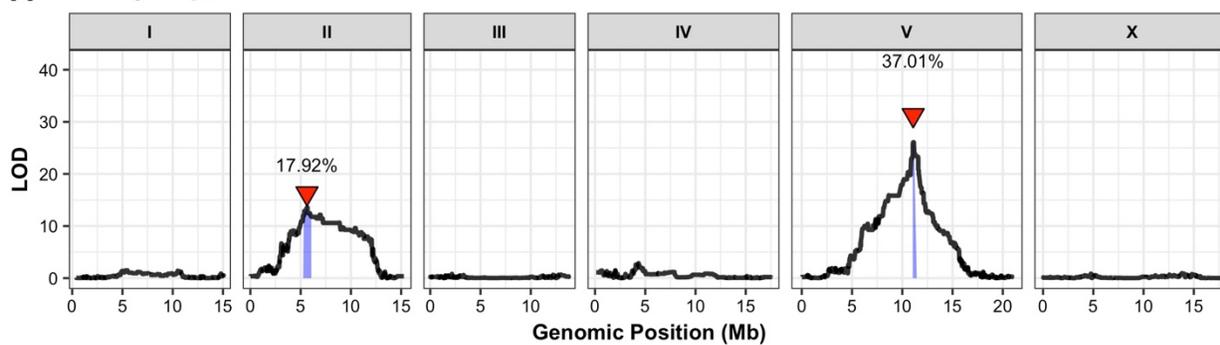
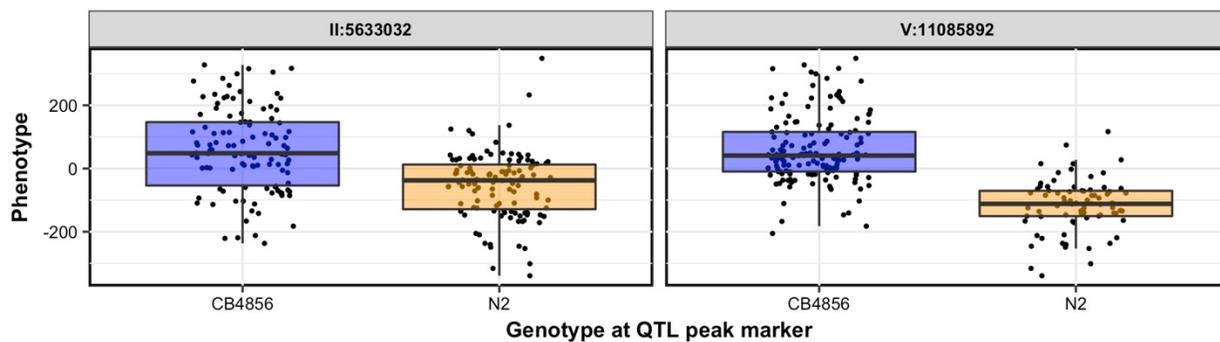
**A** Bleomycin *iqr.EXT***B****A** Bleomycin *iqr.TOF***B**

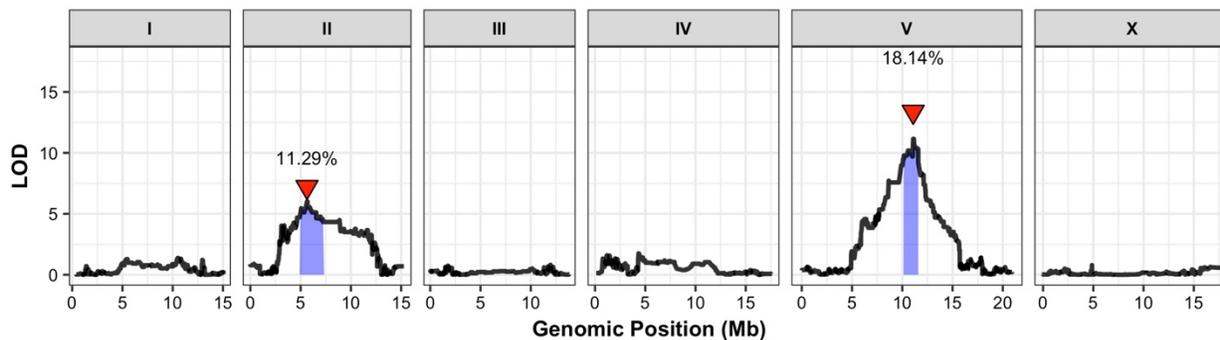
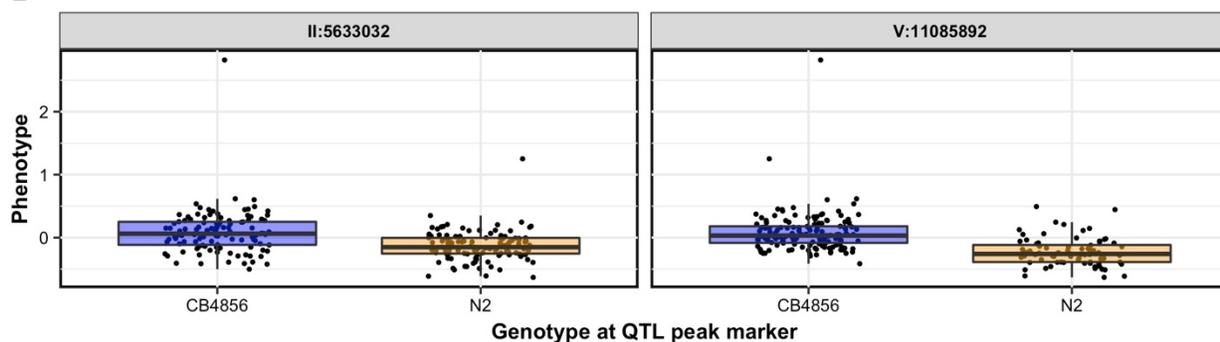
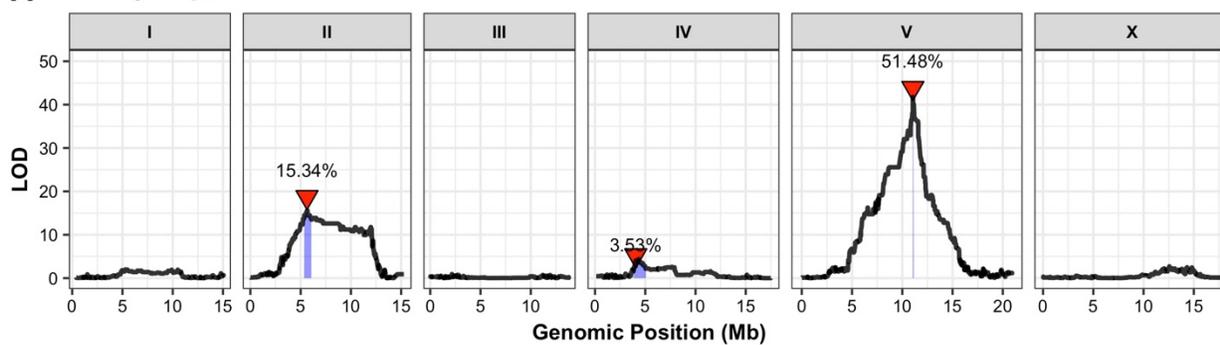
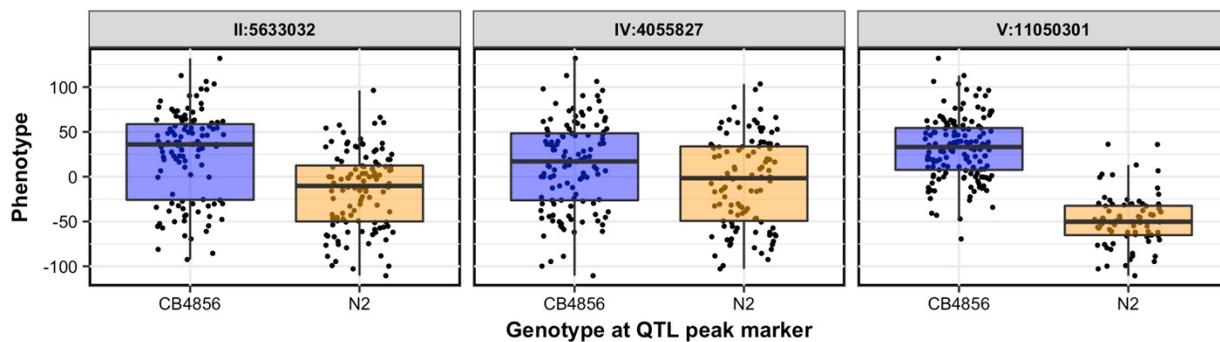
**A** Bleomycin mean.EXT**B****A** Bleomycin q10.norm.EXT**B**

**A** Bleomycin q10.TOF**B****A** Bleomycin q25.EXT**B**

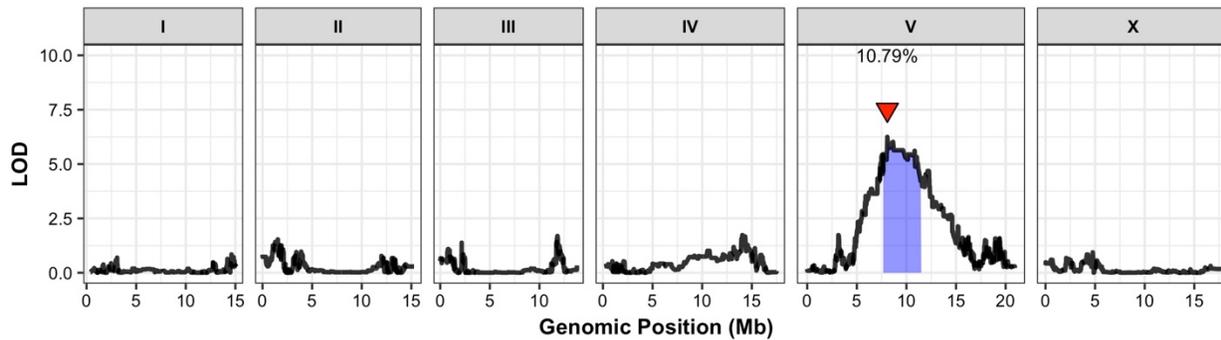
**A** Bleomycin q25.norm.EXT**B****A** Bleomycin q25.TOF**B**

**A** Bleomycin q75.EXT**B****A** Bleomycin q75.norm.EXT**B**

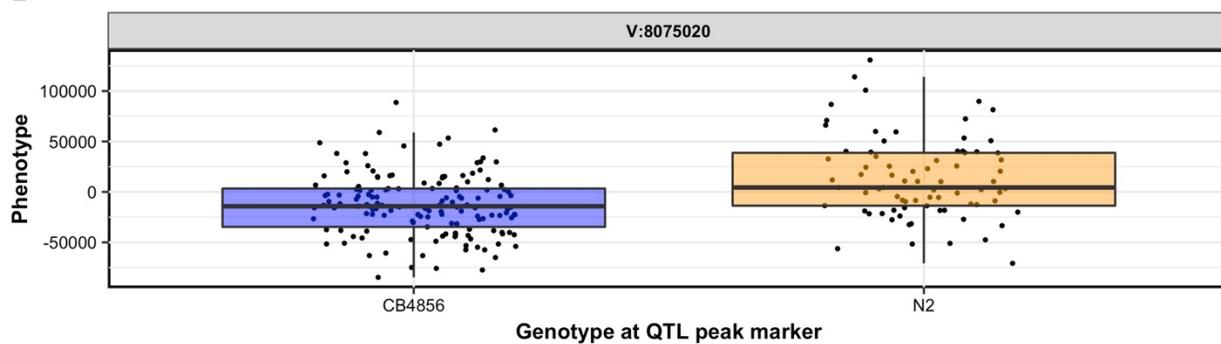
**A** Bleomycin q75.TOF**B****A** Bleomycin q90.EXT**B**

**A** Bleomycin q90.norm.EXT**B****A** Bleomycin q90.TOF**B**

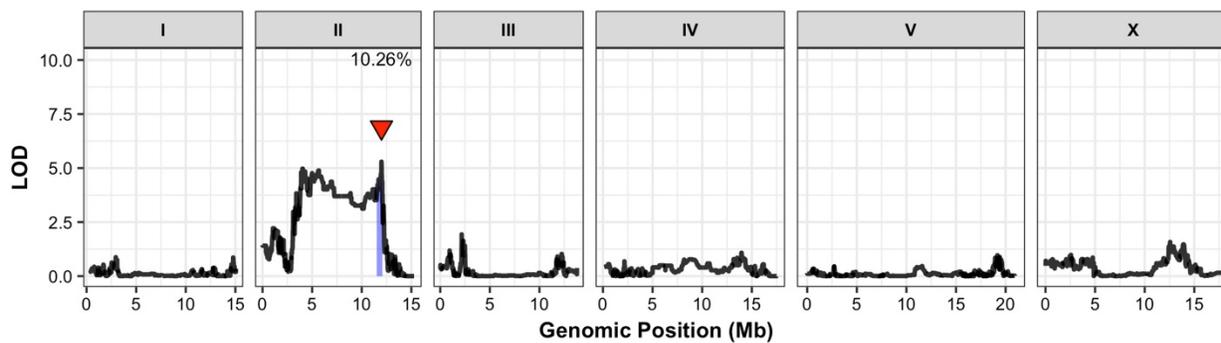
**A** Bleomycin var.EXT



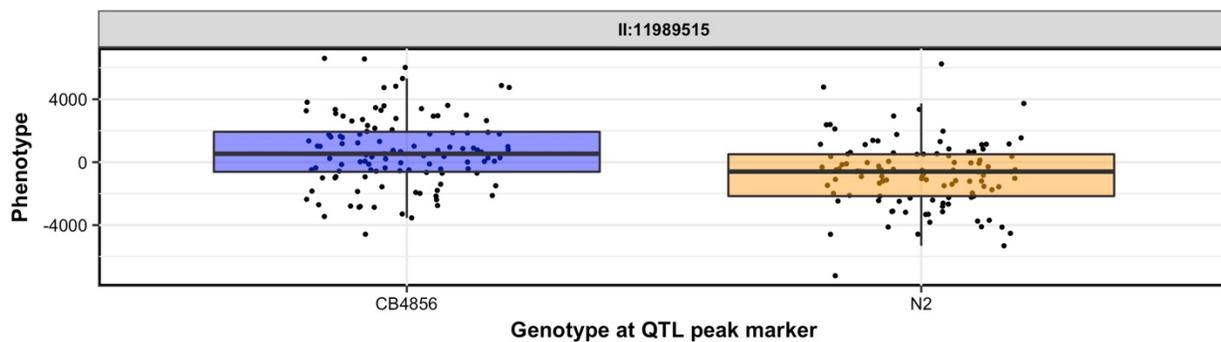
**B**



**A** Bleomycin var.TOF

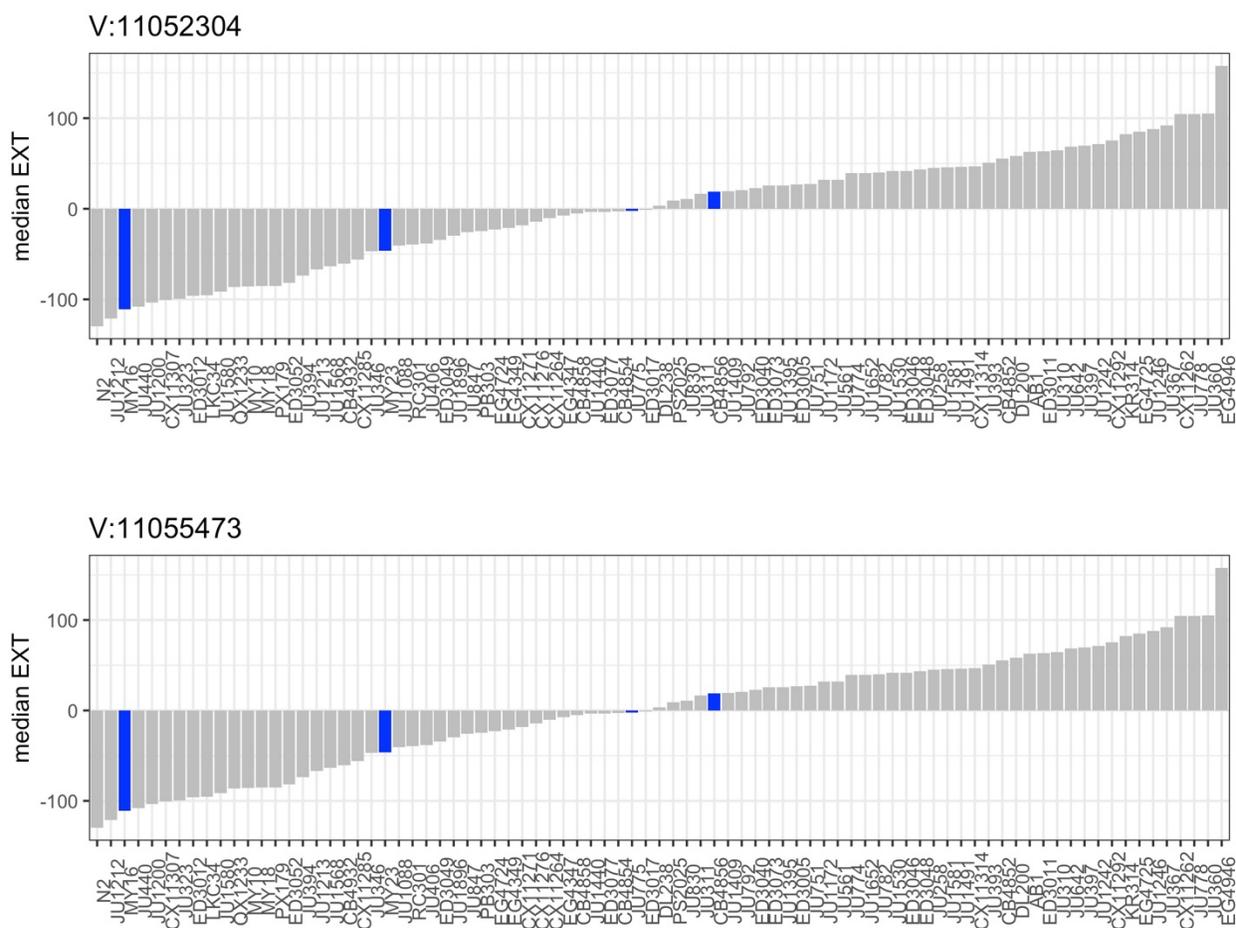


**B**

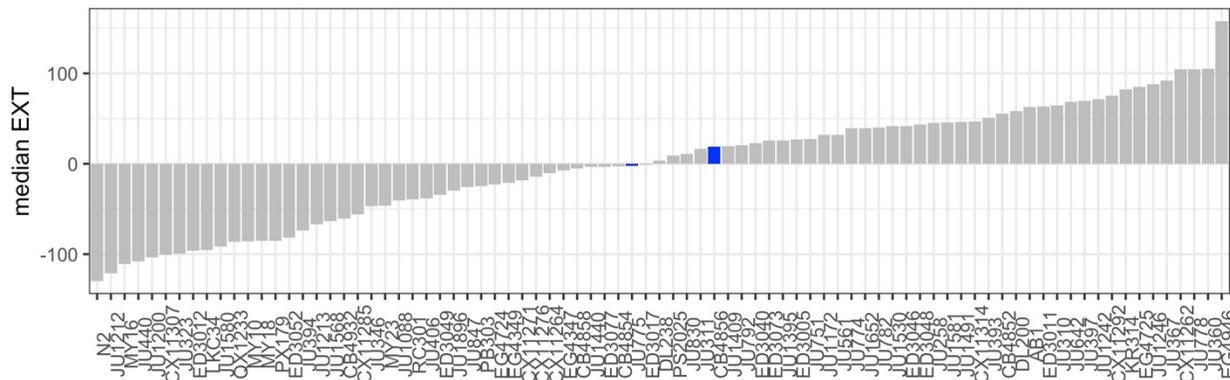


**Figure S2-1 All linkage mapping results for bleomycin-response variation**

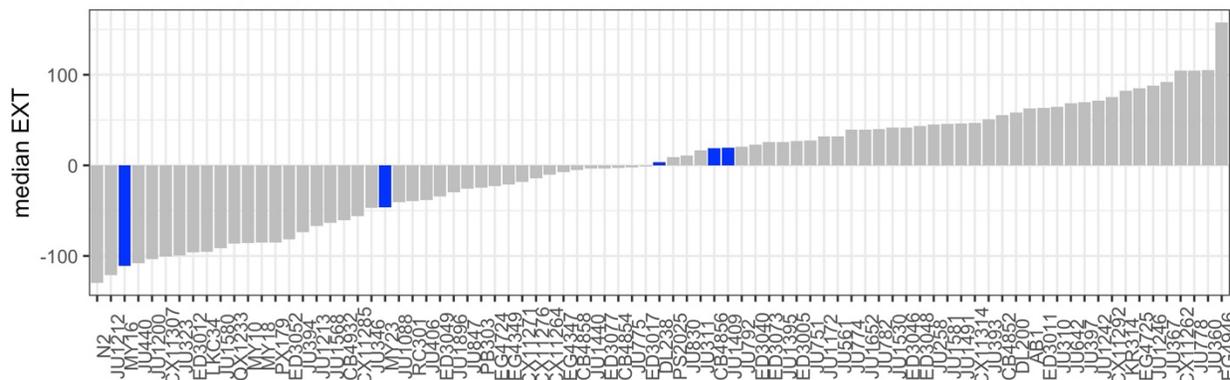
Linkage-mapping analysis of bleomycin-response variation for all 26 high-throughput traits are shown. **A.** On the x-axis, each of 13,003 genomic markers, split by chromosome, were tested for correlation with phenotypic variation across the RIAIL panel. The log of the odds (LOD) score for each marker is reported on the y-axis. Each significant quantitative trait locus (QTL) is indicated by a red triangle at the peak marker, and a blue ribbon shows the 95% confidence interval around the peak marker. The total amount of phenotypic variance across the RIAIL panel explained by the genotype at each peak marker is shown as a percentage. **B.** RIAIL phenotypes (y-axis), split by allele at each QTL peak marker (x-axis) are reported. For each significant QTL, phenotypes of RIAILs containing the N2 allele (orange) are compared to phenotypes of RIAILs containing the CB4856 allele (blue). Phenotypes are shown as Tukey box plots, and each point is the phenotype of an individual strain.



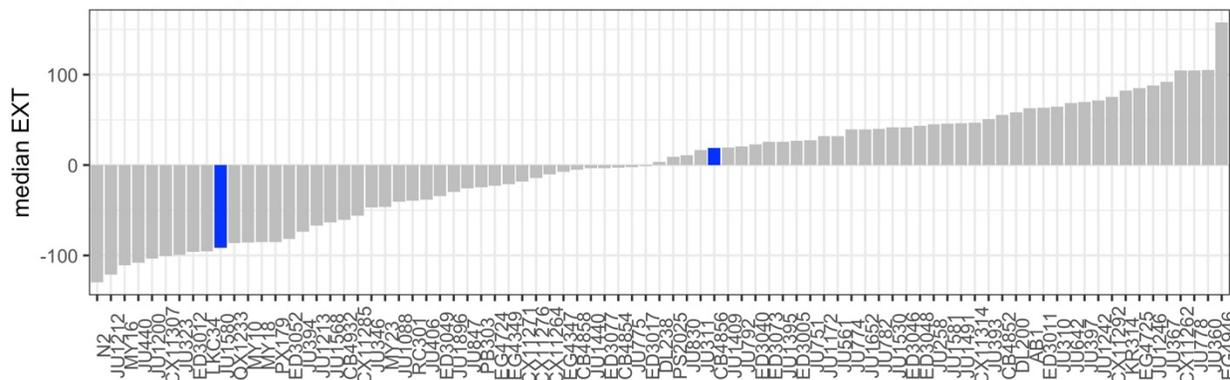
V:11068853



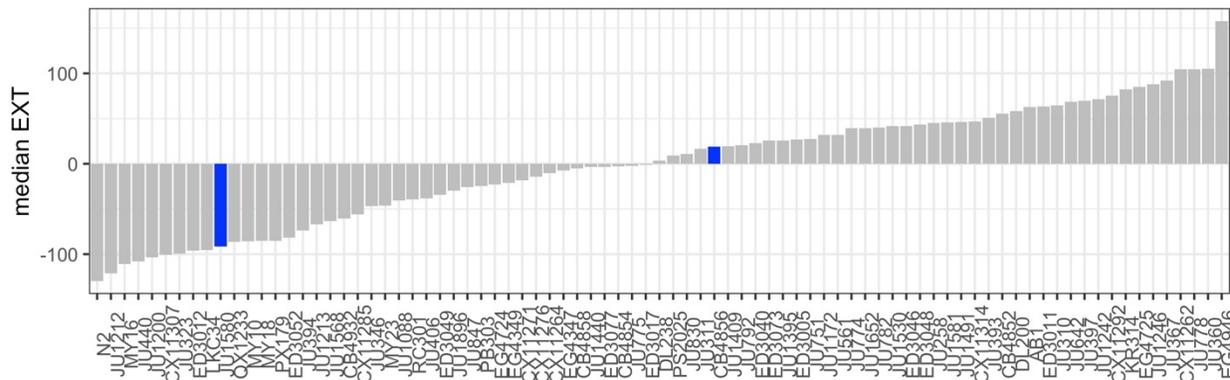
V:11080125



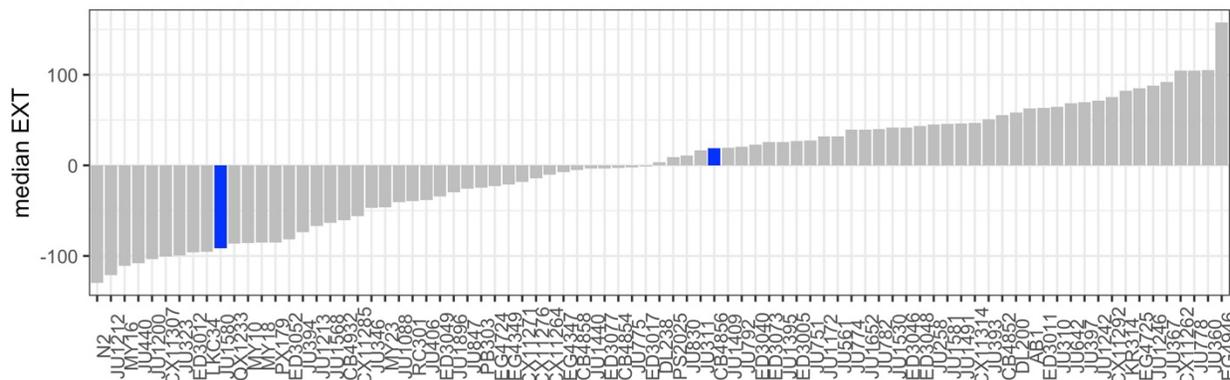
V:11085892



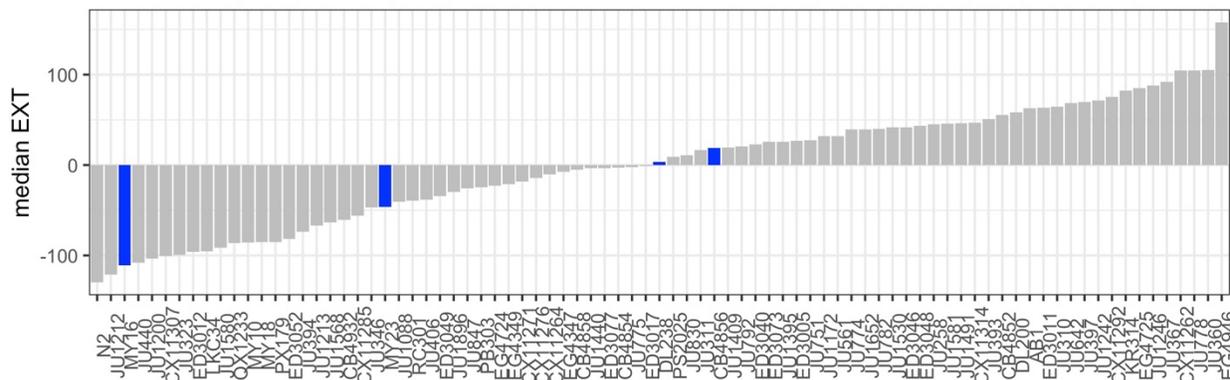
V:11097232



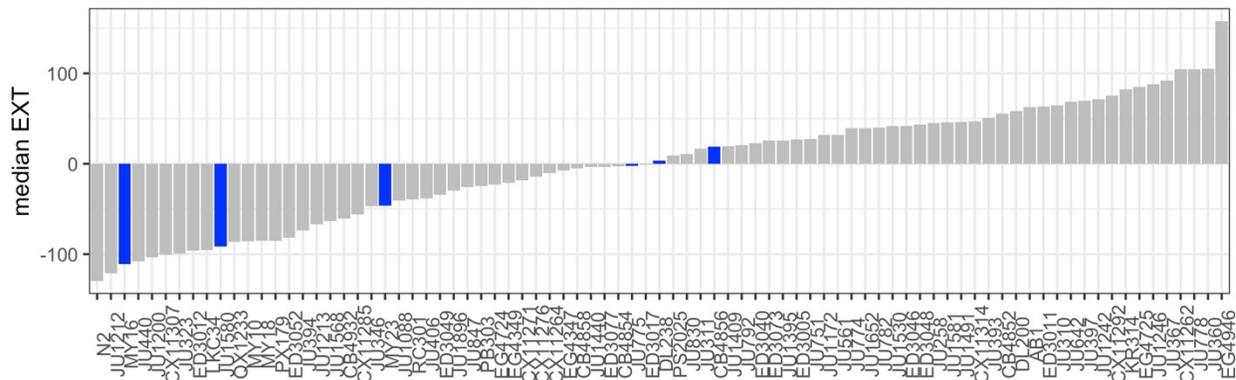
V:11101282



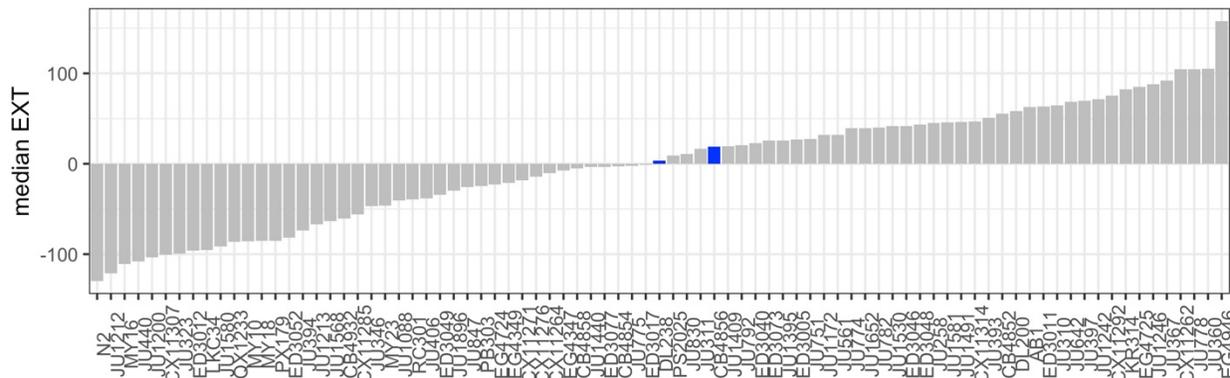
V:11102451



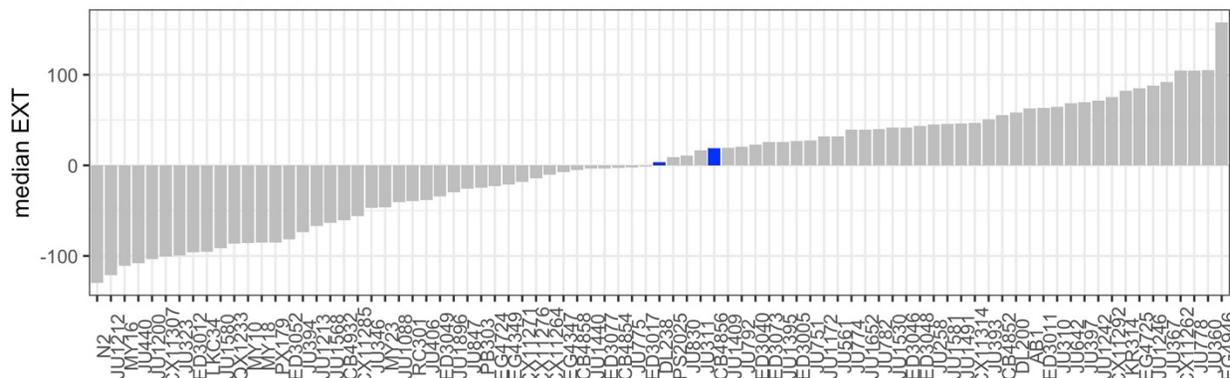
V:11104309



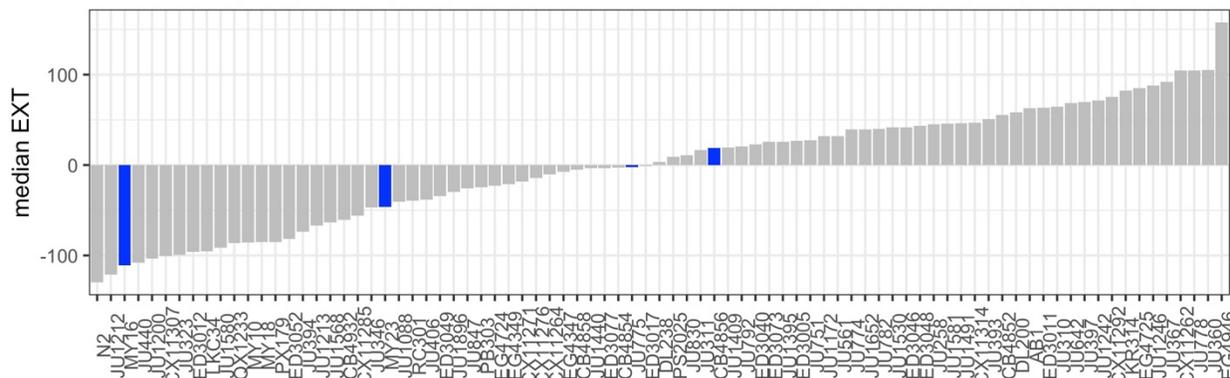
V:11127063



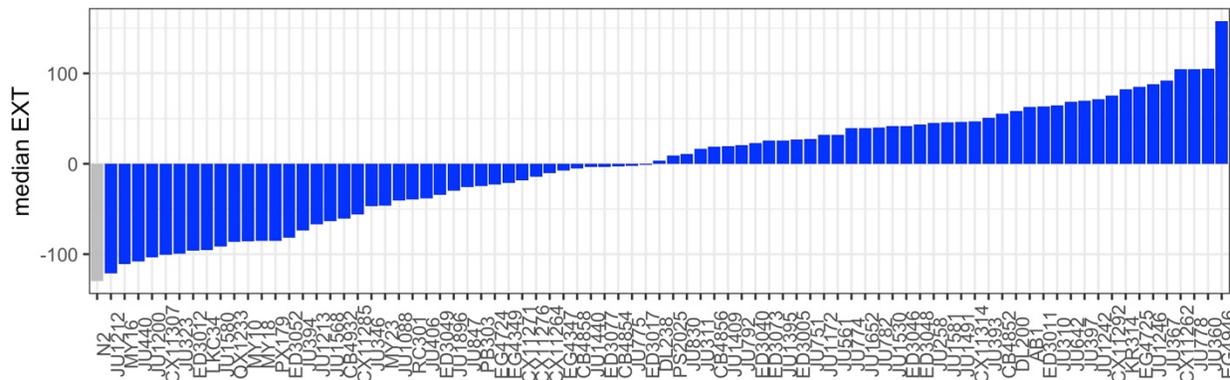
V:11127357



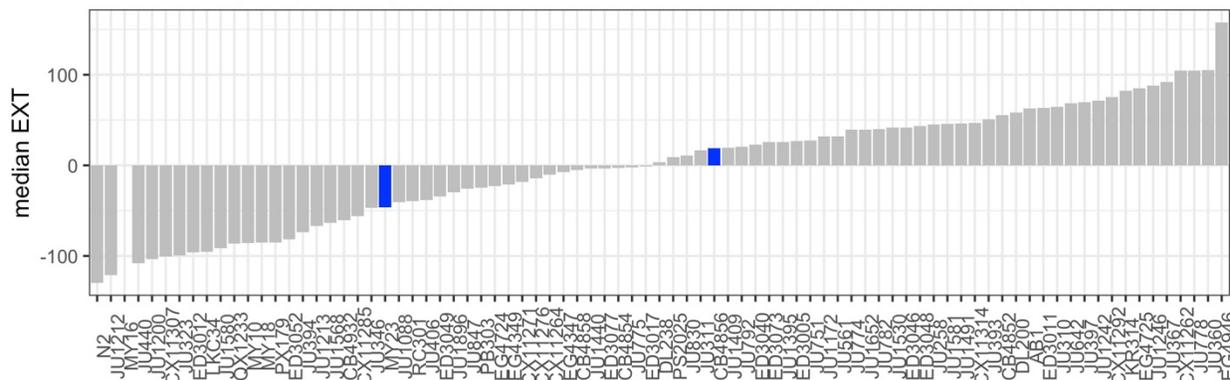
V:11128895



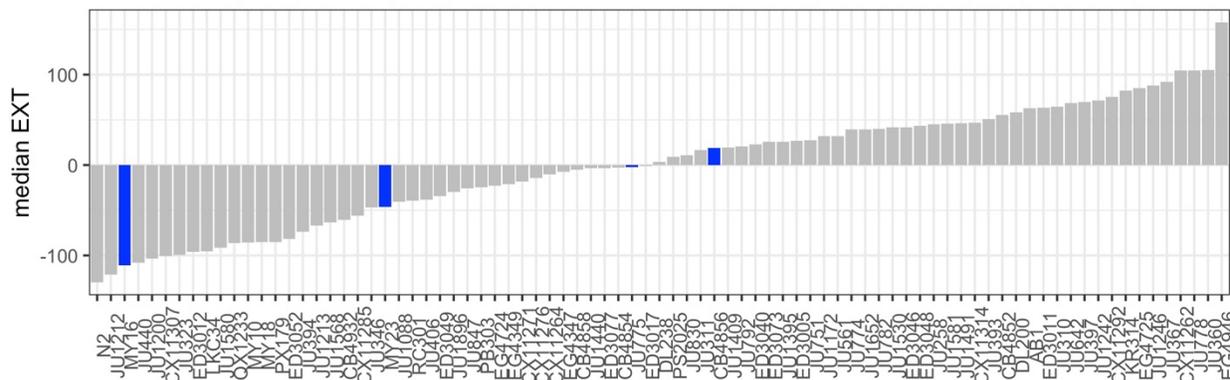
V:11130591



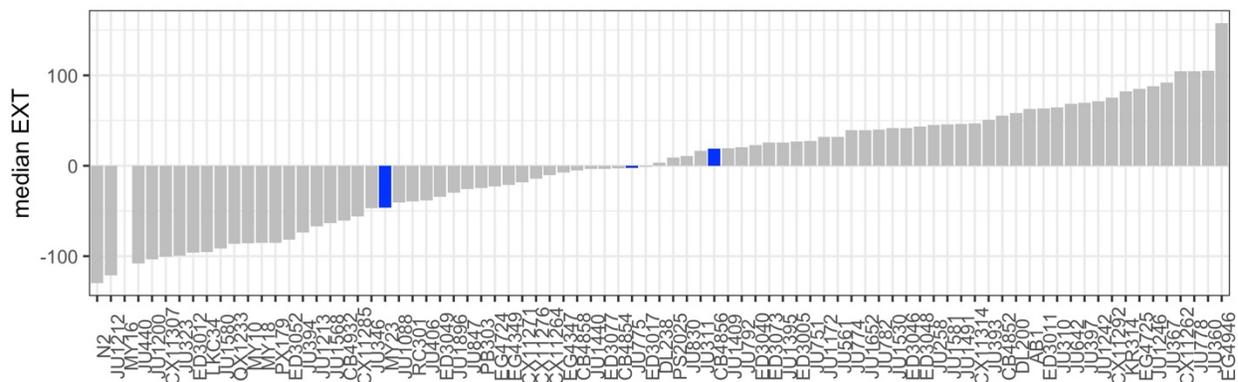
V:11133500



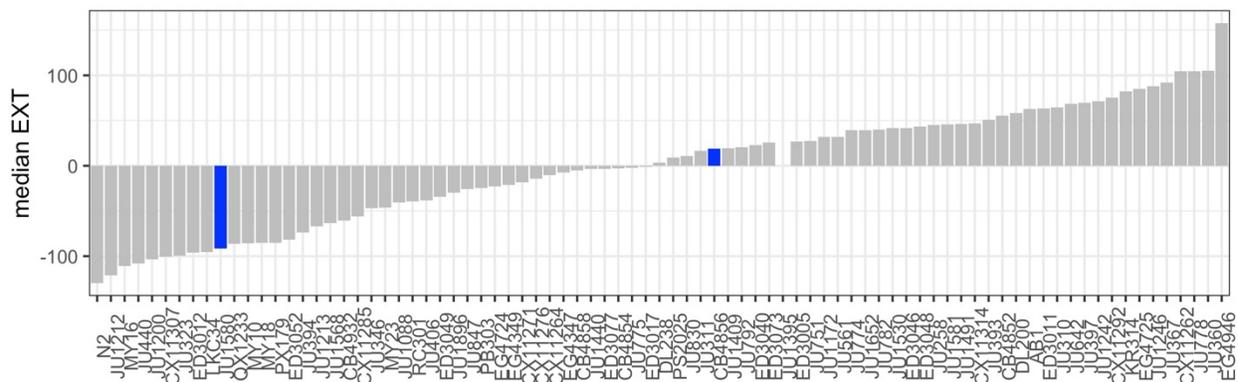
V:11133501



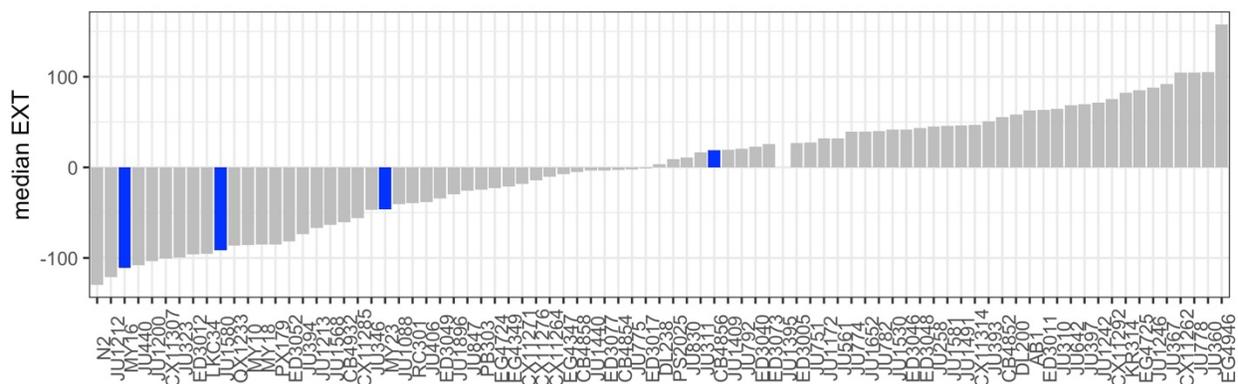
V:11154887



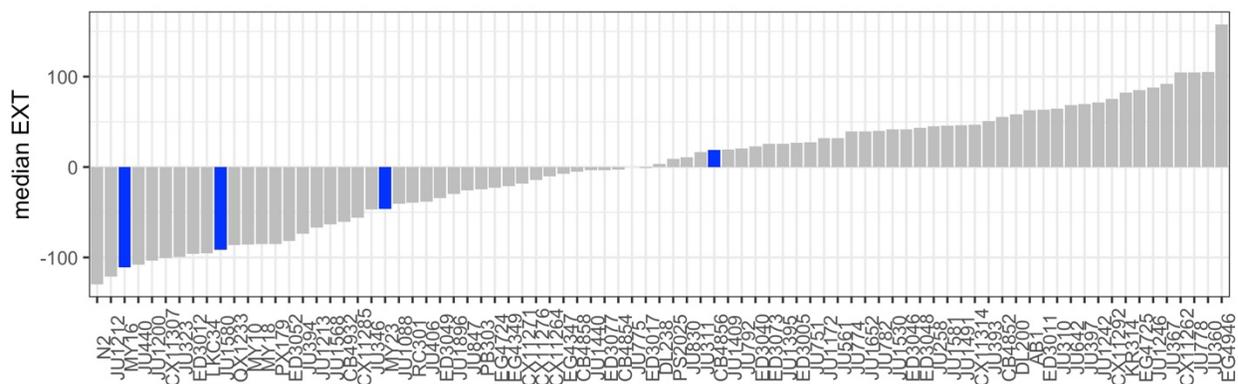
V:11169008



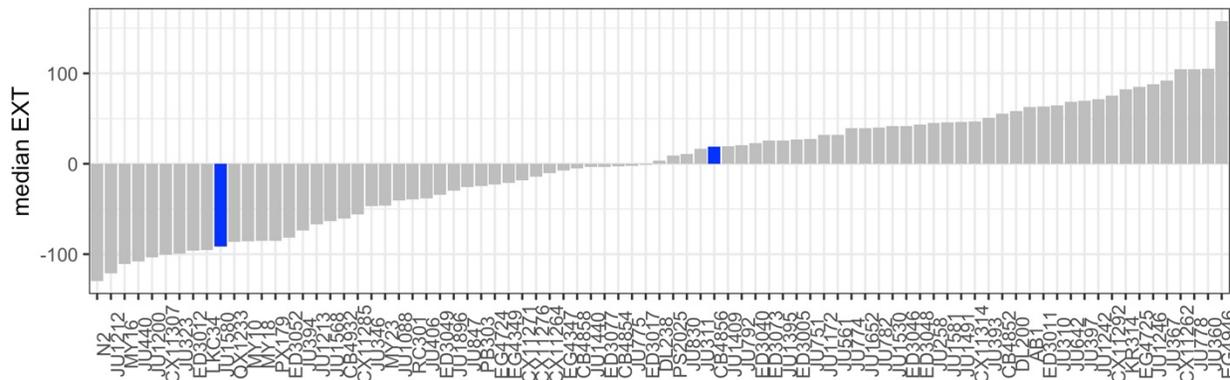
V:11171633



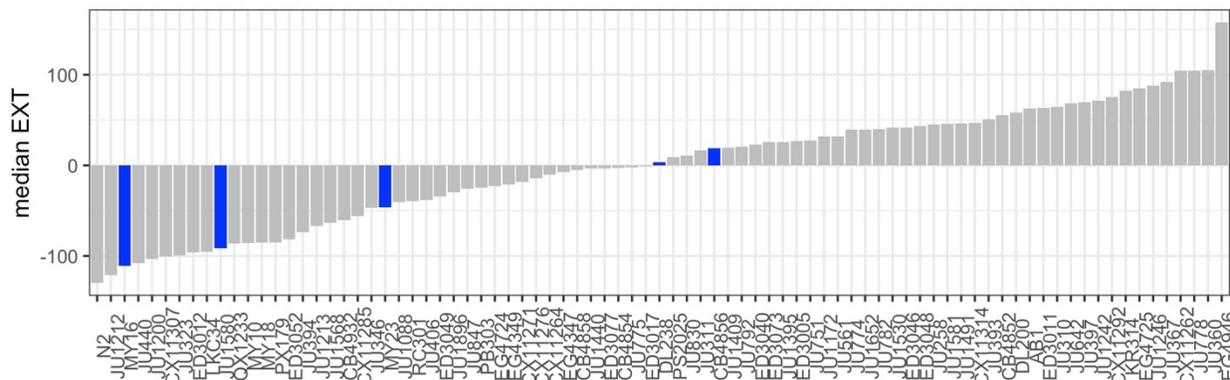
V:11174761



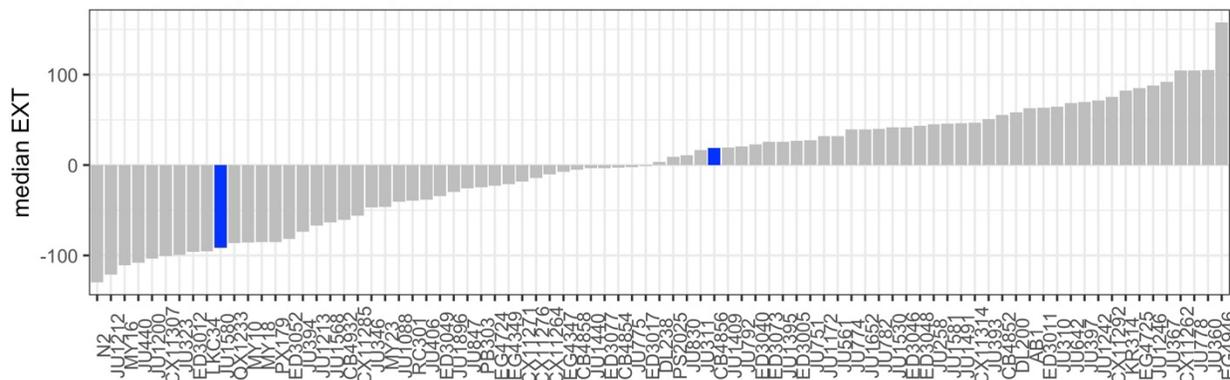
V:11180569

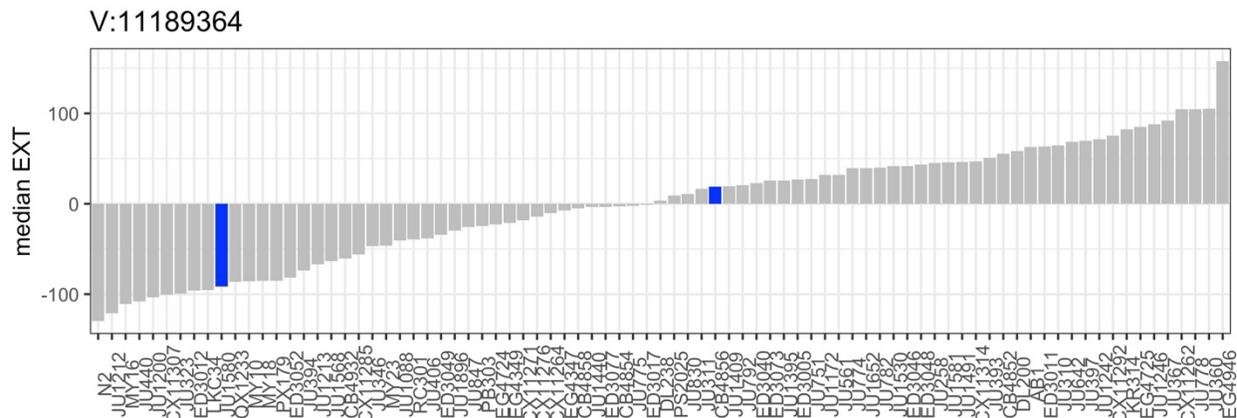


V:11181831



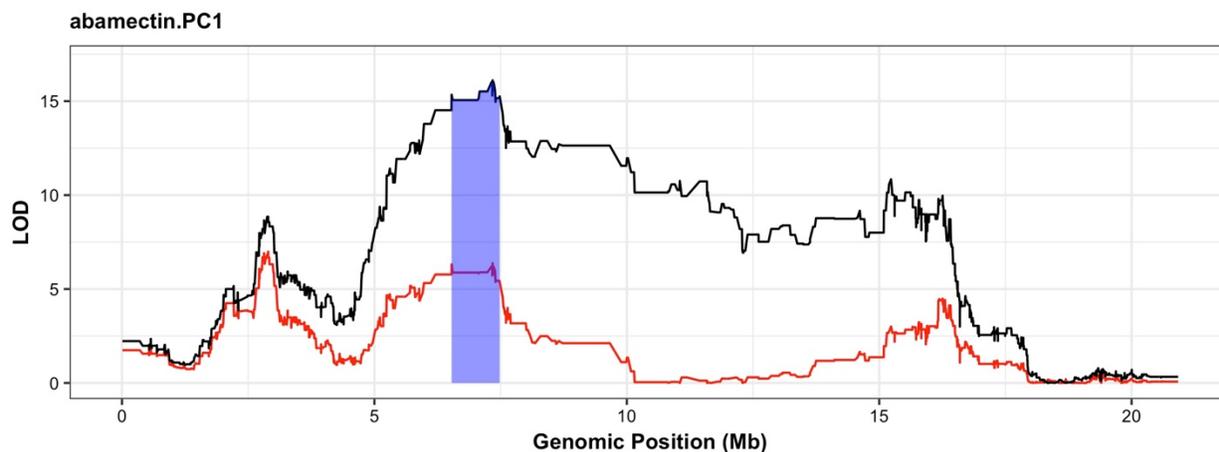
V:11188739

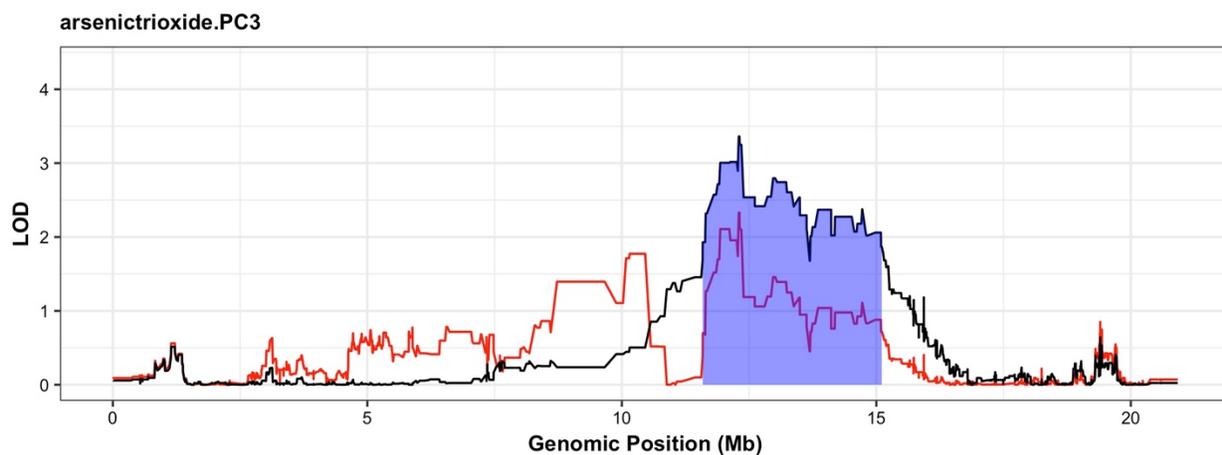
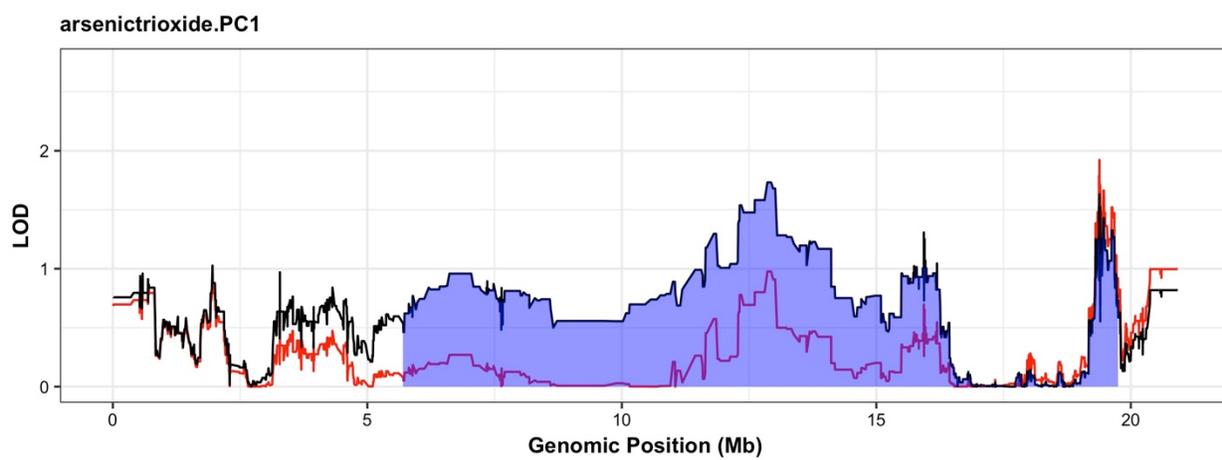
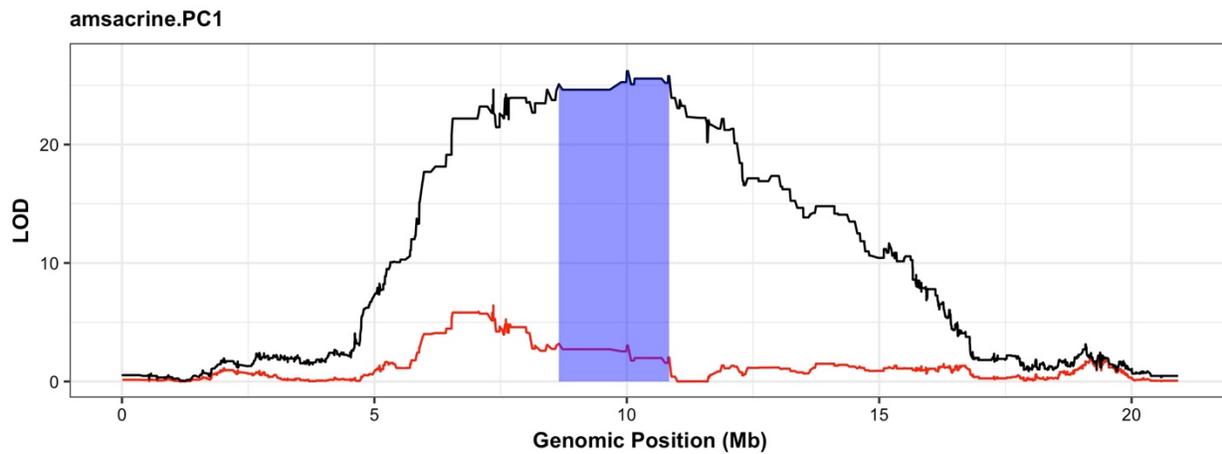


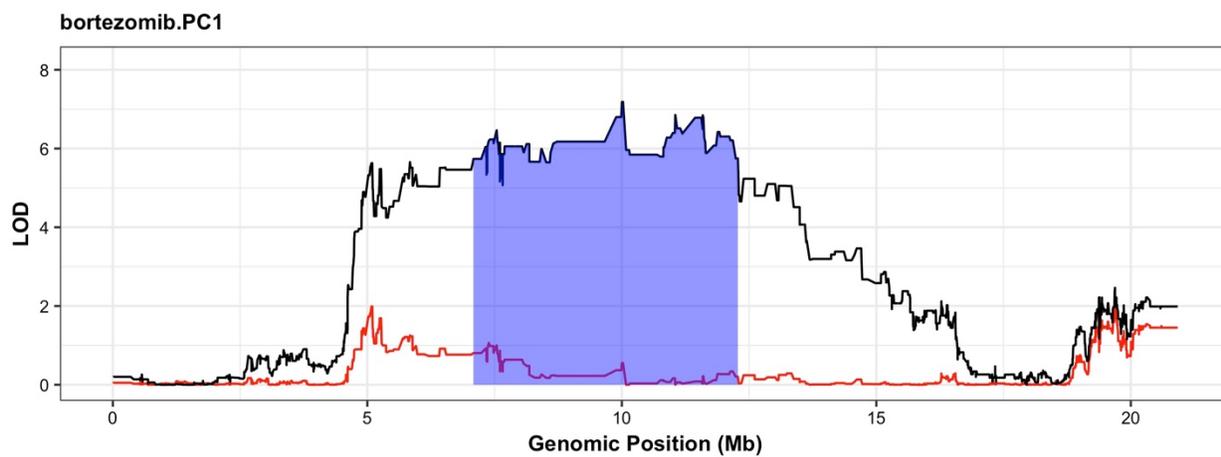
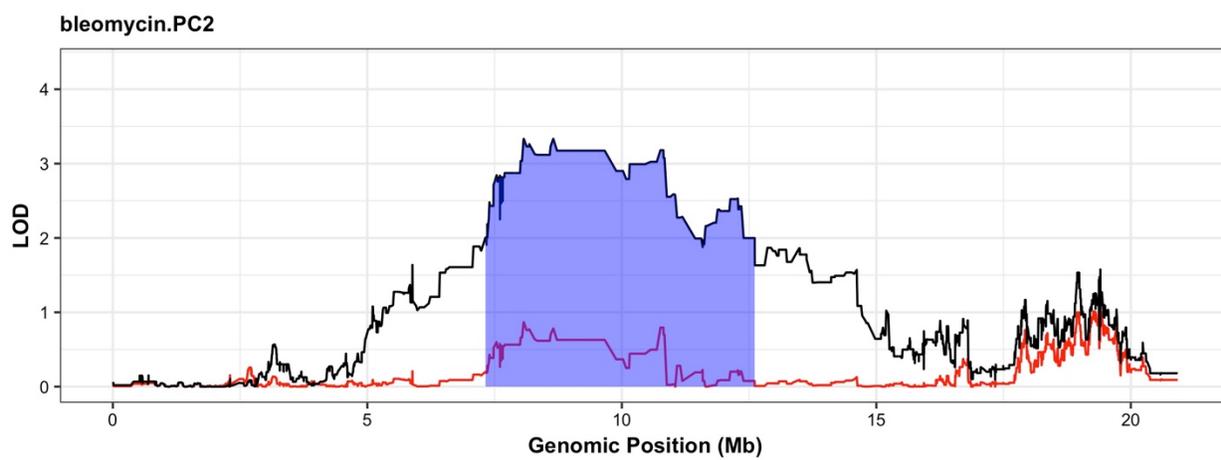
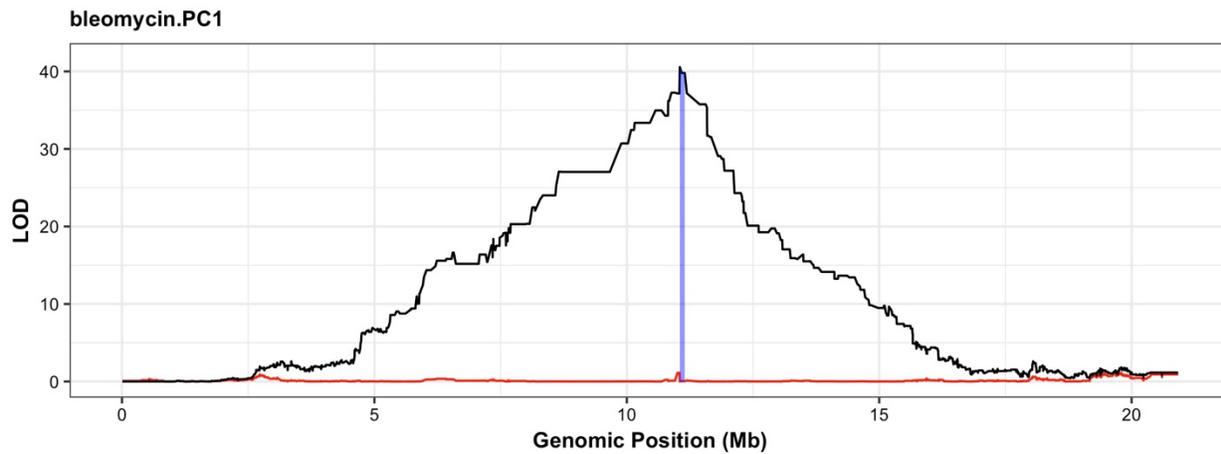


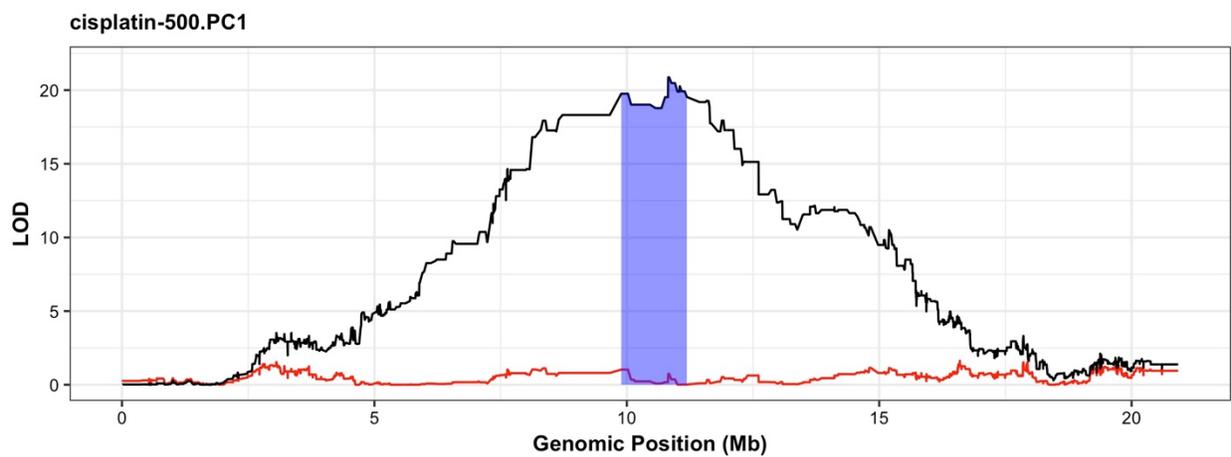
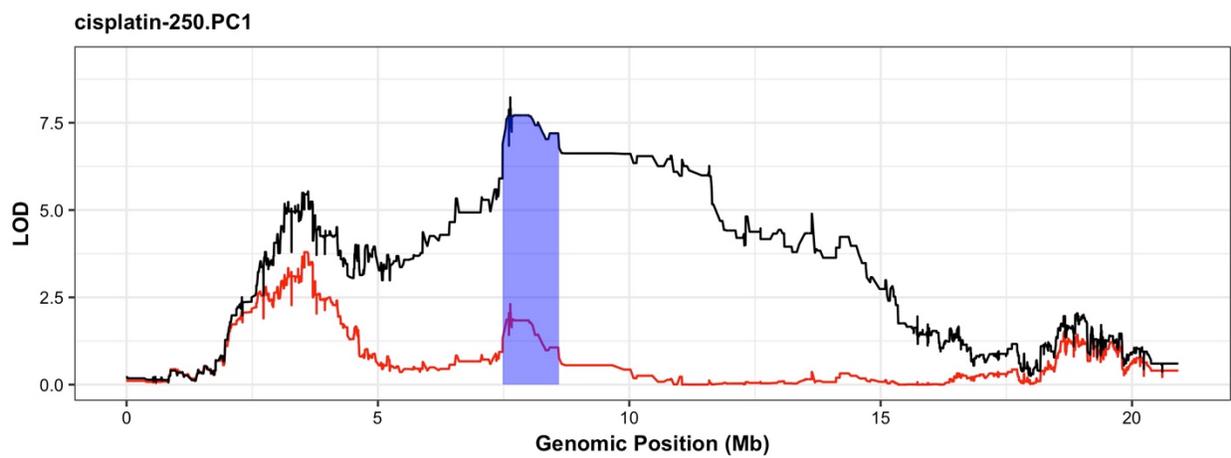
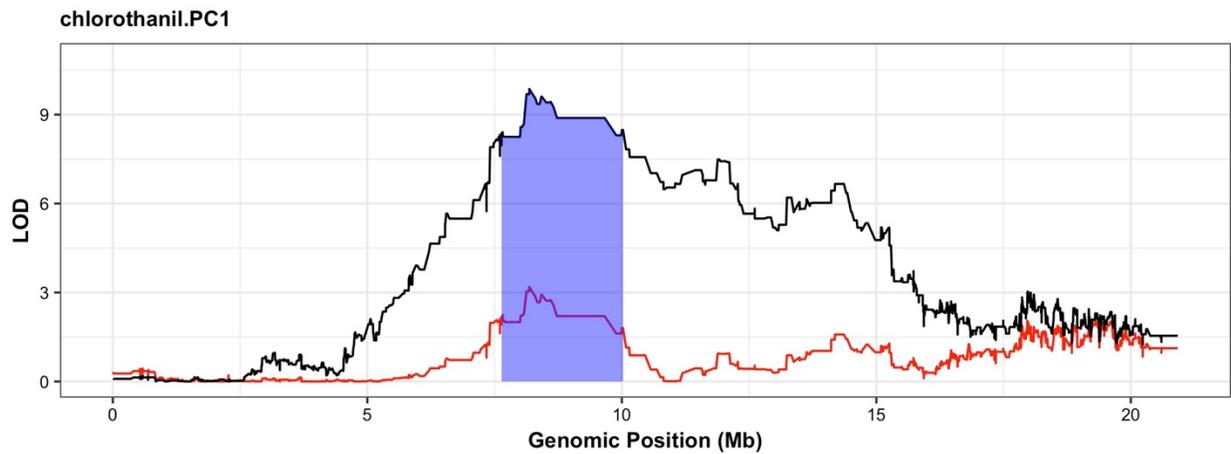
**Figure S2-2 Bleomycin responses for all wild isolates, colored by the genotype at each of the rare alleles around *scb-1***

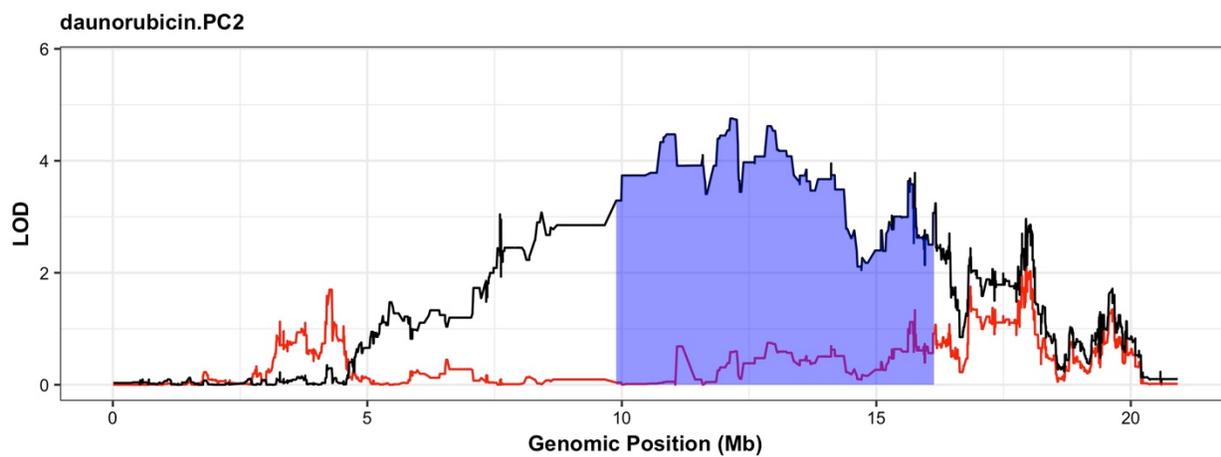
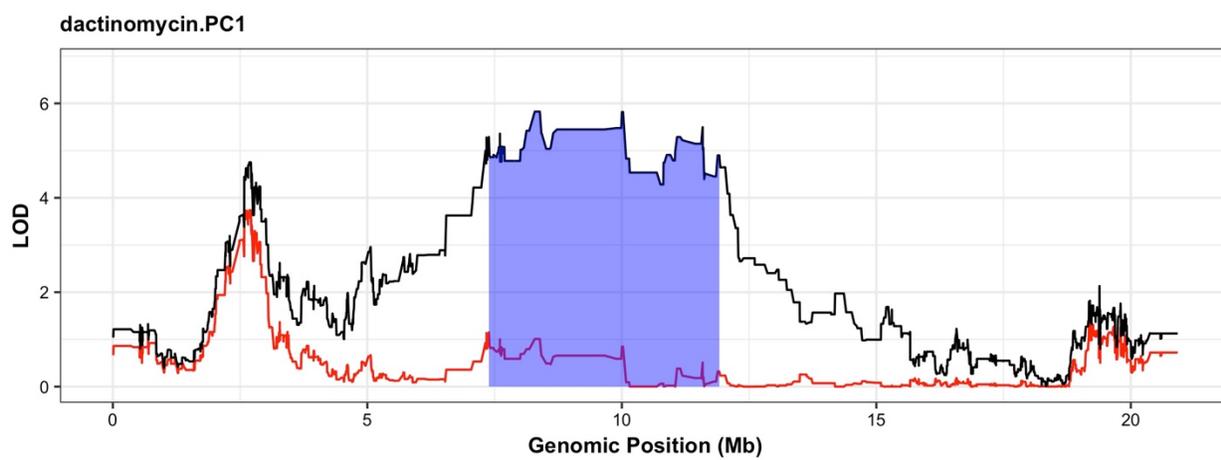
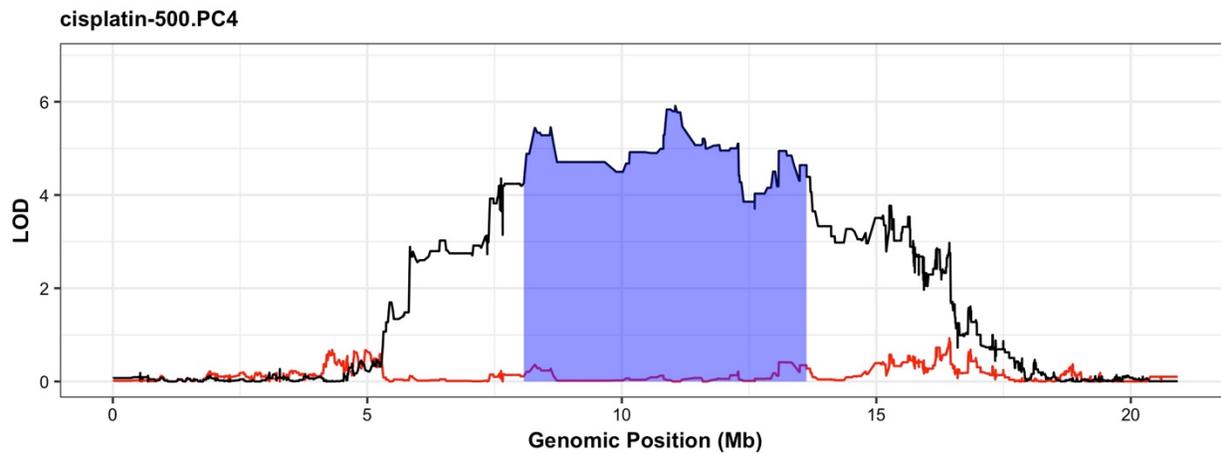
Phenotypes of all wild isolates assayed in bleomycin are shown for each of the 28 variants in the QTL region that are rare (MAF < 0.05) but not unique to CB4856. Each plot shows a bar plot indicating the distribution of residual bleomycin median optical density phenotypes across all 83 wild isolates assayed, arranged by phenotype. The x-axis indicates the strain name, and the y-axis indicates the phenotypic value. For each of the 28 variants, the position is written above the phenotypic distribution. Strains with the alternate allele at that site are colored blue, and strains with the reference allele are colored in grey.

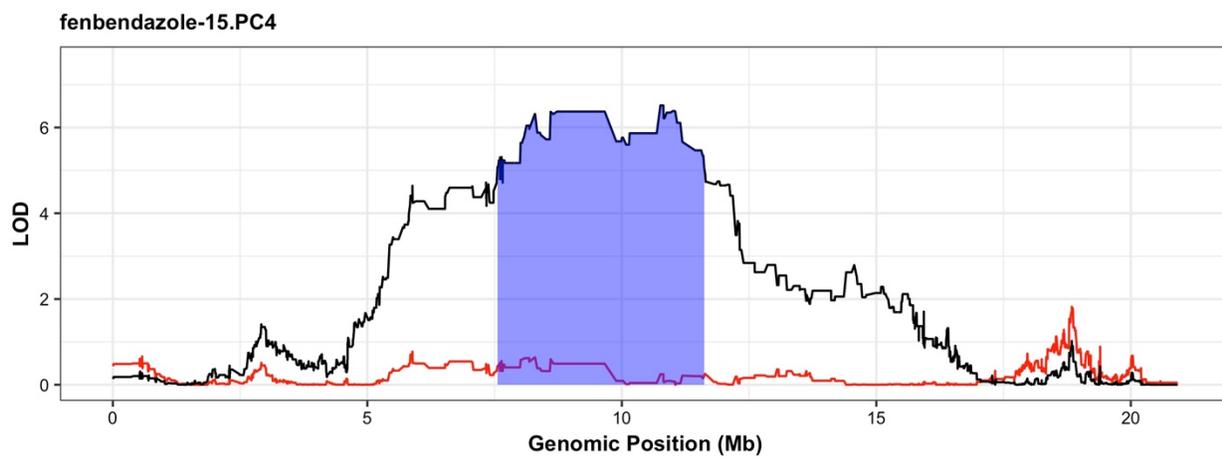
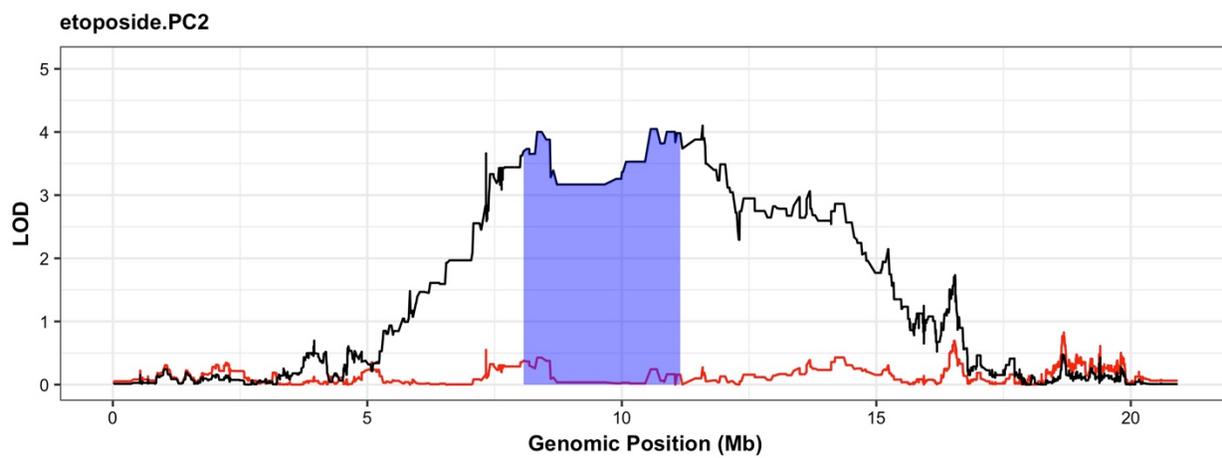
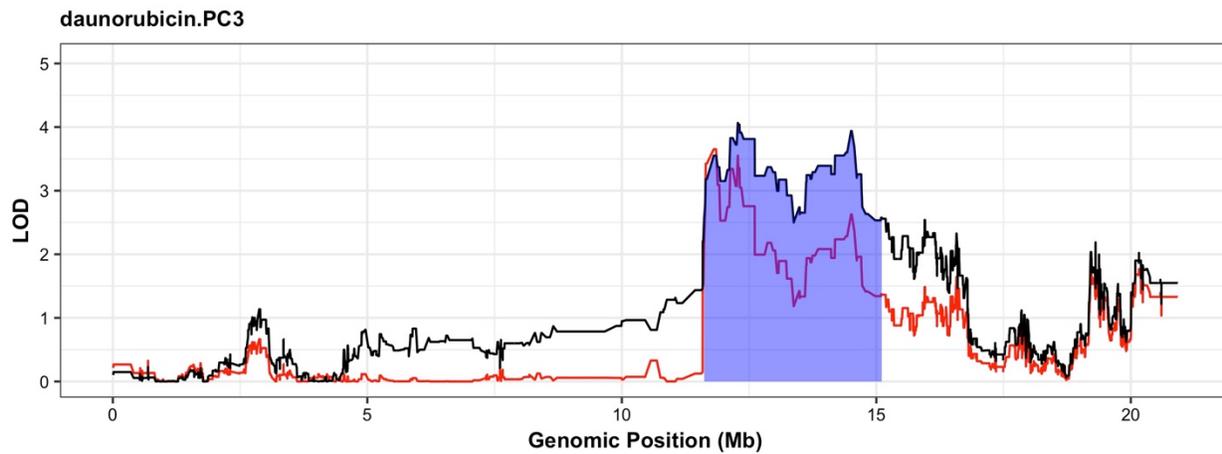


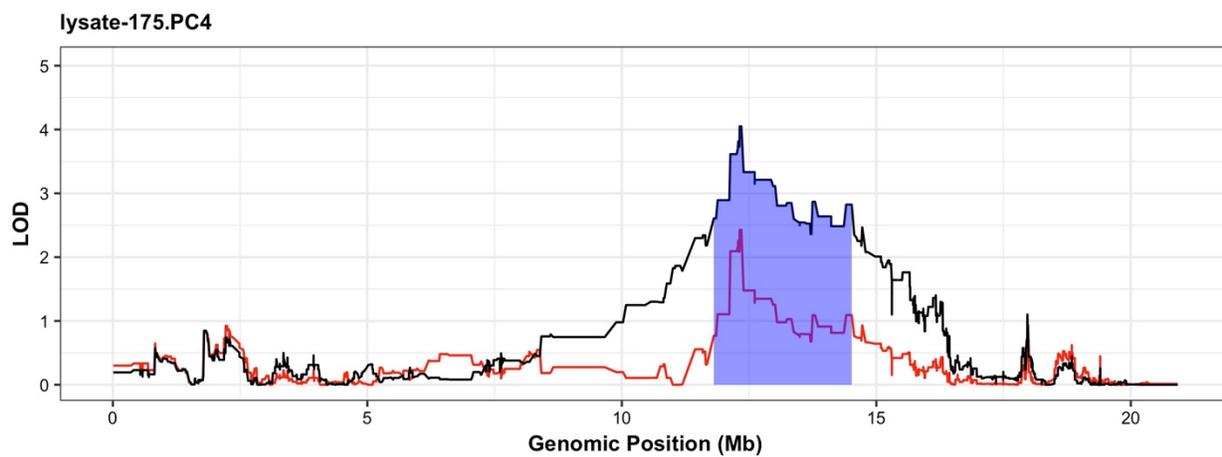
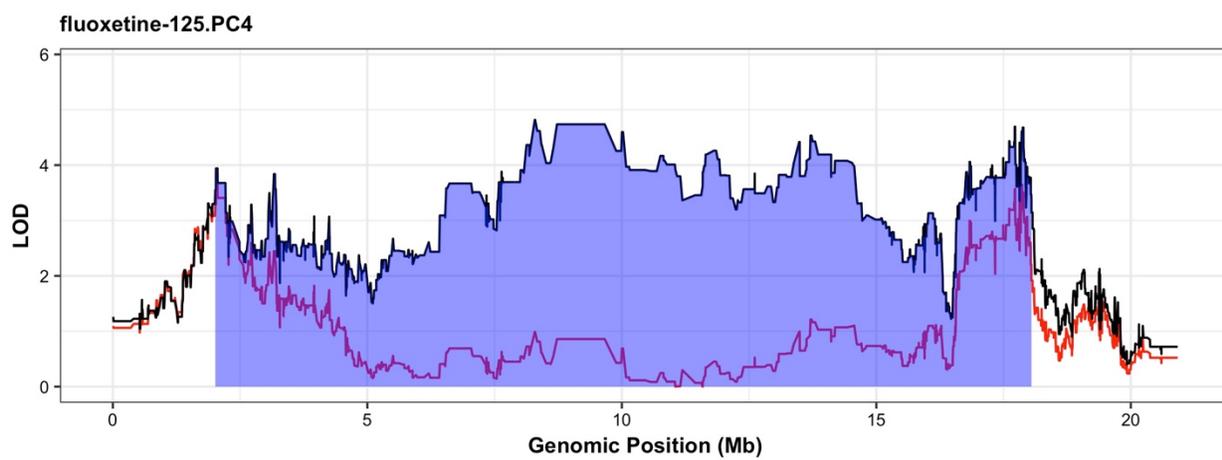
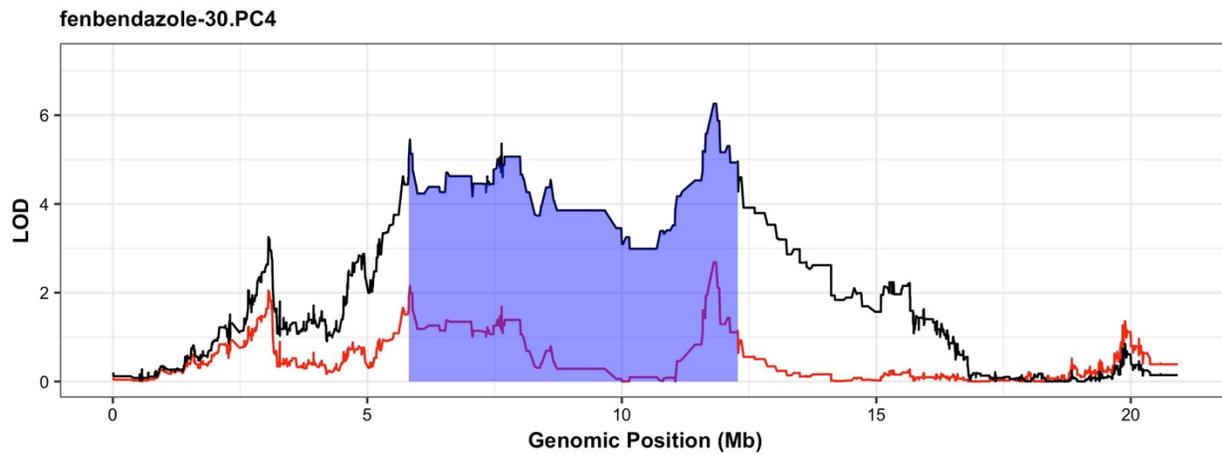


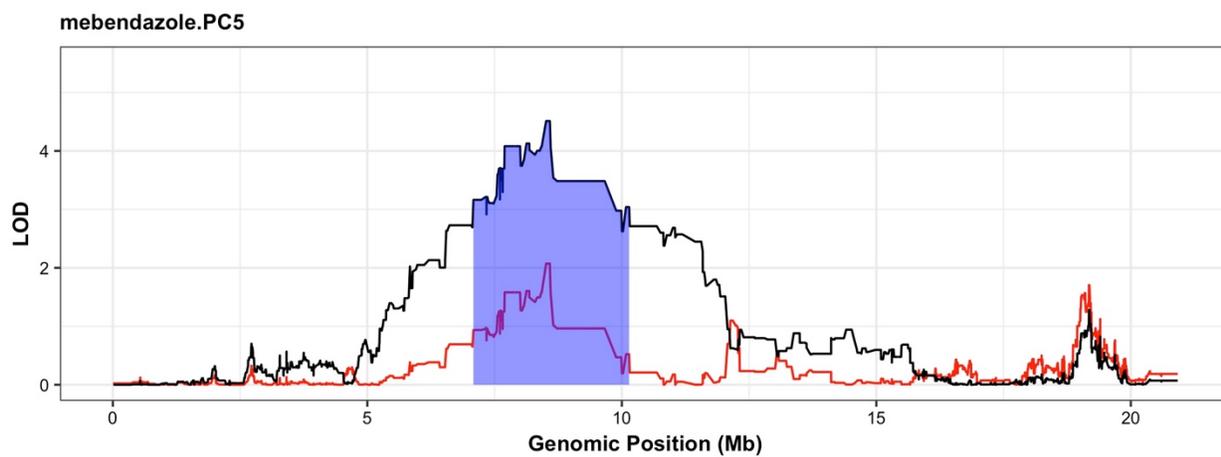
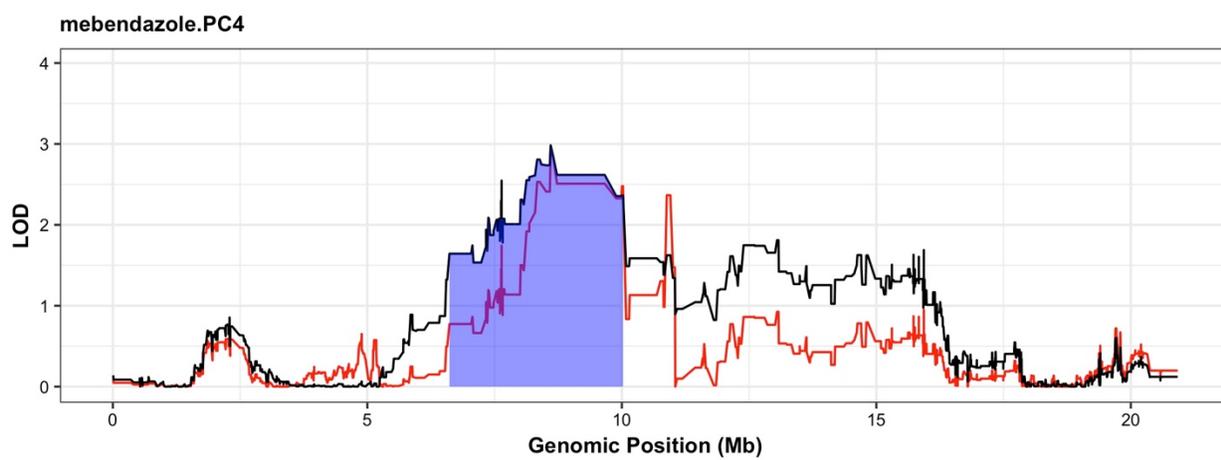
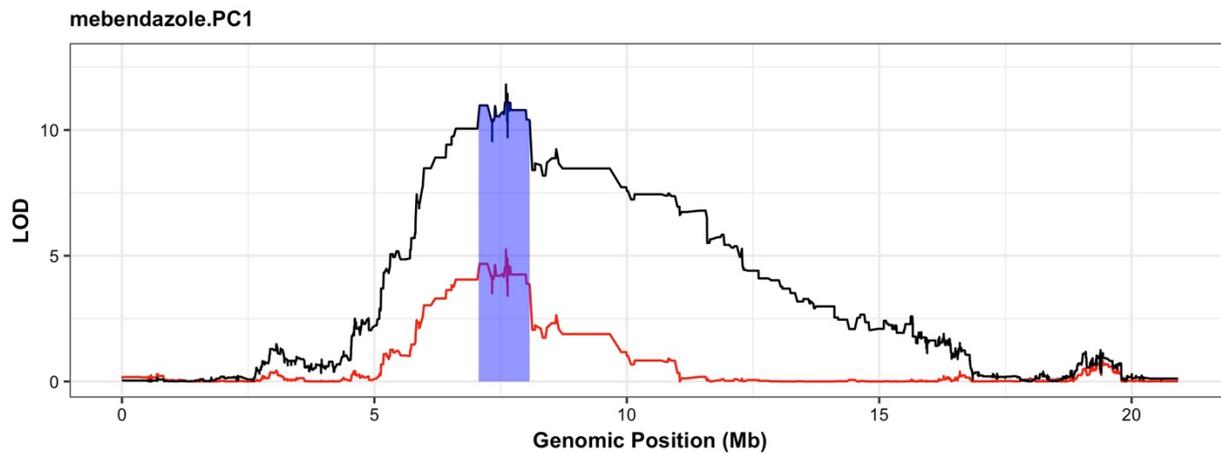


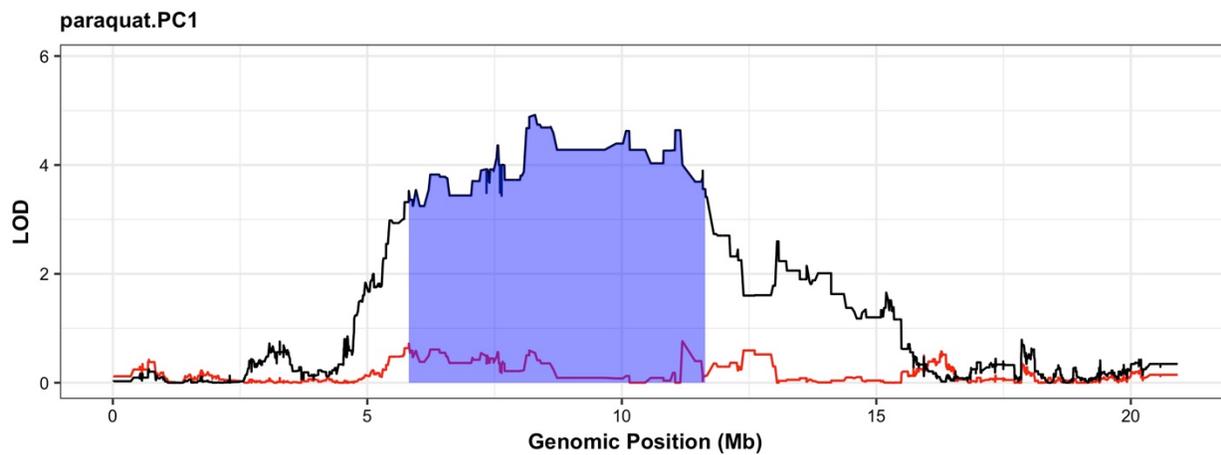
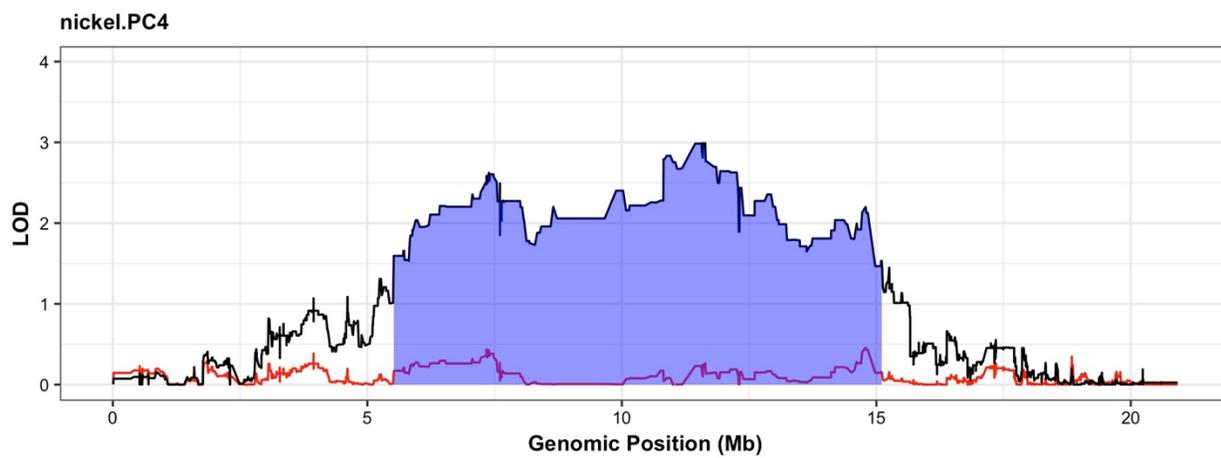
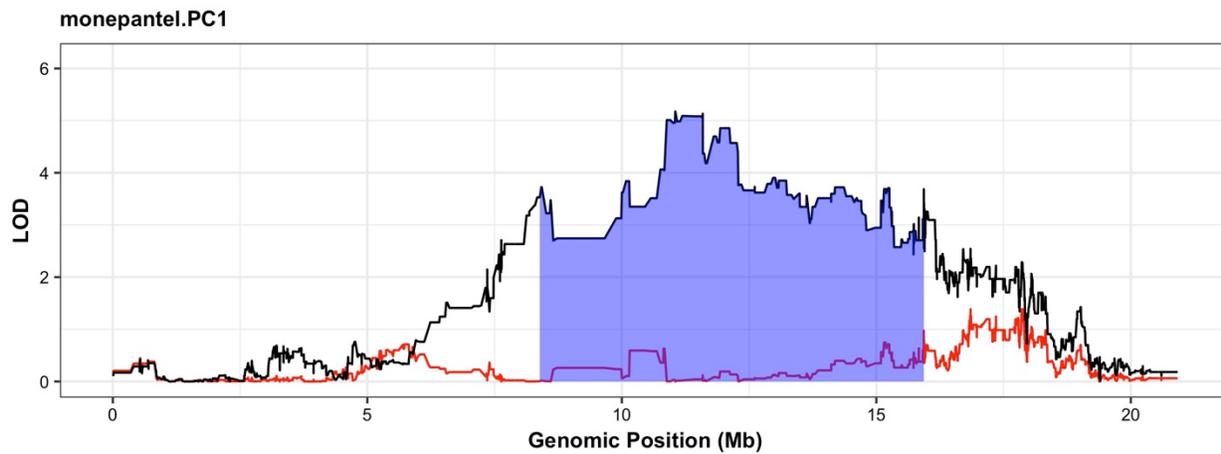


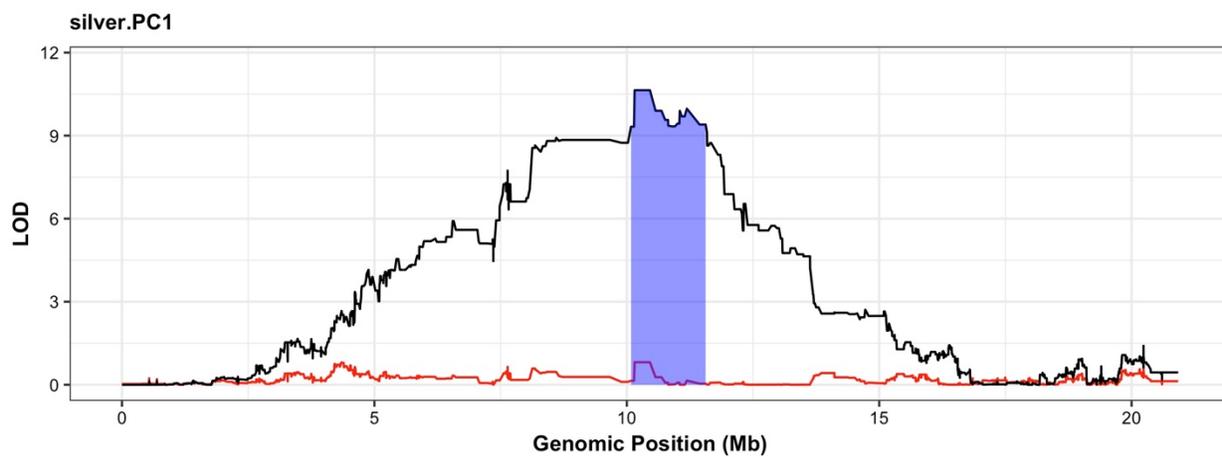
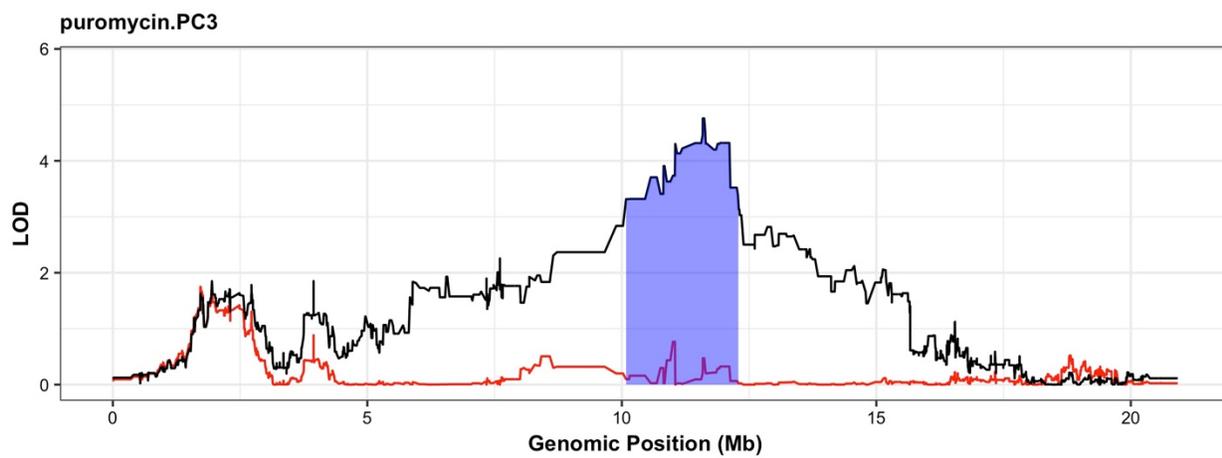
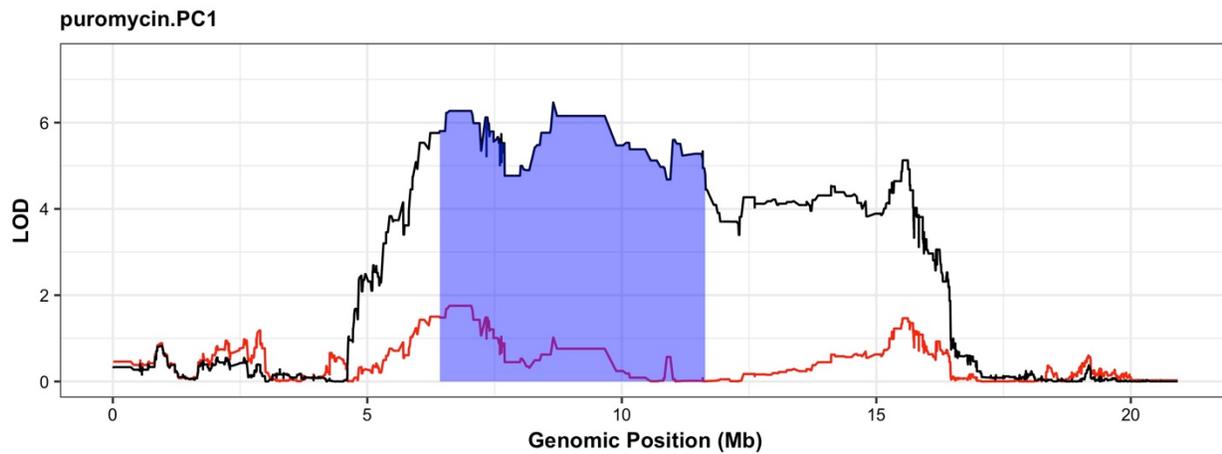


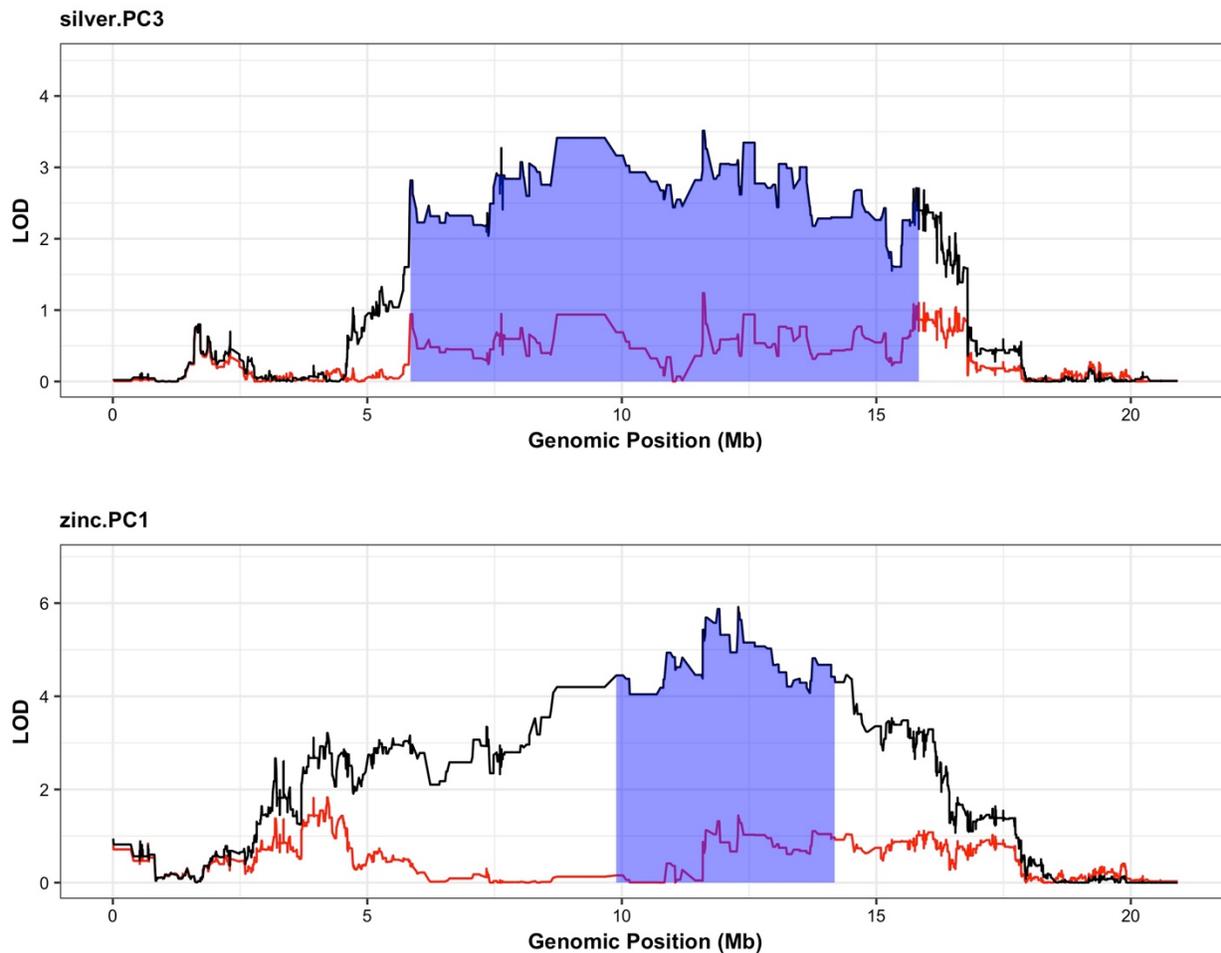












**Figure S2- 3 Chromosome V mappings after bleomycin QTL regression**

The bleomycin QTL peak marker was taken as a cofactor for linkage mapping to assess the impact of *scb-1* variation on other chromosome V mappings. On the x-axis, each of the 4431 markers on chromosome V is tested for correlation with the drug response trait listed at the top of the plot. The y-axis indicates the LOD score of this correlation for each marker, with original mappings in black and mappings after setting *scb-1* as a cofactor in red. The blue region indicates the confidence interval around the QTL from the original mapping.

# Chapter 3 - Shared genomic regions underlie natural variation in diverse toxin responses

## Preface

In my mind, the early years of the Andersen Lab involved each graduate student choosing a drug of interest and “chasing down” its quantitative trait locus in the genome-wide association or linkage mapping. In 2017, we realized that the number of peaks that were identified from the massive 2014 mapping effort greatly outweighed the capacity of a single lab. Katie Evans had just joined the Andersen Lab, and we joined forces to publish some of the data we, as a lab, would probably not pursue further. Katie and I spent many hours together building strains, running high-throughput fitness assays, analyzing data, and writing a manuscript. Given that these data did not comprise the most favorable traits to dissect, the story of the “QTL hotspot paper” took an immense amount of brain power, cooperation, and perseverance. The following chapter is based on the 2018 *Genetics* manuscript, for which Katie and I were equal contributors.

## Abstract

Environmental factors and genetic variants underlie the phenotypic complexity within and among species. Even when environmental variables are tightly controlled, some of the heritability of many complex phenotypes remains elusive. This missing heritability can be attributed to small-effect additive loci, as many yeast and *Arabidopsis* studies find. However, many heritable components of quantitative traits remain undetected in metazoan models. The missing heritability in animals

could be additive, small-effect loci, as in yeast and plants, or undetected interacting loci might be to blame. To obtain sufficient statistical power to detect these elusive additive or interacting loci, high replication of precise phenotypic measurements are required. Here, we use a panel of 296 recombinant strains of the animal model *Caenorhabditis elegans* and a precise fitness assay to detect loci underlying responses to 16 toxins. We use linkage mapping to identify 82 quantitative trait loci (QTL) that contribute to variation in toxin responses and predict the relative contributions of additive and interacting loci toward animal growth. We located three regions of the genome that underlie many toxin responses and call these regions “QTL hotspots”. These hotspots could be enriched for genes in common toxin-response pathways. We validated these QTL hotspots with near-isogenic lines and chromosome substitution strains, implicating additive and interactive loci that underlie toxin-response variation in a metazoan model.

## Introduction

As genome sequencing becomes more feasible and less expensive, high-quality genomic data for a multitude of species rapidly accumulate [218]. When paired with precise, high-throughput phenotypic assays, researchers can use these genomic data to understand how genetic variation impacts complex phenotypes. Linkage mapping has been used to identify many quantitative trait loci (QTL) that underlie phenotypes important to human health [219–221], agriculture and livestock [222–225], and basic biology [109,226,227]. However, some of the heritability of trait variation often cannot be explained by these QTL [228]. Undetected small-effect additive or interacting loci might underlie this missing heritability [229]. Some studies find that interacting loci might explain missing heritability [230–234], whereas others argue that undetected small-effect additive loci are to blame, in cases of low statistical power [235–238]. Quantitative geneticists

have leveraged large numbers of recombinant strains in both yeast and *Arabidopsis* to overcome power limitations and concluded that, when power is sufficient, small-effect additive components can be identified that account for nearly all of the heritability of a given trait [192,229,239]. We require a metazoan system with high statistical power to determine whether this predominantly additive QTL model remains broadly applicable in animals.

*Caenorhabditis elegans* is a tractable model organism that offers well characterized genomic diversity and many methods for precise measurement of quantitative traits. A panel of recombinant inbred advanced intercross lines (RIAILs) has been generated between the N2 and CB4856 strains of *C. elegans* [109,110], and this panel has been used for many linkage-mapping analyses [40,111,112,141–143,156,158,161–167,188,208,240–247]. Additionally, a high-throughput phenotyping platform that rapidly and accurately measures animal fitness has been developed, and it could provide the replication and precision required to detect small-effect additive loci and to determine the relative contributions of additive and/or epistatic loci to trait variation [40,141]. Notably, the combination of these RIAILs and this high-throughput phenotyping platform have facilitated linkage mappings of multiple distinct fitness parameters, and studies have detected a single gene that underlies several fitness-related traits [40,141,142]. This example of pleiotropy suggests that large-scale studies could reveal additional pleiotropic effects.

Studies have used this panel of RIAILs to map pleiotropic QTL that underlie variation in transcript abundance across a wide range of genes [158,248–250]. Variation in one master regulator that lies within an expression QTL hotspot can affect many traits - in this case, expression of many different genes. Other QTL hotspots could impact multiple traits, such as responses to various conditions. For example, in yeast, genetic variants in particular genomic regions have pleiotropic

effects across many conditions [251–253]. Although QTL underlying responses to individual conditions have been identified across multiple animal models [254–258], the existence of QTL hotspots that influence multiple condition responses has yet to be observed broadly in metazoans.

Here, we identified genomic regions implicated in toxin-response variation across a panel of recombinant lines. We found three QTL hotspots that underlie responses to multiple toxins. Additionally, we showed that high replication and a precise fitness assay can enable the identification and validation of even small-effect QTL. We analyzed relative contributions of additive and epistatic genetic loci in various toxin responses in a metazoan model. Finally, we discovered evidence for interactions between loci of the N2 and CB4856 strains that impact several toxin responses and could suggest how large regions of the genome were swept across the species.

## Materials and Methods

### Strains

Animals were grown at 20°C on 6 cm plates of modified nematode growth medium (NGMA), which contained 1% agar and 0.7% agarose, spotted with OP50 bacteria [147]. The two parental strains used in this study were N2 and CB4856. N2 is the canonical laboratory strain of *C. elegans* that has been extensively studied [104]. CB4856 is a well studied Hawaiian wild isolate that is genetically divergent from N2 and has a characterized genome [150–152]. The N2 and CB4856 strains were crossed for several generations to create a panel of recombinant inbred advanced intercross lines (RIAILs) that contain regions of the genome derived from each parental strain.

These RIALs were constructed previously [109,110] and have well characterized genotypes and allele frequencies, and we used this panel of RIALs in our study to identify regions of the genome correlated with drug response. The construction of near-isogenic lines (NILs) and chromosome substitution strains (CSSs) is detailed below. All strains and reagents used in strain constructions are listed in **Table S3-1**.

### High-throughput toxin-response assay

We used the high-throughput assay (HTA) described above. Populations of each strain were passaged on 6 cm plates for four generations to amplify animal numbers and reduce the effects of starvation [141]. Gravid adults were bleached for stage synchronization, and approximately 25 embryos from each strain were aliquoted into 96-well plates at a final volume of 50  $\mu$ L of K medium [189]. The following day, arrested L1 larvae were fed 5 mg/mL HB101 bacterial lysate in K medium (Pennsylvania State University Shared Fermentation Facility, State College, PA; [190] and were grown for 48 hours at 20°C with constant shaking. A large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston, MA) was used to sort three L4 larvae into each well of a 96-well plate that contained 50  $\mu$ L K medium plus HB101 lysate at 10 mg/mL, 50  $\mu$ M kanamycin, and either 1% distilled water (control) or 1% distilled water and a particular toxin (drug). The sorted L4 larvae were grown and propagated for 96 hours at 20°C with constant shaking. The population of parents and progeny were treated with sodium azide (50 mM in M9) and quantified by the BIOSORT for several fitness parameters, including brood size (n), animal length (time of flight, TOF), and optical density (extinction time, EXT).

## Toxin-response trait calculations

Phenotypic measurements collected by the BIOSORT were processed using the R package *easysorter* [144]. Using this package, *read\_data* imported measurements from the BIOSORT and *remove\_contamination* was used to remove contaminated wells from analysis. The *sumplate* function then calculated normalized measurements (norm.n -- brood size normalized to number of animals sorted, norm.EXT -- EXT normalized by TOF measurements) and summary statistics (mean, median, 10<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup> percentile, interquartile range, covariance, and variance) of each trait for the population of animals. A total of 26 HTA traits were measured. When strains were phenotyped across multiple days, the *regress(assay=TRUE)* function was used to fit a linear model with the formula (*phenotype ~ assay*) to account for variation among assay days. Next, the *prune\_outliers()* function removed phenotypic values that were beyond two standard deviations of the mean (unless at least 5% of the strains were outside this range in the case of RIAIL assays). Finally, toxin-specific effects were calculated using the *regress(assay=FALSE)* function from *easysorter*, which fits a linear model with the formula (*phenotype ~ control phenotype*). The residual phenotypic values account for differences among strains that were present in control conditions.

## Dose-response assays

For each toxin, a dose-response experiment was performed using quadruplicates of four genetically diverged strains (N2, CB4856, DL238, and JU258). Animals were assayed using the high-throughput fitness assay and toxin-response trait calculations were performed as described above. The concentration of each toxin that provided a highly reproducible toxin-specific effect with variation between N2 and CB4856 across three distinct traits (brood size, norm.n; mean

length, mean.TOF; and mean optical density, mean.norm.EXT) was selected for linkage mapping experiments. The chosen concentrations and diluents of each toxin are as follows: cadmium 100  $\mu\text{M}$  in water, carmustine 250  $\mu\text{M}$  in DMSO, chlorothalonil 250  $\mu\text{M}$  in DMSO, chlorpyrifos 1  $\mu\text{M}$  in DMSO, cisplatin 250  $\mu\text{M}$  in water, copper 250  $\mu\text{M}$  in water, diquat 250  $\mu\text{M}$  in water, fluoxetine 250  $\mu\text{M}$  in DMSO, Floxuridine (FUdR) 50  $\mu\text{M}$  in water, irinotecan 125  $\mu\text{M}$  in DMSO, mechlorethamine 200  $\mu\text{M}$  in DMSO, paraquat 500  $\mu\text{M}$  in water (50  $\mu\text{M}$  was used for the CSS and NIL assays), silver 150  $\mu\text{M}$  in water, topotecan 400  $\mu\text{M}$  in water, tunicamycin 10  $\mu\text{M}$  in DMSO, and vincristine 80  $\mu\text{M}$  in water. The concentration of paraquat differs to the concentration used previously [109], suggesting why the genetic architectures are different between the two studies. Toxins assayed in this manuscript were purchased from Fluka Chemical (Buchs, Switzerland) (chlorothalonil, #36791-250MG; chlorpyrifos, #45395-250MG; and diquat dibromide monohydrate, #45422-250MG-R), Sigma ([sigma Chemical], St. Louis, MO) (vincristine sulfate salt, #V8879-25MG; cisplatin, #479306-1G; silver nitrate, #209139; carmustine, #1096724-75MG; and topotecan hydrochloride, #1672257-350MG), Calbiochem (San Diego, CA) (tunicamycin, #654380), Aldrich Chemical (Milwaukee, WI) (mechlorethamine hydrochloride, #122564-5G and cadmium chloride #01906BX), Alfa Aesar (irinotecan hydrochloride trihydrate, #AAJ62370-MD), Bioworld (5-fluoro-2'- deoxyuridine, #50256011), Enzo Life Sciences (fluoxetine, #89160-860), Mallinckrodt (cupric sulfate, #4844KBCK), and Chem Service (paraquat, #ps-366).

## Linkage mapping

A total of 296 RIALs were assayed in the high-throughput fitness assay in the presence of each of the 16 toxins listed above as well as control conditions (water or DMSO). Linkage mapping was performed on each of the 384 toxin-response traits (16 toxins x 24 population parameters per toxin) using the R package *linkagemapping* [159]. The genotypic data (WS245) and residual

phenotypic data were merged using the *merge\_pheno* function with the N2xCB4856\_cross object. QTL were detected using the *fsearch* function, which scaled phenotypes to have a mean of zero and variance of one, then calculated logarithm of odds (LOD) scores for each marker and each trait as  $-n(\ln(1-r^2)/2\ln(10))$ , where  $r$  is the Pearson correlation coefficient between RIAL genotypes at the marker and trait values [192]. We noted that this scaling of the data did not impact mappings because scaled and unscaled mappings were identical. The phenotypic values of each RIAL were then permuted randomly while maintaining correlation structure among phenotypes 1,000 times to calculate a significance threshold based on a genome-wide error rate of 5%. The marker with the highest LOD score was then set as a cofactor and mapping repeated iteratively until no significant QTL were detected. Finally, the *annotate\_lods* function was used to calculate the fraction of variation in RIAL phenotypes explained by each QTL. The 95% confidence intervals were defined by markers within a 1.5-LOD drop from the marker with the maximum LOD score.

## Principal component analysis of RIALs

Because some of the 24 population parameters measured by the BIOSORT are highly correlated, a principal component analysis (PCA) was performed. For each growth-response trait, RIAL phenotypic measurements were scaled to have a mean of zero and a standard deviation of one. The *princomp* function within the *stats* package in R [195] was used to run a PCA for each toxin. For each toxin, the minimum number of principal components (PCs) that explained at least 90% of the total phenotypic variance in the RIALs was mapped through linkage mapping, totaling 97 PCs across all toxins. We additionally performed a two-dimensional genome scan using the *scantwo* function in the *qtl* package [259] for all 47 significantly mapped PCs. Significant

interactions were determined by permuting the phenotype data for each PC 1,000 times and determining the 5% genome-wide error rate.

## Heritability estimates

Broad-sense heritability was estimated for each of the 97 PCs using the formula  $H^2 = (\sigma_R^2 - \sigma_P^2) / \sigma_R^2$ , where  $\sigma_R^2$  and  $\sigma_P^2$  are the variance among the RIAIL and parental (N2 and CB4856) phenotypic values, respectively [260]. A variance-component model using the R package *regress* was used to estimate the fraction of phenotypic variation explained by additive genetic factors (“narrow-sense” heritability) [229,261,262]. The additive relatedness matrix was calculated as the correlation of marker genotypes between each pair of strains. In addition, a two-component variance component model was calculated with both an additive and pairwise interaction effect. The pairwise interaction relatedness matrix was calculated as the Hadamard product of the additive relatedness matrix.

## Calculation of hotspots

We estimated centiMorgan distances from recombination events in the RIAIL panel to account for nonuniform distribution of genetic diversity across the genome. The genome was divided into 65 total bins with each bin containing 26 cM. To determine if the 82 QTL significantly clustered around particular genomic regions, we set a threshold for significant QTL hotspots based on the 99<sup>th</sup> percentile of a Poisson distribution with a mean of 1.2 QTL (total QTL/total bins).

## Generation of NILs

Males and hermaphrodites of the desired RIAIL and parental background were crossed in bulk, then male progeny were crossed to the parental strain in bulk for another generation. For each NIL, eight single-parent crosses were performed followed by six generations of propagating isogenic lines to ensure homozygosity of the genome. For each cross, PCR was used to select non-recombinant progeny genotypes within the introgressed region by amplifying insertion-deletion (indel) variants between the N2 and CB4856 genotypes on the left and right side of the introgressed region. NIL strains were whole-genome sequenced as described above to confirm their genotypes. Reagents used to generate all NIL strains and a summary of each introgressed region are detailed in **Table S3-1**. A statistical power calculation was used to determine the minimal number of technical replicates required to observe the predicted phenotypic effect of each QTL at 80% power. The number of technical replicates tested per assay for any given toxin did not exceed 100 because of experimental timing constraints. The PCs that mapped to each NIL region are those with a QTL with a confidence interval that overlaps with or spans the entire introgressed region in the NILs.

## Whole-genome sequence library preparation and analysis

Whole-genome sequencing was performed on recombinant advanced intercross lines (RIAILs) and near-isogenic lines (NILs) using low-coverage sequencing. DNA was isolated from 100-300  $\mu\text{L}$  of packed worms using Omega BioTek's EZ 96 Tissue DNA Kit (catalog no. D1196-01). All samples were diluted to 0.2 ng/ $\mu\text{L}$  and incubated with diluted Illumina transposome (catalog no. FC-121-1031). Tagmented samples were amplified with barcoded primers. Unique libraries (192) were pooled by adding 8  $\mu\text{L}$  of each library. The pooled material was size-selected by separating

the material on a 2% agarose gel and excising the fragments ranging from 400-600 bp. The sample was purified using Qiagen's Gel Extraction Kit (catalog no. 28706) and eluted in 30  $\mu$ L of buffer EB. The concentration of the purified sample was determined using the Qubit dsDNA HS Assay Kit (catalog no. Q32851). RIAILs and NILs were sequenced at low coverage (mean = 2.13x) using the Illumina HiSeq 2500 platform with a paired-end 100 bp reaction lane. RIAIL and NIL genotypes were imputed using VCF-kit [191]. To determine genotypes, a list of filtered, high-quality sites ( $n = 196,565$ ) where parental strains possess different genotypes was extracted from a previously established variant dataset [132].

## Generation of chromosome substitution strains (CSS)

CSSs were generated by crossing N2 and CB4856 parental strains, and mating cross progeny, to each parental genotype. For each CSS, eight crosses were performed followed by six generations of propagating isogenic lines to ensure homozygosity of the genome. For each cross, PCR amplicons for indels on the left and right of the introgressed region were used to confirm progeny genotypes and select non-recombinants within the introgressed region. CSSs were whole-genome sequenced as described above to confirm their genotype. Reagents used to generate CSSs are detailed in **Table S3-1**. As described for NIL assays, power calculations were performed to determine the number of technical replicates required to observe the predicted phenotypic effect of the CSSs.

## Selection of traits to categorize in CSS and NIL assays

Pairwise correlations of RIAIL phenotypes among the 24 growth-response traits measured by the BIOSORT were calculated using the *cor* function within the *stats* package in R, with the *use*

argument set to “pairwise.complete.obs.” For each toxin, hierarchical clustering was performed using the function *hclust* from the *stats* package [195]. *Cutree* was then used to group the resulting dendrogram into *k* groups, where *k* is equal to the minimum number of PCs that explained at least 90% of the phenotypic variance in the RIALs. For each PC that mapped to a hotspot, the growth-response trait that was most correlated to that PC, as well as all growth-response traits within that cluster of the dendrogram, were assayed in NIL and CSS experiments.

## Categorization of CSS and NIL results

Toxin responses for NILs and CSSs were tested using the high-throughput fitness assay for traits correlated with mapped PCs as described above. Complete pairwise statistical analyses of strains was performed for each trait tested in all CSS and NIL assays (Tukey’s honest significant difference test). A *P* value of  $P < 0.05$  was used as a threshold for statistical significance. NIL recapitulation was defined by the significance and direction of effect of the NIL compared to the parental strains. Six categories were defined: (1) “no parental difference,” (2) “recapitulation,” (3) “no QTL effect,” (4) “bidirectional interaction,” (5) “unidirectional interaction,” and (6) “miscellaneous”. Traits for which N2 and CB4856 phenotypes were not statistically different comprise the “no parental difference” category and were not further categorized. Traits in the “recapitulation” category must satisfy the following criteria: significant difference between the parental strain phenotypes, significant difference between phenotypes of each NIL and the parent that shares its background genotype, and both NILs must display the expected direction of effect of the introgressed genotype. Traits with “no QTL effect” displayed a significant parental phenotypic difference and the phenotype of each NIL was not statistically different from the phenotype of the parent sharing its background genotype. Traits that have a “bidirectional interaction” must display a significant parental phenotypic difference, the phenotypes of both NILs

must be significantly different from phenotypes of both parents, and the phenotypes of both NILs must be transgressive (lie beyond the phenotypic range of the parental strains). Lastly, traits with a “unidirectional interaction” were categorized similarly to the bidirectional interaction, except only one NIL must display a transgressive phenotype, and the other NIL either shows no QTL effect or recapitulation. Traits that did not fit these descriptions were categorized as “miscellaneous”.

Traits in the chromosome V hotspot were further categorized using the combined data from both the CSS and NIL assays. Seven categories were defined: (1) “no parental difference”, (2) “recapitulation”, (3) “no QTL effect”, (4) “external interchromosomal interaction” (uni- or bidirectional), (5) “internal interchromosomal interaction” (uni- or bidirectional), (6) “intrachromosomal interaction” (uni- or bidirectional), and (7) “miscellaneous”. No parental difference was defined by traits in which the parental strains were either not significantly different from each other or did not have the same direction of effect in both the CSS and NIL assays. “Recapitulation” and “no QTL effect” traits were defined by traits that were classified as either recapitulating or no QTL effect, respectively, in both assays. Traits displaying an “external interchromosomal interaction” show evidence for interaction in the CSS but no interaction (either recapitulating or no QTL effect) in the NIL. On the other hand, traits displaying an “internal interchromosomal interaction” showed evidence of the same interaction for both the CSS and the NIL assays. Finally, traits displaying an “intrachromosomal interaction” showed evidence of an interaction in the NIL but not in the CSS assay. All other traits that did not fit these descriptions were categorized as “miscellaneous”.

## Statistical analysis

All statistical tests of phenotypic differences in the NIL and CSS assays were performed in R (version 3.3.1) using the *TukeyHSD* function [195] on an ANOVA model with the formula (*phenotype ~ strain*). The *P* values of individual pairwise strain comparisons were reported and a *P* value of  $P < 0.05$  was deemed significant. The direction of effect of each NIL was determined by comparing the median phenotypic value of the NIL replicates to that of each parental strain. NILs whose phenotypes were significantly different from both parents and whose median lied outside of the range of the parental phenotype medians were considered hypersensitive or hyper-resistant. Comparing LOD scores, and variance explained between traits with “no parental effect” and traits with a significant parental effect in the NIL assays, was performed using a Wilcoxon rank sum test with continuity correction using the *wilcox.test* function in R [195].

## Results

### Many QTL underlie responses to 16 diverse toxins

We used a high-throughput fitness assay (**Figure 1-1**) to measure the responses of four genetically divergent strains to various concentrations of 16 toxins, comprising chemotherapeutics, heavy metals, pesticides, and neuropharmaceuticals (**Figure S3-1, Table 3-1**). We selected the concentration of each toxin with maximal broad-sense heritability between two strains, N2 (the laboratory strain) and CB4856 (a Hawaiian wild isolate) (**Table 3-1**). For the selected concentration of each toxin, we measured 24 growth-response traits for a panel of 296 RIALs generated between these two divergent strains [109]. We then performed linkage mapping for each of the 24 traits across each of the 16 toxins, for a total of 384 toxin-trait mappings, and

we identified 462 QTL across 247 traits (data available at Figshare: <https://doi.org/10.25386/genetics.7158911>). Because the high correlation among the 24 traits measured could increase the false-positive rate of QTL detection (**Figure S3-2**), we performed principal component analysis (PCA) for each toxin. We selected the minimum number of principal components (PCs) that explained at least 90% of the total phenotypic variance within each toxin, for a total of 97 PCs across all toxins (minimum of five PCs and a maximum of eight PCs per toxin, **Table 3-2**). We then performed linkage mapping to identify QTL that underlie variation in these 97 PCs.

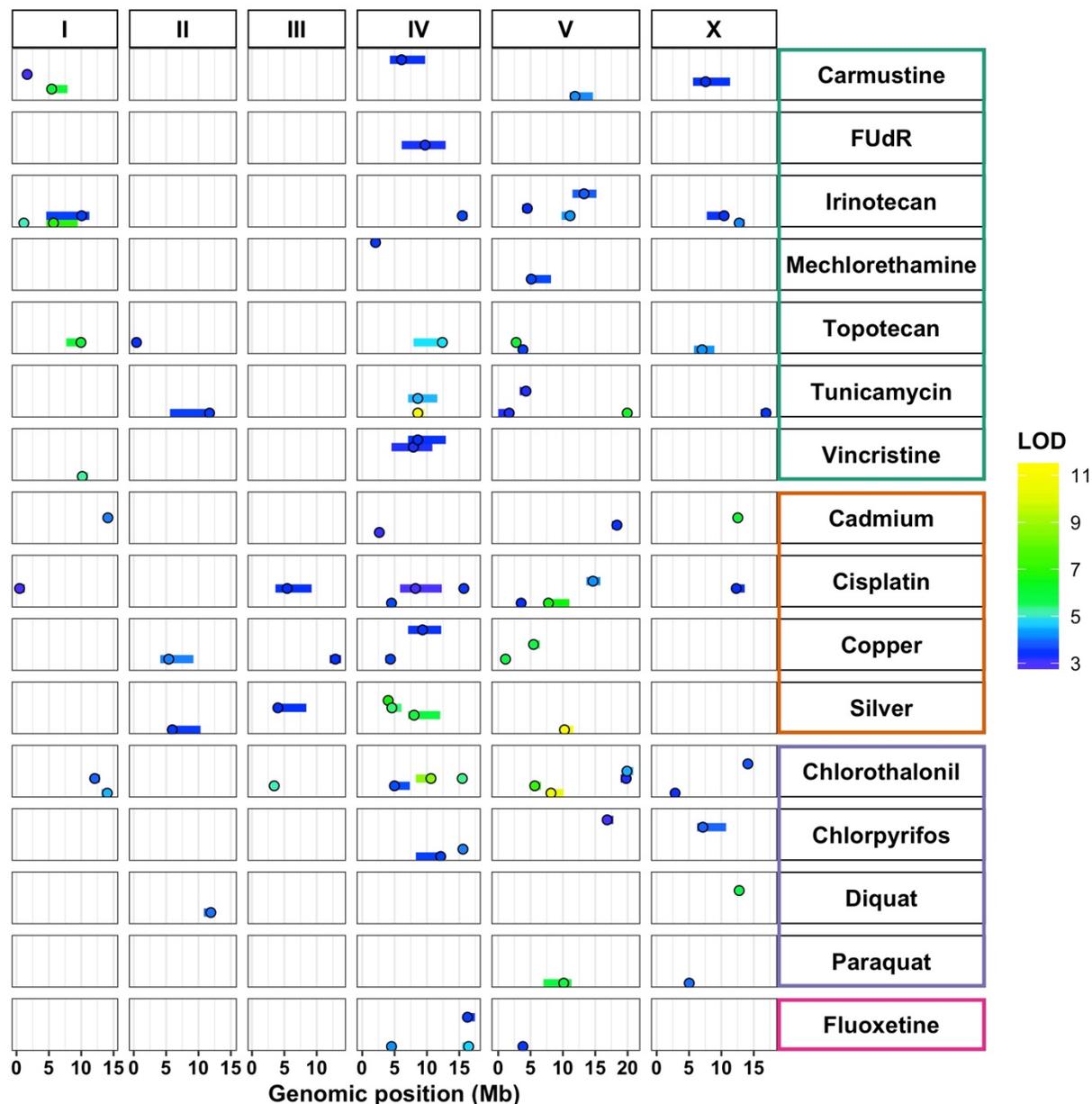
**Table 3-1** List of doses tested for each toxin

Toxin	Class	Doses Tested ( $\mu\text{M}$ )	Mapping Concentration ( $\mu\text{M}$ )	Diluent
Cadmium	Heavy Metal	100, 200, 300, 400	100	Water
Carmustine	Chemotherapeutic	125, 250, 500, 1000	250	DMSO
Chlorothalonil	Pesticide	125, 250, 500, 1000	250	DMSO
Chlorpyrifos	Pesticide	0.25, 0.5, 1, 2	1	DMSO
Cisplatin	Chemotherapeutic	125, 250, 500, 1000	250	Water
Copper	Heavy Metal	625, 125, 250, 500	250	Water
Diquat	Pesticide	250, 500, 1000, 2000	250	Water
Fluoxetine	Neuropharmaceutical	625, 125, 250, 500	250	DMSO
FUdR	Chemotherapeutic	37.5, 50, 75, 100	50	Water
Irinotecan	Chemotherapeutic	625, 125, 250, 500	125	DMSO
Mechlorethamine	Chemotherapeutic	200, 300, 400, 500	200	DMSO
Paraquat	Pesticide	500, 1000, 2000, 4000	500	Water
Silver	Heavy Metal	75, 150, 300, 500	150	Water
Topotecan	Chemotherapeutic	50, 100, 200, 400	400	Water
Tunicamycin	Chemotherapeutic	5, 10, 15, 20	10	DMSO
Vincristine	Chemotherapeutic	20, 40, 60, 80	80	Water

**Table 3-2** Number of principal components that explain at least 90% of the variance in the RIALs for each toxin

Toxin	Number of Principal Components	Cumulative Variance Explained
Cadmium	6	90.93%
Carmustine	6	90.26%
Chlorothalonil	6	91.98%
Chlorpyrifos	7	91.60%
Cisplatin	6	90.75%
Copper	8	91.64%
Diquat	6	90.95%
Fluoxetine	7	90.09%
FUdR	7	91.47%
Irinotecan	5	92.78%
Mechlorethamine	7	91.93%
Paraquat	5	90.50%
Silver	5	92.41%
Topotecan	5	90.39%
Tunicamycin	5	93.28%
Vincristine	6	92.60%

We detected a total of 82 significant QTL (across 47 PCs) from the 97 PCs tested (**Figure 3-1**, Figshare <https://doi.org/10.25386/genetics.7158911>). Although none of these toxin-response QTL were shared robustly across all of the toxins tested, we did find that the majority of chromosome I QTL were detected in responses to chemotherapeutics. Interestingly, almost every toxin (with the exception of FUdR) had PCs that mapped to at least two different chromosomes, highlighting the diverse architectures implicated across traits, even within a single toxin. In general, we found that the majority of the QTL (61%) mapped to chromosomes IV and V.

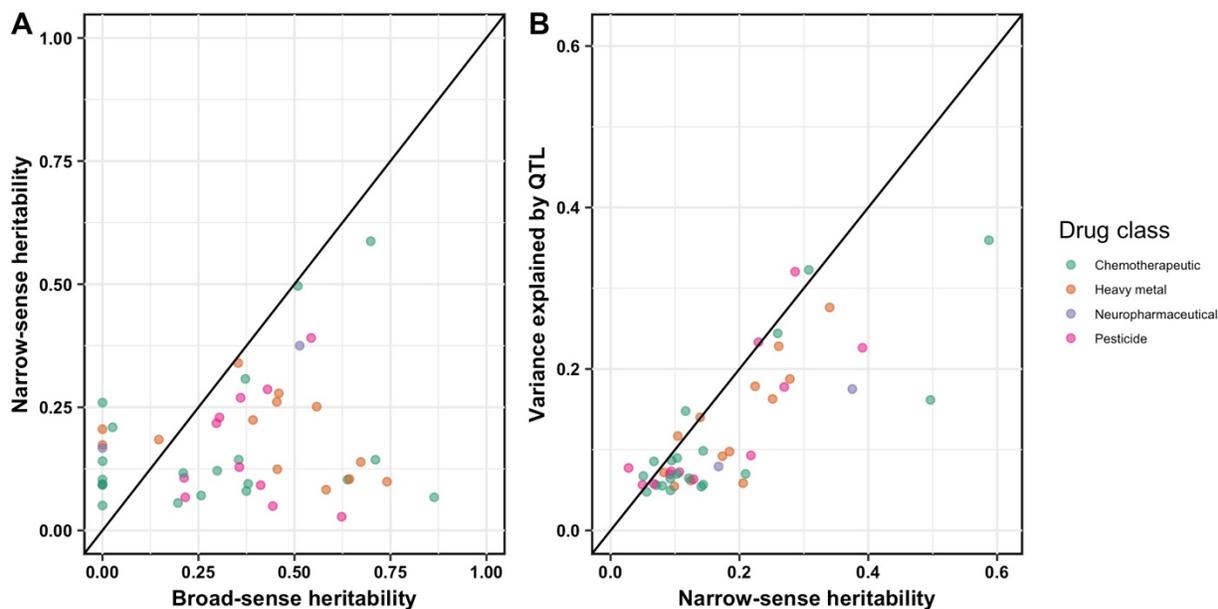


**Figure 3-1 Toxin-response variation maps to many regions of the genome**

Diverse genetic architectures are implicated in responses to 16 toxins. Linkage mapping results for principal components that represent 82 QTL across 16 toxins, comprising chemotherapeutics (teal), heavy metals (orange), pesticides (purple), and neuropharmaceuticals (pink) are plotted. Genomic position (Mb) is shown along the x-axis, split by chromosome, and each of the 47 principal components with a significant QTL is plotted along the y-axis. Each QTL is plotted as a point at the location of the most significant genetic marker and a line indicating the 95% confidence interval. QTL are colored by the logarithm of the odds (LOD) score, increasing in significance from blue to green to yellow.

## Both additive and interactive QTL underlie toxin responses

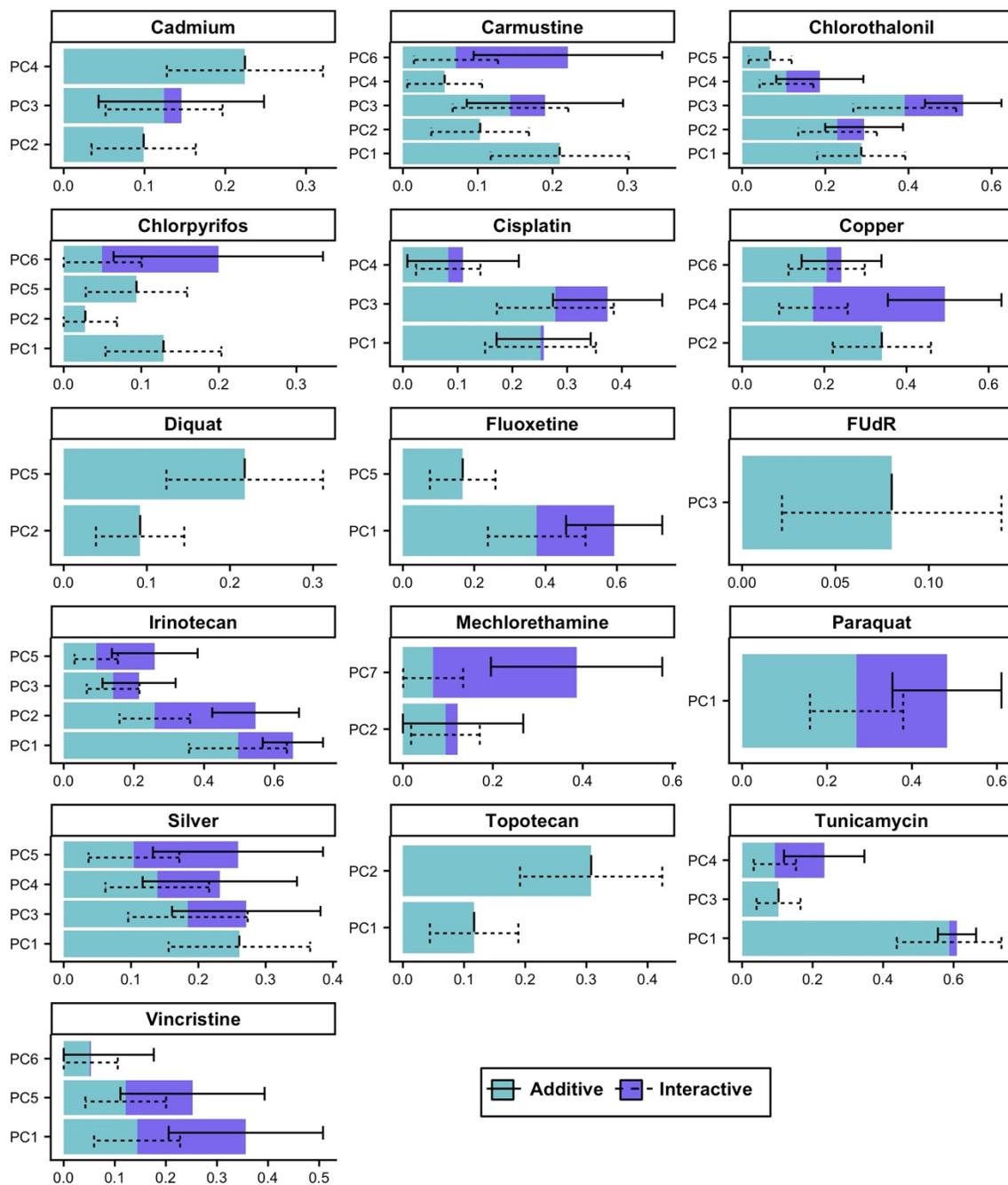
We calculated the broad-sense heritability, narrow-sense heritability, and the proportion of narrow-sense heritability attributed to detected QTL for each of the PCs that mapped to the 82 QTL identified using linkage mapping (**Figure 3-2**, Materials and Methods). In many cases, additive genetic components (narrow-sense heritability) could not explain all of the phenotypic variation predicted to be caused by genetic factors (broad-sense heritability). These results suggest that small-effect additive loci or genetic interactions remain undetected with our assay [192]. This missing heritability could be explained by achieving higher statistical power with larger sample sizes.



**Figure 3-2 Both additive and interacting loci underlie toxin response variation**

Additive genetic components identified by linkage mapping do not explain all heritable contributions to toxin-response variation. For 47 principal components representing the 82 QTL, we compared **A.** the broad-sense heritability (x-axis) calculated from the RIAIL phenotypic data versus the narrow-sense heritability (y-axis) estimated by a mixed model and **B.** the narrow-sense heritability (x-axis) versus the variance explained by all QTL detected by linkage mapping (y-axis). In both plots, each principal component is plotted as a point whose color indicates drug class (chemotherapeutic, heavy metal, neuropharmaceutical, or pesticide). The diagonal line represents  $y = x$  and is shown as a visual guide.

To determine the proportion of phenotypic variance that derives from additive or interacting genetic components, we fit a linear mixed-effect model to the RIAL phenotype data for the 47 PCs controlled by the 82 QTL. Different toxin classes displayed a range of additive and epistatic components (**Figure 3-2, Figure 3-3**). On average, the heritability of responses to cisplatin, topotecan, and FUdR was primarily underlied by additive loci (**Figure 3-3**). On the other hand, variation in responses to paraquat, irinotecan, vincristine, and mechlorethamine was more attributable to genetic interactions than additive effects (**Figure 3-3**). We scanned the genome for interactions between pairs of markers that might affect the phenotypic distribution of the RIAL panel and identified three significant interactions (Materials and Methods). However, the two-factor genome scan did not localize all epistatic components that were identified by the linear mixed-effect model (**Figure 3-2**), perhaps because of missing small-effect additive loci in the model and/or insufficient statistical power to identify small-effect interactions.

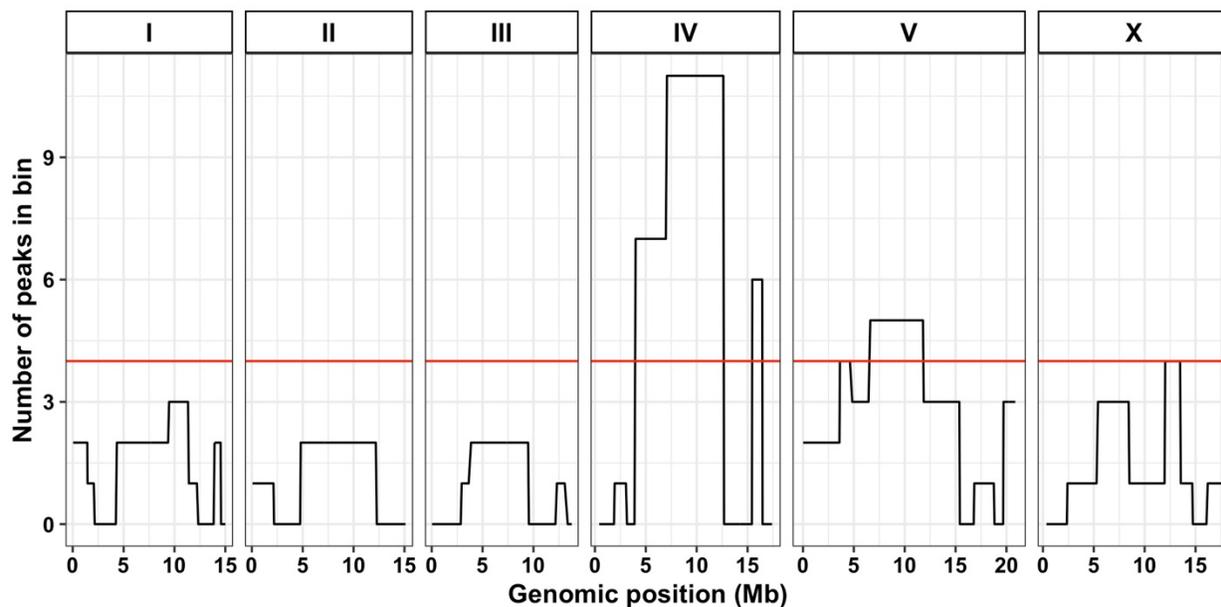


**Figure 3-3 The proportion of additive and interactive loci that underlie toxin responses differs among drugs**

Additive and interactive genetic components. The proportion of phenotypic variation predicted to be caused by additive and interactive genetic components is shown for each of the 47 principal components that represent the 82 QTL identified by linkage mapping. For each toxin, the fraction of phenotypic variation (x-axis) in a given principal component (y-axis) that is attributable to additive (light blue) versus interactive (dark blue) genetic components is shown as a stacked bar plot. Solid and dashed error bars show the standard error around the mean of additive and interactive components, respectively.

## Three QTL hotspots underlie variation in responses to diverse toxins

The majority of toxin-response QTL clustered on chromosomes IV and V (**Figure 3-1**). This clustering could be a chance occurrence, or it could indicate hotspots of toxin-response QTL. To test these two hypotheses, we first accounted for the higher rate of recombination, and thus more genetic diversity, on the chromosome arms [110] by dividing the genome evenly into 65 bins and then calculated the number of QTL that mapped to each bin (**Figure 3-4, Materials and Methods**). We identified three bins with more QTL than expected based on a Poisson distribution [260] and classified these bins as QTL hotspots. These three hotspots are located on the center of chromosome IV, the right of chromosome IV, and the center of chromosome V, and are hereby denoted as IVL, IVR, and V, respectively. We ran the same analysis with the full set of 247 growth-response traits (as opposed to the 47 PCs) with significant QTL and located the same three hotspots. Importantly, these hotspots are not driven by multiple PCs within a single toxin; instead, they comprise multiple QTL across a variety of PCs and toxins. In fact, 14 of the 16 toxins tested (all but cadmium and diquat) have a PC that maps to at least one of the three hotspots (**Table 3-3**). Of the 82 QTL that underlie PCs, 18 mapped to IVL, eight mapped to IVR, and nine mapped to V. In total, 33 QTL map to a hotspot (two QTL have confidence intervals that span both chromosome IV hotspots). We sought to experimentally validate the predicted additive and epistatic effects on toxin responses for QTL that mapped to the three hotspots.



**Figure 3-4 Three regions of the genome are enriched for toxin-response QTL**

Three QTL hotspots impact toxin responses. Each chromosome is divided into equal bins of 26 cM, resulting in a total of 65 bins across the genome. The x-axis shows the genomic position (Mb), and the y-axis shows the number of QTL that lie within the corresponding bin. The red line indicates the 99<sup>th</sup> percentile of a Poisson distribution with a mean of 1.26 QTL (total QTL/total bins).

**Table 3-3 Toxin responses map to three QTL hotspots**

Toxin	Class	PCs in IVL	PCs in IVR	PCs in V
Cadmium	Heavy Metal	0	0	0
Carmustine*	Chemotherapeutic	1*	0	1*
Chlorothalonil*	Pesticide	2*	1*	1*
Chlorpyrifos	Pesticide	1	1	0
Cisplatin*	Chemotherapeutic	2*	1	2*
Copper	Heavy Metal	2	0	0
Diquat	Pesticide	0	0	0
Fluoxetine*	Neuropharmaceutical	1	2*	0
FUdR	Chemotherapeutic	1	1	0
Irinotecan*	Chemotherapeutic	0	1*	2
Mechlorethamine	Chemotherapeutic	0	0	1
Paraquat*	Pesticide	0	0	1*
Silver*	Heavy Metal	3*	0	1*
Topotecan	Chemotherapeutic	1	0	0
Tunicamycin	Chemotherapeutic	2*	0	0
Vincristine	Chemotherapeutic	2	1	0

\*Denotes a toxin tested in NIL/CSS experiments

## NILs confirm some of the predicted QTL effects

We created and assayed NILs for the IVL, IVR, and V hotspots to validate the QTL identified from linkage mapping. Each NIL contains a genomic region introgressed from one parental strain into the genome of the opposite parental strain and were whole-genome sequenced to confirm their genotypes (Materials and Methods). We measured NIL responses to a subset of the toxins that map to a given hotspot. To test our ability to recapitulate QTL of various effect sizes by using

NILs, we chose to test toxins with QTL of small, medium, and large effect sizes (**Table 3-4**). We tested five toxins (10 QTL) with the IVL NILs, three toxins (four QTL) with the IVR NILs, and five toxins (six QTL) with the V NILs. In total, we tested 20 QTL across eight toxins for recapitulation using the NILs.

**Table 3-4 List of each toxin that maps to each QTL hotspot; bold denotes a trait tested in the NIL/CSS assays**

IV Left			IV Right			V		
Toxin	Trait	%VE	Toxin	Trait	%VE	Toxin	Trait	%VE
<b>Carmustine</b>	<b>PC6</b>	<b>5.60</b>	<b>Chlorothalonil</b>	<b>PC3</b>	<b>10.88</b>	<b>Carmustine</b>	<b>PC1</b>	<b>7.00</b>
<b>Chlorothalonil</b>	<b>PC2</b>	<b>4.31</b>	Chlorpyrifos	PC2	7.74	<b>Chlorothalonil</b>	<b>PC1</b>	<b>15.35</b>
<b>Chlorothalonil</b>	<b>PC3</b>	<b>12.90</b>	Cisplatin	PC3	2.68	<b>Cisplatin</b>	<b>PC1</b>	<b>10.25</b>
Chlorpyrifos	PC1	6.34	<b>Fluoxetine</b>	<b>PC1</b>	<b>10.73</b>	<b>Cisplatin</b>	<b>PC4</b>	<b>7.18</b>
<b>Cisplatin</b>	<b>PC1</b>	<b>6.05</b>	<b>Fluoxetine</b>	<b>PC5</b>	<b>7.92</b>	Irinotecan	PC2	6.78
<b>Cisplatin</b>	<b>PC3</b>	<b>4.78</b>	FUdR	PC3	5.54	Irinotecan	PC5	6.43
Copper	PC2	4.85	<b>Irinotecan</b>	<b>PC2</b>	<b>5.49</b>	Mechlorethamine	PC2	8.67
Copper	PC6	5.86	Vincristine	PC6	6.75	<b>Paraquat</b>	<b>PC1</b>	<b>9.98</b>
Fluoxetine	PC1	6.65				<b>Silver</b>	<b>PC1</b>	<b>17.81</b>
FUdR	PC3	5.54						
<b>Silver</b>	<b>PC3</b>	<b>9.76</b>						
<b>Silver</b>	<b>PC4</b>	<b>9.28</b>						
<b>Silver</b>	<b>PC5</b>	<b>11.70</b>						
Topotecan	PC2	9.70						
<b>Tunicamycin</b>	<b>PC1</b>	<b>15.90</b>						
<b>Tunicamycin</b>	<b>PC3</b>	<b>6.70</b>						
Vincristine	PC5	6.47						
Vincristine	PC6	6.75						

For each of these 20 QTL, we identified the toxin-response trait that is most correlated with the PC controlled by that QTL. We then assayed the NILs for that toxin-response trait as well as all toxin-response traits within its same trait cluster, because each PC comprises multiple toxin-response traits (**Table 3-5**, Materials and Methods). We tested 42 toxin-response traits with the

IVL NILs, 12 toxin-response traits with the IVR NILs, and 45 toxin-response traits with the V NILs (**Table 3-5**, data available at Figshare: <https://doi.org/10.25386/genetics.7158911>). Together, we performed 99 tests of recapitulation of QTL effects for toxin-response traits. We then sorted the recapitulation results into six different categories: “no parental effect”, “recapitulation”, “no QTL effect”, “unidirectional transgressive,” “bidirectional transgressive,” or “miscellaneous” (**Figure 3-5, Table 3-6**).

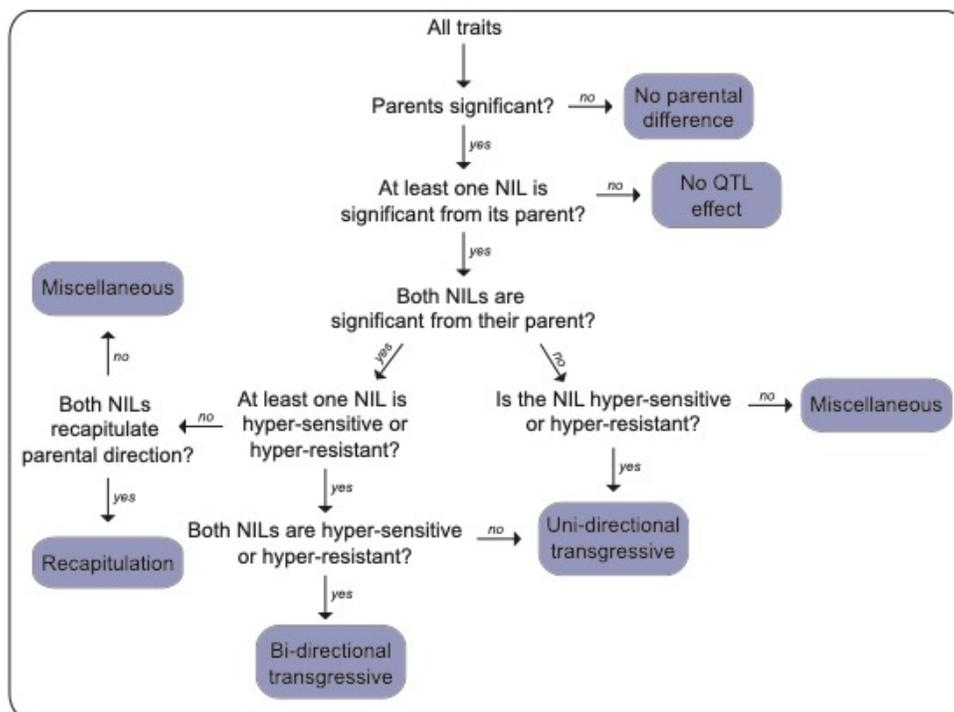
**Table 3-5 All PCs assayed with NILs/CSSs and the traits that underlie each PC**

PC	Hotspot	Correlated Traits	Correlation Range
carmustine.PC1	V	mean.EXT, mean.TOF, q75.EXT, median.EXT, median.TOF, q75.TOF, median.norm.EXT, q90.TOF, q90.EXT	0.72-0.95
carmustine.PC6	IVL	q25.norm.EXT, q10.norm.EXT	0.33-0.39
chlorothalonil.PC1	V	mean.EXT, q75.EXT, mean.TOF, median.EXT, median.TOF, q75.TOF	0.73-0.95
chlorothalonil.PC2	IVL	cv.TOF, cv.EXT	0.72-0.90
chlorothalonil.PC3	IVL, IVR	mean.norm.EXT, q75.norm.EXT, q90.norm.EXT, median.norm.EXT	0.50-0.65
cisplatin.PC1	IVL, V	mean.EXT, mean.TOF, median.EXT, median.TOF, q75.TOF, q75.EXT, q90.EXT, q90.TOF	0.78-0.97
cisplatin.PC3	IVL	var.TOF, var.EXT	0.38-0.54
cisplatin.PC4	V	norm.n, n	0.76-0.80
fluoxetine.PC1	IVR	mean.norm.EXT, q75.norm.EXT, mean.EXT, q75.EXT, q90.norm.EXT, q90.EXT	0.79-0.96
fluoxetine.PC5	IVR	q90.norm.EXT, q75.norm.EXT, mean.norm.EXT, q75.EXT, mean.EXT, q90.EXT	0.07-0.40
irinotecan.PC2	IVR	cv.TOF, cv.EXT	0.57-0.84

paraquat.PC1	V	median.EXT, mean.EXT, q25.EXT, q75.EXT, mean.TOF, q75.TOF, q10.EXT, q90.EXT, q90.TOF, median.TOF, q25.TOF, q10.TOF	0.75-0.95
silver.PC1	V	mean.EXT, median.EXT, q75.EXT, mean.TOF, q90.EXT, q90.TOF, median.TOF, q75.TOF	0.77-0.96
silver.PC3	IVL	q10.norm.EXT, q25.norm.EXT, mean.norm.EXT, median.norm.EXT, q75.norm.EXT, q90.norm.EXT	0.32-0.64
silver.PC4	IVL	n, norm.n	0.84-0.84
silver.PC5	IVL	n, norm.n	0.41-0.41
tunicamycin.PC1	IVL	median.EXT, q75.EXT, mean.TOF, q75.TOF, median.TOF, median.norm.EXT, q90.EXT, q90.TOF, mean.EXT, q75.norm.EXT, mean.norm.EXT, q25.norm.EXT, q90.norm.EXT, q10.norm.EXT	0.69-0.96
tunicamycin.PC3	IVL	norm.n, n	0.47-0.50

Twenty-three of the 99 tests did not show a significant phenotypic difference between the parental strains (N2 and CB4856) in the NIL assay, and these were categorized as “no parental effect” (Materials and Methods, **Figure 3-5, Table 3-6**). The remaining 76 tests in which a significant parental difference was observed were classified further. If a single QTL in the introgressed region contributed to the parental phenotypic difference, one would expect each NIL to have a phenotype significantly different from the parental strain with the same genetic background. Furthermore, one would expect each NIL to have a phenotype reminiscent of the parental strain of its introgressed genomic region. Four tests fit this “recapitulation” model (**Figure 3-5, Table 3-6**). For example, the NILs on the center of chromosome V displayed a normalized brood size in cisplatin (cisplatin.norm.n in cisplatin PC4) that recapitulated the expected parental phenotype (**Figure 3-6**). For 11 of the remaining 72 tests, the phenotype of each NIL was not significantly different from the phenotype of the parental strain sharing its background genotype (**Figure 3-5, Table 3-6**), which would indicate that the introgressed region does not affect the toxin-response phenotype.

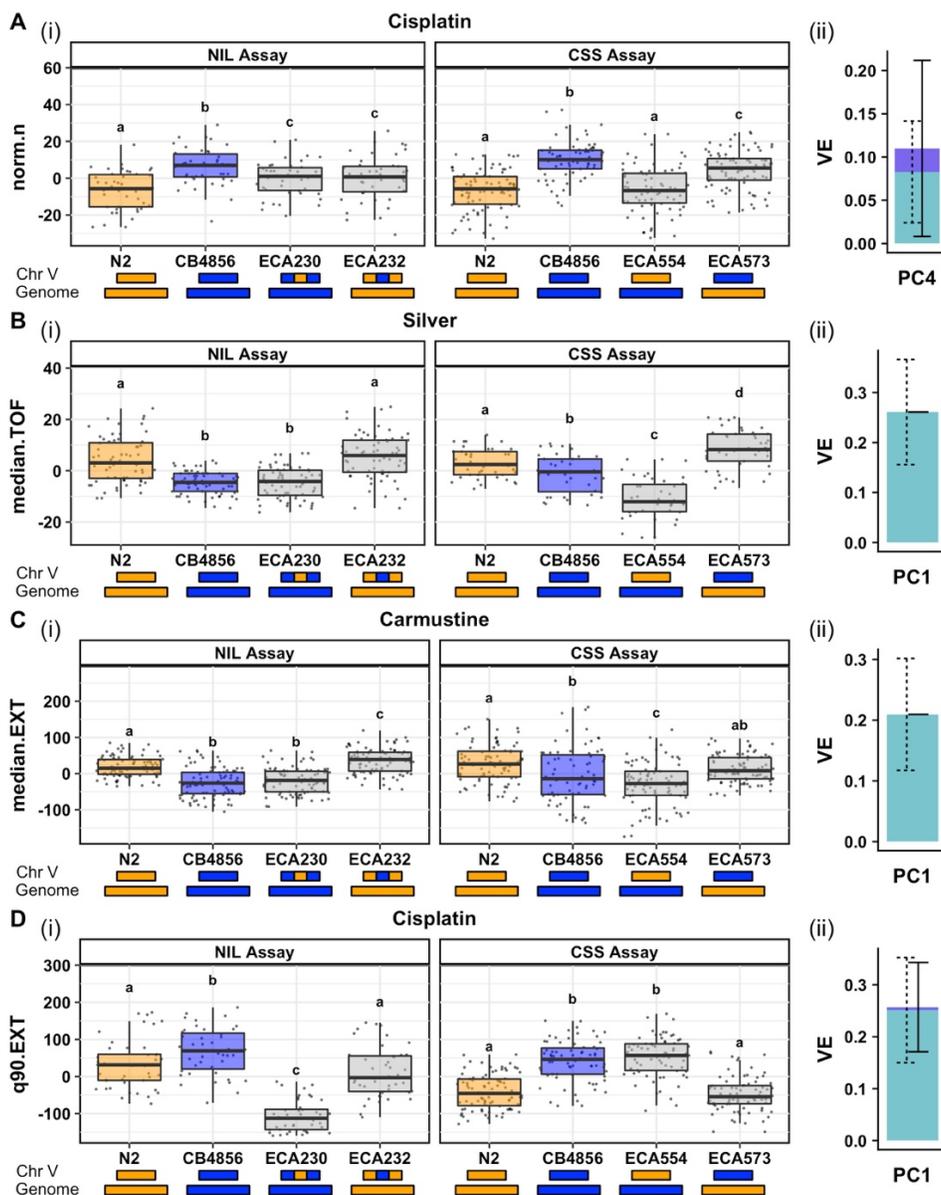
This lack of QTL effect might suggest that the genetic architecture is more complex than the mapping predicted, we lacked sufficient statistical power to detect the QTL effect (especially likely in cases of small QTL effect sizes), or the real QTL lies outside of the introgressed region and therefore the mapping was imprecise. The NILs on the center of chromosome V showed this result for median animal length in silver (silver.median.TOF in silver PC1) (**Figure 3-6**).



Category	CSS	NIL
Recapitulation	Recapitulation	Recapitulation
No parental effect	No parental difference	No parental difference
No QTL effect	No QTL effect	No QTL effect
Inter-chromosomal external	Uni-directional or bi-directional transgressive phenotype	Recapitulation or no QTL effect
Inter-chromosomal internal	Uni-directional or bi-directional transgressive phenotype	Uni-directional or bi-directional transgressive phenotype
Intra-chromosomal	Recapitulation or no QTL effect	Uni-directional or bi-directional transgressive phenotype

**Figure 3-5 NIL and CSS results can be placed into distinct categories**

Top: Flowchart for categorizing traits from the NIL or CSS tests of recapitulation of QTL effects. Bottom: Six potential categories for chromosome V traits tested in both NIL and CSS assays with a significant and consistent parental phenotypic split across both assays. The miscellaneous category is not depicted, but encompasses any other combination of NIL and CSS assay results.



**Figure 3-6 NIL and CSS results represent several categories**

Results from near-isogenic line (NIL) and chromosome-substitution strain (CSS) tests of recapitulation of QTL effects are categorized based on potential genetic mechanisms implicated in toxin responses. A trait contributing to a mapped principal component for each category is reported: **A**. Recapitulation (cisplatin norm.n, PC4), **B**. Inter-chromosomal external bidirectional loci (silver median.TOF, PC1), **C**. Inter-chromosomal internal unidirectional loci (carmustine median.EXT, PC1), and **D**. Intra-chromosomal unidirectional loci (cisplatin q90.EXT, PC1). In each case, we show results from (i) the NIL assay (left) and CSS assay (right) plotted as Tukey box plots. The y-axis indicates residual phenotypic values for the given trait. Different letters (a-d) above each Tukey box plot represent significant differences ( $p < 0.05$ ) whereas the same letter represents non-significant differences between two strains (Tukey HSD). The genotype of each strain on the x-axis is modeled by the colored rectangles beneath the plots (N2 genotypes are orange, CB4856 genotypes are blue). (ii) A stacked bar plot shows the proportion of phenotypic variation attributable to additive (light blue with dashed error bars) and interactive (dark blue with solid error bars) genetic factors of the principal component represented by each trait, based on a mixed model.

**Table 3-6 List of primary categorizations from NIL assay results**

<b>Primary Category</b>	<b>Number of Tests (99)</b>
No Parental Effect	23
Recapitulation	4
No QTL Effect	11
Unidirectional Transgressive	38
Bidirectional Transgressive	7
Miscellaneous	16

The phenotypes of the NILs for the remaining 61 tests cannot be explained by a single-QTL model. For many of these tests, we observed NIL phenotypes that are more sensitive or more resistant than both parental strains, suggesting that loci of opposite genotypes act additively or interact in the NILs to create transgressive phenotypes [263]. This category was supported by the results of the mixed-effects model, which suggested that both additive and interacting QTL remained undetected by linkage mapping (**Figure 3-2**). We further explored the results of these 61 tests by characterizing them based on the patterns of the transgressive phenotypes we observed.

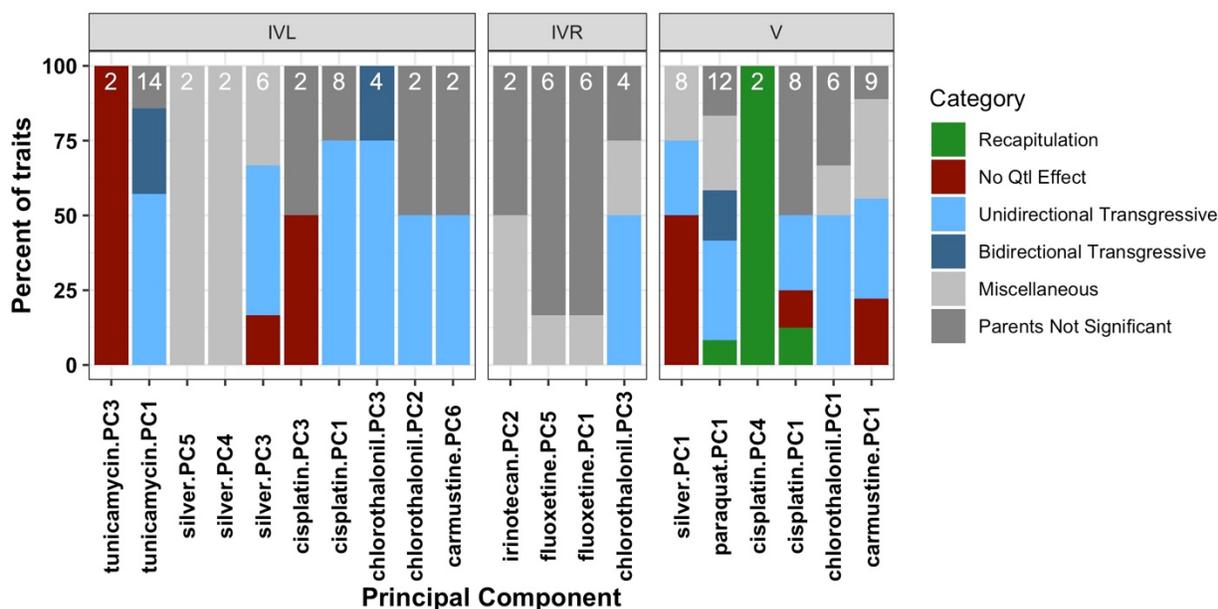
Only one of the two reciprocal NILs showed a transgressive phenotype for 38 of these 61 tests (**Figure 3-5, Table 3-6**). Some of these 38 “unidirectional transgressive” phenotypes seem to show an antagonism that counteracted the effect of the introgressed region (a predicted sensitive phenotype becomes hyper-resistant or a predicted resistant phenotype becomes hypersensitive, *e.g.*, carmustine.median.EXT in carmustine PC1, **Figure 3-6**). Other phenotypes displayed synergy that increased the effect of the introgressed region (a predicted sensitive phenotype

becomes a hypersensitive phenotype or a predicted resistant phenotype becomes a hyper-resistant phenotype, *e.g.*, cisplatin.q90.EXT in cisplatin PC1, **Figure 3-6**). Interestingly, in most cases (82%), the transgressive phenotype was observed in the strain with the N2 genotype introgressed into the CB4856 background.

We identified seven tests with suggested “bidirectional transgressive” phenotypes in which both NILs showed an extreme phenotype compared to the parental strains (**Figure 3-5, Table 3-6**). Some of these “bidirectional transgressive” phenotypes were suggestive of purely antagonistic effects (*e.g.*, tunicamycin.mean.norm.EXT, Figshare <https://doi.org/10.25386/genetics.7158911>), whereas others suggested an antagonistic effect in one NIL and a synergistic effect in the other (*e.g.*, paraquat.median.TOF, Figshare <https://doi.org/10.25386/genetics.7158911>). No cases of bidirectional synergistic effects were identified. The remaining 16 tests of the 76 with a parental difference did not fall into any of the above categories and were classified as “miscellaneous” (**Table 3-6**).

As mentioned above, we selected particular toxin-response traits to represent the mapped PCs. Because each mapped PC comprised several toxin-response traits, we sought to analyze the overall QTL effect of each PC by comparing the NIL assay categorizations for the toxin-response traits that it comprises (**Figure 3-7**). For example, two traits, n and norm.n, were selected to represent cisplatin PC4 (**Table 3-5**). Both of these toxin-response traits were placed into the “recapitulation” category from the NIL assay results (**Figure 3-5, Figure 3-7**). These results suggest that a single additive QTL underlies the brood-size variation captured by PC4. Fourteen tunicamycin-response traits were selected to represent tunicamycin PC1 (**Table 3-5**). Eight of these 14 traits displayed “unidirectional transgressive phenotypes”, four traits displayed

“bidirectional transgressive phenotypes”, and the remaining two traits did not have a significant parental phenotypic difference (**Figure 3-5, Figure 3-7**). However, we see the same trend of resistance (ECA231 > N2 > CB4856 > ECA229) across 11 of the 14 traits representing this PC, regardless of test categorization. Therefore, some phenotypes might have been miscategorized because of our strict significance cutoffs (usually into the “miscellaneous” or “no parental/QTL effect” categories). The fact that 12 out of 14 tunicamycin-response traits showed evidence of transgressive phenotypes suggests that multiple QTL, acting additively or interacting, might impact tunicamycin responses.



**Figure 3-7 HTA traits that comprise each principal component are placed into similar categories**

The x-axis shows the principal component tested in NIL assays, split by hotspot. Each color represents the NIL assay categorization for each trait within a correlation cluster - either recapitulation (green), no QTL effect (red), unidirectional transgressive phenotype (light blue), bidirectional transgressive phenotype (dark blue), miscellaneous (light grey) or no significant parental difference (dark grey). The percent of all traits within the correlation cluster for each principal component that falls within a given category is shown on the y-axis. Numbers on the top of each bar indicate the number of traits within a correlation cluster.

We next asked if QTL effect sizes affected our ability to categorize toxin-response traits. The QTL underlying cisplatin PC4 explains a small percentage (7%) of the total phenotypic variance (**Table 3-4**). The traits selected to represent cisplatin PC4 were placed into the recapitulation category, despite the small effect size of the QTL (**Figure 3-5, Figure 3-7**). On the other hand, the QTL underlying tunicamycin PC1 explains a large amount (almost 16%) of the total phenotypic variance, which is one of the highest effect sizes mapped in this study (**Table 3-4**). The toxin-response traits selected to represent this PC showed mostly transgressive phenotypes, indicating undetected additive or interacting QTL despite the seemingly large-effect additive QTL identified in linkage mapping (**Figure 3-5, Figure 3-7**).

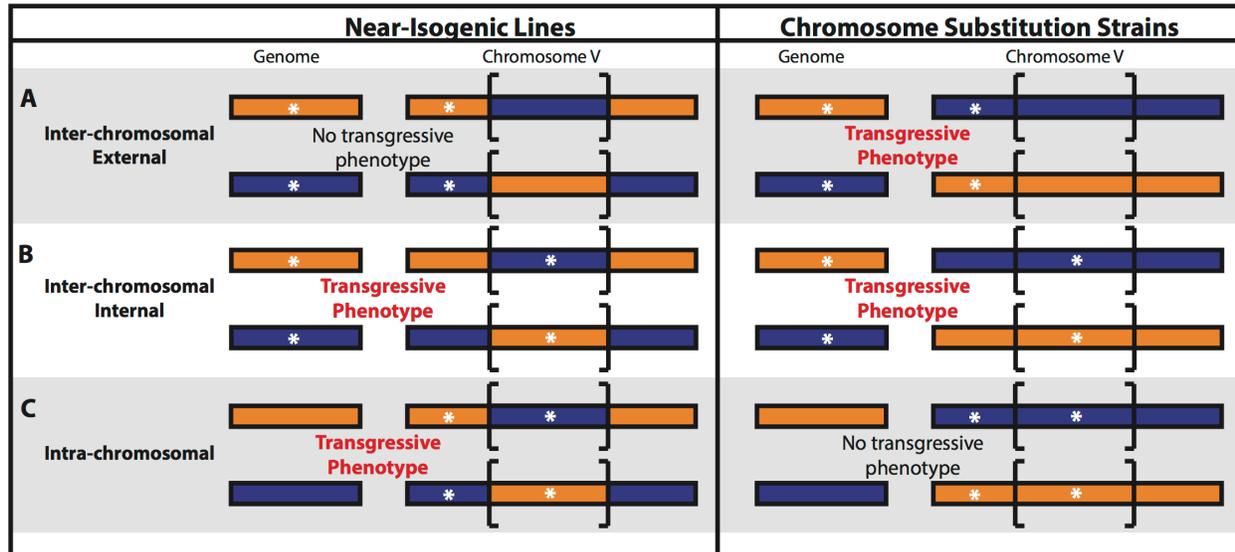
### CSSs localize QTL underlying transgressive phenotypes

Because we found evidence of additive or interacting loci that cause transgressive phenotypes, we attempted to further characterize these loci (**Figure 3-5, Figure 3-7**). We first sought to identify each set of loci as “intrachromosomal” or “interchromosomal”. We chose to dissect transgressive phenotypes of QTL on the chromosome V hotspot to isolate the effects of one hotspot and avoid complications arising from traits whose confidence intervals might lie within both of the hotspots on chromosome IV. We built reciprocal CSSs for the chromosome V hotspot that had the entire chromosome V introgressed from one parental strain into the genome of the opposite parental strain (Materials and Methods). The CSSs were whole-genome sequenced and found to have the expected genotype at all markers (Materials and Methods), except for the chromosome I incompatibility locus [166,167]. We performed tests of recapitulation of QTL effects with the CSSs for each of the 45 toxin-response traits across the five toxins tested with the chromosome V NILs (**Figure 3-7, Table 3-4, Table 3-6**).

NIL and CSS phenotypes could be compared across assays for traits in which the parental phenotypic difference was significant and consistent across both sets of tests. Eight traits across five toxins fit this criterion (**Table 3-7**). One trait (cisplatin.norm.n) displayed phenotypic recapitulation of the introgressed region in both the NIL and the CSS tests, supporting a single-QTL model (**Figure 3-6, Table 3-7**). Alternatively, transgressive phenotypes are indicative of a multi-QTL model, and by comparing the pattern of phenotypes in the NIL and the CSS tests, we can surmise the locations of these additive or interacting QTL. One of these loci must exist on chromosome V, where both the NILs and CSSs share introgressed genomic regions. The other locus can exist on the same chromosome (intrachromosomal) or on a different chromosome (interchromosomal). We further divided the interchromosomal class into two categories: “interchromosomal external,” in which the chromosome V locus is outside the region introgressed in the NILs, and “interchromosomal internal,” in which the chromosome V locus is within the region introgressed in the NILs. **Figure 3-8** is a visual depiction of each type of multi-QTL model, and each of these models is described below.

**Table 3-7** Categorizations from combining the NIL and CSS assay results

Secondary Category	Number of Traits (8)	Traits
Recapitulation	1	cisplatin.norm.n
Inter-chromosomal (external)	1	silver.median.TOF (bidirectional)
Inter-chromosomal (internal)	1	carmustine.median.EXT (unidirectional)
Intra-chromosomal	2	cisplatin.q90.EXT (unidirectional), cisplatin.q90.TOF (unidirectional)
Miscellaneous	3	cisplatin.n, paraquat.q10.TOF, silver.median.EXT



**Figure 3-8 NIL and CSS assay results can be compared to localize potential regions of epistasis**

A model for potential locations of two loci is shown, according to toxin-response phenotypes of near-isogenic lines (NILs) and chromosome-substitution strains (CSSs). The NILs are represented on the left, and the CSSs are represented on the right. The strain genotype is indicated by colored rectangles. N2 is orange, and CB4856 is blue. Brackets indicate the genomic region that is introgressed in the NILs. White asterisks represent a potential location for additive or epistatic loci underlying transgressive phenotypes. Although bidirectional transgressive phenotype models are shown, each model could be bidirectional (both reciprocal introgressed strains show transgressive phenotypes) or unidirectional (only one reciprocal introgressed strain shows a transgressive phenotype). Models showing **A.** inter-chromosomal external effects between a locus outside of the introgressed region in the NILs and a locus on another chromosome, **B.** inter-chromosomal internal effects between a locus within the introgressed region in the NILs and a locus on another chromosome, and **C.** intra-chromosomal effects between a locus within and a locus outside of the introgressed region in the NILs are drawn.

For an “interchromosomal external” model, we expect only the CSSs to display hypersensitivity or hyper-resistance, because both loci share the same genotype in the NILs (**Figure 3-8**) and would therefore not cause a more extreme phenotype than both parents. We found one such trait that fits a “bidirectional interchromosomal external” loci model (silver.median.TOF) (**Figure 3-6, Table 3-7**). For an “interchromosomal internal” model, we expected both the CSSs and the NILs to display the same hypersensitivity or hyper-resistance, because the two loci are of opposite genotypes in both the NILs and the CSS (**Figure 3-8**). We identified one such trait that fits a

“unidirectional interchromosomal internal” loci model (carmustine.median.EXT) (**Figure 3-6, Table 3-7**).

To identify intrachromosomal loci that underlie transgressive phenotypes in the remaining 10 traits, we searched for traits that display evidence of either a uni- or bidirectional transgressive phenotype in the NILs but not in the CSSs (**Figure 3-8**). This result would suggest that two loci of opposite genotypes on chromosome V, one within and one outside the region introgressed in the NILs, act additively or epistatically to cause transgressive phenotypes. Because the CSSs have a consistent genotype across the entirety of chromosome V, we would not expect these strains to show transgressive phenotypes. We found two examples of such “unidirectional intrachromosomal” loci models (*e.g.*, cisplatin.q90.EXT, **Figure 3-6, Table 3-7**). The remaining three traits could not be characterized beyond their NIL assay characterization based on the results of the CSS assay (**Table 3-7**).

We compared the findings of the empirical NIL/CSS comparison and the computational two-factor genome scan for each of these eight empirically classified traits. None of the traits with significant interaction terms in the NIL/CSS comparison were identified by the two-factor genome scan. Although many other pairs of loci show suggestive evidence of additive or interacting effects, the statistical power of this computational approach might be too limited to identify the effects we observed empirically. Overall, this study highlights the benefits of leveraging both experimental and computational strategies to further dissect genetic components that underlie quantitative traits in a metazoan model.

## Discussion

Here, we show that three QTL hotspots underlie differences in responses to 16 diverse toxins. We used computational and empirical approaches to characterize these QTL. By testing the toxin responses of NILs and CSSs, we confirmed small-effect QTL and attempted to localize genomic regions that cause transgressive phenotypes. Finally, we used statistical analyses to computationally identify loci that might support some of our empirical findings. Although the number of biological replicates and recombinant strains in this study increased our power to detect QTL compared to previous studies, we are still too underpowered to definitively assess if missing heritability is composed of small, additive effects or genetic interactions.

### Pleiotropic regions underlie QTL shared between and among toxin classes

We performed PCA on toxin-response phenotypes collected for a panel of RIALs and used linkage mapping to identify 82 toxin-response QTL. Although some of these QTL are unique to one particular toxin or one particular PC within a toxin, others suggest the existence of pleiotropic QTL that underlie responses to a diverse set of toxins. In particular, three QTL hotspots, IVL, IVR, and V, were enriched for toxin-response QTL and were investigated further. The notion that a single gene in each hotspot is regulating the response to several toxins is unlikely, given the diversity of molecular mechanisms implicated in response to each toxin. However, a single gene involved in drug transport could conceivably underlie one or several of these hotspots. More likely, multiple genes in close proximity, each regulating a process important for cellular proliferation and survival, might underlie these hotspots. Notably, two of the three QTL hotspots are in chromosomal regions with lower genetic diversity at the species level [131,132,264]. The laboratory strain, N2, has experienced each of the selective chromosomal sweeps, whereas

CB4856 has not. Linkage mapping using a panel of RIALs built between these two strains could identify QTL that underlie phenotypic differences not only between N2 and CB4856, but more broadly between swept and non-swept strains. Moreover, identifying QTL in these swept regions that underlie variation in fitness-related traits, such as toxin responses, might indicate selective pressures that could have led to these chromosomal sweeps. For example, N2 is more resistant than CB4856 to tunicamycin (Figshare <https://doi.org/10.25386/genetics.7158911>), an antibiotic and chemotherapeutic produced by the soil bacterium *Streptomyces clavuligerus* [265]. This result might suggest that selective pressure toward responses to antibiotic compounds played a role in driving resistance-conferring alleles, such as those present in N2, to a high frequency. Alternatively, climate conditions could also impact local niche environments to sensitize toxin responses [266]. We observed that N2 is more resistant than CB4856 in responses to the majority of conditions, which could indicate that alleles present in swept strains confer robustness in responses to many conditions. This result emphasizes the importance of genetic background when considering toxin effects [185].

In addition to the three QTL hotspots, pleiotropic QTL across toxins within certain classes are suggested by our linkage mapping results. QTL from the chemotherapeutic class are enriched on chromosome I, which could be representative of variants that affect a common mechanism targeted by these toxins, such as DNA damage or cell-cycle control. However, because many of these chemotherapeutics have distinct mechanisms of action, this enrichment is likely caused by an overrepresentation of chemotherapeutics in our study. Direct comparisons of toxins with similar cellular mechanisms could provide more insights. For example, irinotecan and topotecan are both chemotherapeutics that cause DNA damage by inhibiting topoisomerase I [267], and share a QTL

on the center of chromosome I. However, each of these chemotherapeutics also maps to distinct regions of the genome. For example, the irinotecan-response QTL on the right arm of chromosome V is not mapped for topotecan response and the topotecan-response QTL on the left arm of chromosome II is not mapped for irinotecan response. Vincristine also maps to this same region; however, its mechanism of action is distinct from irinotecan and topotecan. The combination of overlapping and distinct genetic architectures underlying these highly similar compounds suggest that although some genetic variation implicated in responses to irinotecan and topotecan is shared, other QTL are specific to each compound and not representative of a general topoisomerase I inhibition mechanism. We have also observed this phenomenon of distinct genetic architectures underlying similar compounds for benzimidazole responses [143].

### Undetected epistatic loci might impact toxin responses

We constructed NILs for the three hotspots and assayed them in responses to multiple toxins to determine if we had sufficient power to experimentally validate even small-effect QTL. NILs showed a significant phenotypic effect for some of these tests of recapitulation, even in cases of small-effect QTL. One such example is cisplatin.norm.n and cisplatin.n, which represent the QTL mapped by cisplatin.PC4 that only explain 7% of the phenotypic variance. Our ability to recapitulate such a small effect suggests that our assay had sufficient power to detect small phenotypic effects in at least some cases. We postulated that our inability to recapitulate other QTL effects could be attributed to either insufficient power or additional additive or epistatic QTL that were undetected by linkage mapping. Particularly in cases where the NILs displayed transgressive phenotypes, undetected loci of opposite genotypes, acting additively or epistatically, likely caused these effects. Therefore, we investigated these interactions and found evidence for additional QTL that interact with the originally detected loci. However, we must note

that whole-genome sequence data revealed that three of our NILs had a portion of the genome from the background of the starting RIAIL. Although we do not believe that these small regions are responsible for the unexpected phenotypes observed, this explanation could be a consideration for certain silver, cisplatin, carmustine, and chlorothalonil PCs, as they have significant QTL in these identified regions. This example emphasizes the importance of whole-genome sequencing NILs to verify the expected genotypes before making conclusions about phenotypic effects of a targeted QTL.

We categorized the results from the NIL assays into genetic models that might underlie NIL phenotypes. Mostly, categorizations were consistent across traits representing a PC, with most of these traits falling into one or a few categorizations. This widespread consistency suggests that similar genetic architectures underlie phenotypes for these correlated traits. Furthermore, this consistency highlights the reproducibility of our high-throughput toxin-response assay, because results from independent assays (trait correlations, linkage mappings from RIAIL assays, and phenotype classifications from NIL assays) often align to support the same conclusion obtained from the individual experiments.

The majority of cases of transgressive phenotypes occur when the N2 genotype is introgressed into the CB4856 genome. This trend might indicate allele-specific unidirectional incompatibilities between the two strains, and localizing these interactions could improve our understanding of the evolutionary processes driving such incompatibilities. However, it is difficult to identify the loci that underlie these unidirectional transgressive phenotypes using a mixed-effect model or a two-factor genomic scan because only a small number of the RIAILs have the required allelic combinations to quantify such an effect. For example, cisplatin.q90.EXT, a trait chosen to represent cisplatin

PC1, fits a unidirectional intrachromosomal model. The results of the NIL and CSS assays show that, although the CSSs seem to display no QTL effect, the NIL with the N2 genotype introgressed into the CB4856 genome displays strong hypersensitivity (**Figure 3-6**). All of the narrow-sense heritability for cisplatin PC1 (25%) predicted by the mixed-effect model is explained by the three QTL identified through linkage mapping (the variance explained estimates of these three QTL add up to 26%). This finding suggests that most of the additive loci have been identified through linkage mapping, so the intrachromosomal loci are likely acting epistatically to cause a unidirectional transgressive phenotype. However, we do not find a significant interaction component for cisplatin PC1 using our mixed-model approach. On the other hand, a two-dimensional genome scan for multiple loci that underlie this PC provides suggestive evidence for a two-QTL model over a one-QTL model, with or without interaction between the loci. These two loci are located on the left of chromosome V (outside the NIL interval) and in the center of chromosome V (inside the NIL interval), and match our empirical evidence of two intrachromosomal loci underlying the transgressive phenotype observed (**Figure 3-6**). Because the transgressive phenotype is unidirectional, RIALs without the allelic combination that causes extreme phenotypes, including RIALs with the reciprocal alleles at these two loci, could dilute our power to detect the loci. For this reason, combining both computational models and empirical investigation facilitates the detection of loci that control transgressive phenotypes. Additionally, future studies should include even larger RIAL panels than what we used here to empower approaches to investigate the contributions of interactive loci.

Although our modeling approaches are statistically underpowered to identify some small-effect additive and interacting loci, the combination of three methods of searching for potential interactions suggests that not all fitness traits in *C. elegans* are composed of additive effects. Our

two computational methods were used to identify additive and epistatic loci underlying many toxin responses, but their power was limited in cases of unidirectional transgressive phenotypes. Alternatively, the NIL and CSS phenotypic assays were able to identify unidirectional transgressive phenotypes, but they were restricted by their inability to distinguish between additive and epistatic loci. Double CSS strains or multi-region NILs, in which pairwise combinations of two genomic regions are introgressed within the opposite genotype, could help to further define loci underlying transgressive phenotypes. However, we must isolate each locus to determine if the two loci act additively or epistatically. The results from the two-dimensional genome scan might provide insights into where to begin these future steps. We are fairly confident that QTL are not purely additive in cases where all three of our techniques suggested epistasis. Testing a larger number of biological replicates and using an even larger panel of recombinant strains might allow us to further address the debate about how heritable loci contribute to trait variation in metazoans.

## Future directions

As mentioned in the text, future studies for this project could leverage a larger panel of RIALs to increase statistical power. In the Andersen Lab, we have accumulated a large number of recombinant strains that we could assay in response to many toxins. This larger panel would increase the number of statistical tests we could perform between various allelic combinations and phenotypic outputs. Additionally, a larger set of recombinant strains could introduce more breakpoints between genotypes, ultimately improving our resolution of QTL confidence intervals.

Additionally, we can test the hypotheses of multi-QTL models by creating double CSS strains and multi-region NILs. The CSSs and NILs that we used in this study each contain a single region of introgression in a genetic background of the opposite genotype. Double CSSs and multi-region NILs would allow us to test each pairwise combination of alleles at two loci (either two whole chromosomes, double CSSs, or two regions of a single chromosome, multi-region NILs) in a controlled genetic background. Depending on the phenotypes of these strains, we can deduce whether two loci act additively or epistatically to impact toxin responses.

## Contributions

This project was a team effort between Katie Evans and me. We both contributed equally to the experiments, analysis, and writing of this manuscript. Josh Bloom performed the mixed-modeling approach to identify additive and epistatic loci that underlie toxin responses. Robyn Tanny and Dan Cook created DNA libraries and analyzed the whole-genome sequence data for all strains we used for the project. Sarah Giuliani, Stephen Hippleheuser, and Mostafa Zamanian contributed near-isogenic lines to this work. Erik Andersen supervised the project and facilitated key discussions regarding the direction of the work. Tyler Shimko, Bryn Gaertner, and Sam Rosenberg assisted with mapping drug sensitivities. During this project, I was funded by the Biotechnology Training Program grant.

## Supplemental Tables

Table S3-1 Reagents used to generate NILs and CSSs

Chromosome IVL NILs		
<b>ECA229</b> <i>eanIR149</i> (IV: 3,684,741 - 9,045,991, N2>CB4856)	Constructed from QX275 x CB4856	N2 into CB4856
<b>ECA231</b> <i>eanIR151</i> (IV: 4,475,146 - 9,334,865, CB4856>N2)	Constructed from QX591 x N2	CB4856 into N2
Left indel primers (IV: 5,110,734)		
oECA781	GAGCACTTTGGCGACTTTTCG	
oECA782	TCCGGGCAAATTAGTGTGGC	
Right indel primers (IV: 8,212,089)		
oECA857	CCACACGTCTACGCTTTGGA	
oECA858	AATCGTGGCATTGGTGGACA	

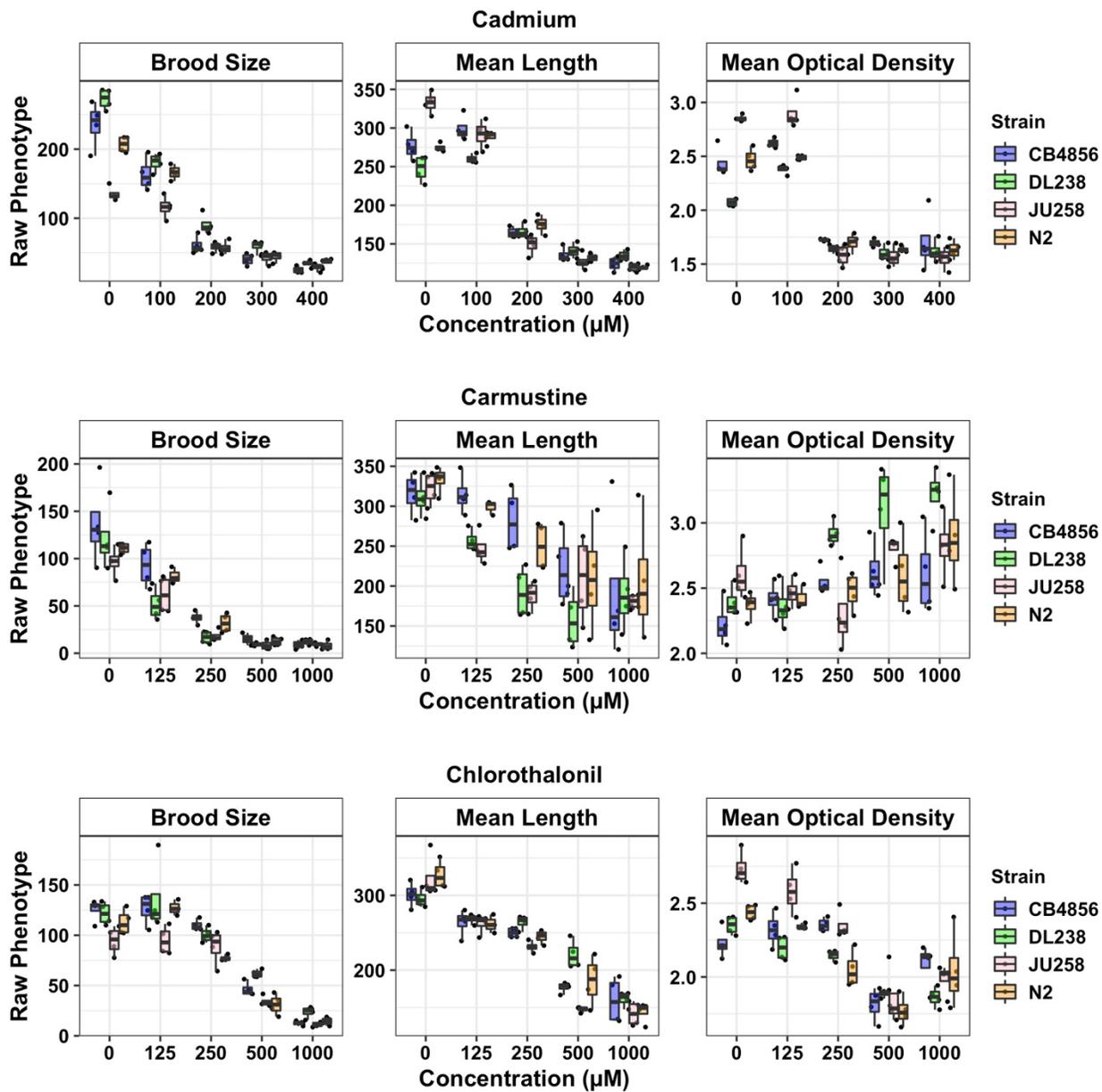
Chromosome IVR NILs		
<b>ECA240</b> <i>eanIR160</i> (IV:12,865,211-17,493,829, CB4856>N2)	Constructed from QX349 x N2	CB4856 into N2
<b>ECA241</b> <i>eanIR161</i> (IV:13,016,066-17,493,829, N2>CB4856)	Constructed from QX375 x CB4856	N2 into CB4856
Left indel primers (IV: 13,207,120)		
oECA904	AACAGATACTCGCCGTTGCT	
oECA905	ATTTGTACCACGCGTGACCT	
Right indel primers (IV: 17,356,993)		
oECA910	GACAACGCCCACTACGACAA	
oECA911	ACCCAACCAAGTTGAGCACAT	

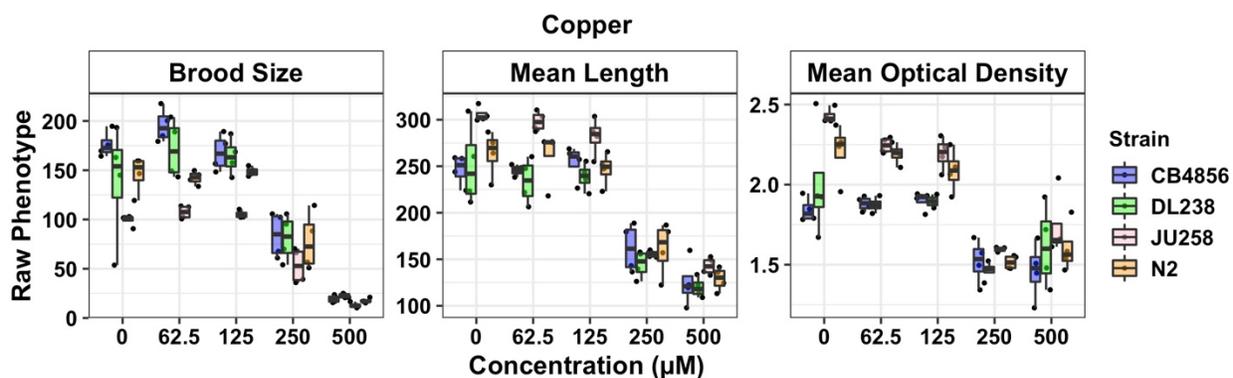
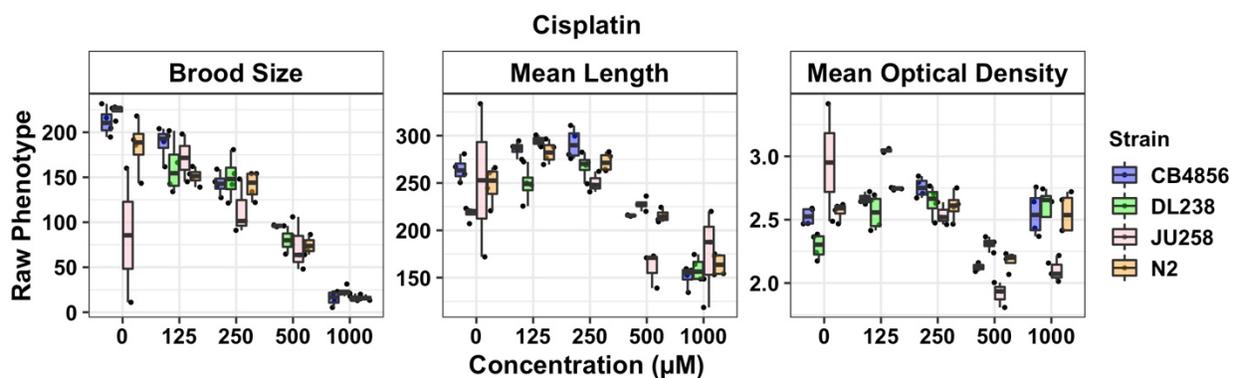
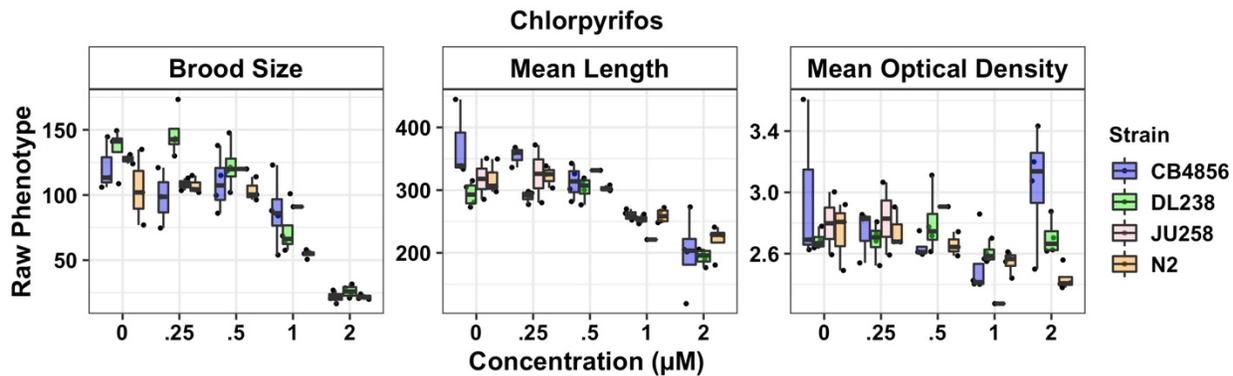
Chromosome V NILs		
<b>ECA230</b> <i>eanIR150</i> (V: 7,082,839-13,839,858, N2>CB4856)	Constructed from QX131 x CB4856	N2 into CB4856
<b>ECA232</b> <i>eanIR152</i> (V: 7,667,158-	Constructed from Q450 x N2	CB4856 into N2

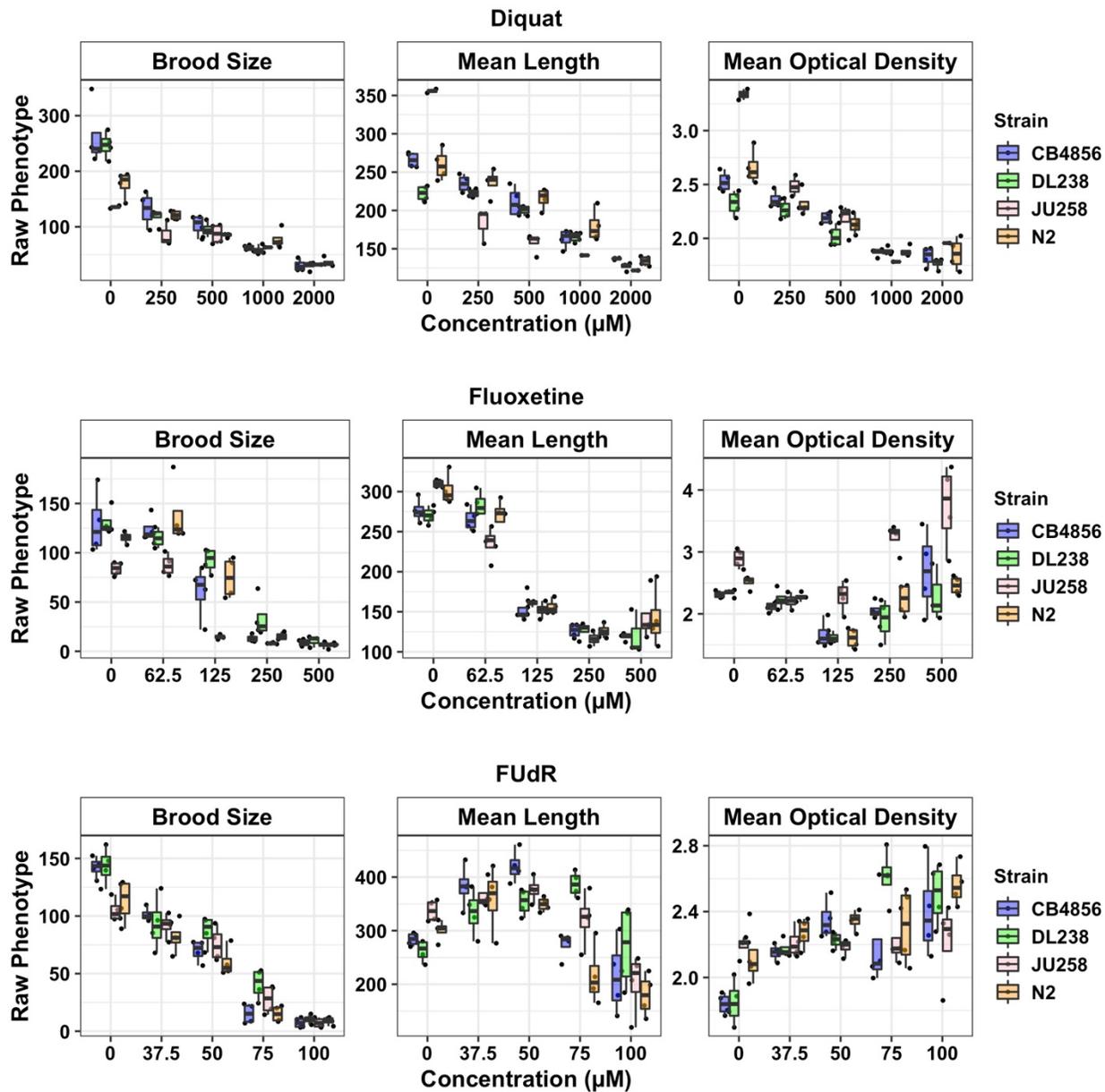
<i>13,678,801, CB4856&gt;N2</i>		
Left indel primers (V: 7,862,556)		
oECA799	TTCTCGCTACTGGAACACGC	
oECA800	TCAAGAAGCGTTGGGAAGTCT	
Right indel primers (V: 13,110,045)		
oECA745	TGCAGAGGTGGAGTAACCCT	
oECA746	CTCGGTCTCTCCCCCACTAA	

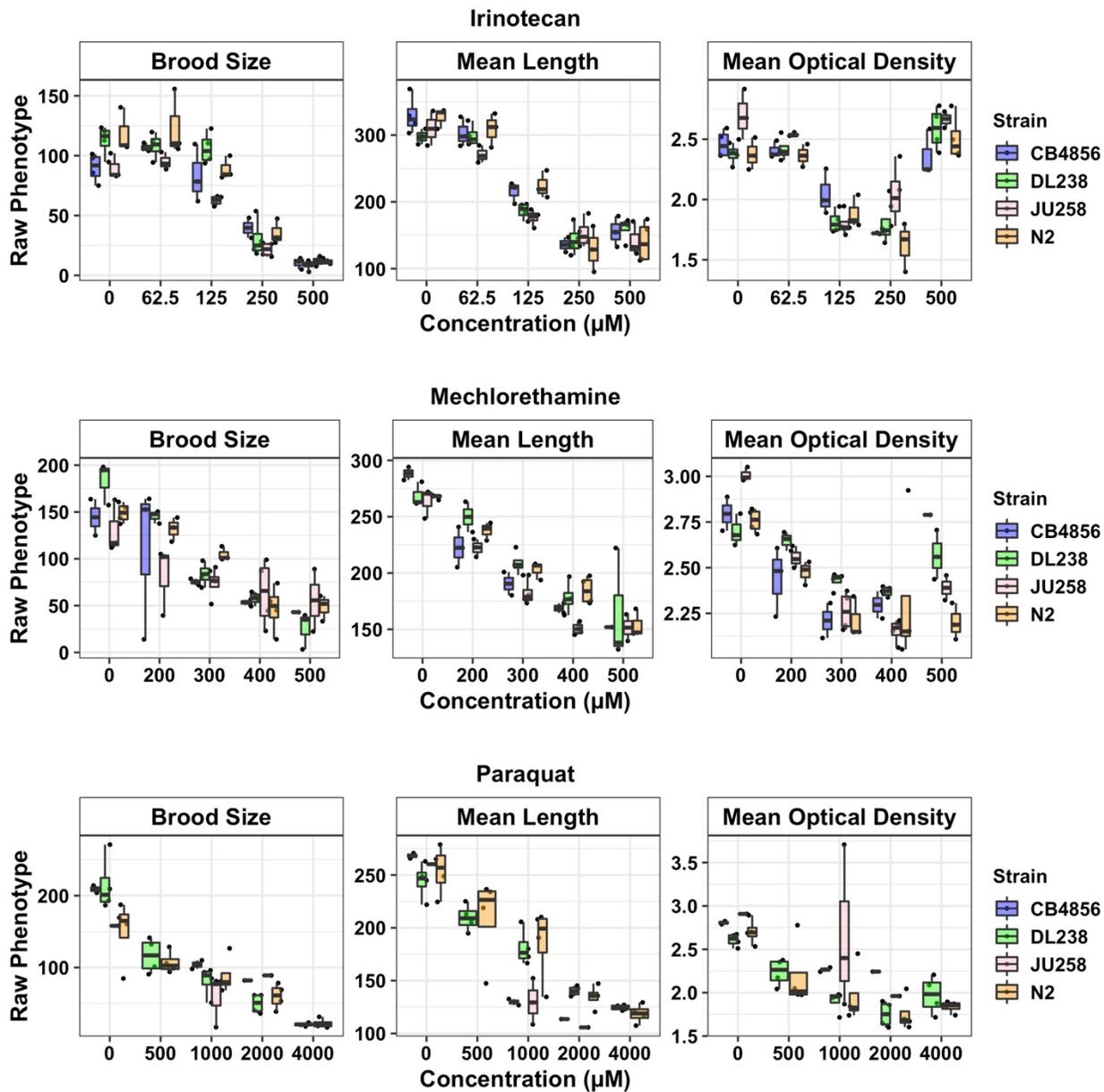
Chromosome V CSS		
<b>ECA554</b> <i>eanIR321(V:1-20,924,180, N2&gt;CB4856)</i>	Constructed from N2 x CB4856	N2 into CB4856
<b>ECA573</b> <i>eanIR322(V:1-20,923,490, CB4856&gt;N2)</i>	Constructed from N2 x CB4856	CB4856 into N2
Left indel primers (V: 144,547)		
oECA1141	CTCATGGGAGTAACCTGGGC	
oECA1142	CGGTGACAACGGAGAATCCA	
Right indel primers (V: 20,622,851)		
oECA1147	GTTTAGTACCAGCGGGGCAT	
oECA1148	TGCATTCCGACCCAAGAGAC	

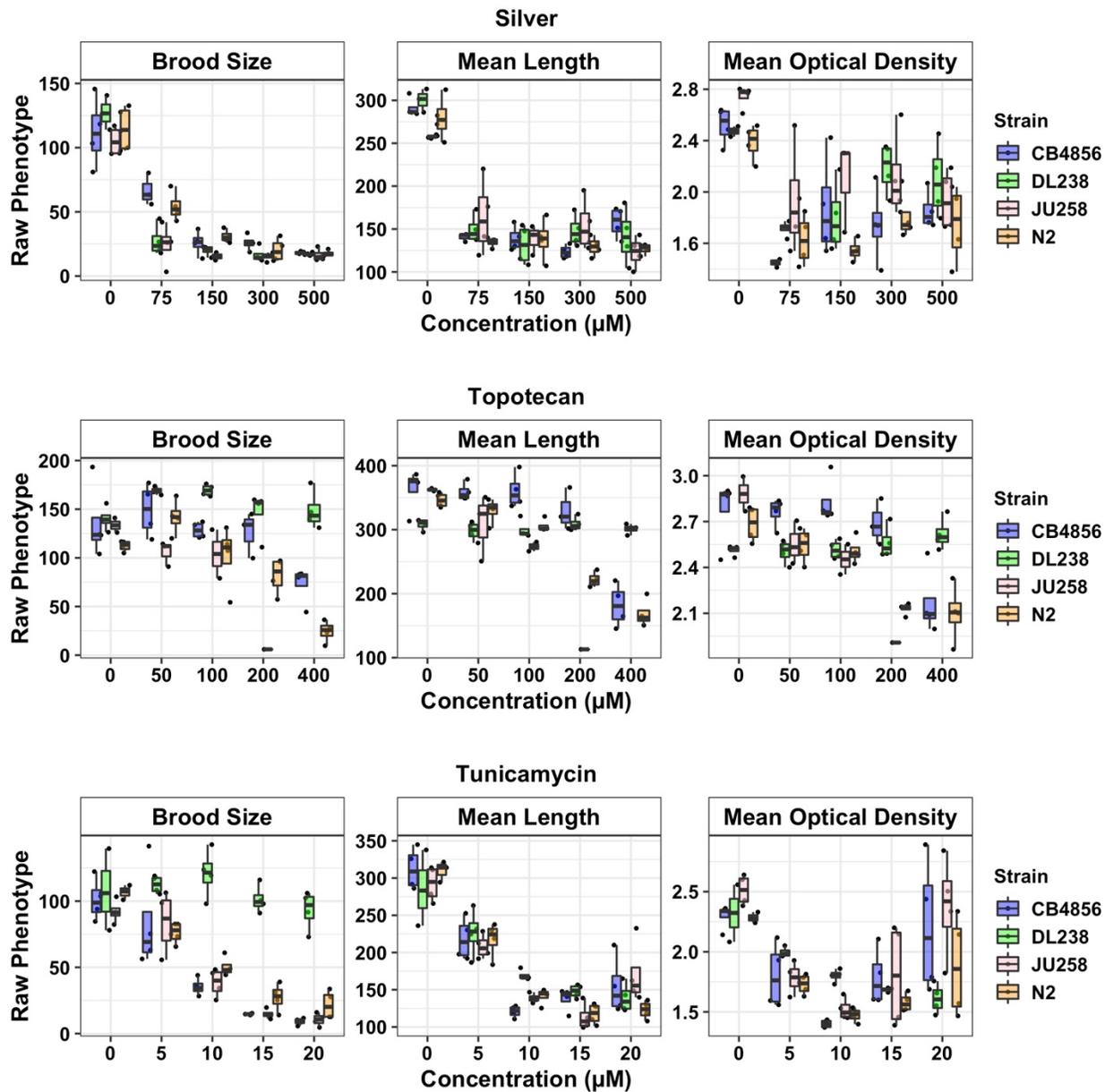
## Supplemental Figures

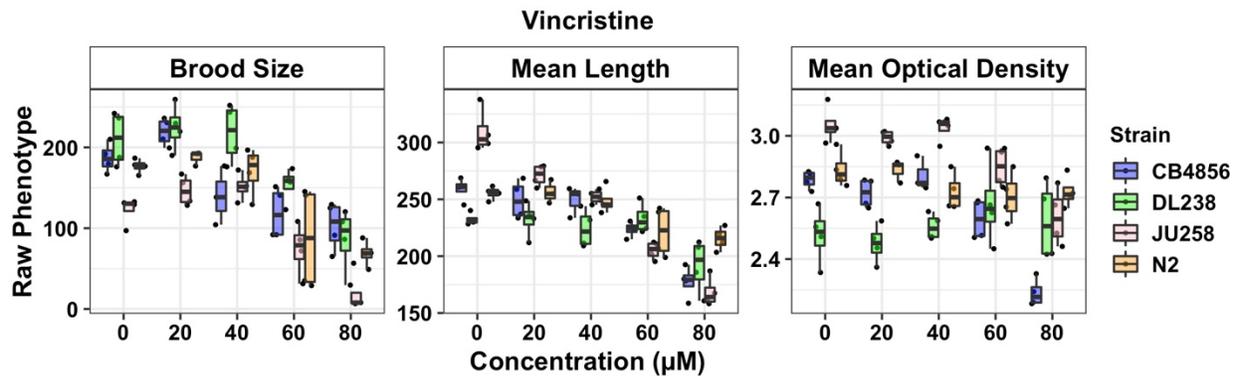








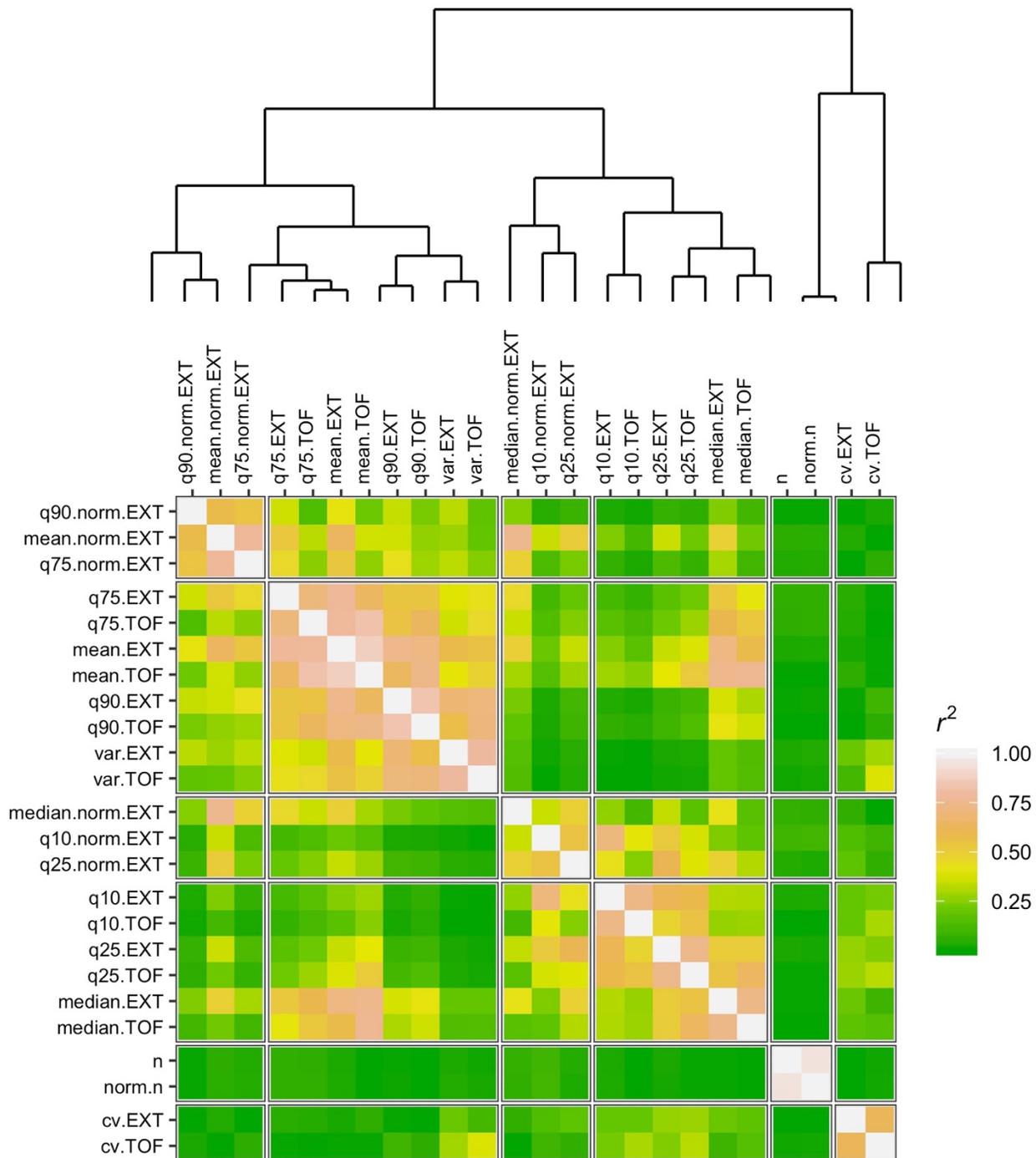




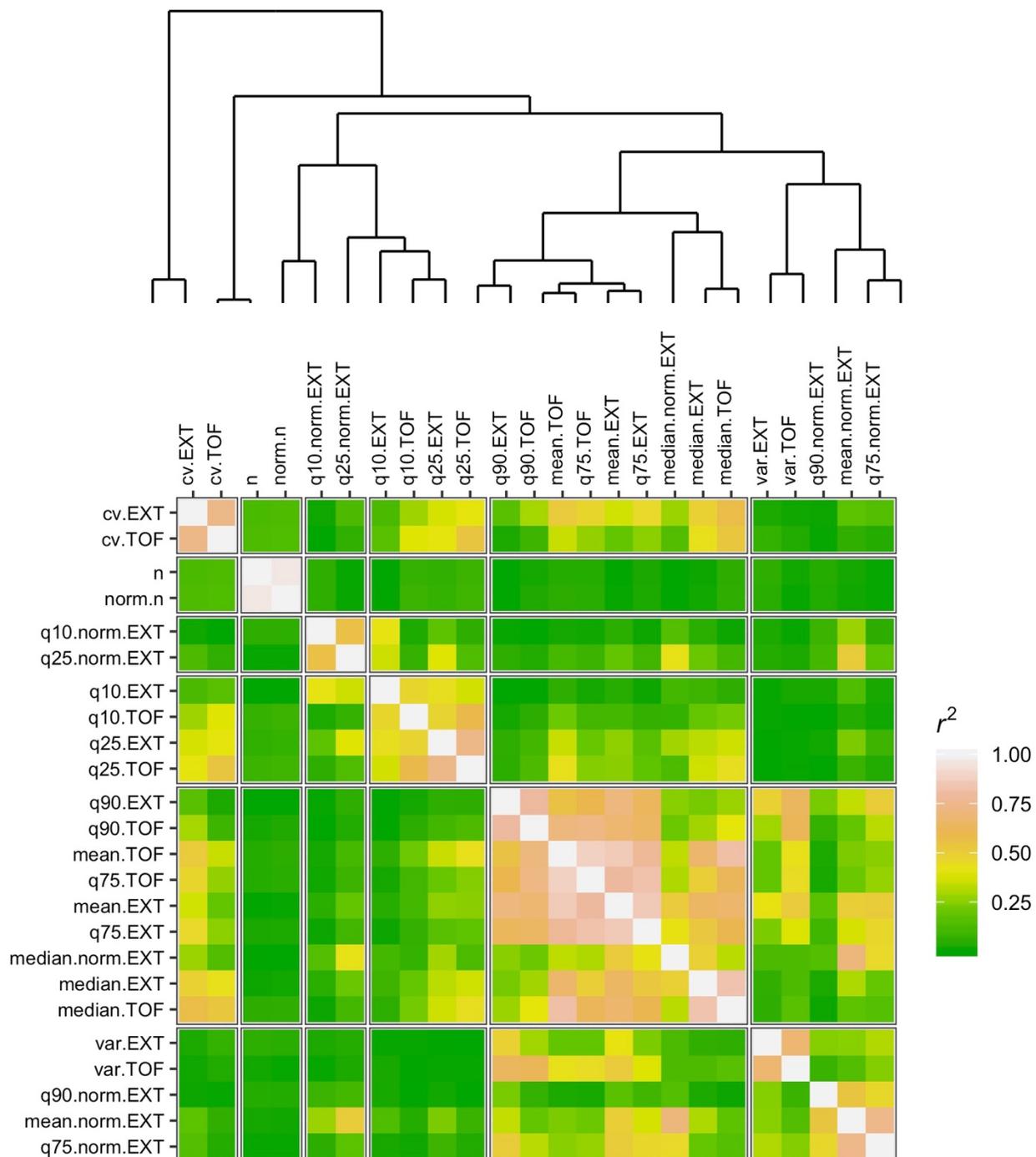
**Figure S3-1 Results of a dose-response assay for all toxins are shown**

Raw phenotypic values for four divergent strains, CB4856 (blue), DL238 (green), JU258 (pink), and N2 (orange), in response to various concentrations of each toxin are plotted as Tukey box plots. Toxin responses for traits representing three distinct population parameters- brood size (norm.n), mean animal length (mean.TOF), and mean optical density (mean.norm.EXT) are shown. The x-axis indicates either the control (water or DMSO) or the concentration ( $\mu\text{M}$ ) of the toxin, and the y-axis shows the raw phenotypic value for each trait.

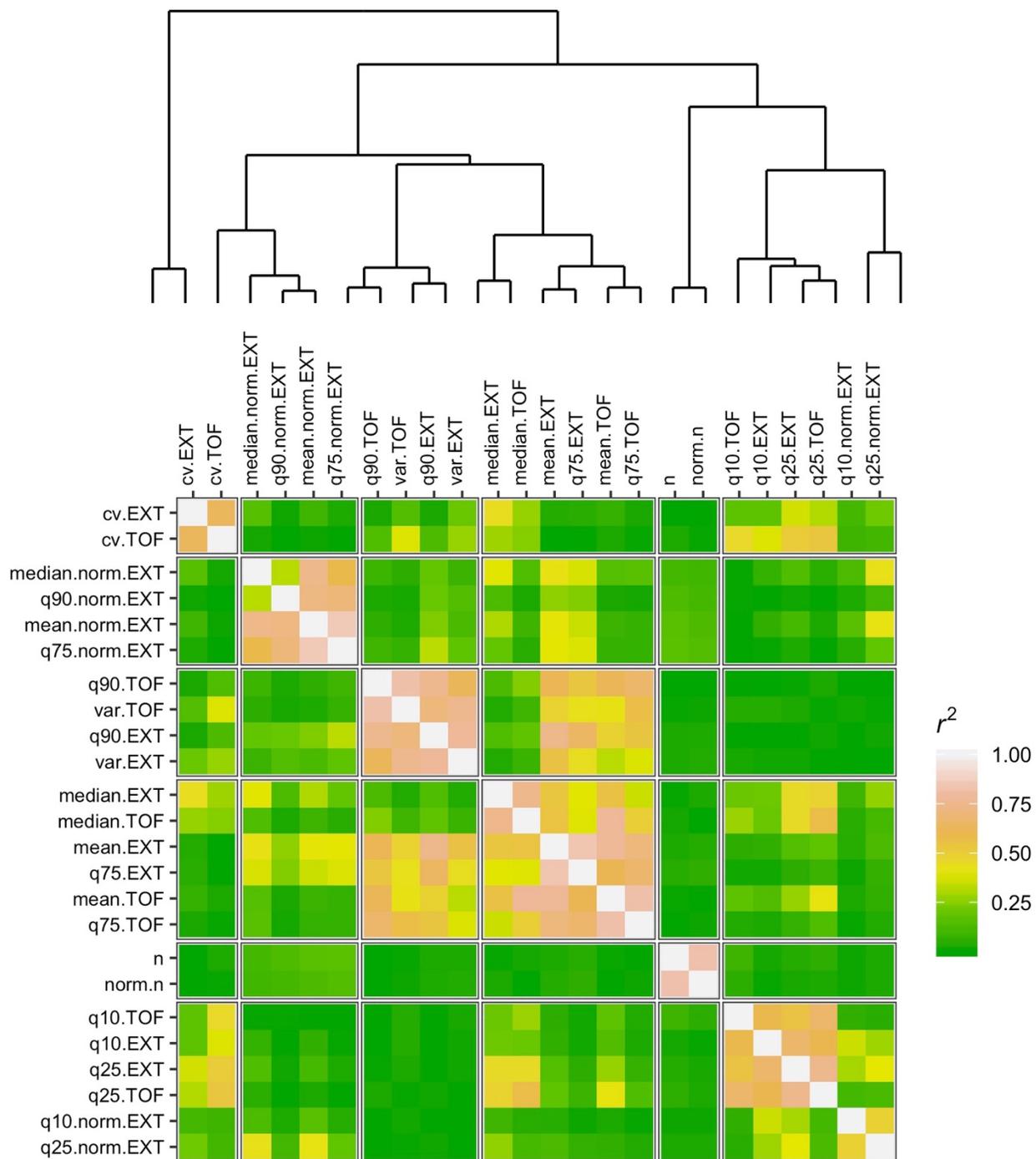
# Cadmium



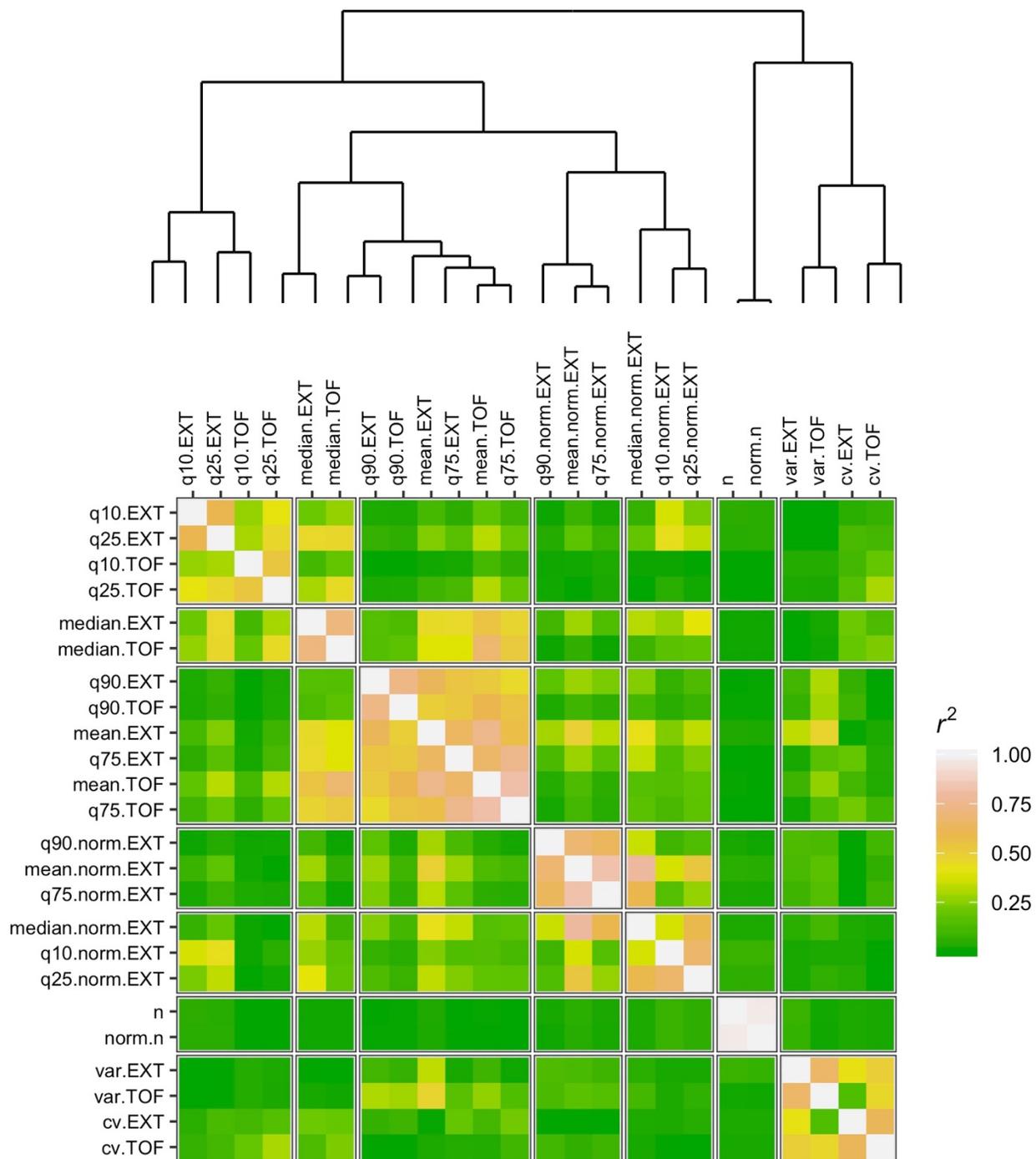
## Carmustine



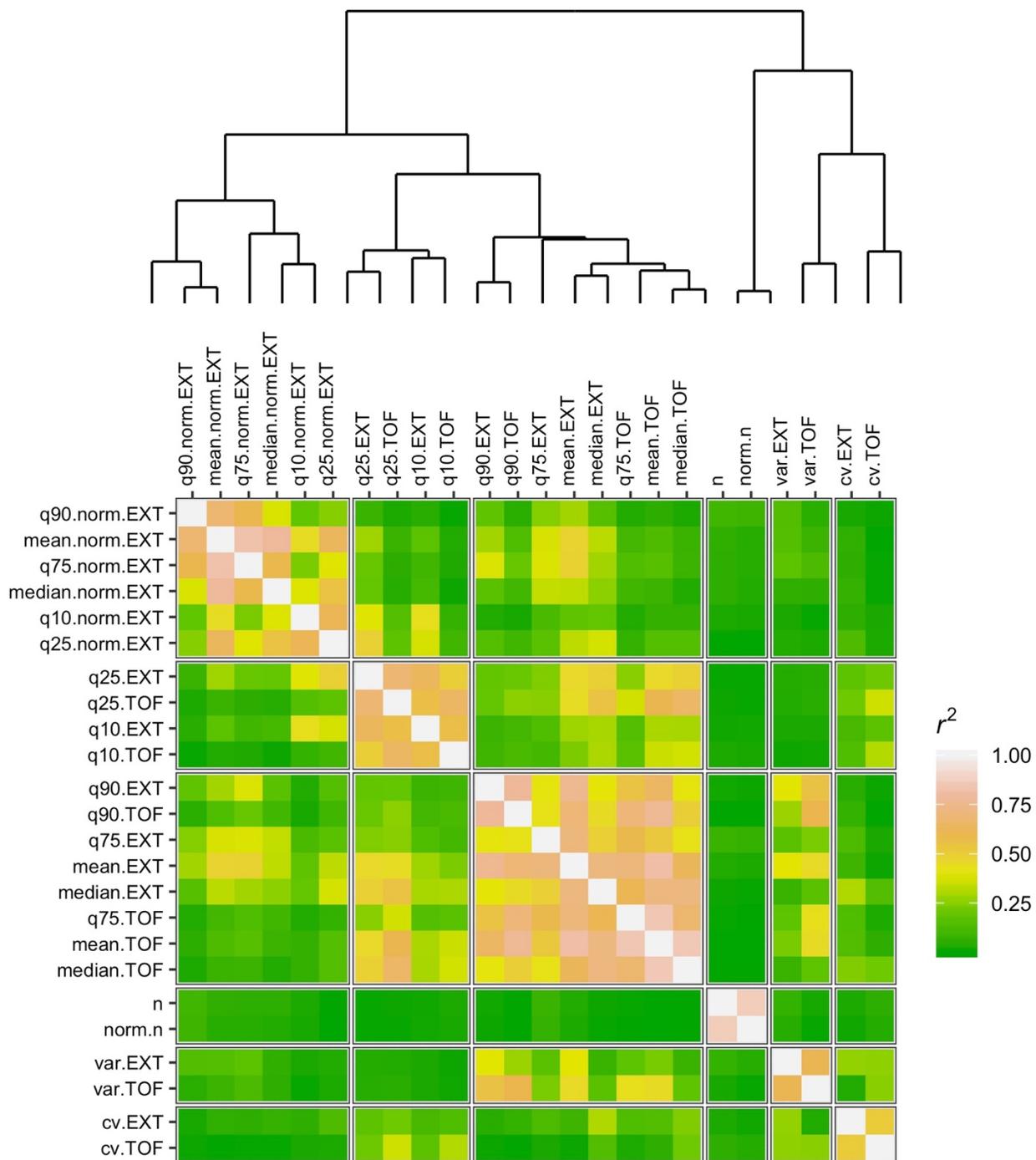
## Chlorothalonil



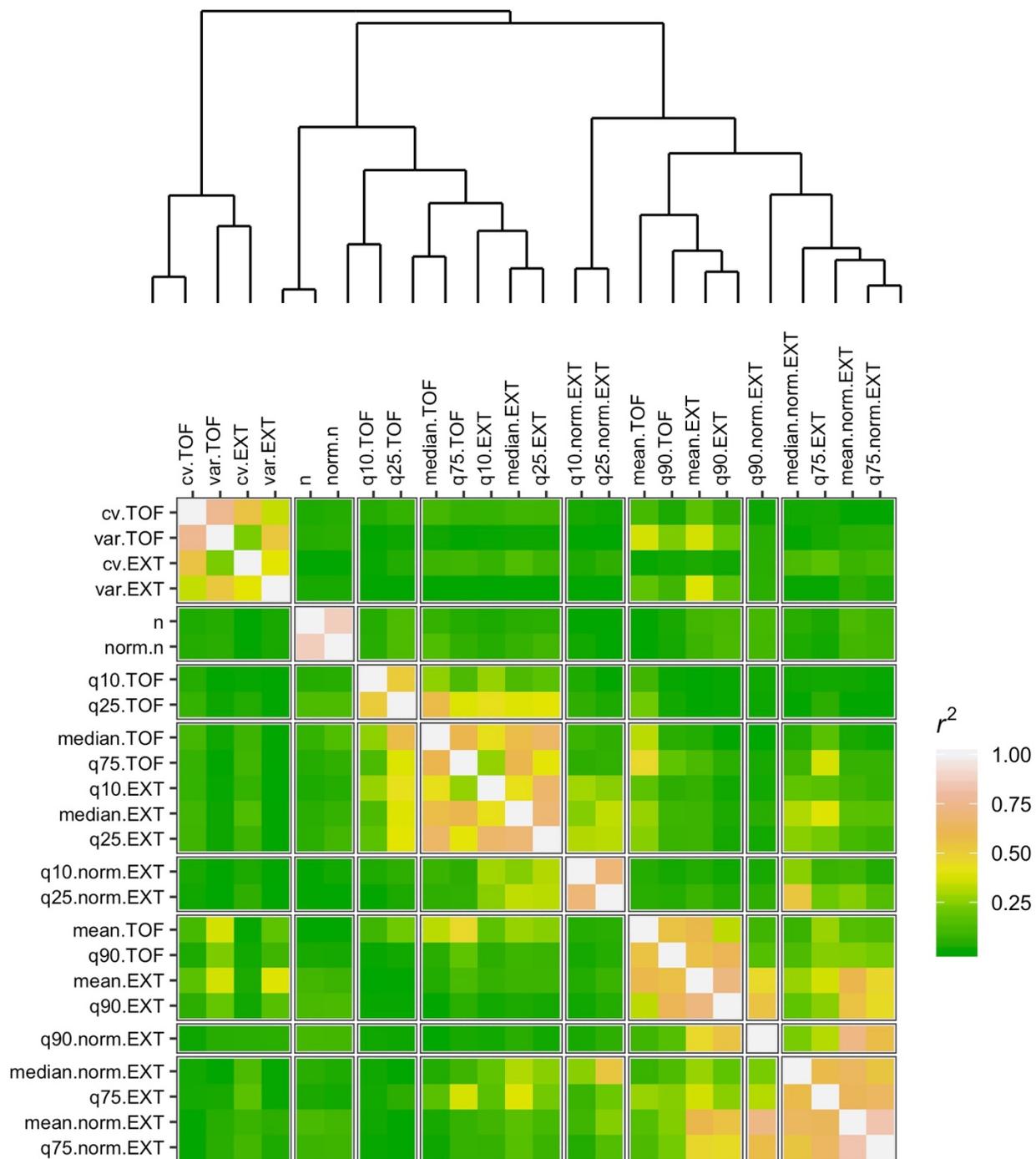
## Chlorpyrifos



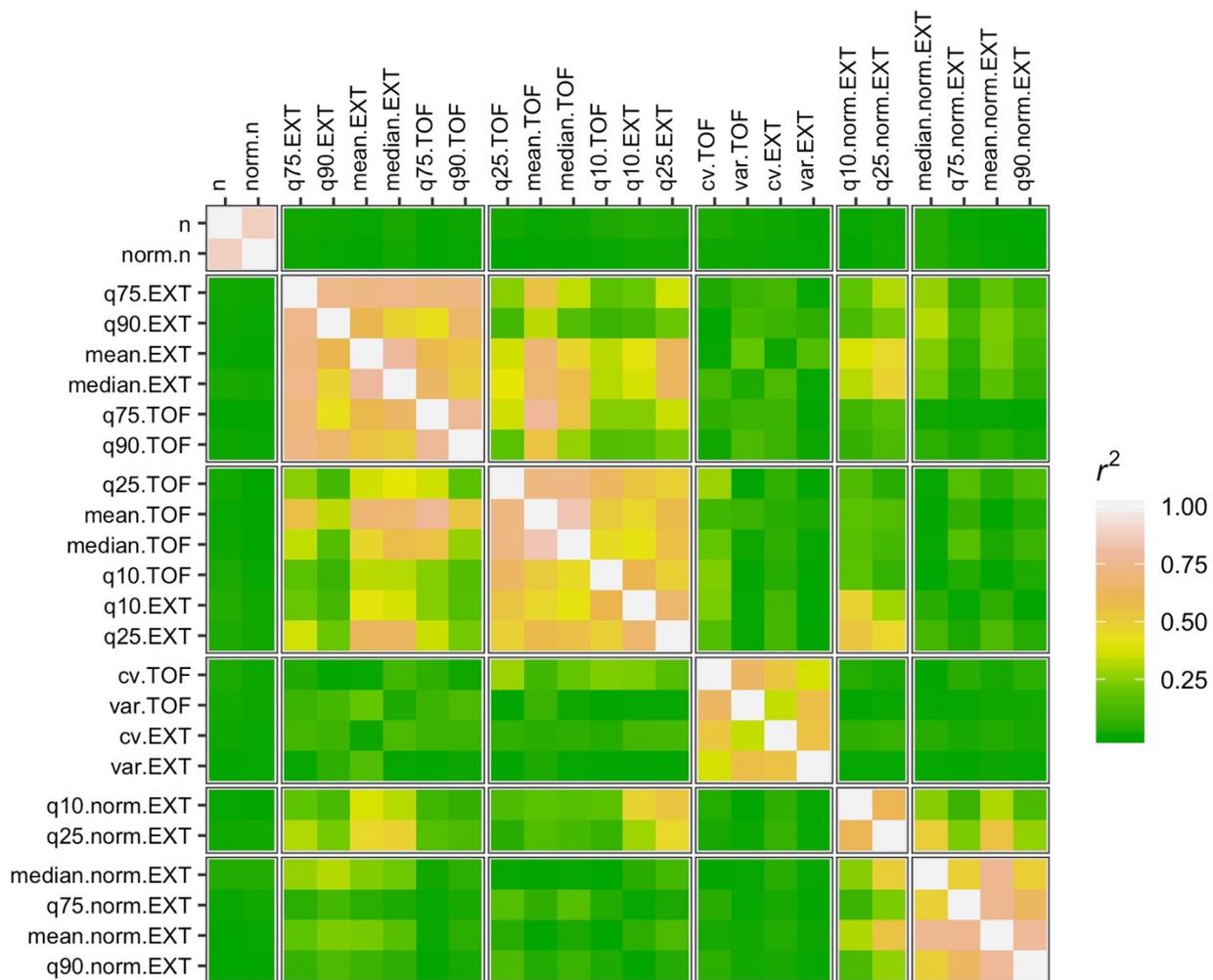
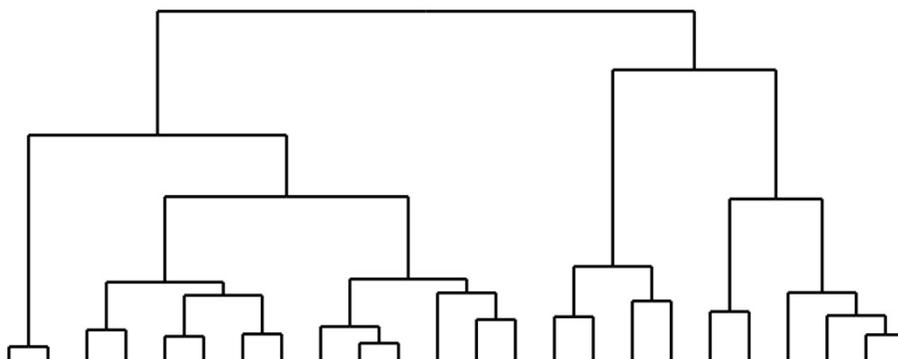
## Cisplatin



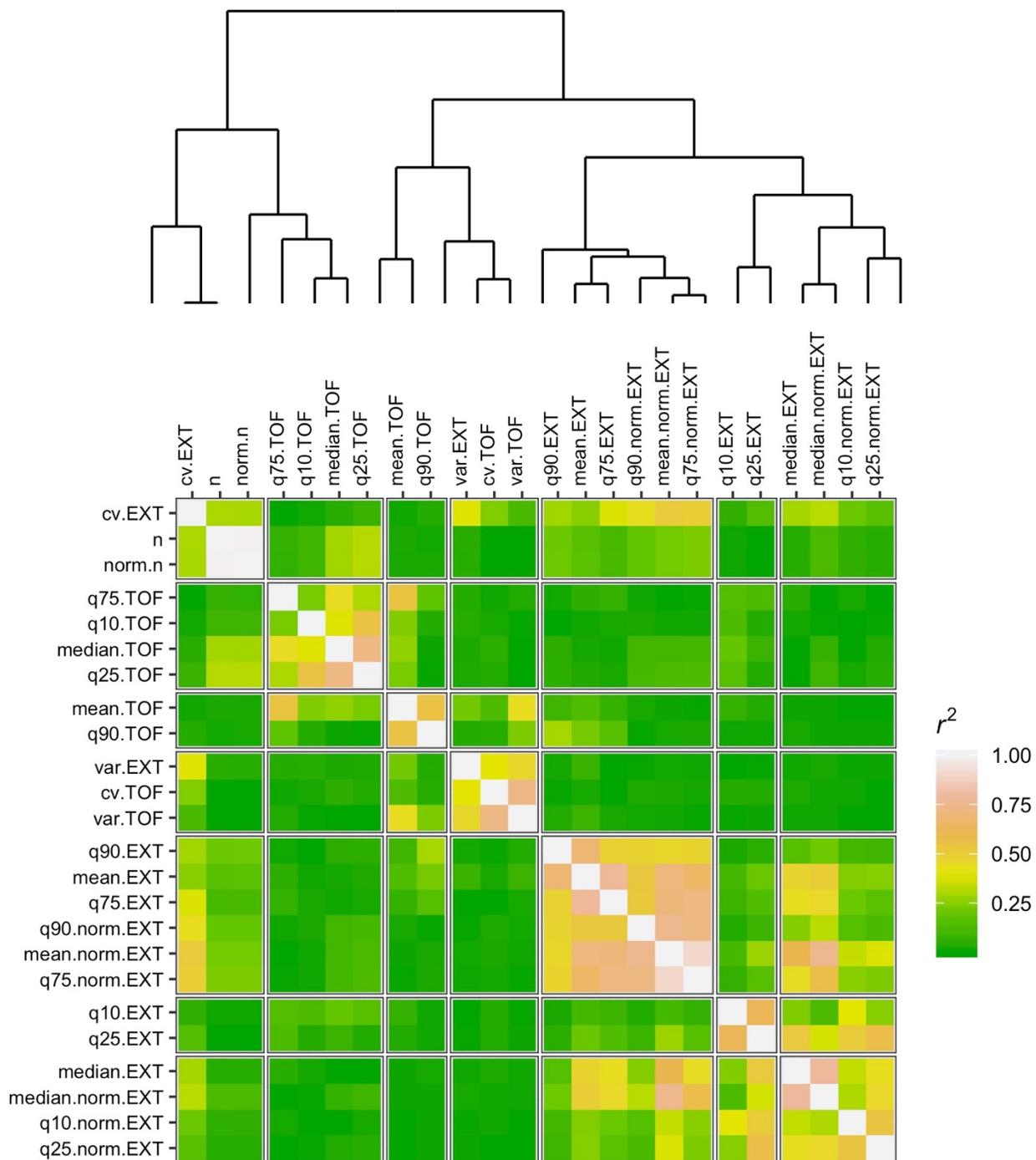
## Copper



# Diquat

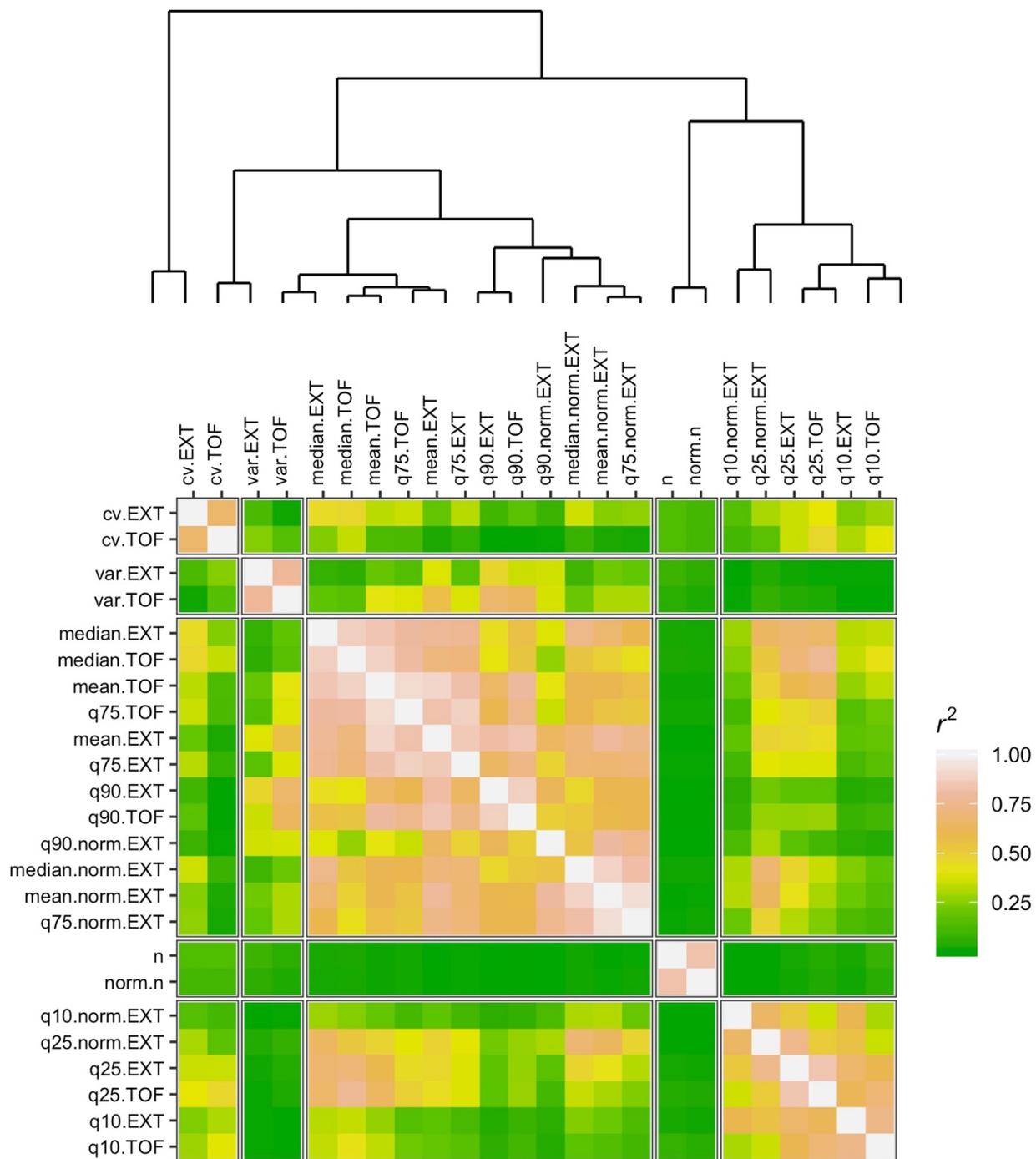


## Fluoxetine

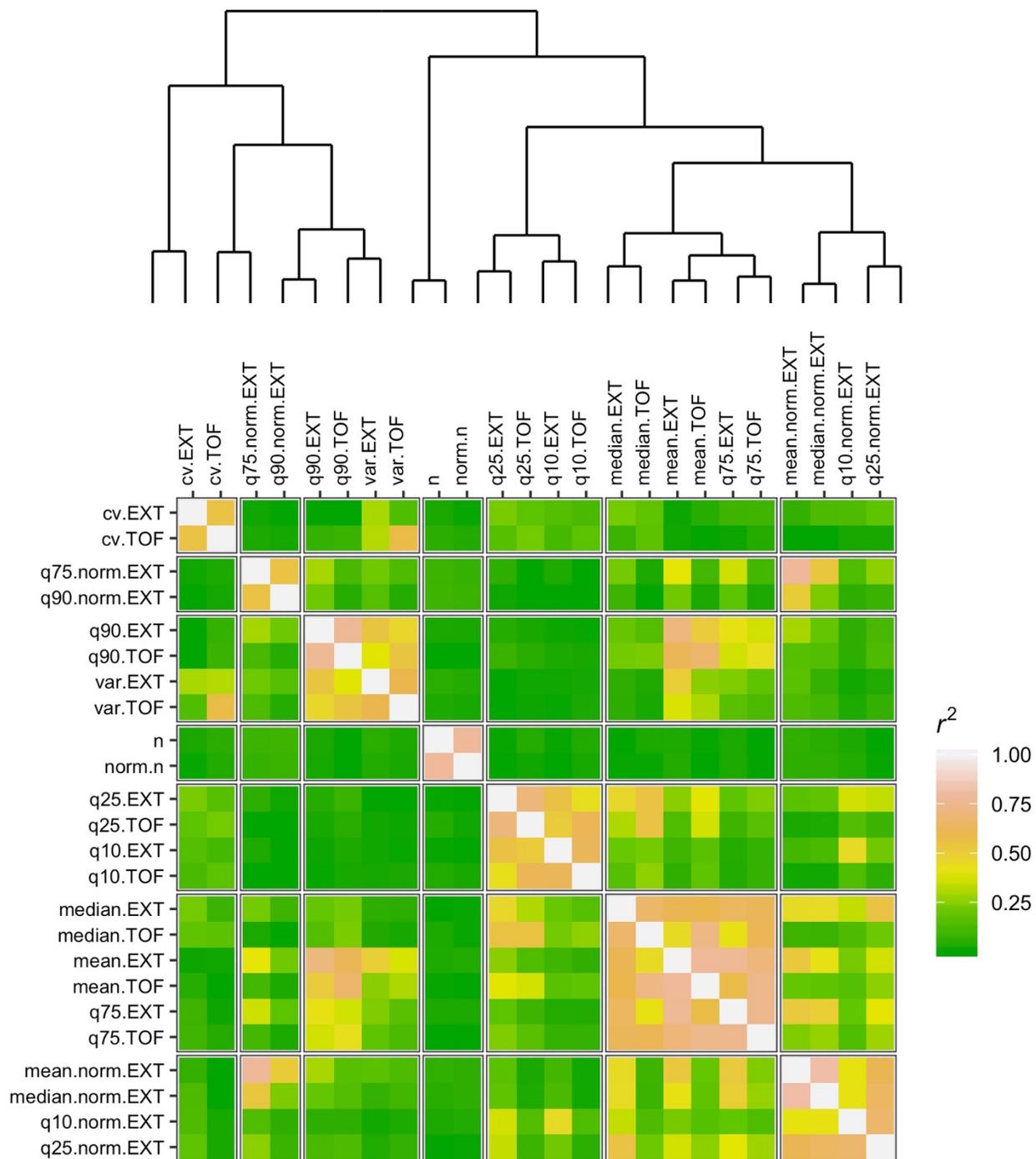




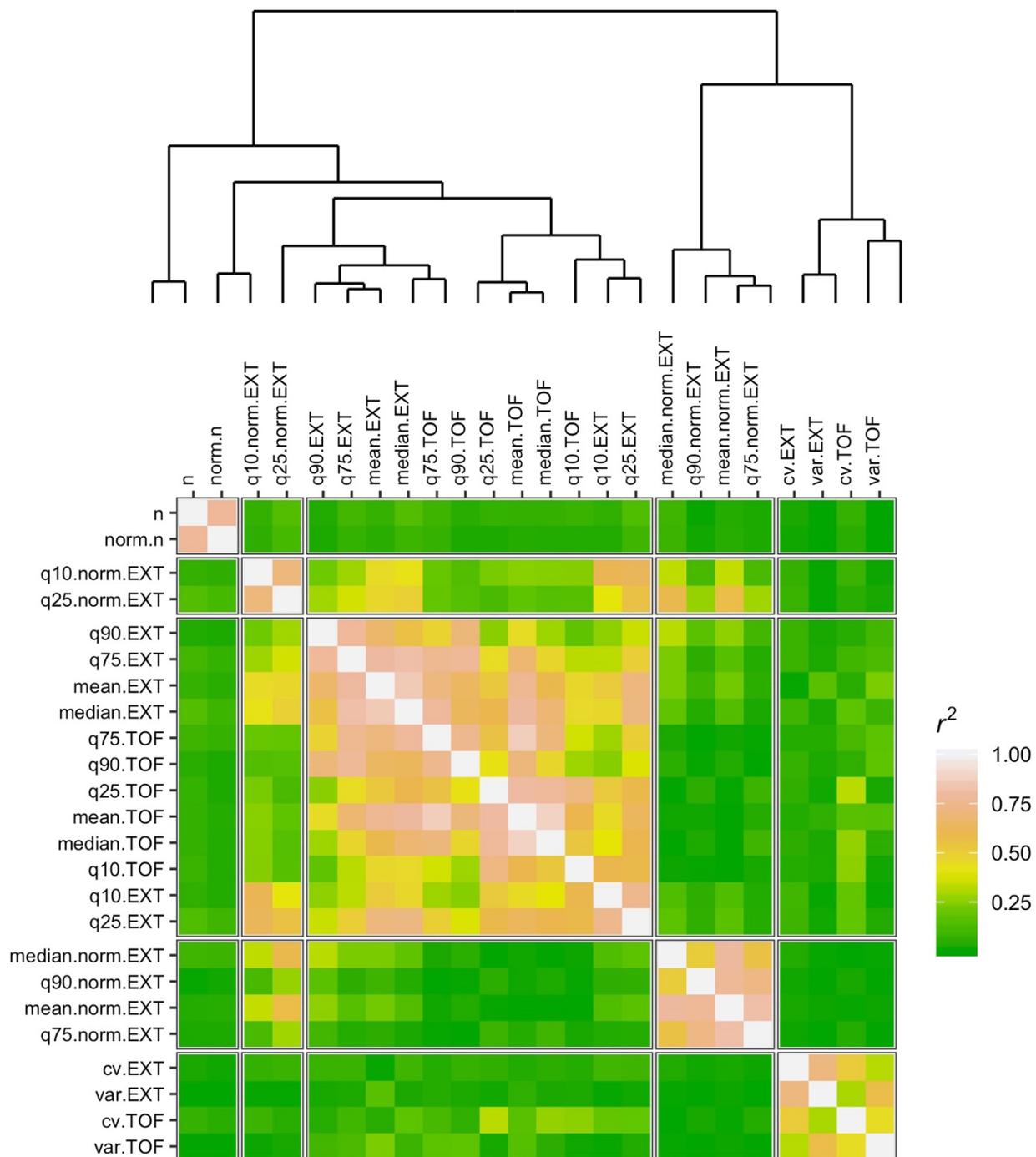
## Irinotecan



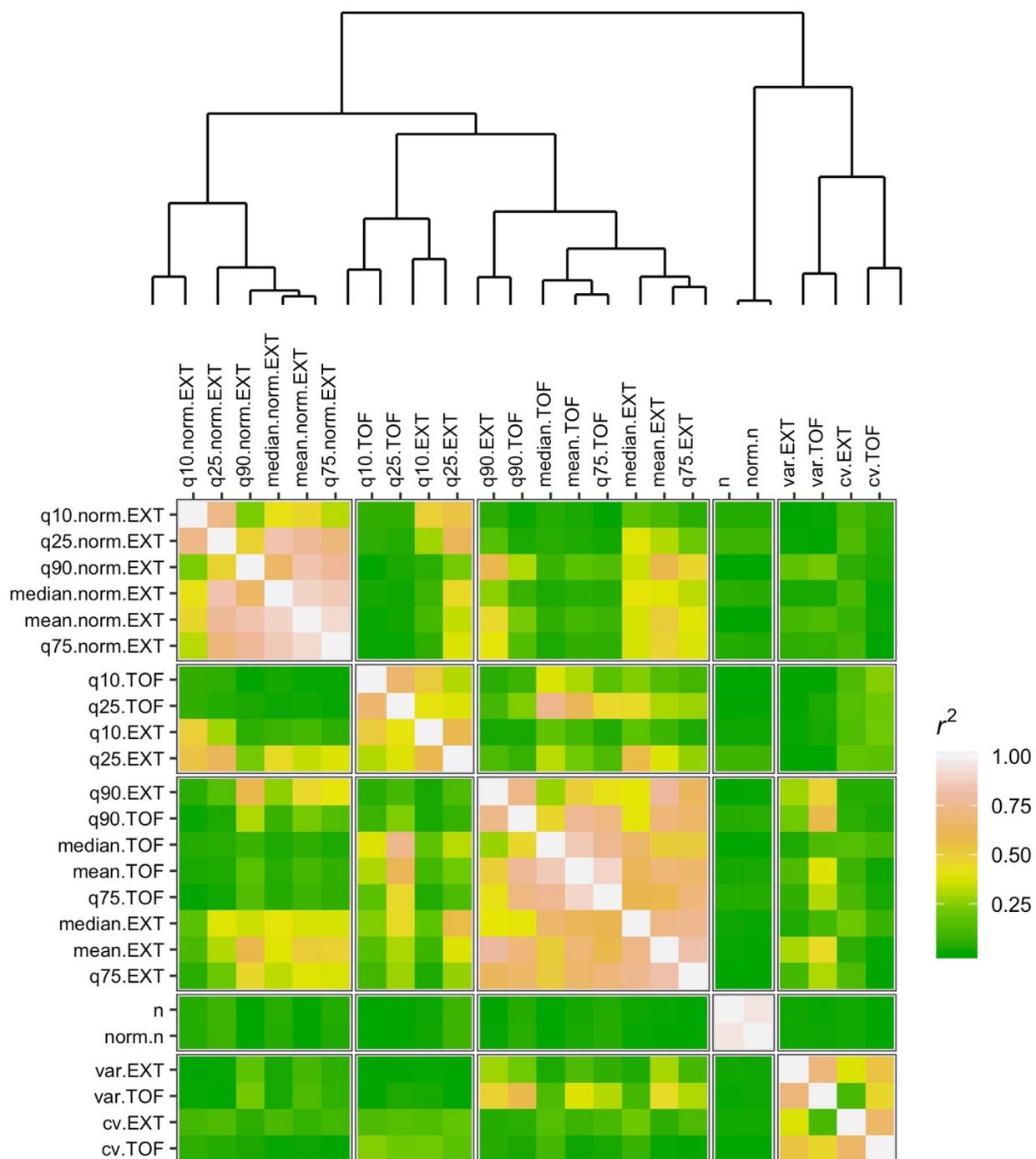
## Mechlorethamine



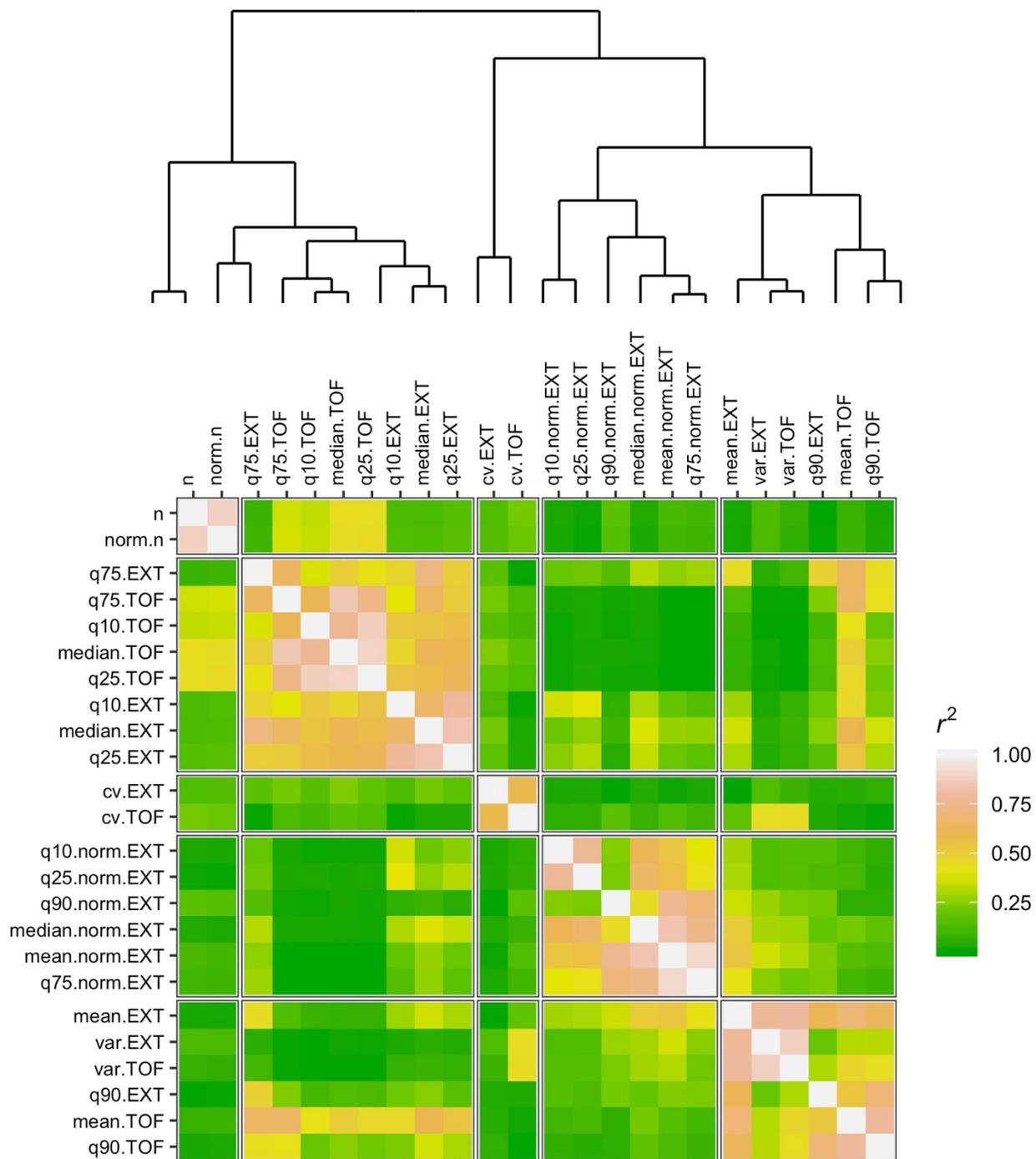
## Paraquat



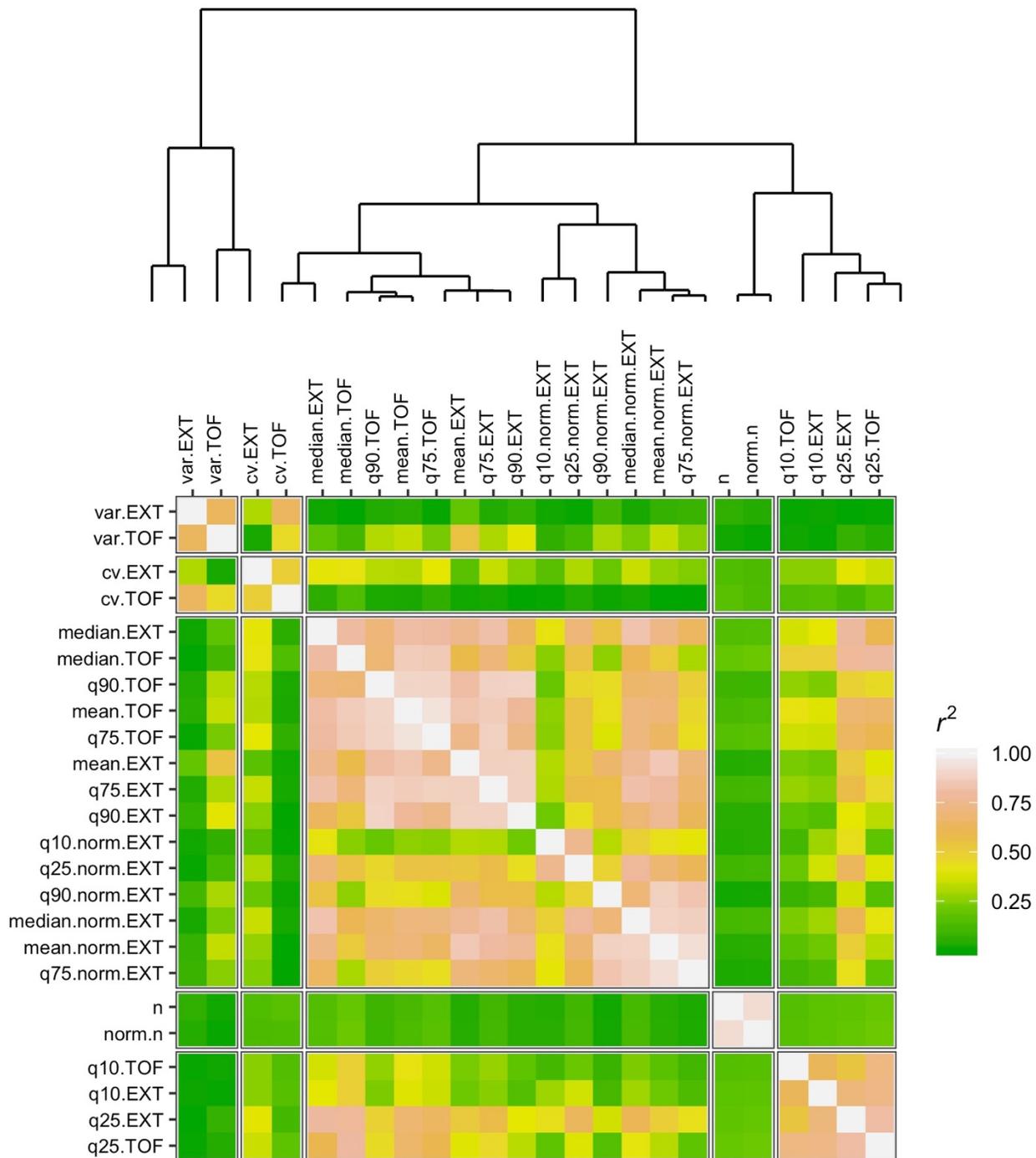
## Silver



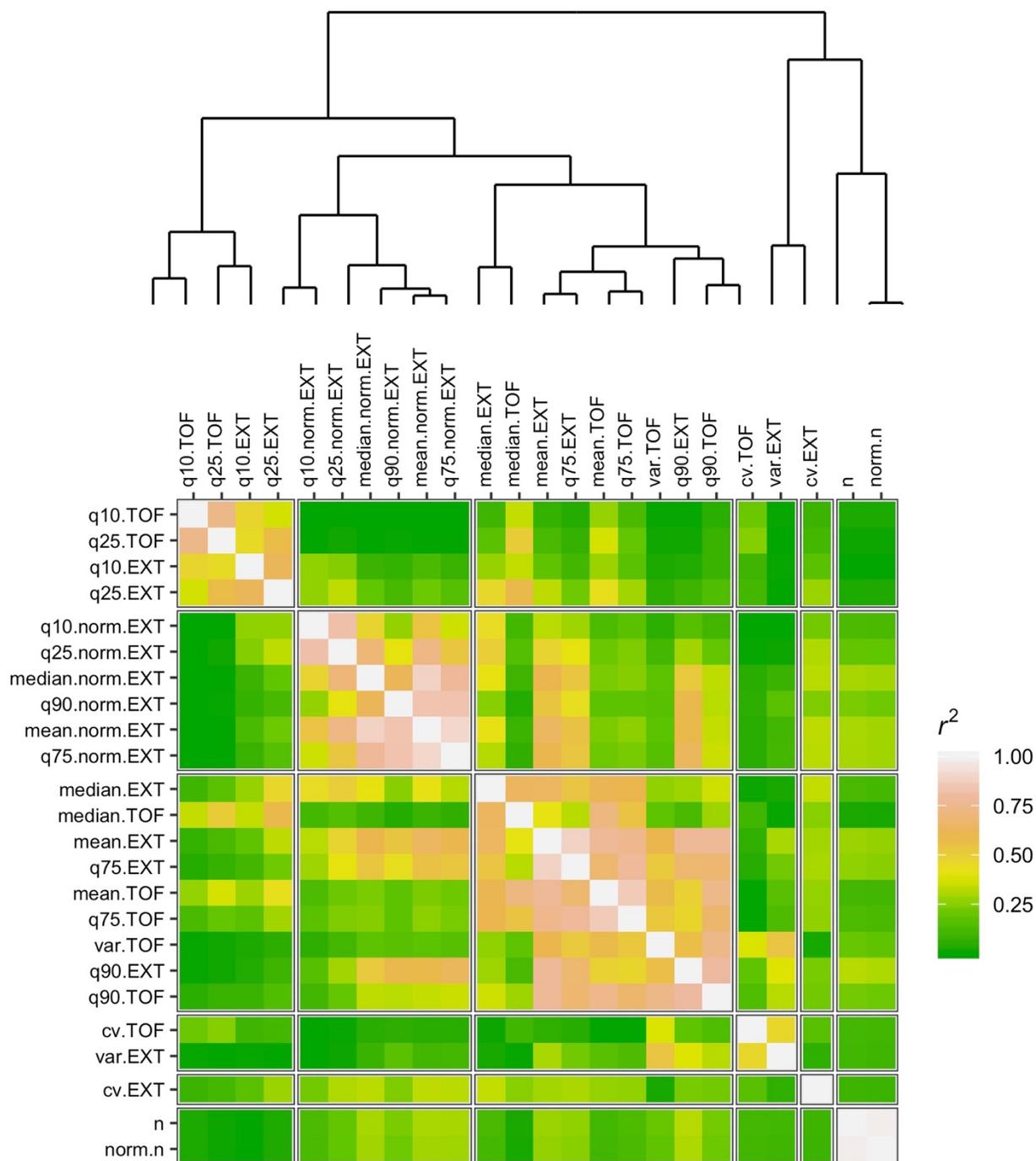
## Topotecan



# Tunicamycin



## Vincristine



**Figure S3-2 Traits measured by the HTA are correlated**

Correlation coefficients ( $r^2$ ) between all pairwise combinations of traits based on the RIAL phenotypes for the indicated concentration of the toxin are plotted as a heat map. Traits are labeled on the x- and y-axes, organized by hierarchical clustering (top) of the correlation matrix (bottom) for each toxin. Colors represent correlation coefficients increasing from green to yellow to pink to white.

# **Chapter 4 - Escape-room inspired learning enrichment**

## **Preface**

During my tenure as a graduate student, I was responsible for being a teaching assistant for two courses. For one of my TA assignments, I worked with Erik on his 2017 genetic analysis course. I enjoyed designing problem sets and holding office hours, during which I could share my passion for genetics with the undergraduate students. At the time, I had recently designed a science-themed escape room for the Biotechnology Training Program orientation, which turned out to be a successful team-building activity. I had the idea of designing an escape room for the genetic analysis course that would allow the students to review for the final exam in a unique manner. Using the theme of firefly genetics, I created a storyline that would walk the students through each of the major steps of genetic analysis. The escape room was enjoyable and effective for the students, so Erik and I facilitated the review game in 2018 and 2019. I presented a poster regarding the escape-room review game at TeachX 2018, where many of the conference attendees encouraged me to publish the game and its results. The following chapter describes the game and our plan to assess its impact, and we plan to submit this work to an educational journal.

## Abstract

Active learning strategies engage students in instruction by making them an integral part of the teaching process. Gamification is one type of active learning approach that incentivizes student participation by incorporating gaming elements into learning. Although gamification might be an effective strategy for the introduction of new material, incorporating elements of games into a review activity can motivate students to study, help them organize material, and identify material that they should review further before an exam. Escape rooms are collaborative problem-solving challenges and aspects of escape-room experiences can be applied to active learning. Here, we describe an escape-room inspired review game for an advanced genetic analysis class at Northwestern University. We provide a detailed account of each puzzle within the review game and describe our plans for assessing the impact of the game on learning outcomes.

## Introduction

Students comprehend and remember material most effectively when they are engaged in the learning process. Learner-centered approaches to teaching involve the use of active, cooperative, and problem-based learning techniques [268]. Many studies have identified positive effects of learning-centered instruction as opposed to traditional teaching-based lectures, especially in science, technology, engineering, and math (STEM) disciplines [268–273]. These learning-centered approaches offer students a chance to apply critical thinking skills to understand the applications of a lesson at a greater depth.

Gamification, or the use of gaming elements in non-gaming contexts [274], is an increasingly popular method for implementing active learning in the classroom. Although the use of gamification to improve learning outcomes has mixed results, proper implementation of game-based learning can increase student engagement and motivation [275–278]. Therefore, the combined use of active learning and gamification strategies to review concepts covered in a course, as opposed to introduce new concepts, can incentivize studying, help students organize material, and assist students in evaluating their depth of knowledge [279,280]. The challenge for educators is to determine a game format that works for their students and for the material taught in the course.

Escape rooms are interactive experiences that challenge participants to use critical thinking and cooperation to achieve a goal [281]. Although many escape rooms are recreational, the use of escape-room inspired experiences for education has rapidly grown, especially in medical education [282–289]. Students report that the escape room games in the classroom improved their motivation to study [282,283,287], helped them conceptualize and retain curriculum [282,283,285,289], revealed topics they should spend more time reviewing [285], and were overall enjoyable experiences [282–288]. The hands-on challenges and time sensitivity of an escape room has obvious parallels to some of the practical skills needed for medical professions. Therefore, the use of escape rooms in medical education seems to be a clear fit. However, escape rooms might not be limited in this capacity. Indeed, they might be effective as review games for a wide array of subjects outside of the medical field.

Here, we present the design and implementation of an escape room game for BIOL\_SCI 393: Genetic Analysis at Northwestern University. This activity supplemented study guides and review

sessions for the final exam. During the game, students collaborated to solve genetic puzzles, ultimately unlocking the “Nobel Prize”. Each puzzle in the game covered a core concept that was taught in the course and would appear on the final exam. We discuss the workflow of the game, assess student responses, and offer suggestions for designing future escape room review activities.

## Methodology

### *Participants:*

Students enrolled in BIOL\_SCI 393, Genetic Analysis, were offered the option to participate in the escape room exam review session. Eighteen students were enrolled in the course and all participated in the escape room activity.

### *Location and set up:*

The escape room review session took place in a standard lecture room on campus. Students were split into two groups of nine each to offer each student a more active role in the game. The course instructor and teaching assistant acted as facilitators for the activity, where they could deliver clues when appropriate but did not offer solutions to challenges in the room.

Each activity was placed in a wooden box with some form of lock and the boxes were scattered throughout the room (**Supplement A**). When the students entered the game room, they noticed six locked boxes (**Supplement A**), a list of tools for a mutagenesis screen written on the board (**Supplement A**), a chalkboard drawing labeled “GFP plasmid” (**Supplement A**), a chalkboard

drawing of a cell lineage (**Supplement A**), and a mysterious poster board with hooks and elastic ID badge holders (**Supplement A**).

## Workflow

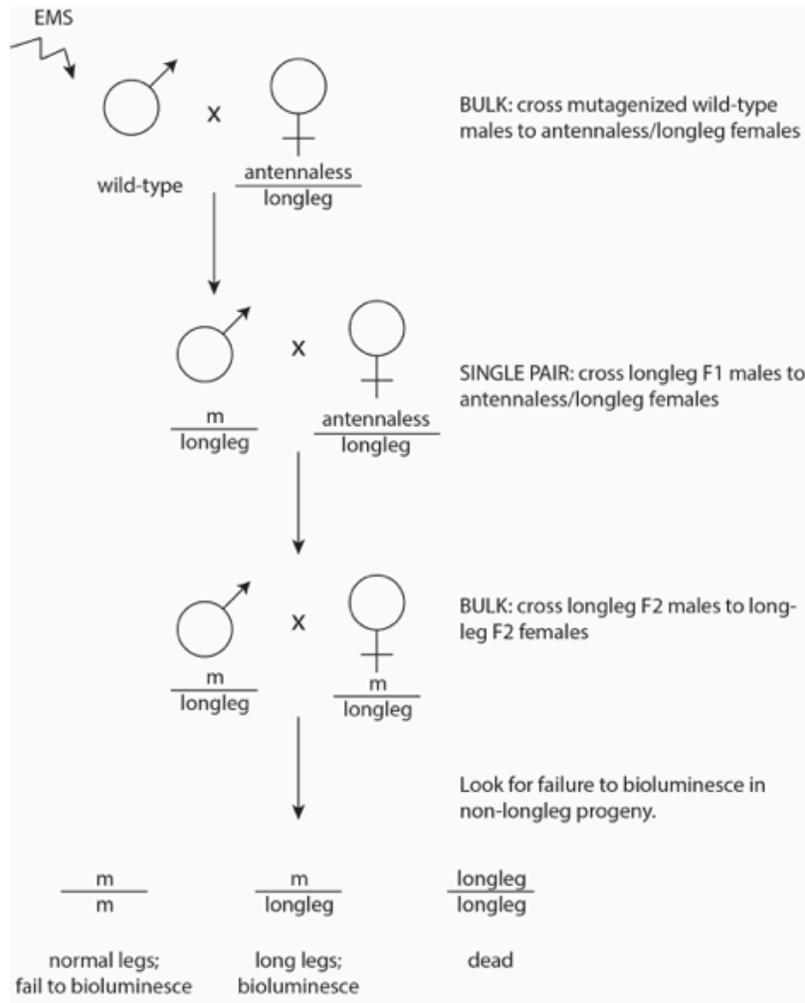
Before the escape room review game took place, the facilitators printed all prompts (**Supplement B**), tested all puzzles and locks, and set up the room according to the setup guide (**Supplement A**). Facilitators followed the workflow in **Supplement C** to ensure the game was consistent across groups. A summary of the escape room workflow is described below.

### Introduction

After all students had entered the room, they were presented with the **Introductory Prompt (Supplement B)**: “Congratulations! You have won the Nobel prize for your work in *Photinus pyralis*, which, as you obviously know, is the scientific name for the big dipper firefly. You performed flawless genetic experiments and identified several genes involved in the bioluminescence of this wonderful organism and found genetic tests for diseases in the species, eventually earning you the highest honor in the field. However, your jealous lab mate doubts your knowledge of genetics and has hidden your Nobel prize in this room. To prove that you have earned this award and claim what is rightfully yours, you must complete a series of tasks, unlock several boxes, and find your Nobel prize before you give your big acceptance speech in one hour. Good luck!”

### Step 1: Mutagenesis

The facilitator handed **Prompt 1 (Supplement B)** to the students and started the timer that would expire after one hour. The prompt indicated that the students needed to draw a mutagenesis screen on the board (**Figure 4-1**).



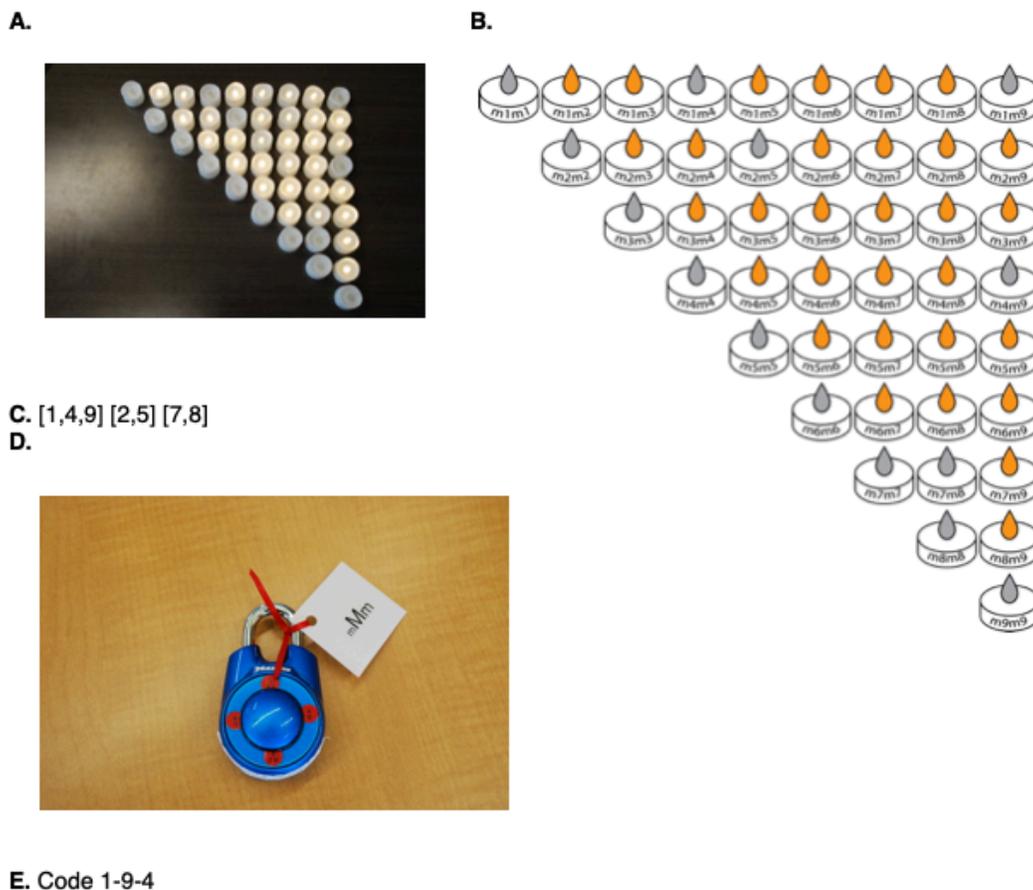
**Figure 4-1 Mutagenesis screen answer**

The solution to **Prompt 1** is shown.

### Step 2: Complementation

When the mutagenesis screen was correct, the facilitator handed **Key 1 (Supplement A)** to the students, which they used to unlock **Box 1 (Supplement A)**. Inside, they found **Prompt 2 (Supplement B)**, and a set of tea lights that were each labeled with a combination of two

mutations (**Figure 4-2**). The students tested the ability of each combination of mutants to glow by turning on each tea light and used this information to identify the largest non-complementation group (**Figure 4-2**). The numbers of the mutants in the largest non-complementation group were used to unlock **Box 2 (Supplement A)** once they were arranged according to the clue on the lock (**Figure 4-2**).



**Figure 4-2 Complementation puzzle**

**A.** Electronic tea lights represent double-mutant fireflies for pairwise combinations of nine mutations. Each tea light is labeled with the two mutations it represents. Tea lights are arranged with m1-m9 from left to right paired with m1-m9 from top to bottom. **B.** An orange flame indicates that the tea light has batteries and will therefore light when turned on, whereas a grey flame indicates that the battery was removed and would therefore fail to light when turned on. **C.** Tea lights that failed to light indicate that the given mutations fail to complement. The non-complementation groups are indicated. **D.** A hint written on the lock of Box 2 is shown. **E.** The largest non-complementation group from **C** is placed into the order indicated by **D** and can be used to unlock Box 2.

### Step 3: Epistasis

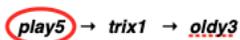
Inside **Box 2 (Supplement A)**, the students found **Prompt 3 (Supplement B)** and the **Epistasis Circuit (Figure 4-3)**. The students flipped the switches on the circuit board on and off to understand how the presence or absence of certain gene products affected the output of the phenotype (the color of the light). Using this information, the students organized the biosynthetic pathway of these modeled genes and identified the most upstream gene in the pathway (**Figure 4-3**). This gene name was used to unlock **Box 3 (Supplement A)**.

A.



B.

<i>trix1</i>	<i>oldy3</i>	<i>play5</i>	Light Color
ON	ON	ON	Green
ON	ON	OFF	White
ON	OFF	ON	Yellow
ON	OFF	OFF	White
OFF	ON	ON	Red
OFF	ON	OFF	White
OFF	OFF	ON	Red
OFF	OFF	OFF	White

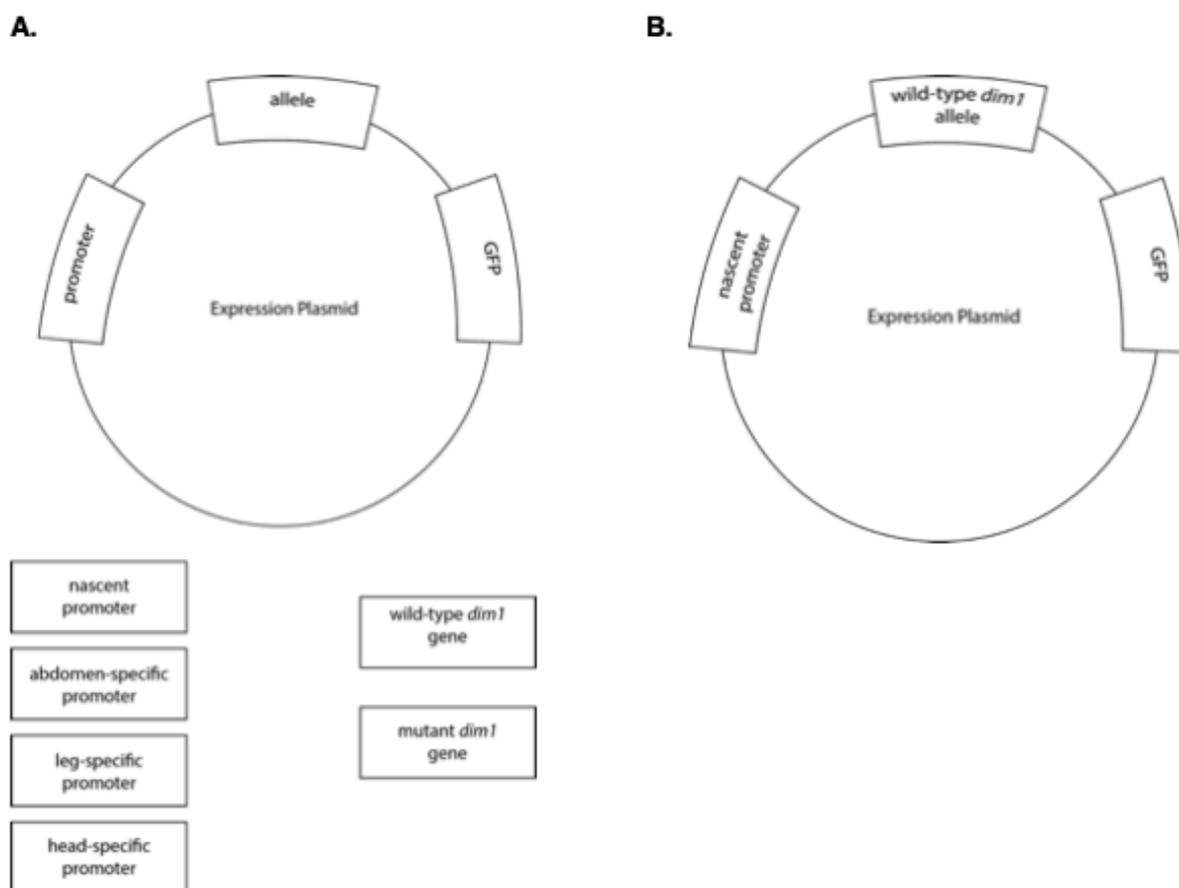


**Figure 4-3 Epistasis circuit**

**A.** A circuit board modeling the biochemical pathway that determines the color of light emission is shown. Three switches represent gene products of the *trix1*, *oldy3*, and *play5* genes. **B.** The table indicates the color of the light output when each gene product is turned on or off. Given the information in the table, the biochemical pathway is modeled below, with the most upstream gene product circled in red. This gene name can be used to unlock Box 3.

### Step 4: Designing plasmids

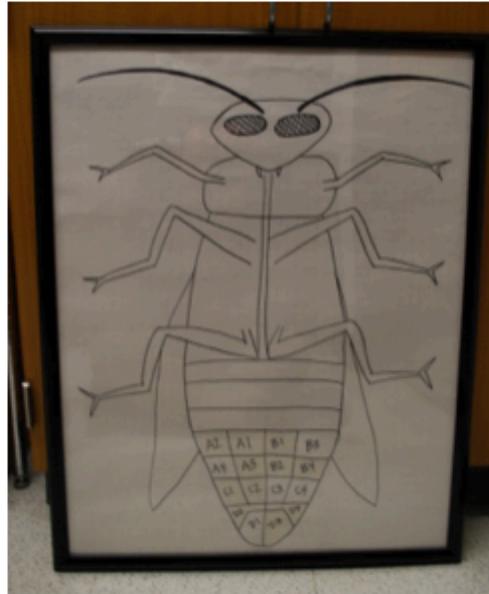
In **Box 3 (Supplement A)**, the students found **Prompt 4 (Supplement B)**, magnets labeled with promoter types and alleles of genes (**Figure 4-4**), and a blacklight. The students placed the magnets into the **Expression Plasmid** on the blackboard (**Supplement A, Figure 4-4**). When they selected the correct magnets (**Figure 4-4**), the facilitator handed them **Gene Expression Poster (Figure 4-5)**.



**Figure 4-4 Expression plasmid**

**A.** The expression plasmid to be drawn on the board before the activity is shown at top. Below, rectangles represent magnets that are labeled with promoter options and alleles of genes. **B.** The correct selection of magnets to complete the promoter is shown.

A.



B.



**Figure 4-5 Gene expression poster**

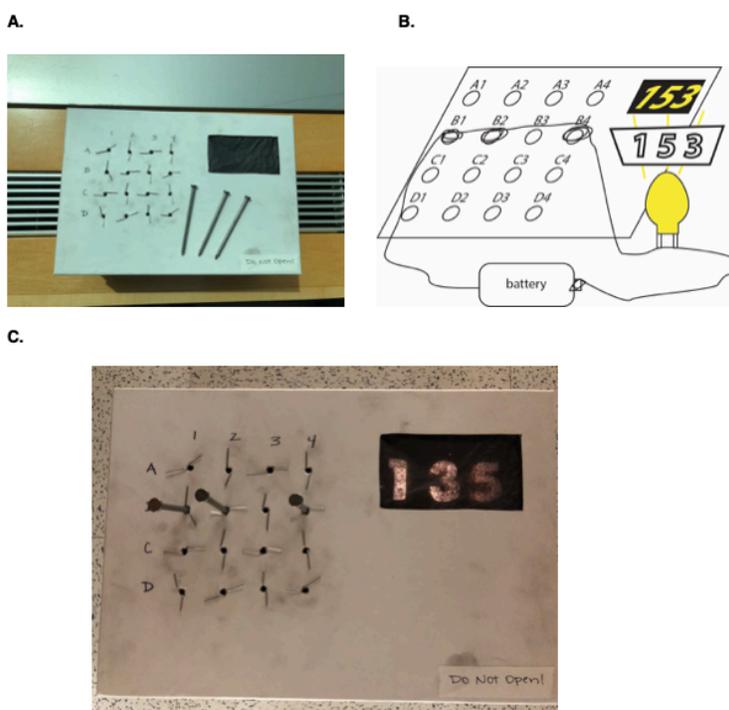
**A.** A poster board with a drawing of a firefly is shown. Each of the abdominal cells are labeled. **B.** When a blacklight is shown onto the drawing, certain cells illuminate. This represents which cells fluoresce when the GFP plasmid from Figure 4B is placed into a mutant animal. Cells A2, A3, B1, B2, B4, C1, C2, D2, D3, and D4 illuminate under the blacklight.

### Step 5: Nascent expression

The students used the blacklight to look for expression of the gene product on the **Gene Expression Poster**. Once they correctly identified which cells were fluorescing (**Figure 4-5**), the facilitator handed them **Prompt 5 (Supplement B)**.

### Step 6: Designing experiments

The students selected the experimental tools they would need to test whether the gene function was required in certain cells. When they selected the correct combination of tools (**Supplement C**), the facilitator handed them **Prompt 6 (Supplement B)** and the **Mutant Firefly Box (Figure 4-6)**.



**Figure 4-6 Mutant firefly box**

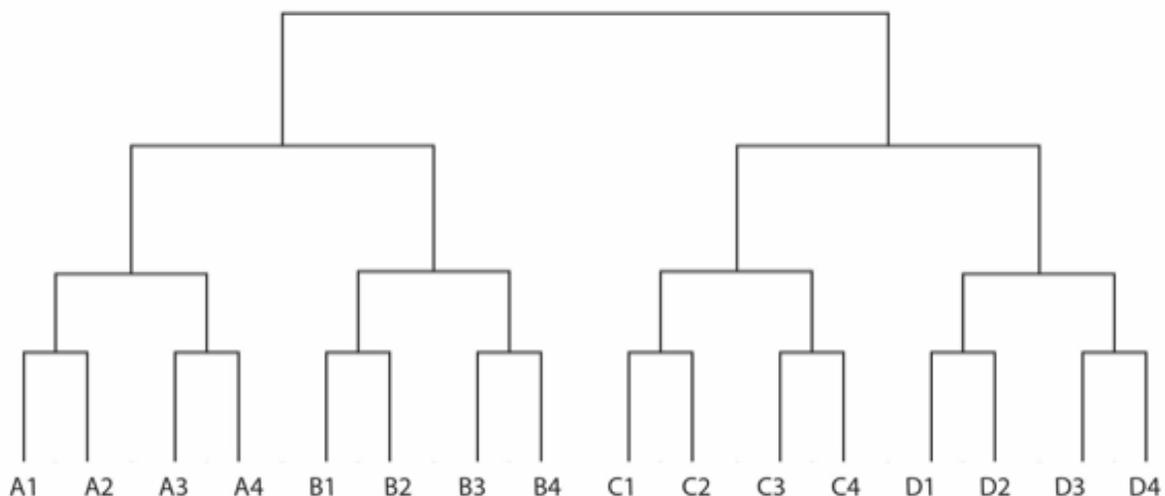
**A.** A circuit box representing the abdominal cells of a mutant firefly is shown. **B.** A diagram of the circuit inside the box is shown. **C.** When iron nails are inserted into the proper locations (B1, B2, and B4), the circuit is completed and the box illuminates. The code shown can be used to unlock Box 4.

### Step 7: Cell autonomy

Following **Prompt 6**, students plugged the three nails into the given holes on the **Mutant Firefly Box**. When they did not see the box light up, they told the facilitator what this means about the cell autonomy of the gene product (**Supplement C**). When their answer was correct (**Supplement C**), the facilitator handed them **Prompt 7 (Supplement B)**.

### Step 8: Rescue and Ablation

Combining the information on **Prompt 7 (Supplement B)**, the **Cell Lineage** drawn on the blackboard (**Supplement A, Figure 4-7**), and information from Steps 6 and 7, the students deduced where the function of the gene product is required to make the animal light up. They then placed the three iron nails into the corresponding slots in the **Mutant Firefly Box (Figure 4-6)**. When correct, the **Mutant Firefly Box** lit up and displayed the code to unlock **Box 4 (Figure 4-6)**.



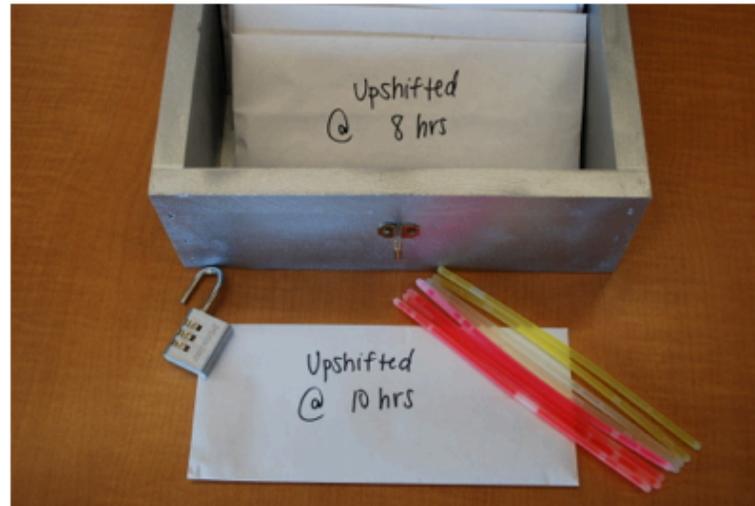
**Figure 4-7 Abdominal cell lineage**

A diagram of the cell lineage for all abdominal cells is shown.

### Step 9: Time of Activity

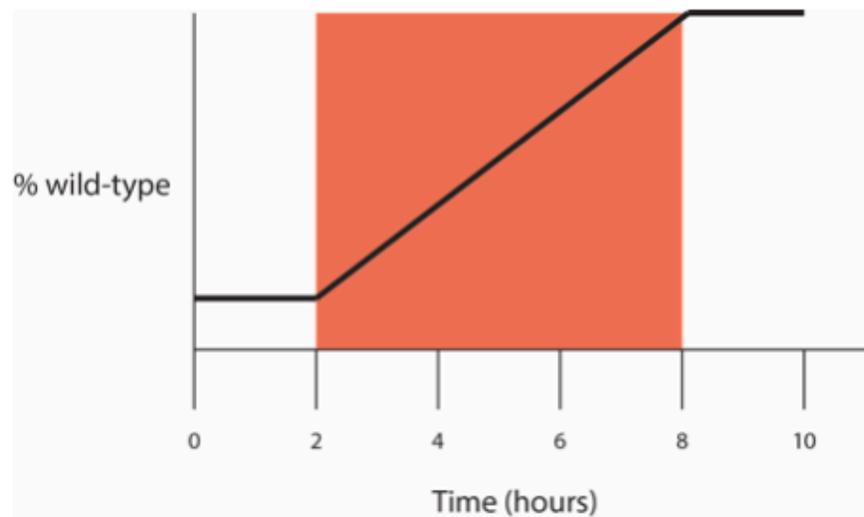
In **Box 4**, the students found **Prompt 8 (Supplement B)** and six **Time of Gene Action** envelopes labeled “0”, “2”, “4”, “6”, “8” and “10” that each contained glow sticks (**Supplement A, Figure 4-8**). The students tested the glow sticks for number of “animals” in the population that had the wild-type glowing phenotype. If the students did not know how to use glow sticks, the facilitator instructed them to crack them. The students then stated when the gene function was required. If correct (**Figure 4-8**), the facilitator handed them the **Modes of Inheritance Pedigrees (Figure 4-9)** and **Prompt 9 (Supplement B)**.

**A.**



**B.**

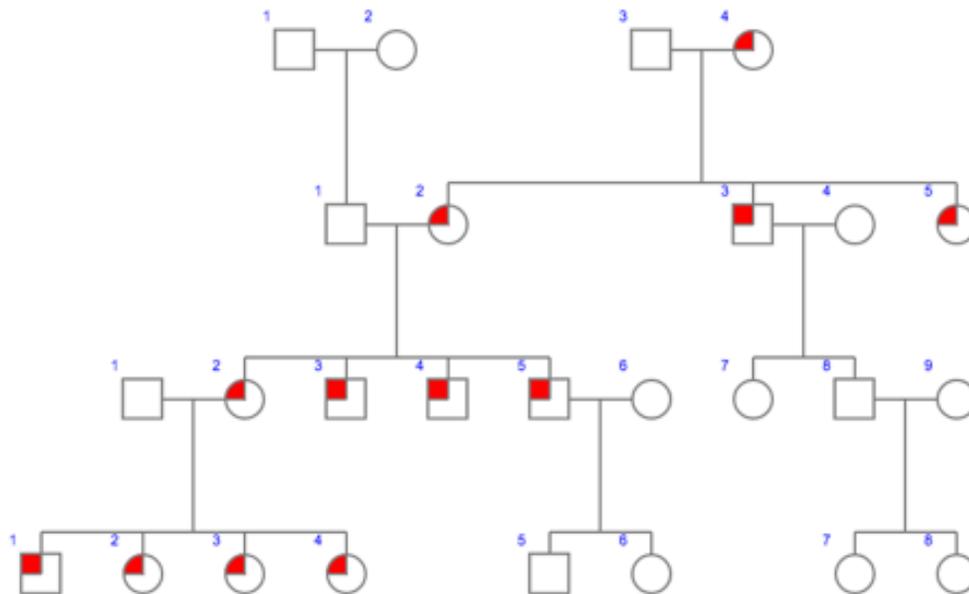
0 hours: 0/8 glowing  
 2 hours: 0/8 glowing  
 4 hours: 3/8 glowing  
 6 hours: 5/8 glowing  
 8 hours: 8/8 glowing  
 10 hours: 8/8 glowing



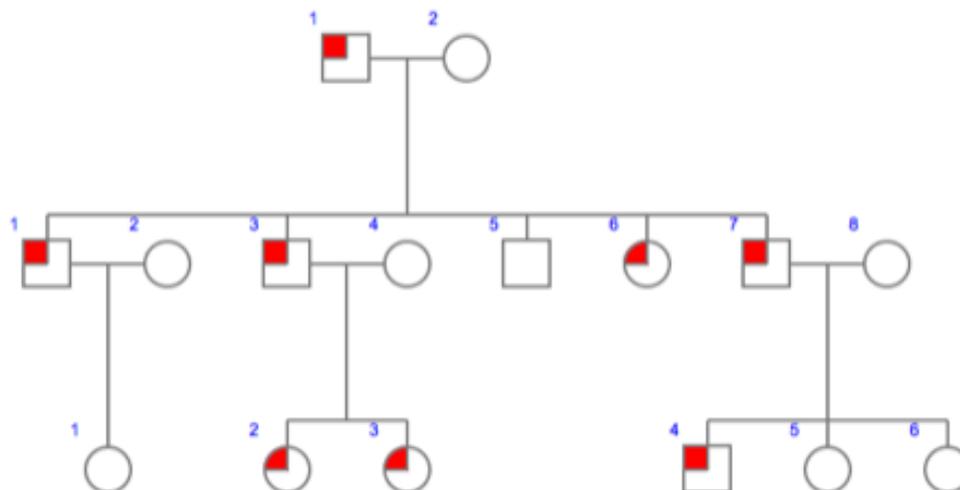
**Figure 4-8 Time of gene action**

**A.** Envelopes containing glow sticks represent temperature-sensitive mutant fireflies. Each envelope contains eight glow sticks that were shifted from the permissive to the restrictive temperature at the indicated time point. **B.** After cracking the glow sticks in each envelope, the number of animals that glow after each temperature shift are shown. This information can be used to generate the time of gene action diagram (below).

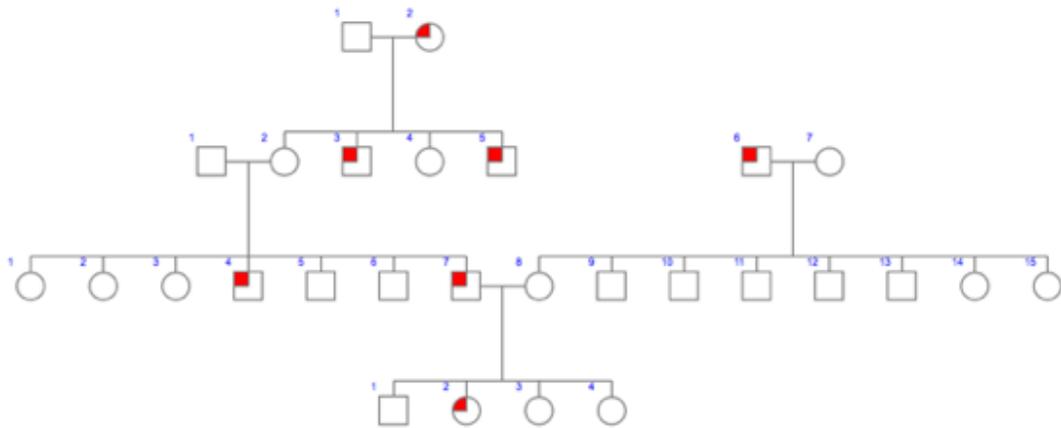
Pedigree 1:



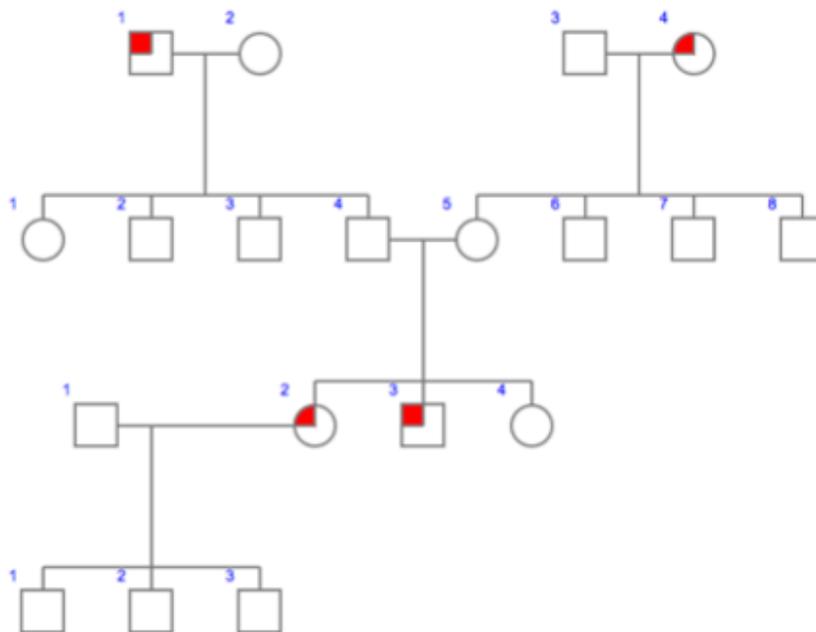
Pedigree 2:



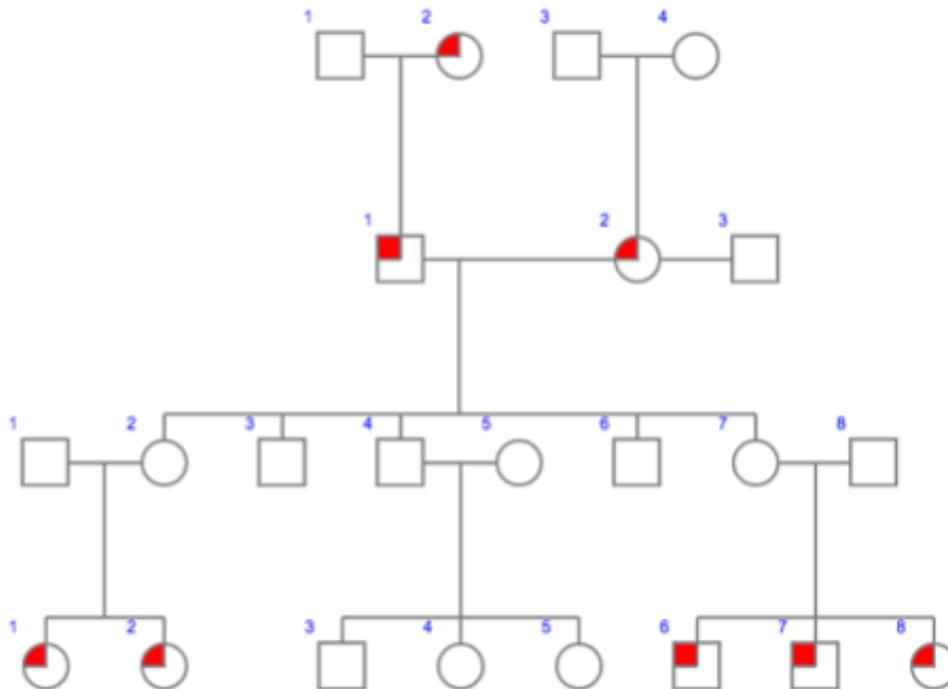
Pedigree 3:



Pedigree 4:



Pedigree 5:

**Figure 4-9 Modes of inheritance pedigrees**

Five pedigrees, each describing a different mode of inheritance are shown. Circles indicate females and squares indicate males. Affected individuals are shown in red.

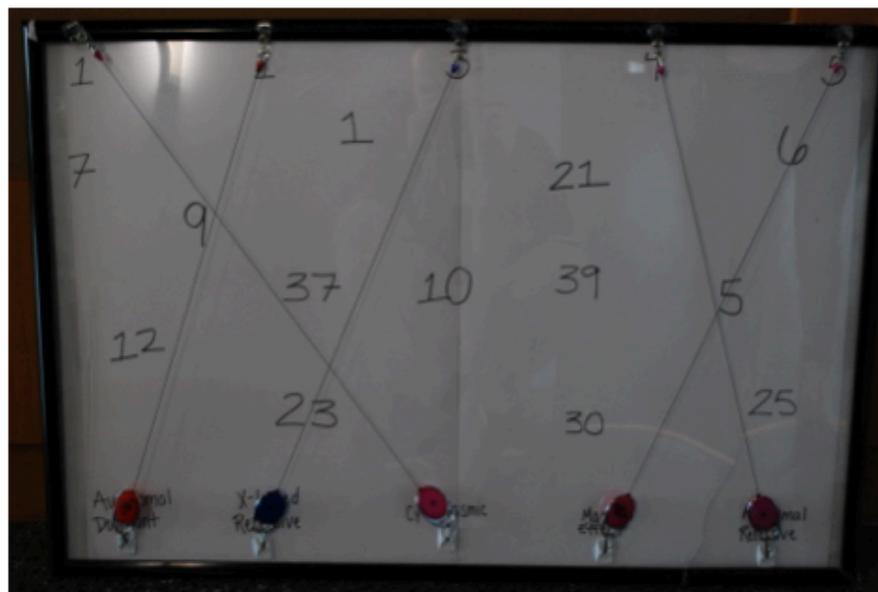
### Step 10: Modes of Inheritance

The students determined the mode of inheritance for each of the **Modes of Inheritance Pedigrees**. They then matched each pedigree to its proper mode of inheritance on the **Mysterious Poster Board** by stretching the ID tags from top hooks to bottom hooks. When correct, the strings from the ID tags made a pattern that crossed over three numbers on the **Mysterious Poster Board (Figure 4-10)**. These three numbers were used to unlock **Box 5 (Supplement A)**.

A.



B. 9-23-5

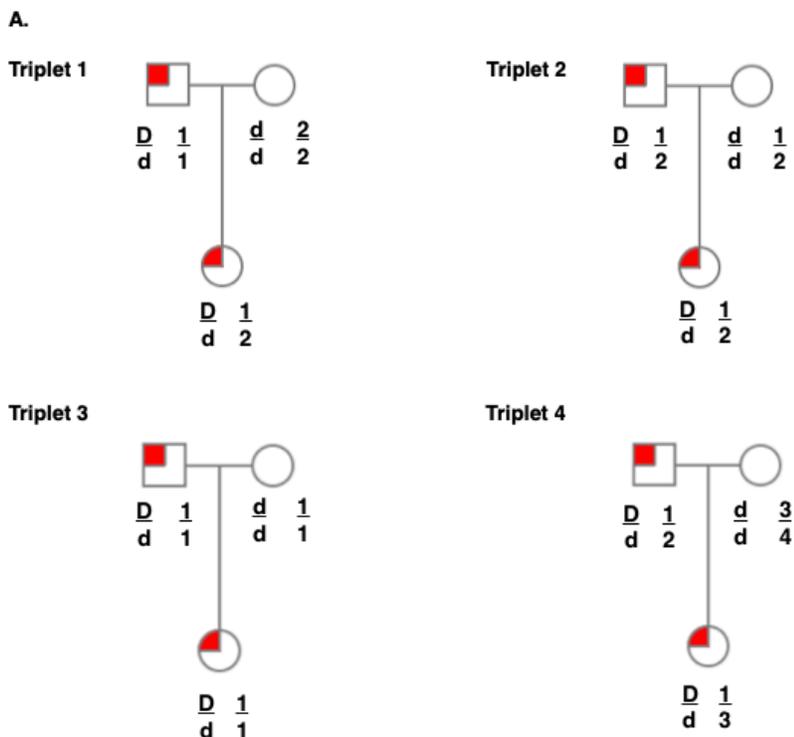


**Figure 4-10 Mysterious poster board**

**A.** A mysterious poster board that is found in the escape room is pictured. **B.** When each pedigree is matched to the proper mode of inheritance, the lines intersect over particular numbers, as shown. These numbers can be used to unlock Box 5.

### Step 11: Informative Individuals

In **Box 5 (Supplement A)**, the students found **Prompt 10 (Supplement B)** and four **Triplet Pedigrees (Figure 4-11)**. When the students correctly selected the most informative triplet (**Figure 4-11**), the facilitator handed them a **Linkage Pedigree** derived from the F1 individual of the selected triplet (**Figure 4-12**). The students determined the number of recombinant progeny in the pedigree and calculated the LOD score of the linkage between the disease and the marker loci using a theta score indicated on the **Linkage Pedigree (Figure 4-12)**. The correct LOD score unlocked **Box 6 (Supplement A)**.

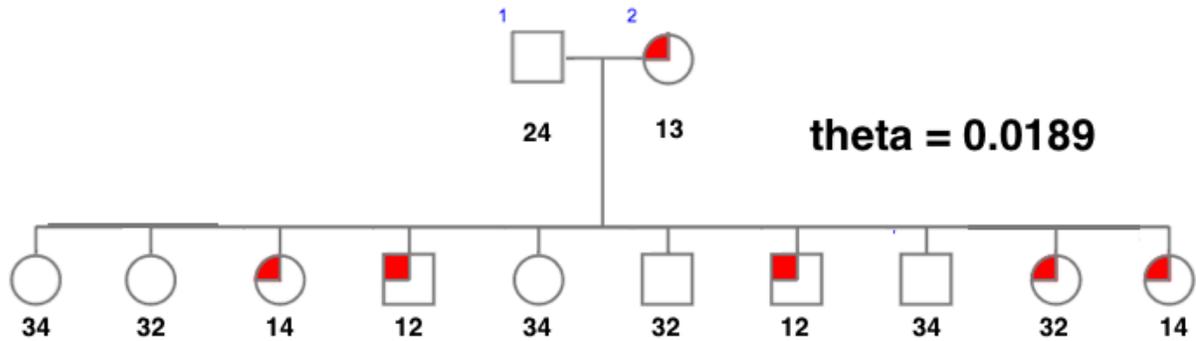


**B.** Triplet 4 is the most informative triplet. The F1 female received the D allele and marker 1 from her father. She received the d allele and marker 3 from her mother.

**Figure 4-11 Triplet pedigrees**

**A.** Pedigrees for four triplets are shown. Circles indicate females and squares indicate males. Affected individuals are shown in red. Beneath each individual, the alleles at the disease locus (D or d) and at a marker locus (1, 2, 3 or 4) are printed. **B.** An explanation for the most informative triplet is written.

A.



B. With 9 parental and 1 recombinant progeny:  $\log_{10} \frac{(0.0189)^1 * (1-0.0189)^9}{(0.5)^{10}} = 1.212$

#### Figure 4-12 Linkage pedigree

**A.** A pedigree derived from triplet 4 from Figure 4-11 is shown. Circles indicate females, squares indicate males, and affected individuals are shown in red. Beneath each individual, their alleles for a marker locus are shown. **B.** A LOD score calculation is shown for the pedigree in **A.** This LOD score can be used to unlock Box 6.

#### Step 12: GWAS

Inside **Box 6 (Supplement A)**, the students found **Prompt 11 (Supplement B)**, and a collection of **GWAS Pipe Cleaners**, folded into shapes (circles or straight lines) that represented the flight patterns of mutant fireflies. Each “flight pattern” was labeled with alleles at three potentially correlated SNPs (**Figure 4-13**). Students identified the SNP most highly correlated with the mutant (circular) flight path (**Figure 4-13**). Then, they used the genetic sequence of the triplet containing the correlated SNP with the putative disease allele (**Figure 4-13**) to unlock **Box 7 (Supplement A)**.

A.



B.

	SNP 1		SNP 2		SNP 3	
	<b>A</b>	T	<b>C</b>	T	<b>G</b>	C
<b>Curly (affected)</b>	13	10	12	11	18	5
<b>Straight (unaffected)</b>	17	14	16	15	4	27
<b>Chi squared</b>	0.015		0.002		23.36	

...ATGTGGTAA...

#### Figure 4-13 GWAS flight patterns

**A.** Pipe cleaners model the flight patterns of wild-type (straight) and mutant (circular) flight patterns of fireflies. The tag on each flight pattern indicates the allele at three SNPs — SNP 1 (A or T), SNP 2 (T or C), and SNP 3 (G or C). **B.** The table indicates the number of affected (circular flight patterns) and unaffected (straight flight patterns) individuals with each allele at each SNP. The allele most correlated with the mutant phenotype is shown in bold. The entire sequence around this SNP is written below the table and can be used to unlock Box 7.

#### Step 13: Receive Nobel Prize

Inside **Box 7 (Supplement A)**, the students found a replica of the Nobel Prize. The timer was stopped to record the amount of time needed to complete the escape room activity.

### Step 14: Debriefing

After each group completed the escape room, the facilitators walked through each of the activities briefly to highlight the topics that were reviewed and the correct answers. At this time, the facilitators answered questions regarding how the escape room activities modeled the concepts covered in the class. If a group did not complete the room in time, the facilitators would walk through the missed concepts with the group to guide them through the complete experience.

## Impact and Applications

This escape-room review game was successful in many facets. The students expressed enthusiasm for the game before the review session took place. We were pleased with the level of collaboration that was displayed by most students, even some that were more reserved during other class discussions. In addition to the general excitement in the room, we were able to observe the thought process of the participants as they worked through the genetics puzzles in the room. Not only did this help us evaluate student comprehension, but we also observed several situations in which students were clearly learning from each other's thought processes. Although our qualitative assessments of the activity's efficacy suggests that the review game positively impacted learning, results of a formal pre- and post-survey (**Supplement D, Supplement E**) will be more telling.

The development of this game involved several cycles of design, implementation, evaluation, and editing across the three academic years during which we included this activity. We attribute the successes of our escape room to a stepwise process of game design, and below we offer these steps to assist educators in designing a similar review game for other courses.

First, we established specific learning objectives that framed the basis of the escape room. For this course, we selected core concepts that would appear on the final exam and sought to review each of those topics during the game. Future educational escape room designers might benefit from similarly establishing the core concepts they aim to cover during the game.

Next, we conceptualized a theme that would remain consistent during the escape room. Theme selection was a challenging piece of the design, because we wanted a theme that would tie all puzzles together and offer variation in the type of puzzle. We selected the theme of firefly genetics, because we imagined a variety of puzzle types that could model aspects of a bioluminescence phenotype (*e.g.* turning lights on or off, building circuits, assessing variation in light color). In our escape room game, we intentionally used the theme to create a storyline, because we wanted the students to think of genetics as a process rather than a series of disjointed puzzles. However, the game designer can choose a theme as an overarching concept without creating a storyline for the flow of the game, depending on the goal of the escape room.

After selecting a theme, we chose a layout of game design that would fit our purposes. Because we were using the escape room as an exam review, we wanted all students to be involved in each step of the game. Therefore, we selected a linear path that requires students to solve one puzzle before they gain access to the next puzzle. One alternative layout we considered was a parallel pathway, where students split into groups and each group solves an independent track of puzzles, eventually leading each group to the final challenge that they would complete together. This type of design could be beneficial in other contexts, especially where the escape room is used as a

tool to assess individual student aptitude. Of course, a mixture of the linear and parallel layouts could be designed, depending on the goal of the game.

Finally, with our learning objectives, theme, and layout in hand, we began designing puzzles. This part of the game design took the most amount of time, effort, and editing. We first brainstormed types of light-related phenomena that could be used during the game and identified which learning objectives could be modeled by each of those phenomena. For example, using a blacklight to expose invisible ink could represent fluorescent protein expression, which is invisible unless excited with a light of the proper wavelength. We then researched types of padlocks and brainstormed ways in which solutions to our genetics puzzles could be used to unlock boxes. The cycle of implementation, evaluation, and editing is an ongoing process that continues to improve the design of our escape room.

In conclusion, we designed an escape room game for the BIOL\_SCI 393: Genetic Analysis final exam review. We found that students were enthusiastic about the activity and we were able to gauge student comprehension of particular topics during the activity. Additionally, we believe that students were able to learn from experiencing the problem-solving process with their peers, but the formal assessment of impact will be more telling. The steps we used to design our escape room can be applied to other topics outside of genetics, and even outside of STEM. Because learning-centered practices encourage deep understanding of concepts, escape room review activities offer a unique method by which studying for exams can become not only more enjoyable but more effective.

## Contributions

We would like to thank members of the Searle Learning Center at Northwestern University for their guidance during the development of this manuscript. In particular, we thank Susanna Calkins for her support in the development of the pre- and post-assessment survey. We are grateful to Emily Mathews from Science in Society at Northwestern University for her advice on the structure of the manuscript. We thank the students of BIOL\_SCI 393, Genetic Analysis for participating in the escape room and providing their feedback on the experience. Lastly, we thank Emma Coughlin and Jeff Coughlin for their crucial support for puzzle design, specifically the epistasis circuit, within the escape room.

## Supplement A – Setup instructions

### *Before the event day:*

Crack some glow sticks to ensure they do not glow on the day of the event when activated. Crack all glow sticks in envelopes “0 hours” and “2 hours”, crack 5/8 glow sticks in envelope “4 hours”, 2/8 in envelope “6 hours”, and none from envelopes “8 hours” and “10 hours”.

### *Room set up:*

1. Draw the expression plasmid (**Figure 4-4A**) on a magnetic chalkboard or whiteboard.
2. Draw the abdominal cell lineage (**Figure 4-7**) on a chalkboard or whiteboard.
3. Place the mysterious poster (**Figure 4-10**) somewhere in the room.
4. Write the mutagenesis tools from **Figure 4-1** on a chalkboard or whiteboard:

WT fireflies

*antennaless/longleg* fireflies

*longleg* is a chromosome 5 balancer

*longleg* has a dominant long-legged phenotype

*antennaless* has a dominant lack-of-antennae phenotype

*longleg* and *antennaless* have recessive lethality phenotype

EMS mutagen

5. Print all prompts (**Supplement B**)
6. Place items in the corresponding boxes (below)
7. Place reserved items behind the desk with the facilitator (below)

**Box 1**

Prompt 2

Complementation tea lights (**Figure 4-2A**)

Locked with key-hole lock

(can be unlocked using Key 1)



**Box 2**

Prompt 3

Epistasis circuit (**Figure 4-3**)

Locked with lock labeled "mMm"

(can be unlocked with code 1-9-4)



### Box 3

Prompt 4

Plasmid magnets (**Figure 4-4A**)

Blacklight

3 iron nails

Locked with word lock

(can be unlocked with code PLAY5)



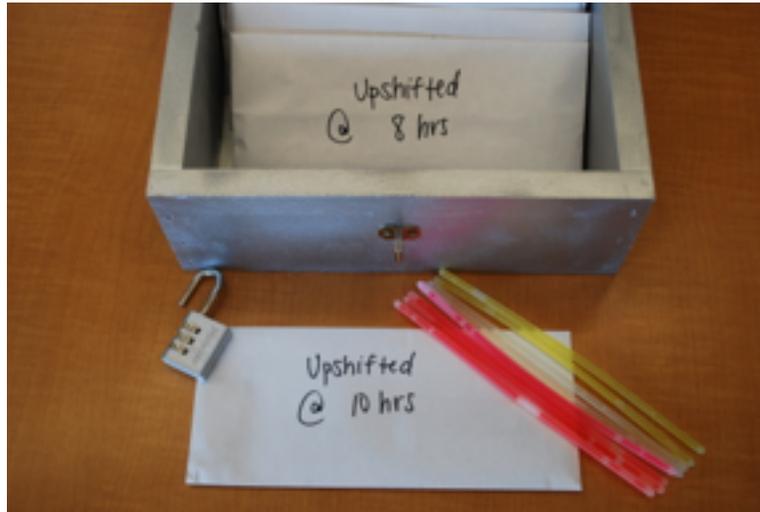
### Box 4

Prompt 8

Envelopes with glow sticks (**Figure 4-8**)

Locked with 3-digit lock

(can be unlocked with code 1-3-5)



**Box 5**  
 Prompt 10  
 Triplet pedigrees (**Figure 4-11**)  
 Locked with combination lock  
 (can be unlocked with code 9-23-5)



**Box 6**  
 Prompt 11  
 Flight pattern pipe cleaners (**Figure 4-13**)  
 Locked with 4-digit lock  
 (can be unlocked with code 1212)



**Box 7**  
Replica of Nobel Prize  
Locked with ACTG lock  
(can be unlocked with code ATGTGGTAA)



***Reserved items (held by facilitator)***

Prompt 1 (given at beginning)

Key 1 (given after correct mutagenesis screen, **Figure 4-1**, is drawn on the board)

Gene expression poster (**Figure 4-5**, given after plasmid magnet is correct)

Prompt 5 (given after students shine the blacklight on the firefly poster)

Mutant firefly box (**Figure 4-6**, given when the students state they want a mutant animal after Prompt 5)

Prompt 6 (given with mutant firefly box)

Prompt 7 (given after students state “cell non-autonomous”)

Prompt 9 (given after students state the correct time of gene action)

Modes of inheritance pedigrees (**Figure 4-9**, given with Prompt 9)

Linkage pedigree (**Figure 4-12**, given when students select the fourth triplet)

## Supplement B – Prompts

### Introductory Prompt

“Congratulations! You have won the Nobel Prize for your work in *Photinus pyralis*, which, as you obviously know, is the scientific name for the Big Dipper Firefly. You performed flawless genetic experiments, identified several genes involved in the bioluminescence of this magical organism, and found genetic tests for diseases in the species, eventually earning you the highest honor in the field. However, your jealous lab mate doubts your knowledge of genetics and has hidden your Nobel Prize in this room. To prove that you have earned this award and claim what is rightfully yours, you must complete a series of tasks, unlock several boxes, and find your Nobel prize before you give your big acceptance speech in one hour! Good luck!”

### Prompt 1

One of your big discoveries was a mutant for a gene on chromosome 5 that causes a recessive inability to bioluminesce. Given the following tools, draw out the cross you performed to identify

this mutant. When you are confident with your answer written on the board, request a check by the facilitator.

Wild-type firefly stock

*antennaless/longleg* stock

*Longleg* is a chromosome 5 balancer

*Longleg* has a dominant long-legged phenotype

*Antennaless* has a dominant lack-of-antennae phenotype

*Longleg* and *antennaless* have recessive lethality phenotypes

EMS

### Prompt 2

From your mutagenesis screen, you identified nine mutants with the recessive inability to bioluminesce phenotype. You called these mutants “*m1*” through “*m9*”. To perform a complementation test, you crossed each of these mutants to each other. The results of those crosses are in this box, represented by tea lights. Use these to identify the largest non-complementation group.

### Prompt 3

When you performed your screen for lack of bioluminescence, you also identified three mutants that are able to shine but show the incorrect color. These genes are called *trix1*, *play5*, and *oldy3*. When all three of these genes are functional, the wild-type color is green. This is represented by switching all three buttons on this epistasis model to “WT”. When the function of *oldy3* is missing

but *play5* and *trix1* are functional (*oldy3* is switched to “mut” and *play5* and *trix1* are switched to “WT”), the animal shines yellow. You suspected that defects in a biosynthetic pathway lead to the build up of colored precursors in these mutants. Identify the most upstream gene in this biosynthetic pathway.

#### Prompt 4

Going back to complementation group [1,9,4], you decided to look further into this gene, that you named *dim1*. Select a combination of the provided promoters and alleles of *dim1* and insert them into the GFP plasmid on the board so you can see where in the animal *dim1* is normally expressed. When you're ready to check for expression, use this 480nm wavelength lightbulb to excite the GFP molecules and find cells that glow. Write the identity of these GFP-positive cells on the board.

#### Prompt 5

The four D cells are the only cells that bioluminesce in the wild-type animal. You suspect *dim1* function is required in some or all of the D cells (probably not required in D1 since *dim1* is not expressed in that cell). Given a promoter that expresses in all D cells, describe to the puzzle master the experiment you would perform to test your hypothesis.

Express \_\_\_(wild-type/mutant)\_\_\_ *dim1* in a \_\_\_(wild-type/mutant)\_\_\_ animal under a \_\_\_(nascent/D-cell specific)\_\_\_ promoter.

**Prompt 6**

This box represents a mutant animal. Notice it does not bioluminesce, because it is lacking a gene product in all 16 of the abdomen cells. Use the nails to represent a wild-type copy of *dim1* to rescue *dim1* function in the D cells (D2, D3, D4). State out loud what this means about the autonomy of *dim1* function.

**Prompt 7**

Because expressing *dim1* in the D cells did not rescue the bioluminescence function, you try other rescue experiments. You have a promoter that drives expression in all A and B cells. You find that using this promoter to drive expression of the wild-type copy of *dim1* causes a rescue of bioluminescence!

You also find that ablating the precursor for all A cells (Ap) in a wild-type embryo results in an animal that is still able to glow. However, ablating the (Bp) cell results in an animal that is unable to glow.

Prove that you have identified the only three cells in which *dim1* function is required for proper bioluminescence using the mutant animal model and the wild-type expression nails.

**Prompt 8**

Now that you know where *dim1* function is required, you want to know when during development it acts. You have a heat-sensitive *dim1* mutant and perform an upshift experiment. Each of these provided populations was upshifted to the restrictive temperature at the indicated time. Use these

to determine the developmental period during which *dim 1* function is required. State your answer out loud when you're ready.

**Prompt 9**

Although your work on bioluminescence was ground-breaking, you are most proud of your work on identifying genetic diseases in firefly populations. You kept track of families of fireflies and drew these pedigrees for five different families. Given each of these pedigrees, determine the mode of inheritance for each firefly disease.

**Prompt 10**

You decided to focus on the third pedigree, with the dominant autosomal disorder known as "diurnal". Unfortunately, this rare disease causes individuals to flash during the daylight instead of the night, ultimately resulting in lower mate attraction abilities. You have isolated some families of fireflies that are affected by this disease and a nearby marker that may be linked to the diurnal disease-causing locus.

Select one family from which you will take the F1 individual and outcross it to determine linkage of the diurnal disease with the marker locus. State your selection out loud when you're ready. Hint: You want your linkage mapping to be as informative as possible.

**Prompt 11**

After you successfully mapped the gene causing the diurnal disease, you focused your attention on a more common disease. This "curlyQ" disorder causes affected individuals to fly in circular patterns instead of straight. Enclosed are the observed flight patterns of a population of

individuals. You identified three SNPs (1, 2, and 3) that may be correlated with the curlyQ disease. Determine the full sequence of the genotype most likely to be found in curlyQ individuals.

SNP1: ...TGCATT(A/T)GG...

SNP2: ...CCGAAG(T/C)AA...

SNP3: ...ATGTG(G/C)TAA...

## Supplement C – Workflow

Key:

Prompt (should be printed and given to students)

Answer (should be deduced by the students)

Facilitator (should be done by the facilitator)

### Introductory Prompt

“Congratulations! You have won the Nobel Prize for your work in *Photinus pyralis*, which, as you obviously know, is the scientific name for the Big Dipper Firefly. You performed flawless genetic experiments, identified several genes involved in the bioluminescence of this magical organism, and found genetic tests for diseases in the species, eventually earning you the highest honor in the field. However, your jealous lab mate doubts your knowledge of genetics and has hidden your Nobel Prize in this room. To prove that you have earned this award and claim what is rightfully yours, you must complete a series of tasks, unlock several boxes, and find your Nobel prize before you give your big acceptance speech in one hour! Good luck!”

Give Prompt 1

Start the timer

### Prompt 1

One of your big discoveries was a mutant for a gene on chromosome 5 that causes a recessive inability to bioluminesce. Given the following tools, draw out the cross you performed to identify this mutant. **When you are confident with your answer written on the board, request a check by the facilitator.**

Wild-type firefly stock

*antennaless/longleg* stock

*Longleg* is a chromosome 5 balancer

*Longleg* has a dominant long-legged phenotype

*Antennaless* has a dominant lack-of-antennae phenotype

*Longleg* and *antennaless* have recessive lethality phenotypes

EMS

Figure 4-1

When correct, give Key 1

Use Key 1 to unlock Box 1

### Prompt 2 (in Box 1)

From your mutagenesis screen, you identified nine mutants with the recessive inability to bioluminesce phenotype. You called these mutants “*m1*” through “*m9*”. To perform a

complementation test, you crossed each of these mutants to each other. The results of those crosses are in this box, represented by tea lights. **Use these to identify the largest non-complementation group.**

Figure 4-2

Unlock Box 2 with code 194

### Prompt 3 (in Box 2)

When you performed your screen for lack of bioluminescence, you also identified three mutants that are able to shine but show the incorrect color. These genes are called *trix1*, *play5*, and *oldy3*. When all three of these genes are functional, the wild-type color is green. This is represented by switching all three buttons on this epistasis model to “WT”. When the function of *oldy3* is missing but *play5* and *trix1* are functional (*oldy3* is switched to “mut” and *play5* and *trix1* are switched to “WT”), the animal shines yellow. You suspected that defects in a biosynthetic pathway lead to the build up of colored precursors in these mutants. **Identify the most upstream gene in this biosynthetic pathway.**

Figure 4-3

Unlock Box 3 with code PLAY5

### Prompt 4 (in Box 3)

Going back to complementation group [1,9,4], you decided to look further into this gene, that you named *dim1*. Select a combination of the provided promoters and alleles of *dim1* and insert them into the GFP plasmid on the board so you can see where in the animal *dim1* is normally

expressed. When you're ready to check for expression, use this 480nm wavelength lightbulb to excite the GFP molecules and find cells that glow. **Write the identity of these GFP-positive cells on the board.**

#### Figure 4-4B

When they use the nascent promoter and the wild-type allele of *dim1*, give them the gene expression poster (Figure 4-5)

Expression in A1, A2, A3, B1, B2, B4, C1, C2, D2, D3, D4

When they write these cells on the board, give them Prompt 5

#### Prompt 5

The four D cells are the only cells that bioluminesce in the wild-type animal. You suspect *dim1* function is required in some or all of the D cells (probably not required in D1 since *dim1* is not expressed in that cell). Given a promoter that expresses in all D cells, describe to the puzzle master the experiment you would perform to test your hypothesis.

Express     (wild-type/mutant)     *dim1* in a     (wild-type/mutant)     animal under a     (nascent/D-cell specific)     promoter.

Express wild-type *dim1* in a mutant animal under a D-cell specific promoter

When they answer correctly, give them the mutant firefly box (Figure 4-6) and Prompt 6.

#### Prompt 6

This box represents a mutant animal. Notice it does not bioluminesce, because it is lacking a gene

product in all 16 of the abdomen cells. Use the nails to represent a wild-type copy of *dim1* to rescue *dim1* function in the D cells (D2, D3, D4). **State out loud what this means about the autonomy of *dim1* function.**

Turn off the room lights when they check if the box is glowing (it is not).

“Cell non-autonomous”

When they say “cell non-autonomous”, give them Prompt 7.

### **Prompt 7**

Because expressing *dim1* in the D cells did not rescue the bioluminescence function, you try other rescue experiments. You have a promoter that drives expression in all A and B cells. You find that using this promoter to drive expression of the wild-type copy of *dim1* causes a rescue of bioluminescence!

You also find that ablating the precursor for all A cells (Ap) in a wild-type embryo results in an animal that is still able to glow. However, ablating the (Bp) cell results in an animal that is unable to glow.

**Prove that you have identified the only three cells in which *dim1* function is required for proper bioluminescence using the mutant animal model and the wild-type expression nails.**

Turn off the room lights when they check if the box is glowing.

Given the cell lineage (Figure 4-7) and expression poster, deduce where function of *dim1* is

required for bioluminescence. Put the nails in B1, B2, and B4 to rescue the function of *dim1* in the necessary cells. Read the numbers from the mutant animal box.

Unlock Box 4 with code 135.

### Prompt 8 (in Box 4)

Now that you know where *dim1* function is required, you want to know when during development it acts. You have a heat-sensitive *dim1* mutant and perform an upshift experiment. Each of these provided populations was upshifted to the restrictive temperature at the indicated time. Use these to determine the developmental period during which *dim1* function is required. **State your answer out loud when you're ready.**

### Figure 8

Function is required somewhere between 2 and 8 hours of development.

When they state the correct time of gene function, give them the modes of inheritance pedigrees (Figure 4-9) and Prompt 9.

### Prompt 9

Although your work on bioluminescence was ground-breaking, you are most proud of your work on identifying genetic diseases in firefly populations. You kept track of families of fireflies and drew these pedigrees for five different families. **Given each of these pedigrees, determine the mode of inheritance for each firefly disease.**

Match the pedigrees to the correct mode of inheritance on the mysterious poster board.

Figure 4-10.

Unlock Box 5 with code 9-23-5.

**Prompt 10 (in Box 5)**

You decided to focus on the third pedigree, with the dominant autosomal disorder known as “diurnal”. Unfortunately, this rare disease causes individuals to flash during the daylight instead of the night, ultimately resulting in lower mate attraction abilities. You have isolated some families of fireflies that are affected by this disease and a nearby marker that may be linked to the diurnal disease-causing locus.

Select one family from which you will take the F1 individual and outcross it to determine linkage of the diurnal disease with the marker locus. **State your selection out loud when you’re ready.**

Hint: You want your linkage mapping to be as informative as possible.

Figure 4-11.

“The fourth triplet”

When they select the correct triplet, give them the linkage pedigree (Figure 4-12).

Theta = 0.019 with 10 progeny. There are nine parental progeny and one recombinant, and we know the phase from the triplet. LOD = 1.212

If they do not have a calculator, give them the answer to the LOD score (1.212).

Use the code 1212 to unlock Box 6.

**Prompt 11 (in Box 6)**

After you successfully mapped the gene causing the diurnal disease, you focused your attention on a more common disease. This “curlyQ” disorder causes affected individuals to fly in circular

patterns instead of straight. Enclosed are the observed flight patterns of a population of individuals. You identified three SNPs (1, 2, and 3) that may be correlated with the curlyQ disease.

**Determine the full sequence of the genotype most likely to be found in curlyQ individuals.**

SNP1: ...TGCATT(A/T)GG...

SNP2: ...CCGAAG(T/C)AA...

SNP3: ...ATGTG(G/C)TAA...

Figure 13.

Use the code ATGTGGTAA to unlock Box 7

When they unlock Box 7, stop the timer.

## Supplement D – Pre-activity survey

Please mark your answers with an “X”.

1 = Strongly disagree    2 = Disagree    3 = Neutral    4 = Agree    5 = Strongly agree

	1	2	3	4	5
I can draw a mutagenesis screen.					
I can draw a suppressor screen.					
I can perform a complementation test.					
I can determine epistasis of a biochemical pathway.					
I can identify where a gene is naturally expressed.					
I can assess cell autonomy of a gene product.					
I can use rescue results to see where a gene’s function is sufficient.					
I can use ablation results to see where a gene’s function is necessary.					
I can assess the developmental time of a gene’s action.					
I can use pedigrees to distinguish among modes of inheritance.					
I can determine phase of individuals in a pedigree.					
I can perform a GWAS given phenotype and genotype data.					
I feel comfortable communicating about genetics with my peers.					
I can follow the steps of genetic analysis from beginning to end.					
I can see how the BIO 393 topics fit into real-world genetic analysis.					
I feel prepared for the BIO 393 final exam.					

1. When wild-type mice are suspended by their tails, they spread their hind legs. You identified four mutants that instead clasp their hind legs together upon tail suspension. You named the four mutants *clasp1* through *clasp4*. You then performed crosses for each pairwise combination of these four genes and recorded the phenotypes of the progeny below. **Identify which clasp mutations complement each other.**

	<i>clasp1</i>	<i>clasp2</i>	<i>clasp3</i>	<i>clasp4</i>
<i>clasp1</i>	clasping	spreading	spreading	clasping
<i>clasp2</i>		clasping	clasping	spreading
<i>clasp3</i>			clasping	spreading
<i>clasp4</i>				clasping

2. Your friend owns a winery and is trying to understand the metabolic process of fermentation. She has collected mutant yeast that are unable to create ethanol, and she has identified certain precursors of the pathway that tend to build up in single and double mutants. Her observations are below. **Please help her draw an epistasis model for the four mutants in the metabolic pathway.**

Wild-type: creates ethanol

P and S double mutant: build-up of sucrose

Mutant P: build-up of pyruvate

G and P double mutant: build-up of glucose

Mutant S: build-up of sucrose

G and S double mutant: build-up of glucose

Mutant G: build-up of glucose

3. Your uncle breeds beagles and notices that some dogs have much wetter noses than others. He has discovered the following: (i) the nose cells of wet-nosed beagles secrete

more mucus than dry-nosed beagles, (ii) the gene sniffer is required for wet noses, because sniffer mutants have dry noses, and (iii) sniffer is expressed in the nose, toenails, and tips of tails. He wanted to know where the gene product for sniffer is required, so he designed a plasmid that rescues sniffer under a nose-specific promoter. He placed the rescue construct into sniffer mutants and all the pups had dry noses. **What does this mean about the cell autonomy of sniffer in the wet nose phenotype?**

## Supplement E – Post-activity survey

Please mark your answers with an “X”.

1 = Strongly disagree    2 = Disagree    3 = Neutral    4 = Agree    5 = Strongly agree

	1	2	3	4	5
I can draw a mutagenesis screen.					
I can draw a suppressor screen.					
I can perform a complementation test.					
I can determine epistasis of a biochemical pathway.					
I can identify where a gene is naturally expressed.					
I can assess cell autonomy of a gene product.					
I can use rescue results to see where a gene’s function is sufficient.					
I can use ablation results to see where a gene’s function is necessary.					
I can assess the developmental time of a gene’s action.					
I can use pedigrees to distinguish among modes of inheritance.					
I can determine phase of individuals in a pedigree.					
I can perform a GWAS given phenotype and genotype data.					
I feel comfortable communicating about genetics with my peers.					
I can follow the steps of genetic analysis from beginning to end.					
I can see how the BIO 393 topics fit into real-world genetic analysis.					
I feel prepared for the BIO 393 final exam.					

1. When wild-type mice are suspended by their tails, they spread their hind legs. You identified four mutants that instead clasp their hind legs together upon tail suspension. You named the four mutants *clasp1* through *clasp4*. You then performed crosses for each pairwise combination of these four genes and recorded the phenotypes of the progeny below. **Identify which clasp mutations complement each other.**

	<i>clasp1</i>	<i>clasp2</i>	<i>clasp3</i>	<i>clasp4</i>
<i>clasp1</i>	clasping	spreading	spreading	clasping
<i>clasp2</i>		clasping	clasping	spreading
<i>clasp3</i>			clasping	spreading
<i>clasp4</i>				clasping

2. Your friend owns a winery and is trying to understand the metabolic process of fermentation. She has collected mutant yeast that are unable to create ethanol, and she has identified certain precursors of the pathway that tend to build up in single and double mutants. Her observations are below. **Please help her draw an epistasis model for the four mutants in the metabolic pathway.**

Wild-type: creates ethanol

P and S double mutant: build-up of sucrose

Mutant P: build-up of pyruvate

G and P double mutant: build-up of glucose

Mutant S: build-up of sucrose

G and S double mutant: build-up of glucose

Mutant G: build-up of glucose

3. Your uncle breeds beagles and notices that some dogs have much wetter noses than others. He has discovered the following: (i) the nose cells of wet-nosed beagles secrete

more mucus than dry-nosed beagles, (ii) the gene sniffer is required for wet noses, because sniffer mutants have dry noses, and (iii) sniffer is expressed in the nose, toenails, and tips of tails. He wanted to know where the gene product for sniffer is required, so he designed a plasmid that rescues sniffer under a nose-specific promoter. He placed the rescue construct into sniffer mutants and all the pups had dry noses. **What does this mean about the cell autonomy of sniffer in the wet nose phenotype?**

Please mark your answers with an "X".

1 = Strongly disagree    2 = Disagree    3 = Neutral    4 = Agree    5 = Strongly agree

	1	2	3	4	5
I enjoyed the escape-room activity.					
The escape room increased my interest in genetic analysis.					
My contributions were valued by my peers during the escape room.					
I valued the contributions that my peers made during the escape room.					
The escape-room activity helped me understand the process of genetic analysis.					
Listening to my peers articulate their problem-solving process during the escape room improved my ability to solve problems.					
The escape room helped me see how the BIO 393 topics fit into real-world genetic analysis.					
The escape room helped me identify topics that I should review more before the final exam.					
The escape room helped me feel prepared for the BIO 393 final exam.					
The escape room helped me retain information I learned in BIO 393.					

What skills, if any, did you learn in the escape room?

Any other comments you would like us to know about the escape room?

## Chapter 5 - Discussion

Chapters two and three present two projects that took up a large portion of my time as a Ph.D. candidate. In this section, I discuss how the results of these two projects might impact *C. elegans* research and, more broadly, the field of quantitative genetics. Additionally, I suggest future experiments for each of these projects that might answer some of the lingering questions.

### Power of linkage mapping in *C. elegans*

Both the bleomycin project and the QTL hotspot project used the *C. elegans* model system to identify genetic variants that underlie complex traits. In both cases, we performed linkage mapping to localize regions of the genome that underlie differences in toxin responses. The extensive panel of RIALs that were previously constructed offered a powerful amount of genetic variants for linkage mappings [109,110]. Additionally, a high-throughput assay for measuring trait responses facilitated efficient and precise measurements of quantitative traits across these RIALs [109]. Linkage mapping of those data detected both large and small-effect QTL that underlie different toxin responses, explaining as much as 51.6% (bleomycin mean.TOF) and as little as 2.3% (tunicamycin PC1) of variation in the RIAL phenotypes. The range of QTL effect sizes identified by these mappings highlights the power of our system to detect genetic components that underlie drug responses.

Why is the ability to detect small-effect QTL important? Studies of human height variation have concluded that this complex trait is influenced by many small-effect genetic loci [290,291]. Immense sample sizes are required to identify loci that underlie human height differences, and

only 10% of total height variation is explained by combining the effects of these detected variants [290]. Many complex traits in humans might also be explained by small-effect loci, but these variants remain undetected because of power limitations of human GWA and linkage studies [96,292,293]. Complex traits in other animals, such as milk production in dairy cattle [294], growth rate in chickens [295], and many phenotypes in mice [296] and dogs [297], are also heavily influenced by small-effect QTL, but many of these loci remain elusive, likely because of insufficient statistical power to detect them in these species. Although studies in yeast and *Arabidopsis* have achieved sufficient statistical power to detect most of the small-effect loci that influence quantitative traits [192,229,239,298], much of the narrow-sense heritability of complex traits remains unaccounted for in metazoans. We can generate large panels of recombinant strains in *C. elegans* more easily than in other animal models because of its fast generation time and simplicity of genetic crosses. Additionally, high-throughput assays developed in this animal model offer methods by which complex phenotypes can be measured more quickly and accurately than in other metazoans. Therefore, *C. elegans* is a uniquely powerful animal model with which we can correlate small-effect genetic loci to quantitative phenotypes. Our capacity to detect small-effect variants that underlie toxin responses emphasizes the impact of leveraging *C. elegans* to understand complex traits in animals.

Some missing heritability can be explained small-effect QTL, but genetic interactions can also impact quantitative phenotypes [230,231,293,299–302]. These interactions often go undetected because of insufficient statistical power [192,303,304]. In chapters two and three, we identified evidence of epistasis that affects toxin responses, and in chapter two we found particular genomic regions that might contain these interacting loci. The ability of our system to identify evidence of

epistasis and to localize candidate epistatic regions again highlights the immense power of our model.

Further exemplifying the power of linkage mapping in *C. elegans* is our ability to go from quantitative trait locus (QTL) to quantitative trait gene (QTG). We showed that *scb-1* underlies bleomycin-response differences between the N2 and CB4856 strains. In many linkage mapping studies, the gene that underlies a QTL is not detected [228]. In fact, only 1% of QTL discovered in a mouse model have been resolved to a single gene [305]. Dozens of QTL underlying the well characterized quantitative trait of wing shape in *Drosophila* have been identified, but the particular mutants causing these associations have not been detected [306–308]. Our ability to identify *scb-1* as the causal gene was facilitated by the narrow confidence interval surrounding the bleomycin-response QTL. The precision of this linkage mapping is partially attributed to the large amount of recombination in the RIAIL panel as well as the high-quality genomic data available for these strains. The expanded set of 13,003 markers that was used to map bleomycin responses should be used for future linkage mappings.

## Looking toward future linkage mapping studies

Because SNPs that cause amino-acid changes are easily identified by sequencing, these protein-coding variants are often tested for association with quantitative traits. However, many QTL are not underlied by variants in coding sequences; instead, mutations that lie outside of coding regions underlie many QTL [309–311]. We identified *scb-1* as the causal gene underlying bleomycin-response differences, despite the lack of *scb-1* coding variants. A QTL that underlies variation in *scb-1* expression mapped to the same location as the bleomycin-response QTL,

suggesting that *scb-1* expression variation might cause bleomycin-response differences across the RIALs. However, we only have expression data for one panel of RIALs, and that panel shows low heritability for drug-response phenotypes, potentially because of improper strain maintenance. Fortunately, the bleomycin-response QTL mapped to the same position in both panels of RIALs (**Figure 2-3**, **Figure 2-15**), so expression QTL were able to be tested for correlation with bleomycin-response QTL. However, the RIALs for which expression data are available do not reliably map for other drug responses. Therefore, we must collect expression data for the RIALs that reliably map with linkage mapping as well as the wild isolates that are used for GWA mappings. Without these data, expression variants that underlie detected QTL will remain unidentifiable.

Although we now know that *scb-1* underlies variation in bleomycin responses, we spent a significant amount of time pursuing candidate genes with protein-coding variants. What lessons can we learn from this project that will improve future efficiency of identifying the variant that underlies a drug response? Now that we know non-coding variation in *scb-1* underlies natural variation in bleomycin responses, we can retrospectively look for hints that might have suggested the effect of an expression variant rather than a protein-coding variant.

For the expression data that is currently at our fingertips, we can now consider strategies to correlate eQTL with drug-response QTL. We can use *scb-1* as a model to help determine signatures of eQTL that might underlie drug-response QTL. For example, eQTL for both *scb-1* and *cnc-10* mapped to the region in the center of chromosome V. We empirically tested the effect of deleting *cnc-10* and observed no difference in bleomycin responses. Perhaps computational methods could have determined that *scb-1* was more likely than *cnc-10* to underlie the bleomycin-

response QTL. Mediation analysis, for example, identifies evidence of a causal chain by which an input affects a mediator which thereby affects the output (rather than the input directly affecting the output). Computational models that identify signatures of gene-expression variation mediating phenotypic differences could be used to predict which QTL are underlied by expression variants.

Even without performing complex analyses, we might have been able to identify *scb-1* expression variation as the causal factor underlying bleomycin-response differences. The linkage mapping of bleomycin responses showed that 43.58% of the phenotypic variation in the RIALs could be explained by the QTL on chromosome V (**Figure 2-3**). Variation in *scb-1* expression mapped to the center of chromosome V, and that eQTL explained 45.7% of the variation in gene expression (**Figure 2-14**). The similarity in effect sizes of these QTL suggested that expression differences of *scb-1* might modulate bleomycin phenotypes. Furthermore, *scb-1* expression was correlated to bleomycin responses with an  $r^2$  value of 0.65, which suggests that 65% of the variation in bleomycin phenotypes can be explained by *scb-1* expression. Interestingly, this 65% effect size is larger than the effect size of the QTL underlying bleomycin responses or the eQTL underlying *scb-1* expression differences, suggesting two conclusions. First, the difference between the 45% effect size of the eQTL on *scb-1* expression and the 65% effect size of *scb-1* expression on bleomycin responses suggests that undetected trans-eQTL must account for the *scb-1* expression variation across the RIALs. Second, the strong correlation between *scb-1* expression and bleomycin responses suggest that *scb-1* is a major contributor to bleomycin responses. Perhaps this evidence could have indicated that *scb-1* should have been our primary candidate gene. For future mappings, comparing the relative drug-response QTL effect sizes, gene-expression QTL effect sizes, and correlation between drug responses and gene expression might identify cases that are indicative of a causal expression difference rather than protein-coding

change. Gathering expression data for each mapping panel is critical to the success of this strategy.

The bleomycin-response variation discussed in Chapter 2 was affected by epistatic loci. In particular, the transgressive phenotypes apparent in the near-isogenic lines indicated that genetic interactions affect bleomycin responses. Could this presence of genetic interactions have hinted that gene-expression variants, rather than protein-coding variants, underlie bleomycin-response differences? *Cis*-regulatory variants that affect gene expression might cause less pleiotropic, and thereby less deleterious, effects than mutations that change amino-acid residues, and mutations in other regions of the genome could arise to compensate for small gene-expression changes [319]. Therefore, evidence of epistasis might indicate an increased probability of expression variants that underlie a quantitative trait. We need to find more cases of expression variants that underlie quantitative traits before we can determine whether epistasis is a strong predictor of expression rather than protein-coding causal variants.

Aside from the *scb-1* project, two studies have paired linkage mapping and GWA in *C. elegans* to identify a gene that underlies toxin-response variation [40,134]. In both of these cases, the linkage-mapping QTL and GWA QTL overlapped, which indicated that a common variant between the RIAILs and the wild isolates underlied both QTL. On the other hand, the bleomycin-response QTL from linkage mapping did not overlap with QTL identified by GWA. This lack of overlap suggested that a rare variant present in CB4856 underlies bleomycin-response variation in the RIAILs but is not present at high enough allele frequency to be detected in the wild strain set. Nonetheless, the distinct sets of QTL from each mapping strategy eliminated the possibility that variants common between both strain sets could underlie bleomycin responses. Therefore,

whether QTL overlap or not, the combination of linkage mapping and GWA allows for particular variants or regions of the genome to be excluded as causal loci.

## Evolutionary implications of toxin-response QTL

Aside from the potential implications of toxin-response QTL for understanding how patients might vary in treatment outcomes, these QTL can also lead to a deeper understanding of evolutionary processes. By studying toxin-response variation between the laboratory-adapted N2 strain and the Hawaiian wild isolate CB4856, we can identify particular genetic regions that might be differentially selected in the lab versus in nature. For example, resistance to pesticides, such as chlorothalonil, has conceivable fitness benefits for wild *C. elegans* strains, and the QTL that underlie resistance to chlorothalonil might be under selection. Perhaps this selective pressure causes wild isolates to be more resistant to chlorothalonil than the laboratory-adapted N2 strain (**Figure S3-1, Figure S3-5**).

The evolutionary implications of resistance to other classes of toxins might not be as obvious. For example, we found that variation in *scb-1* causes differences in bleomycin sensitivity across strains of *C. elegans*. Although we do not know the molecular mechanism of this gene, *scb-1* is conserved across nematodes. Therefore, bleomycin resistance might be selected in nature, but why would free-living nematodes benefit from resistance to a chemotherapeutic drug? Interestingly, bleomycin is produced by the soil bacteria *S. verticillus* [215,216]. Because *C. elegans* is often found on rotting substrates in association with soil, the idea that *S. verticillus* and this nematode might come into contact with each other in nature is highly likely. Perhaps ancestors of the CB4856 wild isolate adapted to bleomycin levels in the soil around the pineapple

field from which the strain was collected. Other nematodes might also encounter bleomycin, and the *scb-1* gene that offers resistance to bleomycin might be under selective pressure.

To confirm this hypothesis of the evolutionary context of bleomycin resistance, we must first prove that *C. elegans* and other nematodes with the *scb-1* gene encounter the toxin in their natural environment. Each new sample that contains nematodes would need to be tested for the presence of bleomycin. Next, the association between wild isolate expression of *scb-1* and substrate bleomycin presence would need to be tested. However, the GWA of bleomycin responses did not detect a QTL at *scb-1*, which might indicate that CB4856 has a rare allele of *scb-1* that explains its bleomycin resistance. In this case, we would not expect to see an association between *scb-1* expression and substrate bleomycin presence, because *scb-1* expression might not underlie bleomycin resistance in all strains.

On the other hand, the N2 allele of *scb-1* could be the minor allele. If this hypothesis is true, the GWA might not have detected the *scb-1* QTL because N2 was one of few strains to have a genotype that differed from CB4856 at the causal SNP. The variant at V:11,130,591 shows this pattern, where all strains except N2 contain the same allele as CB4856 at the variant (**Figure S2-2**). Additionally, the N2 strain is the most sensitive strain to bleomycin. This rare variant in the N2 strain might have caused decreased expression of *scb-1* and therefore decreased resistance to bleomycin, potentially because bleomycin resistance was not selected in the laboratory environment. However, bleomycin responses range widely across the strains with the CB4856 allele at this SNP, so this variant would not explain much of the phenotypic differences in bleomycin responses across the wild isolates.

The lack of overlapping QTL between linkage mapping and GWA suggests that the variant underlying the linkage mapping peak are not at high enough frequency in the wild isolate panel to be detected in GWA. Although a single variant does not reach the allele frequency required for a significant QTL in GWA to be detected, variation in *scb-1* still might underlie bleomycin-response variation in the wild isolate panel. A previous manuscript from the Andersen Lab observed allelic heterogeneity across the *ben-1* gene, and each of these *ben-1* alleles affected animal responses to benzimidazoles [135]. In this case, *ben-1* did not map with GWA because no single *ben-1* variant reached a high enough allele frequency in the wild isolate panel. Burden testing grouped all variants within a gene together to increase the power of the mapping, and *ben-1* alleles were found to significantly contribute to benzimidazole responses. A similar strategy could be used to group all *scb-1* variants together. However, our hypothesis is that variation in *scb-1* expression causes differences in bleomycin responses, and grouping all variants that impact *scb-1* expression would be more challenging than grouping all variants within the coding region of a given gene. More sophisticated burden tests would be required to confirm whether allelic heterogeneity of *scb-1* expression causes bleomycin-response variation in the wild isolates. Alternatively, reciprocal hemizyosity tests could be used to assess if *scb-1* function varies between wild isolates.

## Potential for future experiments

Although we identified *scb-1* as the gene that underlies the chromosome V bleomycin response QTL, the mechanism by which it impacts bleomycin responses remains unknown. Whereas RIALs show variation in *scb-1* expression that is correlated with bleomycin responses, RNA-seq measurements of *scb-1* expression did not show significant differences between the N2 and

CB4856 parental strains. This result could suggest that the RIALs have a novel variant that arose during strain construction or that epistatic loci cause *scb-1* expression variation in the RIALs. Alternatively, the lack of significant parental difference could be caused by the limited sample number in the RNA-seq experiment. More biological replicates can be measured with qPCR, and this experiment should be performed on the N2 and CB4856 parental strains as well as the ECA411 and ECA528 NILs. Because these NILs recapitulate the linkage-mapping QTL effect, the NILs should show *scb-1* expression differences, if *scb-1* expression indeed underlies the linkage mapping QTL. If parental strains do not show significant expression differences with qPCR, then the RIALs might have a novel variant (and the NILs were created from the RIALs, so they would have the novel variant too) or epistatic loci could cause the expression variation in the RIALs and NILs. These two hypotheses could be differentiated by performing deep sequencing of the ECA411 and ECA528 NILs to identify novel variants that are not present in the N2 and CB4856 parental strains.

To determine if *scb-1* expression is induced upon bleomycin exposure, qPCR of *scb-1* could be performed in bleomycin and control conditions. Ideally, *scb-1* expression upon bleomycin exposure would be compared between the N2 strain and the CB4856 strain. However, bleomycin causes developmental delays, as evidenced by animal size differences in bleomycin and control conditions. Additionally, these developmental delays are more pronounced for the N2 strain than for the CB4856 strain, because N2 is more sensitive to bleomycin than CB4856. After the same amount of exposure time to bleomycin, N2 animals will be at earlier developmental stages than CB4856 animals. Unfortunately, gene expression profiles vary broadly in *C. elegans* depending on developmental stage. Therefore, the comparison between N2 and CB4856 *scb-1* expression

upon bleomycin exposure would be confounded by developmental staging differences between the two test populations.

To minimize the effect of developmental stage on bleomycin-induced expression differences, we could expose animals to lower doses of bleomycin. Lower doses of the drug might elicit a bleomycin-induced *scb-1* expression difference without affecting animal growth and development. Alternatively, animals could be exposed to bleomycin at the adult stage. By exposing animals to bleomycin at the end of their life cycle, all animals assayed for expression differences would be of the same developmental stage, and therefore expression variation driven by developmental differences would not be present. However, *scb-1* function might not be implicated in bleomycin responses in all animal stages, so expression differences that mapped with HTA phenotypes might not be detected with this experiment. Expression differences caused by low doses of bleomycin or by bleomycin exposure at the adult stage might not be representative of expression changes that underlie the bleomycin-response differences that were mapped, so conclusions about the role *scb-1* expression in bleomycin responses should be considered carefully.

In addition to quantifying the levels of *scb-1* expression to confirm its role in bleomycin responses, the mechanism of *scb-1* should be explored further. The effects of *scb-1* overexpression should be tested by injecting plasmids that express *scb-1* under a heat-shock inducible promoter into both the N2 and CB4856 strains. Strains with *scb-1* overexpression should be assayed with the HTA in both control and bleomycin conditions to determine if *scb-1* overexpression causes observable phenotypic effects in the control condition, which might inform its mechanism, or if the bleomycin responses of N2 and CB4856 strains change after *scb-1* overexpression. One might expect an overexpression of *scb-1* to increase bleomycin resistance, but ectopic expression of

the gene could also cause fitness disadvantages. Additionally, exposing animals to heat-shock could affect their responses to bleomycin and confound the results of this experiment.

Bleomycin responses and responses to other toxins are partly underlied by epistatic loci, as evidenced in the NIL results in chapter two and three. We were able to identify three regions that interact to cause bleomycin hypersensitivity. For some of the other toxins, we were able to determine if epistatic loci were acting on the same chromosome or on different chromosomes by comparing NIL and CSS results. Further work to locate these epistatic regions would require the construction of additional NILs and CSSs. In particular, multi-region NILs (for example, two regions from the N2 genotype introgressed into a CB4856 background) and double CSSs (two chromosomes from one parental strain introgressed into the opposite genetic background) would be useful for localizing these interactions. Eventually, CRISPR/Cas9-mediated genome editing could be used to introduce pairwise combinations of candidate variants in multiple genetic backgrounds. Strains with multiple edits should be compared to strains with each individual edit to assess the epistatic versus additive effects of candidate variants.

In summary, this work highlights the power of quantitative genetics in *C. elegans*. Linkage mapping identified small and large-effect QTL that impact toxin responses. NILs and CRISPR/Cas9-mediated genome editing were used to confirm QTL effects, and in one case to locate a gene that impacts bleomycin responses, despite its lack of coding variants. Additionally, this model system was used to locate regions of epistasis, confirming that both additive and interactive effects underlie toxin responses in metazoans. These findings might be informative for patient treatment decisions, and more broadly might indicate how selective pressures shaped the evolution of *C. elegans*.

## Chapter 6 - References

1. World Health Organization [Internet]. 12 Sep 2018. Available: [www.who.int/news-room/fact-sheets/detail/cancer](http://www.who.int/news-room/fact-sheets/detail/cancer)
2. Djulbegovic B, Kumar A, Soares HP, Hozo I, Bepler G, Clarke M, *et al.* Treatment success in cancer: new cancer treatment successes identified in phase 3 randomized controlled trials conducted by the National Cancer Institute-sponsored cooperative oncology groups, 1955 to 2006. *Arch Intern Med.* 2008;168: 632–642.
3. DeVita VT Jr, Chu E. A history of cancer chemotherapy. *Cancer Res.* 2008;68: 8643–8653.
4. Arruebo M, Vilaboa N, Sáez-Gutierrez B, Lambea J, Tres A, Valladares M, *et al.* Assessment of the evolution of cancer treatment therapies. *Cancers* . 2011;3: 3279–3330.
5. National Cancer Institute. NCI Dictionary of Cancer Terms [Internet]. Available: [www.cancer.gov/publications/dictionaries/cancer-terms](http://www.cancer.gov/publications/dictionaries/cancer-terms)
6. Nygren P. What is cancer chemotherapy? *Acta Oncol* . Taylor & Francis; 2001;40: 166–174.
7. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, *et al.* A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2017;16: 19–34.
8. Fu D, Calvo JA, Samson LD. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer.* 2012;12: 104–120.
9. Liu LF. DNA topoisomerase poisons as antitumor drugs. *Annu Rev Biochem.* 1989;58:

- 351–375.
10. Nobili S, Lippi D, Witort E, Donnini M, Bausi L, Mini E, *et al.* Natural compounds for cancer treatment and prevention. *Pharmacol Res.* 2009;59: 365–378.
  11. Ghersi D, Wilcken N, Simes RJ. A systematic review of taxane-containing regimens for metastatic breast cancer. *Br J Cancer.* 2005;93: 293–301.
  12. Bernard-Marty C, Cardoso F, Piccart MJ. Use and abuse of taxanes in the management of metastatic breast cancer. *Eur J Cancer.* 2003;39: 1978–1989.
  13. Dumontet C, Jordan MA. Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat Rev Drug Discov.* 2010;9: 790–803.
  14. Gilman A. Therapeutic applications of chemical warfare agents. *Fed Proc.* 1946;5: 285–292.
  15. Gilman A, Philips FS. The Biological Actions and Therapeutic Applications of the B-Chloroethyl Amines and Sulfides. *Science.* 1946;103: 409–436.
  16. Goodman LS, Wintrobe MM. Nitrogen mustard therapy; use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *J Am Med Assoc.* 1946;132: 126–132.
  17. Chon SY, Champion RW, Geddes ER, Rashid RM. Chemotherapy-induced alopecia. *J Am Acad Dermatol.* 2012;67: e37–47.
  18. Grassi L, Berardi MA, Ruffilli F, Meggiolaro E, Andritsch E, Sirgo A, *et al.* Role of

- psychosocial variables on chemotherapy-induced nausea and vomiting and health-related quality of life among cancer patients: a European study. *Psychother Psychosom.* 2015;84: 339–347.
19. Salihah N, Mazlan N, Lua PL. Chemotherapy-induced nausea and vomiting: exploring patients' subjective experience. *J Multidiscip Healthc.* 2016;9: 145–151.
  20. Kang D-H, Weaver MT, Park N-J, Smith B, McArdle T, Carpenter J. Significant impairment in immune recovery after cancer treatment. *Nurs Res.* 2009;58: 105–114.
  21. Wojcik T, Szczesny E, Chlopicki S. Detrimental effects of chemotherapeutics and other drugs on the endothelium: A call for endothelial toxicity profiling. *Pharmacol Rep.* 2015;67: 811–817.
  22. Soultati A, Mountzios G, Avgerinou C, Papaxoinis G, Pectasides D, Dimopoulos M-A, *et al.* Endothelial vascular toxicity from chemotherapeutic agents: preclinical evidence and clinical implications. *Cancer Treat Rev.* 2012;38: 473–483.
  23. Mardas M, Madry R, Stelmach-Mardas M. Link between diet and chemotherapy related gastrointestinal side effects. *Contemp Oncol.* 2017;21: 162–167.
  24. Duregon F, Vendramin B, Bullo V, Gobbo S, Cugusi L, Di Blasio A, *et al.* Effects of exercise on cancer patients suffering chemotherapy-induced peripheral neuropathy undergoing treatment: A systematic review. *Crit Rev Oncol Hematol.* 2018;121: 90–100.
  25. Kelemen LE, Warren GW, Koziak JM, Köbel M, Steed H. Smoking may modify the association between neoadjuvant chemotherapy and survival from ovarian cancer. *Gynecol Oncol.* 2016;140: 124–130.

26. Li H, He J, Jia W. The influence of gut microbiota on drug metabolism and toxicity. *Expert Opin Drug Metab Toxicol.* 2016;12: 31–40.
27. Nicholson JK, Holmes E, Wilson ID. Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol.* 2005;3: 431–438.
28. Alexander JL, Wilson ID, Teare J, Marchesi JR, Nicholson JK, Kinross JM. Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat Rev Gastroenterol Hepatol.* 2017;14: 356–365.
29. Robb C, Lee A, Jacobsen P, Dobbin KK, Extermann M. Health and personal resources in older patients with cancer undergoing chemotherapy. *J Geriatr Oncol.* 2013;4: 166–173.
30. Bloom JR, Petersen DM, Kang SH. Multi-dimensional quality of life among long-term (5+ years) adult cancer survivors. *Psychooncology.* 2007;16: 691–706.
31. da silva Monica E. Jarrett ALSCMMHGPALPDRM. Social Support Is a Predictor of Lower Stress and Higher Quality of Life and Resilience in Brazilian Patients with Colorectal Cancer. *Cancer Nursing.* 2017;40: 352–360.
32. Lyman GH, Lyman CH, Agboola O. Risk models for predicting chemotherapy-induced neutropenia. *Oncologist.* 2005;10: 427–437.
33. Lyman GH, Abella E, Pettengell R. Risk factors for febrile neutropenia among patients with cancer receiving chemotherapy: A systematic review. *Crit Rev Oncol Hematol.* 2014;90: 190–199.
34. Watters JW, Kraja A, Meucci MA, Province MA, McLeod HL. Genome-wide discovery of

- loci influencing chemotherapy cytotoxicity. *Proc Natl Acad Sci U S A*. 2004;101: 11809–11814.
35. Dolan ME, Newbold KG, Nagasubramanian R, Wu X, Ratain MJ, Cook EH Jr, *et al*. Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity. *Cancer Res*. 2004;64: 4353–4356.
  36. Huang RS, Duan S, Kistner EO, Hartford CM, Dolan ME. Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. *Mol Cancer Ther*. 2008;7: 3038–3046.
  37. Motsinger-Reif A, Brown C, Havener T, Hardison N, Peters E, Beam A, *et al*. Ex-Vivo Modeling for Heritability Assessment and Genetic Mapping in Pharmacogenomics. *Proc State Secr Manage Conf Am Dent Assoc*. 2011;2011: 306–318.
  38. Peters EJ, Motsinger-Reif A, Havener TM, Everitt L, Hardison NE, Watson VG, *et al*. Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. *Pharmacogenomics*. 2011;12: 1407–1415.
  39. Shukla SJ, Duan S, Badner JA, Wu X, Dolan ME. Susceptibility loci involved in cisplatin-induced cytotoxicity and apoptosis. *Pharmacogenet Genomics*. 2008;18: 253–262.
  40. 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: e1006891.
  41. Evans KS, Brady SC, Bloom JS, Tanny RE, Cook DE, Giuliani SE, *et al*. Shared Genomic Regions Underlie Natural Variation in Diverse Toxin Responses. *Genetics*. 2018;210:

- 1509–1525.
42. Kislukhin G, Murphy ML, Jafari M, Long AD. Chemotherapy-induced toxicity is highly heritable in *Drosophila melanogaster*. *Pharmacogenet Genomics*. 2012;22: 285–289.
  43. Gatti DM, Weber SN, Goodwin NC, Lammert F, Churchill GA. Genetic background influences susceptibility to chemotherapy-induced hematotoxicity. *Pharmacogenomics J*. 2018;18: 319–330.
  44. King EG, Kislukhin G, Walters KN, Long AD. Using *Drosophila melanogaster* to identify chemotherapy toxicity genes. *Genetics*. 2014;198: 31–43.
  45. Wolman IJ. Parallel responses to chemotherapy in identical twin infants with concordant leukemia. *J Pediatr*. 1962;60: 91–95.
  46. Zaremba T, Thomas H, Cole M, Plummer ER, Curtin NJ. Doxorubicin-induced suppression of poly(ADP-ribose) polymerase-1 (PARP-1) activity and expression and its implication for PARP inhibitors in clinical trials. *Cancer Chemother Pharmacol*. 2010;66: 807–812.
  47. Imray FP, Smith PJ, Relf W, Kidson C. Wilms' tumour: association with cellular sensitivity to mitomycin C in patients and first-degree relatives. *Lancet*. 1984;1: 1148–1151.
  48. Poot M, Gollahon KA, Rabinovitch PS. Werner syndrome lymphoblastoid cells are sensitive to camptothecin-induced apoptosis in S-phase. *Hum Genet*. 1999;104: 10–14.
  49. Stamatoyannopoulos G, Chen SH, Fukui M. Liver alcohol dehydrogenase in Japanese: high population frequency of atypical form and its possible role in alcohol sensitivity. *Am J Hum Genet*. 1975;27: 789–796.

50. Schroth W, Goetz MP, Hamann U, Fasching PA, Schmidt M, Winter S, *et al.* Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *JAMA*. 2009;302: 1429–1436.
51. Rodriguez-Antona C, Gomez A, Karlgren M, Sim SC, Ingelman-Sundberg M. Molecular genetics and epigenetics of the cytochrome P450 gene family and its relevance for cancer risk and treatment. *Hum Genet*. 2010;127: 1–17.
52. Hoskins JM, Carey LA, McLeod HL. CYP2D6 and tamoxifen: DNA matters in breast cancer. *Nat Rev Cancer*. 2009;9: 576–586.
53. Marsh S, Hoskins JM. Irinotecan pharmacogenomics. *Pharmacogenomics*. 2010;11: 1003–1010.
54. Mega JL, Simon T, Collet J-P, Anderson JL, Antman EM, Bliden K, *et al.* Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. *JAMA*. 2010;304: 1821–1830.
55. Hulot J-S, Bura A, Villard E, Azizi M, Remones V, Goyenvalle C, *et al.* Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood*. 2006;108: 2244–2247.
56. Murphy DL, Lerner A, Rudnick G, Lesch K-P. Serotonin transporter: gene, genetic disorders, and pharmacogenetics. *Mol Interv*. 2004;4: 109–123.
57. Nozawa T, Minami H, Sugiura S, Tsuji A, Tamai I. Role of organic anion transporter OATP1B1 (OATP-C) in hepatic uptake of irinotecan and its active metabolite, 7-ethyl-10-hydroxycamptothecin: in vitro evidence and effect of single nucleotide polymorphisms. *Drug*

- Metab Dispos. 2005;33: 434–439.
58. Kalliokoski A, Niemi M. Impact of OATP transporters on pharmacokinetics. *Br J Pharmacol.* 2009;158: 693–705.
  59. Corless CL, Harrell P, Lacouture M, Bainbridge T, Le C, Gatter K, *et al.* Allele-specific polymerase chain reaction for the imatinib-resistant KIT D816V and D816F mutations in mastocytosis and acute myelogenous leukemia. *J Mol Diagn.* 2006;8: 604–612.
  60. Verma SS, Josyula N, Verma A, Zhang X, Veturi Y, Dewey FE, *et al.* Rare variants in drug target genes contributing to complex diseases, phenome-wide. *Sci Rep.* 2018;8: 4624.
  61. Palmberg C, Koivisto P, Hyytinen E, Isola J, Visakorpi T, Kallioniemi OP, *et al.* Androgen receptor gene amplification in a recurrent prostate cancer after monotherapy with the nonsteroidal potent antiandrogen Casodex (bicalutamide) with a subsequent favorable response to maximal androgen blockade. *Eur Urol.* 1997;31: 216–219.
  62. Sequist LV, Martins RG, Spigel D, Grunberg SM, Spira A, Jänne PA, *et al.* First-line gefitinib in patients with advanced non-small-cell lung cancer harboring somatic EGFR mutations. *J Clin Oncol.* 2008;26: 2442–2449.
  63. Chung W-H, Hung S-I, Hong H-S, Hsieh M-S, Yang L-C, Ho H-C, *et al.* Medical genetics: a marker for Stevens-Johnson syndrome. *Nature.* 2004;428: 486.
  64. Bracq) C, Huyn' NT, Chausalet E, Raffin' C. HLA phenotypes and bullous cutaneous reactions to. *Tissue Antigens.* 1986;28: 251–254.
  65. Hung S-I, Chung W-H, Liou L-B, Chu C-C, Lin M, Huang H-P, *et al.* HLA-B\*5801 allele as a

- genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proc Natl Acad Sci U S A*. 2005;102: 4134–4139.
66. Frueh FW, Amur S, Mummaneni P, Epstein RS, Aubert RE, DeLuca TM, *et al*. Pharmacogenomic biomarker information in drug labels approved by the United States food and drug administration: prevalence of related drug use. *Pharmacotherapy*. 2008;28: 992–998.
67. Table of Pharmacogenomic Biomarkers in Drug Labeling. In: U.S. Food and Drug Administration [Internet]. 26 Mar 2019 [cited Apr 2019]. Available: [www.fda.gov](http://www.fda.gov)
68. Diasio RB, Harris BE. Clinical pharmacology of 5-fluorouracil. *Clin Pharmacokinet*. 1989;16: 215–237.
69. Lee AM, Shi Q, Pavey E, Alberts SR, Sargent DJ, Sinicrope FA, *et al*. DPYD variants as predictors of 5-fluorouracil toxicity in adjuvant colon cancer treatment (NCCTG N0147). *J Natl Cancer Inst*. 2014;106. doi:10.1093/jnci/dju298
70. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, *et al*. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol*. 2012;13: 239–246.
71. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, *et al*. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med*. 2010;362: 2380–2388.
72. Sequist LV, Yang JC-H, Yamamoto N, O'Byrne K, Hirsh V, Mok T, *et al*. Phase III study of

- afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol*. 2013;31: 3327–3334.
73. Lauschke VM, Ingelman-Sundberg M. Prediction of drug response and adverse drug reactions: From twin studies to Next Generation Sequencing. *Eur J Pharm Sci*. 2019;130: 65–77.
  74. Chan EY. Advances in sequencing technology. *Mutat Res*. 2005;573: 13–40.
  75. Levy SE, Myers RM. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet*. 2016;17: 95–115.
  76. Buermans HPJ, den Dunnen JT. Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta*. 2014;1842: 1932–1941.
  77. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, *et al*. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526: 75–81.
  78. International HapMap Consortium. The International HapMap Project. *Nature*. 2003;426: 789–796.
  79. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al*. A global reference for human genetic variation. *Nature*. 2015;526: 68–74.
  80. Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med*. 2009;360: 1759–1768.
  81. Giacomini KM, Yee SW, Mushiroda T, Weinshilboum RM, Ratain MJ, Kubo M. Genome-wide association studies of drug response and toxicity: an opportunity for genome

- medicine. *Nat Rev Drug Discov.* 2017;16: 1.
82. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet.* 2012;13: 135–145.
  83. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet.* 2008;17: R156–65.
  84. Zhao B, Hemann MT, Lauffenburger DA. Intratumor heterogeneity alters most effective drugs in designed combinations. *Proc Natl Acad Sci U S A.* 2014;111: 10773–10778.
  85. Bayat Mokhtari R, Homayouni TS, Baluch N, Morgatskaya E, Kumar S, Das B, *et al.* Combination therapy in combating cancer. *Oncotarget.* 2017;8: 38022–38043.
  86. Blagosklonny MV. Analysis of FDA approved anticancer drugs reveals the future of cancer therapy. *Cell Cycle.* 2004;3: 1035–1042.
  87. Yap TA, Omlin A, de Bono JS. Development of therapeutic combinations targeting major cancer signaling pathways. *J Clin Oncol.* 2013;31: 1592–1605.
  88. Barendse W. The effect of measurement error of phenotypes on genome wide association studies. *BMC Genomics.* 2011;12: 232.
  89. Cobb JN, Declerck G, Greenberg A, Clark R, McCouch S. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theor Appl Genet.* 2013;126: 867–887.
  90. Singer CF, Klinglmüller F, Stratmann R, Staudigl C, Fink-Retter A, Gschwantler D, *et al.*

- Response prediction to neoadjuvant chemotherapy: comparison between pre-therapeutic gene expression profiles and in vitro chemosensitivity assay. *PLoS One*. 2013;8: e66573.
91. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. *Int J Behav Healthc Res*. 2011;2: 320–332.
  92. Bent S, Padula A, Avins AL. Brief communication: Better ways to question patients about adverse medical events: a randomized, controlled trial. *Ann Intern Med*. 2006;144: 257–261.
  93. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, *et al*. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol*. 2006;7: 903–909.
  94. Pearce A, Haas M, Viney R, Pearson S-A, Haywood P, Brown C, *et al*. Incidence and severity of self-reported chemotherapy side effects in routine care: A prospective cohort study. *PLoS One*. 2017;12: e0184360.
  95. Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. *Genome Med*. 2015;7: 16.
  96. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al*. Finding the missing heritability of complex diseases. *Nature*. 2009;461: 747–753.
  97. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet*. 2010;11: 356–366.

98. Knight JC. Approaches for establishing the function of regulatory genetic variants involved in disease. *Genome Med.* 2014;6: 92.
99. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337: 1190–1195.
100. Luizon MR, Ahituv N. Uncovering drug-responsive regulatory elements. *Pharmacogenomics.* 2015;16: 1829–1841.
101. Lauschke VM, Milani L, Ingelman-Sundberg M. Pharmacogenomic Biomarkers for Improved Drug Therapy-Recent Progress and Future Developments. *AAPS J.* 2017;20: 4.
102. Nofziger C, Paulmichl M. Accurately genotyping CYP2D6: not for the faint of heart. *Pharmacogenomics.* 2018;19: 999–1002.
103. Yang Y, Botton MR, Scott ER, Scott SA. Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics.* 2017;18: 673–685.
104. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics.* 1974;77: 71–94.
105. A. K. Corsi, B. Wightman, M. A. Chalfie. A transparent window into biology: A primer on *Caenorhabditis elegans*. In: The *C. elegans* Research Community, editor. *WormBook.* 2015.
106. Sulston J, Hodgkin J. The Nematode *Caenorhabditis elegans*. Wood WB, editor. Cold Spring Harbor; 1988.

107. Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol.* 1983;100: 64–119.
108. Ward S, Carrel JS. Fertilization and sperm competition in the nematode *Caenorhabditis elegans*. *Dev Biol.* 1979;73: 304–321.
109. Andersen EC, Shimko TC, Crissman JR, Ghosh R, Bloom JS, Seidel HS, *et al.* A Powerful New Quantitative Genetics Platform, Combining *Caenorhabditis elegans* High-Throughput Fitness Assays with a Large Collection of Recombinant Strains. *G3* . 2015;5: 911–920.
110. Rockman MV, Kruglyak L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* 2009;5: e1000419.
111. Doroszuk A, Snoek LB, Fradin E, Riksen J, Kammenga J. A genome-wide library of CB4856/N2 introgression lines of *Caenorhabditis elegans*. *Nucleic Acids Res.* 2009;37: e110.
112. Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, *et al.* Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* 2006;2: e222.
113. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007;315: 1709–1712.
114. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337:

- 816–821.
115. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. 2008;321: 960–964.
  116. Chiu H, Schwartz HT, Antoshechkin I, Sternberg PW. Transgene-free genome editing in *Caenorhabditis elegans* using CRISPR-Cas. *Genetics*. 2013;195: 1167–1171.
  117. Cho SW, Lee J, Carroll D, Kim J-S, Lee J. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics*. 2013;195: 1177–1180.
  118. Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat Methods*. 2013;10: 741–743.
  119. Lo T-W, Pickle CS, Lin S, Ralston EJ, Gurling M, Schartner CM, *et al.* Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. *Genetics*. 2013;195: 331–348.
  120. Waaijers S, Portegijs V, Kerver J, Lemmens BBLG, Tijsterman M, van den Heuvel S, *et al.* CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. *Genetics*. 2013;195: 1187–1191.
  121. Katic I, Großhans H. Targeted heritable mutation and gene conversion by Cas9-CRISPR in *Caenorhabditis elegans*. *Genetics*. 2013;195: 1173–1176.
  122. Arribere JA, Bell RT, Fu BXH, Artiles KL, Hartman PS, Fire AZ. Efficient marker-free

- recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics*. 2014;198: 837–846.
123. Paix A, Wang Y, Smith HE, Lee C-YS, Calidas D, Lu T, *et al.* Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics*. 2014;198: 1347–1356.
124. Dickinson DJ, Goldstein B. CRISPR-Based Methods for *Caenorhabditis elegans* Genome Engineering. *Genetics*. 2016;202: 885–901.
125. Paix A, Folkmann A, Rasoloson D, Seydoux G. High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics*. 2015;201: 47–54.
126. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. 1998;282: 2012–2018.
127. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, *et al.* The sequence of the human genome. *Science*. 2001;291: 1304–1351.
128. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A*. 2007;104: 19428–19433.
129. Shaye DD, Greenwald I. OrthoList: a compendium of *C. elegans* genes with human orthologs. *PLoS One*. 2011;6: e20085.
130. Kyriakakis E, Markaki M, Tavernarakis N. *Caenorhabditis elegans* as a model for cancer

- research. *Mol Cell Oncol*. 2015;2: e975027.
131. 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: 285–290.
132. Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, *et al*. The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics*. 2016;204: 371–383.
133. Cook DE, Zdraljevic S, Roberts JP, Andersen EC. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res*. 2017;45: D650–D657.
134. 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. doi:10.7554/eLife.40260
135. Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC. Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog*. 2018;14: e1007226.
136. Kwok TCY, Ricker N, Fraser R, Chan AW, Burns A, Stanley EF, *et al*. A small-molecule screen in *C. elegans* yields a new calcium channel antagonist. *Nature*. 2006;441: 91–95.
137. Leung CK, Wang Y, Malany S, Deonaraine A, Nguyen K, Vasile S, *et al*. An ultra high-throughput, whole-animal screen for small molecule modulators of a specific genetic pathway in *Caenorhabditis elegans*. *PLoS One*. 2013;8: e62166.

138. Leung CK, Deonarine A, Strange K, Choe KP. High-throughput screening and biosensing with fluorescent *C. elegans* strains. *J Vis Exp*. 2011; doi:10.3791/2745
139. Gosai SJ, Kwak JH, Luke CJ, Long OS, King DE, Kovatch KJ, *et al*. Automated high-content live animal drug screening using *C. elegans* expressing the aggregation prone serpin  $\alpha$ 1-antitrypsin Z. *PLoS One*. 2010;5: e15460.
140. O'Reilly LP, Luke CJ, Perlmutter DH, Silverman GA, Pak SC. *C. elegans* in high-throughput drug discovery. *Adv Drug Deliv Rev*. 2014;69-70: 247–253.
141. Andersen EC, Bloom JS, Gerke JP, Kruglyak L. A variant in the neuropeptide receptor *npr-1* is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet*. 2014;10: e1004156.
142. Zdraljevic S, Fox BW, Strand C, Panda O, Tenjo-Castano FJ, Brady SC, *et al*. Natural variation in arsenic toxicity is explained by differences in branched chain amino acid catabolism [Internet]. *bioRxiv*. 2018. p. 373787. doi:10.1101/373787
143. Zamanian M, Cook DE, Zdraljevic S, Brady SC, Lee D, Lee J, *et al*. Discovery of genomic intervals that underlie nematode responses to benzimidazoles. *PLoS Negl Trop Dis*. 2018;12: e0006368.
144. Shimko TC, Andersen EC. COPASutils: an R package for reading, processing, and visualizing data from COPAS large-particle flow cytometers. *PLoS One*. 2014;9: e111090.
145. Shimko T, Zdraljevic S, Brady SC, Evans KS, Crombie TA. easysorter [Internet]. 21 Aug 2015 [cited 4 May 2019]. Available: <https://github.com/AndersenLab/easysorter>

146. Shimko T, Brady SC, Andersen EC. COPASutils package [Internet]. 25 May 2014 [cited 4 May 2019]. Available: <https://github.com/AndersenLab/COPASutils>
147. 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: 574–578.
148. Nicholas WL, Dougherty EC, Hansen EL. Axenic Cultivation of *Caenorhabditis briggsae* (Nematoda: Rhabditidae) with Chemically Undefined Supplements; Comparative Studies with Related Nematodes. *Ann N Y Acad Sci*. 2006;77: 218–236.
149. J. Hodgkin TD. Natural Variation and Copulatory Plug Formation in *Caenorhabditis Elegans*. *Genetics*. Genetics Society of America; 1997;146: 149.
150. Thompson OA, Snoek LB, Nijveen H, Sterken MG, Volkens RJM, Brenchley R, *et al*. Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. *Genetics*. 2015;200: 975–989.
151. Wicks SR, Yeh RT, Gish WR, Waterston RH, Plasterk RH. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat Genet*. 2001;28: 160–164.
152. Swan KA, Curtis DE, McKusick KB, Voinov AV, Mapa FA, Cancilla MR. High-throughput gene mapping in *Caenorhabditis elegans*. *Genome Res*. 2002;12: 1100–1105.
153. Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, *et al*. The million mutation project: a new approach to genetics in *Caenorhabditis elegans*. *Genome Res*. 2013;23: 1749–1762.

154. Vergara IA, Tarailo-Graovac M, Frech C, Wang J, Qin Z, Zhang T, *et al.* Genome-wide variations in a natural isolate of the nematode *Caenorhabditis elegans*. *BMC Genomics*. 2014;15: 255.
155. de Bono M, Bargmann CI. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell*. 1998;94: 679–689.
156. Kammenga JE, Doroszuk A, Riksen JAG, Hazendonk E, Spiridon L, Petrescu A-J, *et al.* A *Caenorhabditis elegans* wild type defies the temperature-size rule owing to a single nucleotide polymorphism in *tra-3*. *PLoS Genet*. 2007;3: e34.
157. Pollard DA, Rockman MV. Resistance to germline RNA interference in a *Caenorhabditis elegans* wild isolate exhibits complexity and nonadditivity. *G3* . 2013;3: 941–947.
158. Rockman MV, Skrovaneck SS, Kruglyak L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science*. 2010;330: 372–376.
159. Shimko T, Brady SC, Cook DE, Zdraljevic S, Andersen EC. linkagemapping package [Internet]. 1 Jun 2014 [cited 4 May 2019]. Available: <https://github.com/AndersenLab/linkagemapping/>
160. Brady SC, Zdraljevic S, Bisaga KW, Tanny RE, Cook DE, Lee D, *et al.* A nematode-specific gene underlies bleomycin-response variation in *Caenorhabditis elegans* [Internet]. *bioRxiv*. 2019. p. 565218. doi:10.1101/565218
161. Bendesky A, Tsunozaki M, Rockman MV, Kruglyak L, Bargmann CI. Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. *Nature*. 2011;472: 313–318.

162. Lee D, Yang H, Kim J, Brady S, Zdraljevic S, Zamanian M, *et al.* The genetic basis of natural variation in a phoretic behavior. *Nat Commun.* 2017;8: 273.
163. Bendesky A, Pitts J, Rockman MV, Chen WC, Tan M-W, Kruglyak L, *et al.* Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in *Caenorhabditis elegans*. *PLoS Genet.* 2012;8: e1003157.
164. Schmid T, Snoek LB, Fröhli E, van der Bent ML, Kammenga J, Hajnal A. Systemic Regulation of RAS/MAPK Signaling by the Serotonin Metabolite 5-HIAA. *PLoS Genet.* 2015;11: e1005236.
165. Reddy KC, Andersen EC, Kruglyak L, Kim DH. A polymorphism in *npr-1* is a behavioral determinant of pathogen susceptibility in *C. elegans*. *Science.* 2009;323: 382–384.
166. Seidel HS, Ailion M, Li J, van Oudenaarden A, Rockman MV, Kruglyak L. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *PLoS Biol.* 2011;9: e1001115.
167. Seidel HS, Rockman MV, Kruglyak L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science.* 2008;319: 589–594.
168. Palopoli MF, Rockman MV, TinMaung A, Ramsay C, Curwen S, Aduna A, *et al.* Molecular basis of the copulatory plug polymorphism in *Caenorhabditis elegans*. *Nature.* 2008;454: 1019–1022.
169. Chen J, Stubbe J. Bleomycins: towards better therapeutics. *Nat Rev Cancer.* 2005;5: 102–112.

170. Blum RH, Carter SK, Agre K. A clinical review of bleomycin--a new antineoplastic agent. *Cancer*. 1973;31: 903–914.
171. de Haas EC, Zwart N, Meijer C, Nuver J, Boezen HM, Suurmeijer AJH, *et al*. Variation in bleomycin hydrolase gene is associated with reduced survival after chemotherapy for testicular germ cell cancer. *J Clin Oncol*. 2008;26: 1817–1823.
172. Relling MV, Dervieux T. Pharmacogenetics and cancer therapy. *Nat Rev Cancer*. 2001;1: 99–108.
173. Cloos J, Nieuwenhuis EJ, Boomsma DI, Kuik DJ, van der Sterre ML, Arwert F, *et al*. Inherited susceptibility to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. *J Natl Cancer Inst*. 1999;91: 1125–1130.
174. Altés A, Paré L, Esquirol A, Xicoy B, Rámila E, Vicente L, *et al*. Pharmacogenetic analysis in the treatment of Hodgkin lymphoma. *Leuk Lymphoma*. 2013;54: 1706–1712.
175. Nuver J, Lutke Holzik MF, van Zweeden M, Hoekstra HJ, Meijer C, Suurmeijer AJH, *et al*. Genetic variation in the bleomycin hydrolase gene and bleomycin-induced pulmonary toxicity in germ cell cancer patients. *Pharmacogenet Genomics*. 2005;15: 399–405.
176. Lazo JS, Humphreys CJ. Lack of metabolism as the biochemical basis of bleomycin-induced pulmonary toxicity. *Proc Natl Acad Sci U S A*. 1983;80: 3064–3068.
177. Gu J, Ye Y, Spitz MR, Lin J, Kiemenev LA, Xing J, *et al*. A genetic variant near the PMAIP1/Noxa gene is associated with increased bleomycin sensitivity. *Hum Mol Genet*. 2011;20: 820–826.

178. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet.* 2014;15: 335–346.
179. Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet.* 2005;6: 287–298.
180. Liu J, Huang J, Zhang Y, Lan Q, Rothman N, Zheng T, *et al.* Identification of gene-environment interactions in cancer studies using penalization. *Genomics.* 2013;102: 189–194.
181. Low S-K, Chung S, Takahashi A, Zembutsu H, Mushiroda T, Kubo M, *et al.* Genome-wide association study of chemotherapeutic agent-induced severe neutropenia/leucopenia for patients in Biobank Japan. *Cancer Sci.* 2013;104: 1074–1082.
182. McClellan J, King M-C. Genetic heterogeneity in human disease. *Cell.* 2010;141: 210–217.
183. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol.* 2018;15: 81–94.
184. Taylor EM, Lehmann AR. Conservation of eukaryotic DNA repair mechanisms. *Int J Radiat Biol.* 1998;74: 277–286.
185. Zdraljevic S, Andersen EC. Natural diversity facilitates the discovery of conserved chemotherapeutic response mechanisms. *Curr Opin Genet Dev.* 2017;47: 41–47.
186. Stein L, Sternberg P, Durbin R, Thierry-Mieg J, Spieth J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.* 2001;29: 82–86.

187. Hillier LW, Coulson A, Murray JI, Bao Z, Sulston JE, Waterston RH. Genomics in *C. elegans*: so many genes, such a little worm. *Genome Res.* 2005;15: 1651–1660.
188. McGrath PT, Rockman MV, Zimmer M, Jang H, Macosko EZ, Kruglyak L, *et al.* Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron.* 2009;61: 692–699.
189. Boyd WA, Smith MV, Freedman JH. *Caenorhabditis elegans* as a model in developmental toxicology. *Methods Mol Biol.* 2012;889: 15–24.
190. García-González AP, Ritter AD, Shrestha S, Andersen EC, Yilmaz LS, Walhout AJM. Bacterial Metabolism Affects the *C. elegans* Response to Cancer Chemotherapeutics. *Cell.* 2017;169: 431–441.e8.
191. Cook DE, Andersen EC. VCF-kit: assorted utilities for the variant call format. *Bioinformatics.* 2017;33: 1581–1582.
192. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. *Nature.* 2013;494: 234–237.
193. Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, *et al.* The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.* 2005;6: R44.
194. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol.* 2016;34: 184–191.
195. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. R

Foundation for Statistical Computing; Available: <http://www.R-project.org/>

196. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11: R106.
197. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods.* 2017;14: 687–690.
198. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34: 525–527.
199. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27: 2987–2993.
200. 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.
201. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81: 559–575.
202. Endelman JB, Jannink J-L. Shrinkage estimation of the realized relationship matrix. *G3 .* 2012;2: 1405–1413.
203. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome.* Madison, WI: Crop Science Society of America; 2011;4: 250–255.

204. Covarrubias-Pazaran G. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLoS One. 2016;11: e0156744.
205. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, *et al.* Efficient control of population structure in model organism association mapping. Genetics. 2008;178: 1709–1723.
206. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32: 1792–1797.
207. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5: 113.
208. Snoek LB, Orbidans HE, Stastna JJ, Aartse A, Rodriguez M, Riksen JAG, *et al.* Widespread genomic incompatibilities in *Caenorhabditis elegans*. G3 . 2014;4: 1813–1823.
209. Zhang L, Li L, Yan L, Ming Z, Jia Z, Lou Z, *et al.* Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus. J Virol. 2018;92. doi:10.1128/JVI.00893-18
210. Riedel CG, Downen RH, Lourenco GF, Kirienko NV, Heimbucher T, West JA, *et al.* DAF-16 employs the chromatin remodeller SWI/SNF to promote stress resistance and longevity. Nat Cell Biol. 2013;15: 491–501.
211. Stubbe J, Kozarich JW. Mechanisms of bleomycin-induced DNA degradation. Chem Rev. American Chemical Society; 1987;87: 1107–1136.
212. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for

- protein modeling, prediction and analysis. *Nat Protoc.* 2015;10: 845–858.
213. Turek M, Besseling J, Spies J-P, König S, Bringmann H. Sleep-active neuron specification and sleep induction require FLP-11 neuropeptides to systemically induce sleep. *Elife.* 2016;5. doi:10.7554/eLife.12499
214. Du L, Sánchez C, Chen M, Edwards DJ, Shen B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem Biol.* 2000;7: 623–642.
215. Shen B, Du L, Sanchez C, Edwards DJ, Chen M, Murrell JM. Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003. *J Nat Prod.* 2002;65: 422–431.
216. Calcutt MJ, Schmidt FJ. Gene organization in the bleomycin-resistance region of the producer organism *Streptomyces verticillus*. *Gene.* 1994;151: 17–21.
217. Samuel BS, Rowedder H, Braendle C, Félix M-A, Ruvkun G. *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proc Natl Acad Sci U S A.* 2016;113: E3941–9.
218. Mardis ER. DNA sequencing technologies: 2006-2016. *Nat Protoc.* 2017;12: 213–218.
219. Easton DF, Bishop DT, Ford D, Crockford GP. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *Am J Hum Genet.* 1993;52: 678–701.

220. Cowley AW Jr. The genetic dissection of essential hypertension. *Nat Rev Genet.* 2006;7: 829–840.
221. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science.* 2008;322: 881–888.
222. Rothschild MF, Hu Z-L, Jiang Z. Advances in QTL mapping in pigs. *Int J Biol Sci.* 2007;3: 192–197.
223. Johnsson M, Jonsson KB, Andersson L, Jensen P, Wright D. Quantitative Trait Locus and Genetical Genomics Analysis Identifies Putatively Causal Genes for Fecundity and Brooding in the Chicken. *G3 .* 2015;6: 311–319.
224. Leal-Bertioli SCM, Moretzsohn MC, Roberts PA, Ballén-Taborda C, Borba TCO, Valdisser PA, *et al.* Genetic Mapping of Resistance to *Meloidogyne arenaria* in *Arachis stenosperma*: A New Source of Nematode Resistance for Peanut. *G3 .* 2015;6: 377–390.
225. Shang L, Ma L, Wang Y, Su Y, Wang X, Li Y, *et al.* Main Effect QTL with Dominance Determines Heterosis for Dynamic Plant Height in Upland Cotton. *G3 .* 2016;6: 3373–3379.
226. Mackay TF. Quantitative trait loci in *Drosophila*. *Nat Rev Genet.* 2001;2: 11–20.
227. Peng W, Xu J, Zhang Y, Feng J, Dong C, Jiang L, *et al.* An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (*Cyprinus carpio*). *Sci Rep.* 2016;6: 26693.
228. Rockman MV. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution.* 2012;66: 1–17.

229. Bloom JS, Kotenko I, Sadhu MJ, Treusch S, Albert FW, Kruglyak L. Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. *Nat Commun.* 2015;6: 8712.
230. Malmberg RL, Held S, Waits A, Mauricio R. Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics.* 2005;171: 2013–2027.
231. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A.* 2012;109: 1193–1198.
232. Nelson RM, Pettersson ME, Carlborg Ö. A century after Fisher: time for a new paradigm in quantitative genetics. *Trends Genet.* 2013;29: 669–676.
233. Lachowiec J, Shen X, Queitsch C, Carlborg Ö. A Genome-Wide Association Analysis Reveals Epistatic Cancellation of Additive Genetic Variance for Root Length in *Arabidopsis thaliana*. *PLoS Genet.* 2015;11: e1005541.
234. Mackay TFC. Epistasis for quantitative traits in *Drosophila*. *Methods Mol Biol.* 2015;1253: 47–70.
235. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 2008;4: e1000008.
236. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42: 565–569.

237. Mäki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. *Genetics*. 2014;198: 355–367.
238. Ehrenreich IM. Epistasis: Searching for Interacting Genetic Variants Using Crosses. *G3* . 2017;7: 1619–1622.
239. Simon M, Loudet O, Durand S, Bérard A, Brunel D, Sennesal F-X, *et al.* Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics*. 2008;178: 2253–2264.
240. Gutteling EW, Riksen JAG, Bakker J, Kammenga JE. Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* . 2007;98: 28–37.
241. Gutteling EW, Doroszuk A, Riksen JAG, Prokop Z, Reszka J, Kammenga JE. Environmental influence on the genetic correlations between life-history traits in *Caenorhabditis elegans*. *Heredity* . 2007;98: 206–213.
242. Viñuela A, Snoek LB, Riksen JAG, Kammenga JE. Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res*. 2010;20: 929–937.
243. Bendesky A, Bargmann CI. Genetic contributions to behavioural diversity at the gene-environment interface. *Nat Rev Genet*. 2011;12: 809–820.
244. Rodriguez M, Snoek LB, Riksen JAG, Bevers RP, Kammenga JE. Genetic variation for stress-response hormesis in *C. elegans* lifespan. *Exp Gerontol*. 2012;47: 581–587.

245. Glater EE, Rockman MV, Bargmann CI. Multigenic natural variation underlies *Caenorhabditis elegans* olfactory preference for the bacterial pathogen *Serratia marcescens*. *G3*. 2014;4: 265–276.
246. Balla KM, Andersen EC, Kruglyak L, Troemel ER. A wild *C. elegans* strain has enhanced epithelial immunity to a natural microsporidian parasite. *PLoS Pathog*. 2015;11: e1004583.
247. Singh KD, Roschitzki B, Snoek LB, Grossmann J, Zheng X, Elvin M, *et al*. Natural Genetic Variation Influences Protein Abundances in *C. elegans* Developmental Signalling Pathways. *PLoS One*. 2016;11: e0149418.
248. Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, *et al*. Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci U S A*. 2007;104: 1708–1713.
249. Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, *et al*. Genetical genomics: spotlight on QTL hotspots. *PLoS Genet*. 2008;4: e1000232.
250. Hasin-Brumshtein Y, Khan AH, Hormozdiari F, Pan C, Parks BW, Petyuk VA, *et al*. Hypothalamic transcriptomes of 99 mouse strains reveal trans eQTL hotspots, splicing QTLs and novel non-coding genes. *Elife*. 2016;5. doi:10.7554/eLife.15614
251. Ehrenreich IM, Bloom J, Torabi N, Wang X, Jia Y, Kruglyak L. Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet*. 2012;8: e1002570.
252. Knoch D, Riewe D, Meyer RC, Boudichevskaia A, Schmidt R, Altmann T. Genetic dissection of metabolite variation in *Arabidopsis* seeds: evidence for mQTL hotspots and a

- master regulatory locus of seed metabolism. *J Exp Bot.* 2017;68: 1655–1667.
253. Singh UM, Yadav S, Dixit S, Ramayya PJ, Devi MN, Raman KA, *et al.* QTL Hotspots for Early Vigor and Related Traits under Dry Direct-Seeded System in Rice (*Oryza sativa* L.). *Front Plant Sci.* 2017;8: 286.
254. Bubier JA, Jay JJ, Baker CL, Bergeson SE, Ohno H, Metten P, *et al.* Identification of a QTL in *Mus musculus* for alcohol preference, withdrawal, and Ap3m2 expression using integrative functional genomics and precision genetics. *Genetics.* 2014;197: 1377–1393.
255. Marriage TN, King EG, Long AD, Macdonald SJ. Fine-mapping nicotine resistance loci in *Drosophila* using a multiparent advanced generation inter-cross population. *Genetics.* 2014;198: 45–57.
256. Najarro MA, Hackett JL, Smith BR, Highfill CA, King EG, Long AD, *et al.* Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in *Drosophila*. *PLoS Genet.* 2015;11: e1005663.
257. Crusio WE, Dhawan E, Chesler EJ, Delprato A. Analysis of morphine responses in mice reveals a QTL on Chromosome 7. *F1000Res.* 2016;5: 2156.
258. Highfill CA, Tran JH, Nguyen SKT, Moldenhauer TR, Wang X, Macdonald SJ. Naturally Segregating Variation at *Ugt86Dd* Contributes to Nicotine Resistance in *Drosophila melanogaster*. *Genetics.* 2017;207: 311–325.
259. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* 2003;19: 889–890.

260. Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A*. 2005;102: 1572–1577.
261. David Clifford PM. The regress function. *R News*. 2006;6: 6–10.
262. David Clifford PM. The regress package. 2013.
263. Dittrich-Reed DR, Fitzpatrick BM. Transgressive Hybrids as Hopeful Monsters. *Evol Biol*. 2013;40: 310–315.
264. Laricchia KM, Zdraljevic S, Cook DE, Andersen EC. Natural Variation in the Distribution and Abundance of Transposable Elements Across the *Caenorhabditis elegans* Species. *Mol Biol Evol*. 2017;34: 2187–2202.
265. Price NPJ, Tsvetanova B. Biosynthesis of the tunicamycins: a review. *J Antibiot* . 2007;60: 485–491.
266. Evans KS, Zhao Y, Brady SC, Long L, McGrath PT, Andersen EC. Correlations of Genotype with Climate Parameters Suggest *Caenorhabditis elegans* Niche Adaptations. *G3* . 2017;7: 289–298.
267. Pommier Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nat Rev Cancer*. 2006;6: 789–802.
268. Prince M. Does Active Learning Work? A Review of the Research. *Journal of Engineering Education*. 2004;93: 223–231.
269. Ruiz-Primo MA, Briggs D, Iverson H, Talbot R, Shepard LA. Impact of undergraduate science course innovations on learning. *Science*. 2011;331: 1269–1270.

270. Springer L, Stanne ME, Donovan SS. Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis. *Rev Educ Res. American Educational Research Association*; 1999;69: 21–51.
271. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, *et al.* Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci U S A.* 2014;111: 8410–8415.
272. Smith KA, Sheppard SD, Johnson DW, Johnson RT. Pedagogies of Engagement: Classroom-Based Practices. *Journal of Engineering Education.* 2005;
273. Udovic D, Morris D, Dickman A, Postlethwait J, Wetherwax P. Workshop Biology: Demonstrating the Effectiveness of Active Learning in an Introductory Biology Course. *Bioscience. Narnia*; 2002;52: 272–281.
274. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining “Gamification.” *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments.* 2011. pp. 9–15.
275. van Roy R, Zaman B. Why Gamification Fails in Education and How to Make It Successful: Introducing Nine Gamification Heuristics Based on Self-Determination Theory. *Serious Games and Edutainment Applications 2.* 2017. pp. 485–509.
276. Dichev C, Dicheva D. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International Journal of Educational Technology in Higher Education.* 2017;14: 9.
277. Landers RN. Developing a Theory of Gamified Learning: Linking Serious Games and

- Gamification of Learning. Simul Gaming. SAGE Publications Inc; 2014;45: 752–768.
278. de Sousa Borges S, Durelli VHS, Reis HM, Isotani S. A systematic mapping on gamification applied to education. Proceedings of the 29th Annual ACM Symposium on Applied Computing. ACM; 2014. pp. 216–222.
279. Gibson B. Research Methods Jeopardy: A Tool for Involving Students and Organizing the Study Session. Teaching of Psychology. 1991;18. Available: [https://journals.sagepub.com/doi/pdf/10.1207/s15328023top1803\\_13](https://journals.sagepub.com/doi/pdf/10.1207/s15328023top1803_13)
280. Bord D. Enhancing Learning and Exam Preparation. APS Obs. 2008;21. Available: <https://www.psychologicalscience.org/observer/enhancing-learning-and-exam-preparation>
281. Nicholson S. Peeking behind the locked door: A survey of escape room facilities. 2015; Available: <http://scottnicholson.com/pubs/erfacwhite.pdf>
282. Kinio AE, Dufresne L, Brandys T, Jetty P. Break out of the Classroom: The Use of Escape Rooms as an Alternative Teaching Strategy in Surgical Education. J Surg Educ. 2019;76: 134–139.
283. Gómez-Urquiza JL, Gómez-Salgado J, Albendín-García L, Correa-Rodríguez M, González-Jiménez E, Cañadas-De la Fuente GA. The impact on nursing students' opinions and motivation of using a "Nursing Escape Room" as a teaching game: A descriptive study. Nurse Educ Today. 2019;72: 73–76.
284. Cain J. Exploratory implementation of a blended format escape room in a large enrollment pharmacy management class. Curr Pharm Teach Learn. 2019;11: 44–50.

285. Morrell BLM, Ball HM. Can You Escape Nursing School?: Educational Escape Room in Nursing Education. *Nurs Educ Perspect*. 2019; doi:10.1097/01.NEP.0000000000000441
286. Richelle Monaghan S, Nicholson S. Bringing Escape Room Concepts to Pathophysiology Case Studies. *Journal of the Human Anatomy and Physiology Society*. 2017;21. Available: <http://scottnicholson.com/pubs/escaperoompatho.pdf>
287. Dietrich N. Escape Classroom: The Leblanc Process—An Educational “Escape Game.” *J Chem Educ*. American Chemical Society; 2018;95: 996–999.
288. Vickie Adams, Stephanie Burger, Kaile Crawford, Robyn Setter. Can you escape? Creating an escape room to facilitate active learning. *Journal for Nurses in Professional Development*. 2018;
289. Eukel HN, Frenzel JE, Cernusca D. Educational Gaming for Pharmacy Students - Design and Evaluation of a Diabetes-themed Escape Room. *Am J Pharm Educ*. 2017;81: 6265.
290. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, *et al*. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010;467: 832–838.
291. Visscher PM. Sizing up human height variation. *Nat Genet*. 2008;40: 489–490.
292. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, *et al*. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn’s disease susceptibility loci. *Nat Genet*. 2010;42: 1118–1125.

293. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010;11: 446–450.
294. Kemper KE, Littlejohn MD, Lopdell T, Hayes BJ, Bennett LE, Williams RP, *et al.* Leveraging genetically simple traits to identify small-effect variants for complex phenotypes. *BMC Genomics.* 2016;17: 858.
295. Jacobsson L, Park H-B, Wahlberg P, Fredriksson R, Perez-Enciso M, Siegel PB, *et al.* Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. *Genet Res.* 2005;86: 115–125.
296. Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet.* 2006;38: 879–887.
297. Hayward JJ, Castelhana MG, Oliveira KC, Corey E, Balkman C, Baxter TL, *et al.* Complex disease and phenotype mapping in the domestic dog. *Nat Commun.* 2016;7: 10460.
298. Lorenz K, Cohen BA. Small- and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency. *Genetics.* 2012;192: 1123–1132.
299. Leamy LJ, Routman EJ, Cheverud JM. An epistatic genetic basis for fluctuating asymmetry of mandible size in mice. *Evolution.* 2002;56: 642–653.
300. Gray-McGuire C, Moser KL, Gaffney PM, Kelly J, Yu H, Olson JM, *et al.* Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and

- epistasis at 4p16-15.2. *Am J Hum Genet.* 2000;67: 1460–1469.
301. Fijneman RJ, de Vries SS, Jansen RC, Demant P. Complex interactions of new quantitative trait loci, *Sluc1*, *Sluc2*, *Sluc3*, and *Sluc4*, that influence the susceptibility to lung cancer in the mouse. *Nat Genet.* 1996;14: 465–467.
302. Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, Ayroles JF, *et al.* Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci U S A.* 2012;109: 15553–15559.
303. Ackermann M, Beyer A. Systematic Detection of Epistatic Interactions Based on Allele Pair Frequencies. *PLOS Genetics.* 2012; Available: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002463>
304. Young AI, Durbin R. Estimation of epistatic variance components and heritability in founder populations and crosses. *Genetics.* 2014;198: 1405–1416.
305. Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet.* 2005;6: 271–286.
306. Weber K, Eisman R, Morey L, Patty A, Sparks J, Tausek M, *et al.* An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. *Genetics.* 1999;153: 773–786.
307. Zimmerman E, Palsson A, Gibson G. Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. *Genetics.* 2000;155: 671–683.
308. Mezey JG, Houle D, Nuzhdin SV. Naturally segregating quantitative trait loci affecting

- wing shape of *Drosophila melanogaster*. *Genetics*. 2005;169: 2101–2113.
309. Li MJ, Yan B, Sham PC, Wang J. Exploring the function of genetic variants in the non-coding genomic regions: approaches for identifying human regulatory variants affecting gene expression. *Brief Bioinform*. 2015;16: 393–412.
310. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS*. 2009;106: 9362–9367.
311. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*. 2011;12: 683–691.
312. Large EE, Xu W, Zhao Y, Brady SC, Long L, Butcher RA, *et al*. Selection on a Subunit of the NURF Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of *Caenorhabditis elegans*. *PLoS Genet*. Public Library of Science; 2016;12: e1006219.
313. Cook DE. Characterization and Applications of Natural Variation in *C. elegans*. Andersen EC, editor. PhD, Northwestern University. 2018.
314. Caswell-Chen EP, Chen J, Lewis EE, Douhan GW, Nadler SA, Carey JR. Revising the standard wisdom of *C. elegans* natural history: ecology of longevity. *Sci Aging Knowledge Environ*. 2005;2005: e30.
315. Barrière A, Félix M-A. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Curr Biol*. 2005;15: 1176–1184.
316. Kiontke KC, Félix M-A, Ailion M, Rockman MV, Braendle C, Pénigault J-B, *et al*. A

- phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol Biol.* 2011;11: 339.
317. McGrath PT, Xu Y, Ailion M, Garrison JL, Butcher RA, Bargmann CI. Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. *Nature.* 2011;477: 321–325.
318. Ross JA, Koboldt DC, Staisch JE, Chamberlin HM, Gupta BP, Miller RD, *et al.* *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS Genet.* 2011;7: e1002174.
319. Rockman MV, Stern DL. Tinker where the tinkering's good. *Trends in Genetics.* 2008; 24: 317-319.

## Appendix A: Co-authored publications

I had the pleasure of collaborating with members of my own laboratory as well as members of other laboratories on several papers. Here, I describe the publications I co-authored.

### Selection on a Subunit of the NURF Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of *Caenorhabditis elegans*

Edward E. Large<sup>1</sup> , Wen Xu<sup>1</sup> , Yuehui Zhao<sup>1</sup>, Shannon C. Brady<sup>2</sup> , Lijiang Long<sup>1</sup> , Rebecca A. Butcher<sup>3</sup> , Erik C. Andersen<sup>2</sup> , Patrick T. McGrath<sup>1</sup> \*

<sup>1</sup>School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA

<sup>2</sup>Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, USA

<sup>3</sup>Department of Chemistry, University of Florida, Gainesville, Florida, USA

\*Corresponding author, [patrick.mcgrath@biology.gatech.edu](mailto:patrick.mcgrath@biology.gatech.edu)

This manuscript was published in *PLoS Genetics* in July 2016 [312]

### Abstract

Evolutionary life history theory seeks to explain how reproductive and survival traits are shaped by selection through allocations of an individual's resources to competing life functions. Although life-history traits evolve rapidly, little is known about the genetic and cellular mechanisms that

control and couple these tradeoffs. Here, we find that two laboratory-adapted strains of *C. elegans* descended from a single common ancestor that lived in the 1950s have differences in a number of life-history traits, including reproductive timing, lifespan, dauer formation, growth rate, and offspring number. We identified a quantitative trait locus (QTL) of large effect that controls 24%–75% of the total trait variance in reproductive timing at various timepoints. Using CRISPR/Cas9-induced genome editing, we show this QTL is due in part to a 60 bp deletion in the 3' end of the *nurf-1* gene, which is orthologous to the human gene encoding the BPTF component of the NURF chromatin remodeling complex. Besides reproduction, *nurf-1* also regulates growth rate, lifespan, and dauer formation. The fitness consequences of this deletion are environment specific—it increases fitness in the growth conditions where it was fixed but decreases fitness in alternative laboratory growth conditions. We propose that chromatin remodeling, acting through *nurf-1*, is a pleiotropic regulator of life history trade-offs underlying the evolution of multiple traits across different species.

## Contributions

I performed the high-throughput fitness assay of N2 and LSJ2 in water, arsenic, zinc, DMSO, albendazole, and abamectin and analyzed these data (Figure 6 of the manuscript).

# Correlations of genotype with climate parameters suggest *Caenorhabditis elegans* niche adaptations

Kathryn S. Evans<sup>\*,†</sup>, Yuehui Zhao<sup>‡</sup>, Shannon C. Brady<sup>\*,†</sup>, Lijiang Long<sup>‡</sup>, Patrick T. McGrath<sup>‡</sup>, and Erik C. Andersen<sup>\*,§,\*\*,1</sup>

<sup>\*</sup>Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208

<sup>†</sup>Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL 60208

<sup>§</sup>Chemistry of Life Processes Institute, Northwestern University, Evanston, IL 60208

<sup>\*\*</sup>Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208

<sup>‡</sup>School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332

<sup>1</sup>Corresponding author, [erik.andersen@northwestern.edu](mailto:erik.andersen@northwestern.edu)

This manuscript was published in *G3* in January 2017 [266]

## Abstract

Species inhabit a variety of environmental niches, and the adaptation to a particular niche is often controlled by genetic factors, including gene-by-environment interactions. The genes that vary in order to regulate the ability to colonize a niche are often difficult to identify, especially in the context of complex ecological systems and in experimentally uncontrolled natural environments. Quantitative genetic approaches provide an opportunity to investigate correlations between

genetic factors and environmental parameters that might define a niche. Previously, we have shown how a collection of 208 whole-genome sequenced wild *Caenorhabditis elegans* can facilitate association mapping approaches. To correlate climate parameters with the variation found in this collection of wild strains, we used geographic data to exhaustively curate daily weather measurements in short-term (3 month), middle-term (one year), and long-term (three year) durations surrounding the date of strain isolation. These climate parameters were used as quantitative traits in association mapping approaches, where we identified 11 quantitative trait loci (QTL) for three climatic variables: elevation, relative humidity, and average temperature. We then narrowed the genomic interval of interest to identify gene candidates with variants potentially underlying phenotypic differences. Additionally, we performed two-strain competition assays at high and low temperatures to validate a QTL that could underlie adaptation to temperature and found suggestive evidence supporting that hypothesis.

## Contributions

I worked with Katie on the competition assay for this manuscript.

## The genetic basis of natural variation in a phoretic behavior

Daehan Lee<sup>1,2</sup>, Heeseung Yang<sup>1</sup>, Jun Kim<sup>1</sup>, Shannon C. Brady<sup>2</sup>, Stefan Zdraljevic<sup>2</sup>, Mostafa Zamanian<sup>2,3</sup>, Heekyeong Kim<sup>4</sup>, Young-ki Paik<sup>4,5,6</sup>, Leonid Kruglyak<sup>7,8</sup>, Erik C. Andersen<sup>2\*</sup>, and Junho Lee<sup>1\*</sup>

<sup>1</sup>Department of Biological Sciences, Institute of Molecular Biology and Genetics, Seoul National University, Seoul 08826, Korea

<sup>2</sup>Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

<sup>3</sup>Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

<sup>4</sup>Yonsei Proteome Research Center, Yonsei University, Seoul 03722, Korea

<sup>5</sup>Department of Integrated OMICS for Biomedical Science, Yonsei University, Seoul 03722, Korea

<sup>6</sup>Department of Biochemistry, Yonsei University, Seoul 03722, Korea

<sup>7</sup>Department of Human Genetics and Biological Chemistry, University of California, Los Angeles, CA 90095, USA

<sup>8</sup>Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

\*Corresponding authors, [erik.andersen@northwestern.edu](mailto:erik.andersen@northwestern.edu) and [elegans@snu.ac.kr](mailto:elegans@snu.ac.kr)

This manuscript was published in *Nature Communications* in August 2017 [162]

### Abstract

Phoresy is a widespread form of commensalism that facilitates dispersal of one species through an association with a more mobile second species. Dauer larvae of the nematode *Caenorhabditis*

*C. elegans* exhibit a phoretic behavior called nictation, which could enable interactions with animals such as isopods or snails. Here, we show that natural *C. elegans* isolates differ in nictation. We use quantitative behavioral assays and linkage mapping to identify a genetic locus (*nict-1*) that mediates the phoretic interaction with terrestrial isopods. The *nict-1* locus contains a Piwi-interacting small RNA (piRNA) cluster; we observe that the Piwi Argonaute PRG-1 is involved in the regulation of nictation. Additionally, this locus underlies a trade-off between offspring production and dispersal. Variation in the *nict-1* locus contributes directly to differences in association between nematodes and terrestrial isopods in a laboratory assay. In summary, the piRNA-rich *nict-1* locus could define a novel mechanism underlying phoretic interactions.

## Contributions

I constructed the CRISPR/Cas9-mediated deletion allele of *prg-1* for this manuscript.

# Discovery of genomic intervals that underlie nematode responses to benzimidazoles

Mostafa Zamanian<sup>1,\*</sup>, Daniel E. Cook<sup>2,3</sup>, Stefan Zdraljevic<sup>2,3</sup>, Shannon C. Brady<sup>2,3</sup>, Daehan Lee<sup>2,4</sup>, Junho Lee<sup>4</sup>, Erik C. Andersen<sup>2,5,6\*</sup>

<sup>1</sup>Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>2</sup>Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, USA

<sup>3</sup>Interdisciplinary Biological Science Program, Northwestern University, Evanston, Illinois, USA

<sup>4</sup>Institute of Molecular Biology and Genetics, Department of Biological Sciences, Seoul National University, Seoul, Korea

<sup>5</sup>Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Chicago, Illinois, USA

<sup>6</sup>Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois, USA

\*Corresponding authors, [mzamanian@wisc.edu](mailto:mzamanian@wisc.edu) and [erik.andersen@northwestern.edu](mailto:erik.andersen@northwestern.edu)

This manuscript was published in *PLoS Neglected Tropical Diseases* in March 2018 [143]

## Abstract

Parasitic nematodes impose a debilitating health and economic burden across much of the world. Nematode resistance to anthelmintic drugs threatens parasite control efforts in both human and veterinary medicine. Despite this threat, the genetic landscape of potential resistance mechanisms to these critical drugs remains largely unexplored. Here, we exploit natural variation

in the model nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* to discover quantitative trait loci (QTL) that control sensitivity to benzimidazoles widely used in human and animal medicine. High-throughput phenotyping of albendazole, fenbendazole, mebendazole, and thiabendazole responses in panels of recombinant lines led to the discovery of over 15 QTL in *C. elegans* and four QTL in *C. briggsae* associated with divergent responses to these anthelmintics. Many of these QTL are conserved across benzimidazole derivatives, but others show drug and dose specificity. We used near-isogenic lines to recapitulate and narrow the *C. elegans* albendazole QTL of largest effect and identified candidate variants correlated with the resistance phenotype. These QTL do not overlap with known benzimidazole target resistance genes from parasitic nematodes and present specific new leads for the discovery of novel mechanisms of nematode benzimidazole resistance. Analyses of orthologous genes reveal conservation of candidate benzimidazole resistance genes in medically important parasitic nematodes. These data provide a basis for extending these approaches to other anthelmintic drug classes and a pathway towards validating new markers for anthelmintic resistance that can be deployed to improve parasite disease control.

## Contributions

I generated the CRISPR/Cas9-mediated deletion alleles of *prg-1* in the N2 and CB4856 backgrounds and measured their responses to albendazole.

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

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

<sup>1</sup>Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA

<sup>2</sup>Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

<sup>3</sup>Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA

<sup>4</sup>The Buck Institute for Research on Aging, Novato, CA 94945, USA

<sup>5</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

<sup>6</sup>Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL 60611, USA

\*Corresponding author, [erik.andersen@northwestern.edu](mailto:erik.andersen@northwestern.edu)

This manuscript was published in *eLife* in April 2019 [134].

## Abstract

We find that variation in the *dbt-1* gene underlies natural differences in *Caenorhabditis elegans* responses to the toxin arsenic. This gene encodes the E2 subunit of the branched-chain  $\alpha$ -keto

acid dehydrogenase (BCKDH) complex, a core component of branched-chain amino acid (BCAA) metabolism. We causally linked a non-synonymous variant in the conserved lipoyl domain of DBT-1 to differential arsenic responses. Using targeted metabolomics and chemical supplementation, we demonstrate that differences in responses to arsenic are caused by variation in iso-branched chain fatty acids. Additionally, we show that levels of branched chain fatty acids in human cells are perturbed by arsenic treatment. This finding has broad implications for arsenic toxicity and for arsenic-focused chemotherapeutics across human populations. Our study implicates the BCKDH complex and BCAA metabolism in arsenic responses, demonstrating the power of *C. elegans* natural genetic diversity to identify novel mechanisms by which environmental toxins affect organismal physiology.

## Contributions

I assisted Stefan in the construction of the CRISPR/Cas9-mediated allele replacement strains.

# Appendix B: Collection of wild isolates on the Hawaiian Islands

## Preface

In the summer of 2017, I and other members of the Andersen Lab went on a sampling trip to the Hawaiian Islands. In this section, I will summarize the purpose of the trip and my contributions to the sampling effort. A more complete description of the collection effort is detailed in Dan Cook's Ph.D. thesis and Tim Crombie is developing a manuscript for publication [313].

## Motivation for collections on the Hawaiian Islands

The species of *Caenorhabditis elegans* harbors immense amount of genetic diversity [131–133]. In particular, strains isolated from the Hawaiian Islands, California, and New Zealand are highly divergent from strains collected in other regions of the globe [131]. In Erik's 2012 paper, he states that "Hawaii and the Pacific Rim may be fruitful ground for discovery of additional highly diverged isolates" [131]. Additionally, collection efforts on the Hawaiian Islands often yield new strains of *C. elegans*. The potential to uncover genomic diversity and the high probability of sampling success encouraged Erik to schedule a sampling effort for the lab.

## Results

One purpose of the sampling effort was to determine the niche preferences of *C. elegans*. Previous sampling efforts suggested that *C. elegans* are most likely found on decaying plant

material and in association with invertebrates [314,315]. We sampled these substrates along many trails of the islands of Kauai, Molokai, Maui, and Hawaii. We preferentially selected trails that traversed a range of elevations to sample across various environmental parameters. On each hike, we would collect samples of fruits, nuts, vegetable, flowers, leaf litter, slugs, and isopods. We collected samples opportunistically, upon finding a suitable substrate, as well as deeply within a three-meter grid. In total, we collected 2,263 samples.

When we collected a sample, we recorded the precise location of the sample, ambient temperature and humidity, and substrate temperature and moisture level. We stored each collection in a plastic bag with a unique barcode. Each evening when we returned to the field house, we plated samples out on 6 or 10 cm agar plates seeded with *E. coli* and labeled with the sample barcode. The following day, we removed the substrate and isolated nematodes onto individual 3 cm plates, each labeled with a unique barcode. We used the Fulcrum app ([fulcrumapp.com](http://fulcrumapp.com)) to facilitate data management of nematode collections and isolations. Of the 2,263 samples we collected, we were able to isolate nematodes from 1,120. Because we isolated multiple nematodes from each productive sample, this collection trip yielded 2,531 nematodes.

We sent each of the 2,531 nematodes to Northwestern University, where members of the Andersen Lab led by Robyn Tanny and Tim Crombie performed lysis and PCR on each of the isolates. The primers used for PCR amplified a section of the ITS2 (internally transcribed spacer) region between the 5.8S and 28S rDNA, and successful amplification indicated that a sample was of the Rhabditid family [316]. We found 427 PCR-positive isolates, and each of these was Sanger sequenced for species identification. Ninety-five of these samples were positive for *C. elegans*

and were whole-genome sequenced, and we identified 26 distinct *C. elegans* isotypes from that analysis.

To identify niche preferences of *C. elegans*, we looked for associations between environmental and substrate parameters to the presence or absence of the species on each sample. One striking observation was that we did not find *C. elegans* at elevations less than 500 meters. Perhaps because of this elevation constraint, we often identified *C. elegans* on cool substrates. We did not notice a clear substrate type preference of the species.

## Contributions

The members of the Andersen Lab that attended the sampling trip in Hawaii were Erik Andersen, Daehan Lee, Dan Cook, Stefan Zdraljevic, Katie Evans, Briana Rodriguez, and me. The whole team spent one week on the island of Kauai, where Erik planned the hikes each day. The first day, the whole team hiked the Alaka'i Swamp Trail. The remaining days, I hiked with Briana and Katie (together, we were team "Moana") along various hikes while Erik, Daehan, Stefan, and Dan (team "Raptors") hiked together in other locations.

After the first week, team Moana and the Raptors separated to sample different islands. I led team Moana and planned the hikes along which we sampled. On the first day, we flew to the island of Molokai where we first sampled in the Pala'au state park, including around Phallic Rock and Kalaupapa Lookout and then drove toward the Kamakou peak until the terrain became too muddy to safely traverse. We then flew to Maui, where we spent the rest of the sampling trip. We hiked on many trails around the island, including the Iao Valley, Wailuku trail, Polipoli cloud forest,

Waihou park, Haleakala Ridge, Hosmer Campground, Pipiwai, Road to Hana, Twin Falls, Garden of Eden, Pua'a Ka'a Falls, and Waihe'e trail. Overall, we found that many of the hikes in Maui were low elevation and/or too dry to contain rotting substrates. Future sampling efforts should focus on the big island, as the Raptors found many productive samplings sites on Hawaii.

## Appendix C: *Easysorter* package

### Preface

The *easysorter* package is a collection of functions in R that allows for easy analysis of data from the COPAS BIOSORT. This package uses many functions from the *COPASutils* package [144]. In this section, I explain the utility of functions within the *easysorter* package as well as my contributions to the package development. A brief description of the package and all of its functions are available at [github.com/AndersenLab/easysorter](https://github.com/AndersenLab/easysorter).

### Explanation of functionality

The COPAS BIOSORT is a large-particle flow cytometer that can measure the length (time of flight, TOF), fluorescence (green, yellow, or red), and optical density (extinction, EXT) of objects in wells of a 96-well plate. The BIOSORT stores information about each object as a tab-separated value (tsv) file, where each row is a unique object. During our standard high-throughput assay, the BIOSORT is used to collect data on both the sort/setup and the score day.

To import data from the BIOSORT, template comma-separated value (csv) files first must be constructed to provide detail about the contents of each well of each plate. The *strains.csv* file should provide the name of each strain in each well. Strain layout should be identical across plates of a given assay, and only one *strain.csv* file is needed per experiment. The *conditions.csv* file should provide the experimental condition present in each well of each plate. Because each experiment likely tests multiple conditions (*e.g.* bleomycin and water), multiple *condition.csv* files

might be included for each experiment. The `controls.csv` file should provide the corresponding control condition that will be used to compare condition versus control responses. If a condition does not have a control, “None” should be used as the control condition. For example, “water” is the control condition for bleomycin, whereas “None” is the control for the water condition. Finally, each plate should contain a `contamination.csv` file, where each well is either TRUE (contaminated, as determined by visual inspection) or FALSE. The user should provide a contamination file for each plate assayed.

After the template csv files and setup and score tsv files are in the proper directories (see [github.com/andersenlab/easysorter](https://github.com/andersenlab/easysorter) for more detail), the *easysorter* package can be used to analyze the data. First, the *read\_data* function creates a list of two data frames, where the first data frame derives from the score day data, and the second data frame is from the sort/setup day data. This function adds strain, condition, control, and contamination columns that incorporate these data for each object of each well. Next, the *remove\_contamination* file removes all wells from the score data frame for which the contamination column is TRUE.

Because each well of an assay contains many animals, the phenotypic distribution of all animals in each well can be summarized in several ways. The *sumplate* function groups all animals within the same well of the same plate and calculates summary statistics for the population distribution. First, the function counts the number of animals in each well and outputs that number as “n” and, if data from the sort/setup day are available, normalizes the number of animals in the well to the number of animals sorted into that well (brood size, `norm.n`). Then, the function normalizes the EXT measurement of each animal by its TOF measurement (normalized extinction, `norm.EXT`) and each of the fluorescence measurements by TOF (normalized green, `norm.green`, normalized

red, norm.red, and normalized yellow, norm.yellow). These normalized traits account for variation in animal lengths that might affect extinction and fluorescence values. The function then calculates summary statistics for the distribution of all animals in each well (mean, median, variance, covariance, interquartile range, and 10th, 25th, 75th, and 90th quantiles of the TOF, EXT, green, red, yellow, norm.EXT, norm.green, norm.red, and norm.yellow). Finally, the function categorizes animals into predicted larval stages, depending on their TOF measurements, and it calculates the fraction of the population of each well that is in the L1, L2/L3, L4, and adult life stages.

After the population of animals in each well is summarized, the *bioprune* function is used to remove data that lie beyond particular biological cutoffs. Wells that contain less than five animals (indicating a lack of progeny and too few animals to accurately score) or more than 1,000 animals (indicating too many animals and likely contamination) are removed. If the norm.n trait was calculated in the previous step, this function also removes wells with a brood size greater than 350 (again, indicating likely contamination).

If assays were run across multiple days, assay normalization can then be performed using the *regress(assay = TRUE)* function. This function groups all data by the condition and trait and then fits the data to a linear model with the formula (*phenotype ~ assay - 1*), which takes phenotype as the response variable and sets assay as a fixed variable. The “-1” part of the model removes the y-intercept from the output. The phenotype output from this function is the residual value after this linear model is performed.

Next, outliers of the assay-regressed data are removed using either the *bamf\_prune* function or the *prune\_outliers()* function. The *bamf-prune* function, or “binned anomaly mitigate and fit” function, groups all data by condition and trait. It then calculates the interquartile range (IQR) for each group and determines how many data points lie beyond two, three, four, five, seven, and ten IQRs above the third or below the first quartile. Each data point is binned based on how many IQRs beyond the range it lies. The function then calculates the proportion of data points that lie in each bin. If less than 5% of all data points lie within a bin, all data within that bin are called outliers and are removed. However, if more than 5% of the data lie within a particular bin, all data points within that bin are kept. The *bamf\_prune* function should be used for pruning mapping data, where data points are individual strains and clusters of outliers could represent a differential phenotype of a group of strains. For assays that are testing many replicates of isogenic lines, these clusters are not expected, and the *prune\_outliers* function should be used. This function simply removes data points that lie beyond two IQR or two standard deviations from the median phenotype of each strain.

Finally, the *regress* function is used to account for differences between strains in the control condition. For each strain, the mean phenotype in the control condition is calculated. The function then performs a linear model with the formula (*phenotype ~ control phenotype -1*), where phenotype is the dependent variable, the mean control phenotype for that strain is the fixed variable, and the y-intercept of the model is ignored. This function outputs residual phenotypes.

## Contributions

This package was initially developed by Tyler Shimko. I contributed to the package by fixing bugs and adding functionality, which are detailed here.

### Fixing Bugs

Occasionally, during a high-throughput assay, the BIOSORT run needs to be interrupted, often to fix a clog in the flow cell. When a plate is interrupted during scoring, the data from that plate are split into different files, and these files need to be stitched together. The *COPASutils* package that works closely with the *easysorter* package contains a function that stitches data from each plate together in the event of a run interruption. However, this function was incorrectly assigning data into particular wells and therefore plate stitching was not working properly. I corrected the error in this function. Another error with the *easysorter* package was in the *remove\_contamination* function. This function was not removing contaminated wells in all cases, so I modified the script to fix this issue.

To account for phenotypic variation between strains that exists in the control condition, we often fit the data to a linear model with the condition phenotype as the dependent variable and control phenotype as the independent variable. The residual values from this model are often used as our control-corrected phenotypes. However, alternative strategies for control condition correction can be used, and I added a function in the *easysorter* package to perform two alternate methods for control correction. The *altregress* function can be used instead of the *regress* function and either “deltapheno” or “fracpheno” can be selected as the methods for control correction. Both of these methods first calculate the mean control phenotype of each strain. Then, the “deltapheno”

method subtracts each condition data point from its control-condition mean. The output for this method is the difference between the condition phenotype and the mean control phenotype. Alternatively, the “fracpheno” method divides each data point by its mean control phenotype. The output for this method is the condition phenotype as a fraction of the control phenotype.

## Updating for compatibility with LP sampler

Union Biometrica has developed other models of large-particle flow cytometers aside from the COPAS BIOSORT. One of these models is the BioSorter, which is a modified version of the BIOSORT with increased capabilities including an increased number of user-defined sorting gates, ability to sort into a conical, electronic pressure valves, and up to four excitation lasers. Data from the BioSorter are imported in a slightly different manner than data from the BIOSORT. In particular, the “sort” data column is denoted as “Sorted.status” instead of “Status.sort”, the “time” data is denoted as “Time” rather than “Time.Stamp”, and the “EXT” measurement is called “Extinction” rather than “EXT”. To allow *easorter* to be compatible with the BioSorter data, I added a *reflx* flag to the *read\_data* function. When BioSorter data are being imported, the user should run *read\_data(reflx = FALSE)* and when BIOSORT data are being imported, the user should use the default TRUE setting of the *reflx* flag.

## Acknowledgements

Tyler Shimko developed the majority of the *COPASutils* and *easysorter* package. Erik Andersen oversaw the development of both packages. Stefan Zdraljevic developed the *bamf\_prune* function and fixed several bugs in the *easysorter* package. Katie Evans developed the *prune\_outliers* function.

## Appendix D: *Linkagemapping* package

### Preface

The *linkagemapping* package is a collection of functions in R that allows users to run linkage mapping of phenotypic data from certain panels of *C. elegans* recombinant lines. Linkage mapping is used to identify regions of the genome that are correlated with phenotypic variation. These genomic regions are called quantitative trait loci, or QTL. The *linkagemapping* package uses many functions from the *qtl* package. In this section, I explain the utility of functions within the *linkagemapping* package as well as my contributions to the package development. A brief description of the package and each of its functions are available at [github.com/AndersenLab/linkagemapping](https://github.com/AndersenLab/linkagemapping).

### Explanation of functionality

The *linkagemapping* package requires a cross object that contains genotype and phenotype data for the mapping analysis. We provide four cross objects in the package that each contain genotype information for particular strain sets. The `N2xCB4856cross` contains genotype information for the 598 recombinant inbred advanced intercross lines generated between the N2 and CB4856 strains at 1,454 genetic markers that were genotyped with the Illumina GoldenGate assay [109,110]. The `N2xCB4856cross_full` object contains the same strains plus an additional set of recombinant lines and an expanded set of 13,003 SNPs derived from whole-genome sequencing data of the panel. The `N2xLSJ2cross` object contains genotype information for 94 recombinant lines generated between the CX12311 and LSJ2 strains at 176 informative SNPs

[317]. The AF16xHK104cross contains genotype information for 167 recombinant inbred lines between the AF16 and HK104 strains of *C. briggsae* at 1,031 SNPs [318]. Each cross object was generated by loading the genotype information and a dummy phenotype file and combining them using the *read.cross* function from the *qtl* package. Cross objects can be accessed by using the *load\_cross\_obj* function with the cross object name in double quotations (e.g. *load\_cross\_obj("N2xCB4856cross\_full")*).

Each cross object contains a “geno” list and a “pheno” data frame. The “geno” list is split into the six chromosomes, and each chromosome component contains a “data”, “map”, “argmax”, and “prob” object. The “data” data frame contains genotype information for each marker on the chromosome. For the N2xCB4856cross and N2xCB4856cross\_full, a genotype of 1 indicates an N2 allele, whereas a 2 indicates a CB4856 allele; for the N2xLSJ2cross, a genotype of 1 indicates an N2 allele, whereas a genotype of 2 indicates an LSJ2 allele; for the AF16xHK104cross, a genotype of 1 indicates an AF16 allele, whereas a genotype of 2 indicates an HK104 allele. The “map” vector lists the centiMorgan position for each of the genetic markers in the cross object, calculated based on recombination information within the strains of the cross object. The “argmax” data frame is a matrix that contains the most likely sequence of genotypes for each strain given the marker data for all strains in the cross object, and the “prob” data frame contains the probabilities of those predicted genotypes. The “pheno” data frame of the cross object contains a dummy dataset that should be replaced by actual phenotype data for a mapping. This data frame contains a “strain” column with the name of all recombinant lines in the cross object and a “set” column that groups strains into particular categories, if applicable (N2xCB4856cross has two sets, the Rockman 2009 set and the Andersen 2012 set, N2xCB4856cross\_full has an additional set of new recombinant strains).

The first step of performing a mapping with the *linkagemapping* package is to load the phenotype data. This data frame must contain columns named “strain” (strain name), “condition” (name of condition to which the strain was exposed), “trait” (phenotypic parameter that was measured, such as mean.TOF), and “phenotype” (numerical value of phenotype measurement) for proper merging of phenotype data into the cross object. The *merge\_pheno* function is used to combine the phenotypic data into the “pheno” element of the cross object. If the user wants to run a mapping on a particular set of strains in a cross object, the *set* flag can be used to denote which strain set to include in the mapping (e.g. *merge\_pheno(set = 2)*).

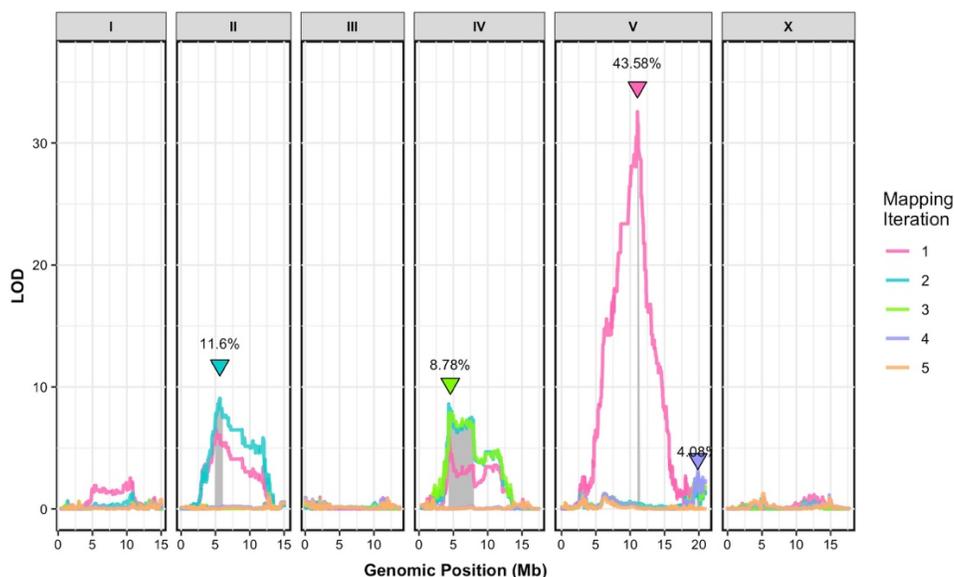
After the user has merged the phenotype data into a given cross object, the linkage mapping can be performed. The *fsearch* function performs a mapping using the package’s internal *map* function. This mapping strategy scales phenotypes to have a mean of zero and a standard deviation of one, then tests the correlation between each genetic marker and the phenotypic variation of the recombinant lines and calculates a log of the odds (LOD) score for each marker. The forward search algorithm runs this mapping to identify the most significant QTL, then sets that QTL as a cofactor and performs another mapping. This process continues iteratively until no significant QTL are found. The *fsearch* function requires a cross object to be specified. Each trait within the cross object will be mapped independently, or the user can specify one particular trait to map using the *phenotype* flag in the *fsearch* function (e.g. *fsearch(cross, phenotype = “bleomycin.mean.TOF”)*). Additionally, the *markerset* flag can be used to convert marker names to genomic positions at the end of mapping. Markersets include “N2xCB4856”, “N2xLSJ2”, “AF16xHK104”, “full” (for the N2xCB4856cross\_full object), and NA if no marker conversion is desired.

The *fsearch* function calculates significance thresholds based on either a false-discovery rate (FDR) or a genome-wide error rate (GWER). In general, the FDR threshold should be used when many traits (>100) are being mapped, whereas GWER should be used when a few traits (<100) are being mapped. Both of these strategies calculate a threshold for significant QTL after performing permutations of the mapping. The type of threshold strategy to be used can be selected with the *thresh* flag and the number of permutations of the mapping to run can be set with the *permutations* flag in the *fsearch* function (e.g. *fsearch(cross, thresh = "FDR", permutations = 1000)*).

The output of the *fsearch* function is a data frame with the LOD score of each marker for each iteration of the mapping. It also reports the significance threshold for each iteration, so significant QTL can be identified. The *annotate\_lod* function is used to define QTL confidence intervals and to calculate their effects on phenotypes. This function identifies the highest LOD score for each mapping iteration and then determines the confidence interval around the QTL using the method specified by the user. When the *bayes* argument is FALSE, the function uses markers within a 1.5-LOD drop from the peak marker to define the confidence interval. When the *bayes* argument is TRUE, the function uses Bayesian probability to identify a 95% confidence interval. Additionally, the *cutoff* argument can be set to "chromosomal", in which all markers on the same chromosome as the QTL that have LOD scores above the threshold are included in the QTL confidence interval, or to "proximal", in which the nearest markers to the left and right of the QTL peak that reach the threshold are used to define the confidence interval. Finally, for each QTL, the *annotate\_LODs* function performs a linear model with the formula (*phenotype ~ genotype - 1*). The slope of the

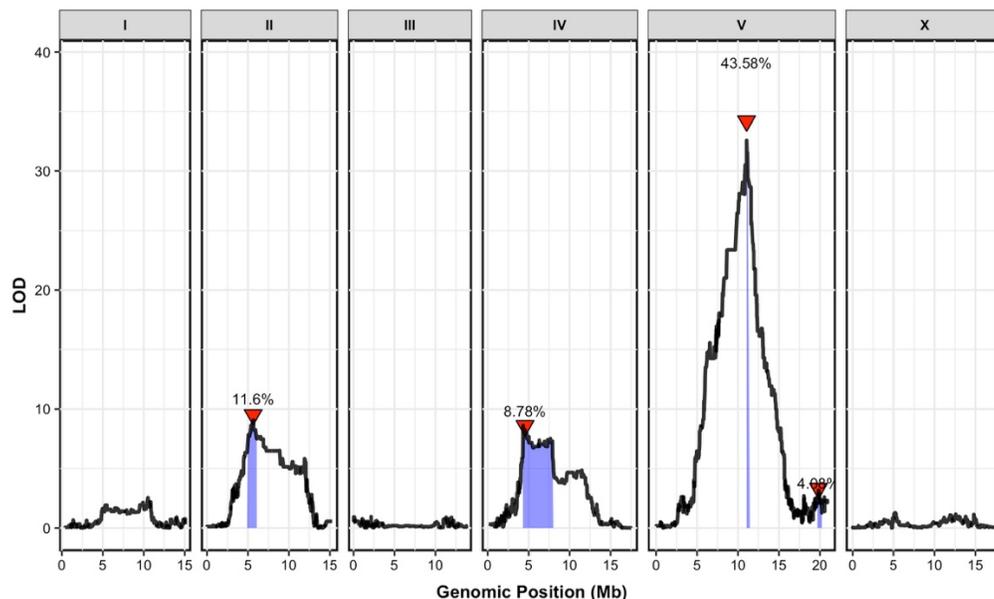
model defines the effect size of a QTL and the  $r^2$  value of the model defines the percent of variance in the RIAIL phenotypes that is explained by the QTL.

To easily visualize the results of linkage mapping, the package includes several plotting functions. The *lodplot* function plots an annotated mapping data frame, with each iteration of the mapping shown as a different colored line. An example of this type of plot is shown in **Figure 10-1**. The *maxlodplot* function combines the data from all iterations and plots the maximum LOD score for each marker across all iterations. Each QTL is shown as a red triangle and the percent of phenotypic variance explained by that QTL is printed as a percentage. An example of a *maxlodplot* output is shown in **Figure 10-2**.



#### Appendix D Figure 1 Example *lodplot* output

An example of a *lodplot* output is shown. On the x-axis, each of 13,003 genomic markers, split by chromosome, were tested for correlation with phenotypic variation across the RIAIL panel. The log of the odds (LOD) score for each marker is reported on the y-axis. Mapping repeated iteratively, and each iteration of the mapping is shown in a different color. The most significant quantitative trait locus (QTL) in each iteration is indicated by a triangle at the peak marker, and a grey ribbon shows the 95% confidence interval around the peak marker. The total amount of phenotypic variance across the RIAIL panel explained by the genotype at each peak marker is shown as a percentage.

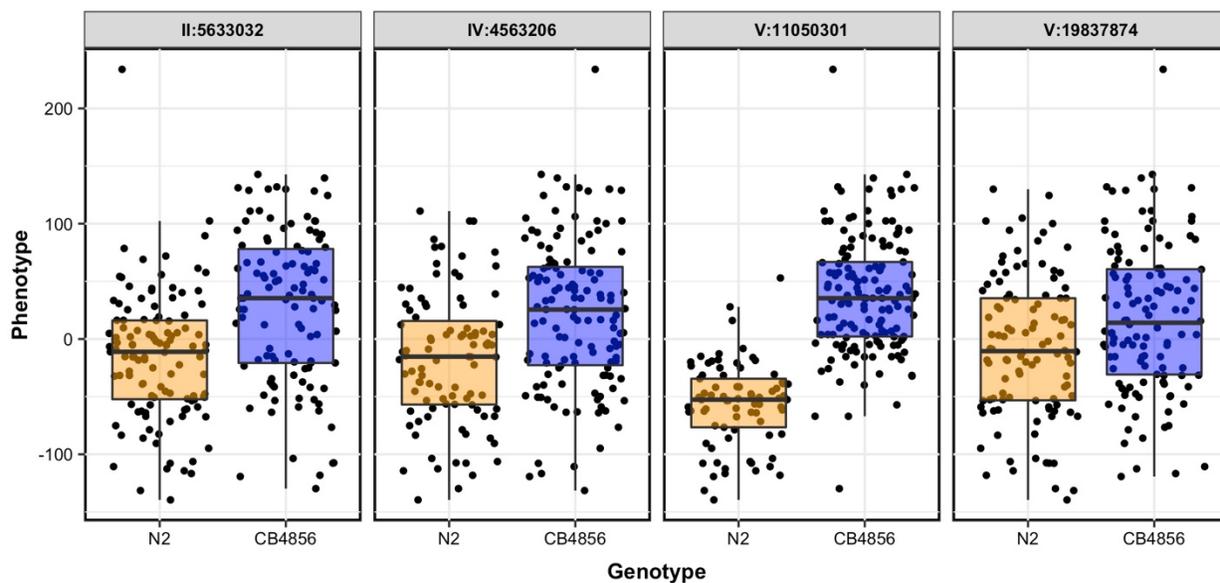


#### Appendix D Figure 2 Example *maxlodplot* output

An example of a *maxlodplot* output is shown. On the x-axis, each of 13,003 genomic markers, split by chromosome, were tested for correlation with phenotypic variation across the RIAL panel. The log of the odds (LOD) score for each marker is reported on the y-axis. Each significant quantitative trait locus (QTL) is indicated by a red triangle at the peak marker, and a blue ribbon shows the 95% confidence interval around the peak marker. The total amount of phenotypic variance across the RIAL panel explained by the genotype at each peak marker is shown as a percentage.

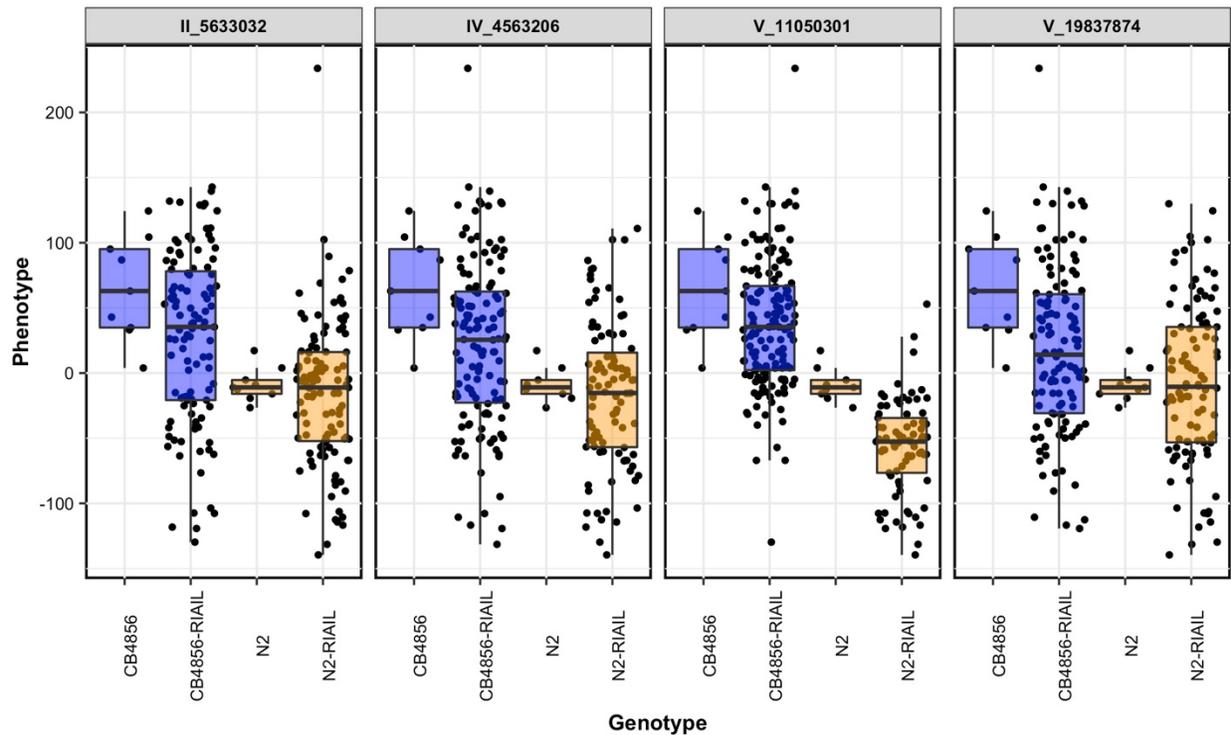
Phenotypic differences between recombinant lines of alternate genotypes at a QTL marker can be visualized with the *pxgplot* function. This function outputs a box plot, where the recombinant lines are split into groups based on their allele at each significant QTL peak marker. Colors of boxplots are determined by the *parent* argument, which defaults to “N2xCB4856” but can be set to “N2xLSJ2” or “AF16xHK104”. This function requires an annotated linkage mapping data frame as well as the cross object with which the mapping was performed. An example of the *pxgplot* output is shown in **Figure 10-3**. The parental strains are automatically removed from mapping, so parent phenotypes will not be found in the cross object. If the user wishes to add the parental phenotypes to the *pxgplot*, the function *pxgplot\_par* can be used instead. This function takes an additional argument, which is the phenotype data frame that was used in the *merge\_pheno*

function and includes parental phenotypes. An example of a *pxgplot\_par* output is shown in **Figure 10-4**.



#### Appendix D Figure 3 Example *pxgplot* output

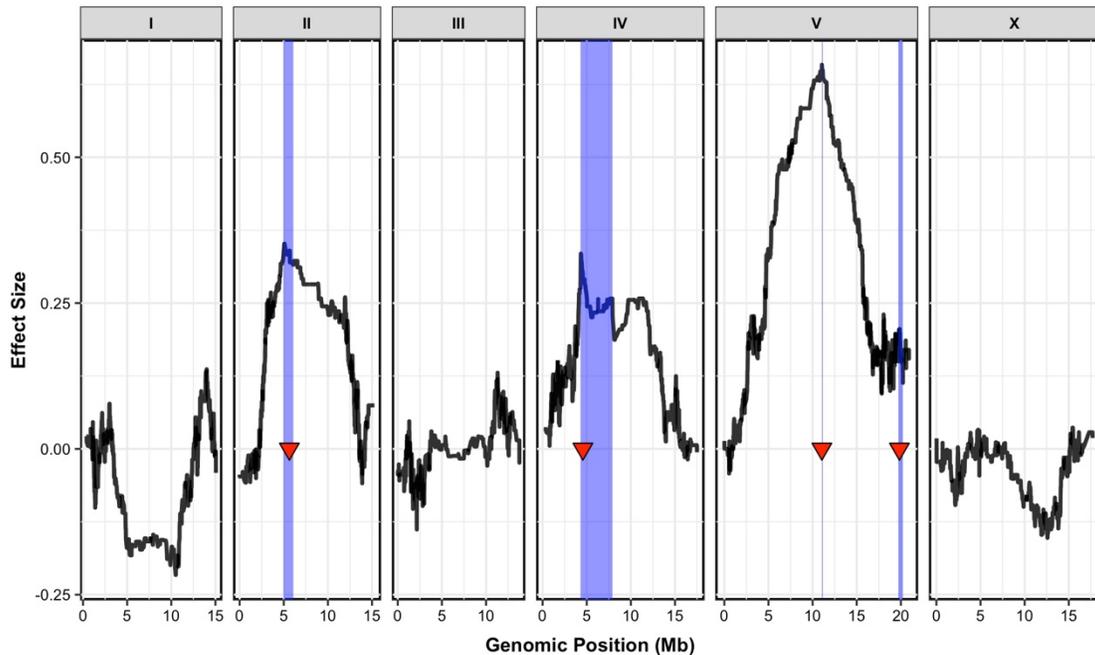
An example of a *pxgplot* output is shown. On the x-axis, RIALs are split into two groups according to their genotype at each QTL peak marker. Phenotypes of the RIALs are shown as Tukey boxplots, where each point is a RIAL. Strains with the N2 allele at the QTL marker are colored orange and strains with the CB4856 allele at the QTL marker are colored in blue.



#### Appendix D Figure 4 Example *pxgplot\_par* output

An example of the *pxgplot\_par* output is shown. On the x-axis, RIAILs are split into two groups according to their genotype at each QTL peak marker, “N2-RIAIL” and “CB4856-RIAIL”. Parental strain phenotypes are shown as “N2” and “CB4856”. Phenotypes of the RIAILs and parents are shown as Tukey boxplots, where each point is a RIAIL. Strains with the N2 allele at the QTL marker are colored orange and strains with the CB4856 allele at the QTL marker are colored in blue.

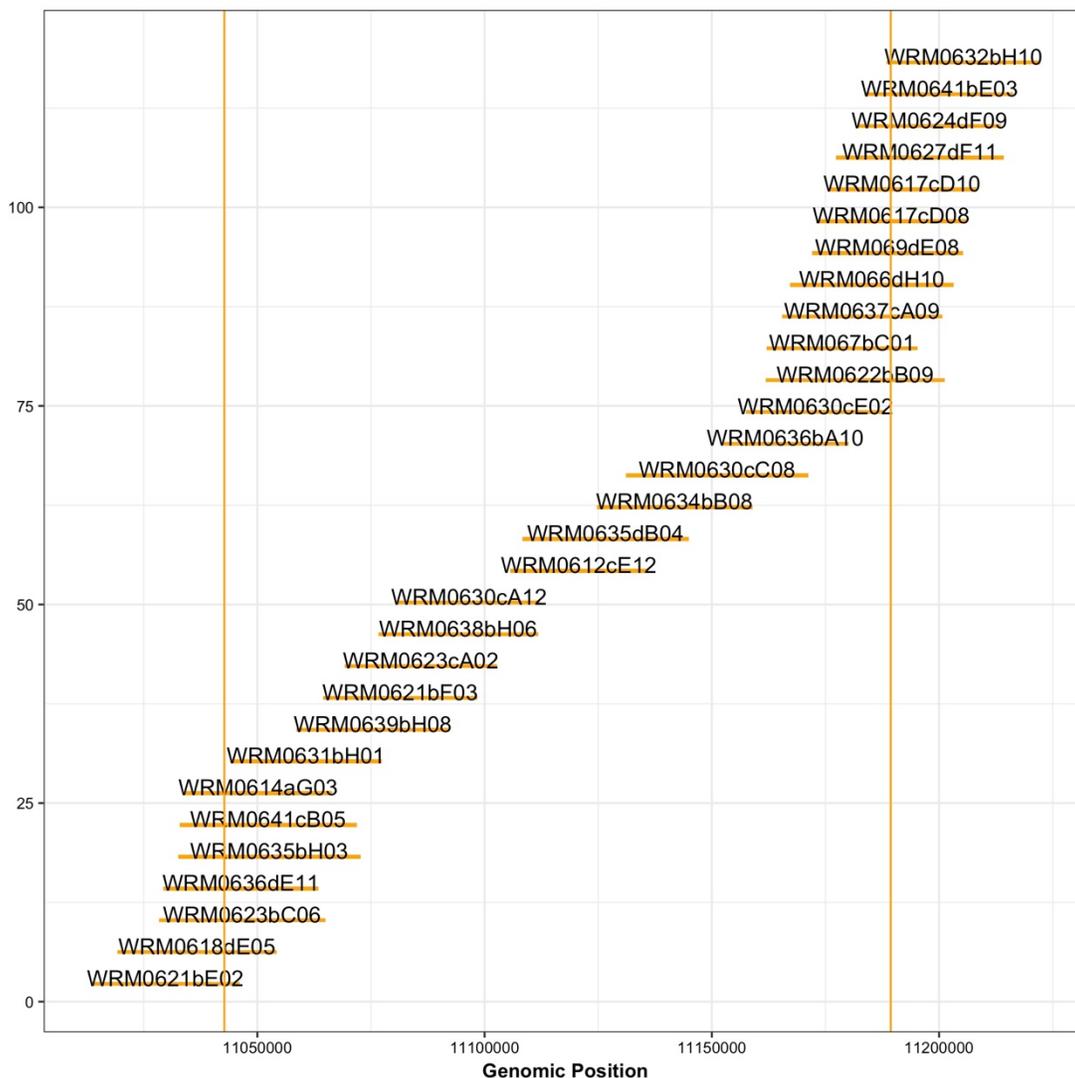
The effect size of each marker can be easily visualized with the *effectplot* function. This function takes the annotated mapping data frame as well as the cross object and determines the direction and magnitude of the phenotypic effect of each genetic marker. In the case of an N2xCB4856cross mapping, positive effect sizes indicate that strains with the CB4856 allele have larger phenotypes. An example of an *effectplot* output is shown in **Figure 10-5**.



#### Appendix D Figure 5 Example *effplot* output

An example of the *effplot* output is shown. On the x-axis, each of the 13,003 genomic markers are shown, split by chromosome. On the y-axis, the maximum effect size for each marker across all mapping iterations is shown. Positive effect sizes indicate that RIALs with the CB4856 genotype have larger phenotypes than RIALs with the N2 genotype, and *vice versa*. Each QTL is indicated by a red triangle at the QTL peak marker and a blue ribbon showing the 95% confidence interval around the QTL.

After a mapping is performed, the user might want to find reagents that can be used to test the validity of the QTL identified by the mapping. Fosmids, which are plasmids transformed with regions of the *C. elegans* genome, that contain a region of interest can be found using the *findN2fosmids* or *findCBfosmids* search. These functions find fosmids with the N2 and the CB4856 genotype, respectively, that tile across a given interval. The names of all fosmids in the region as well as a plot of their genotypes are output. An example of an output from the *findN2fosmids* is shown in **Figure 10-6**.



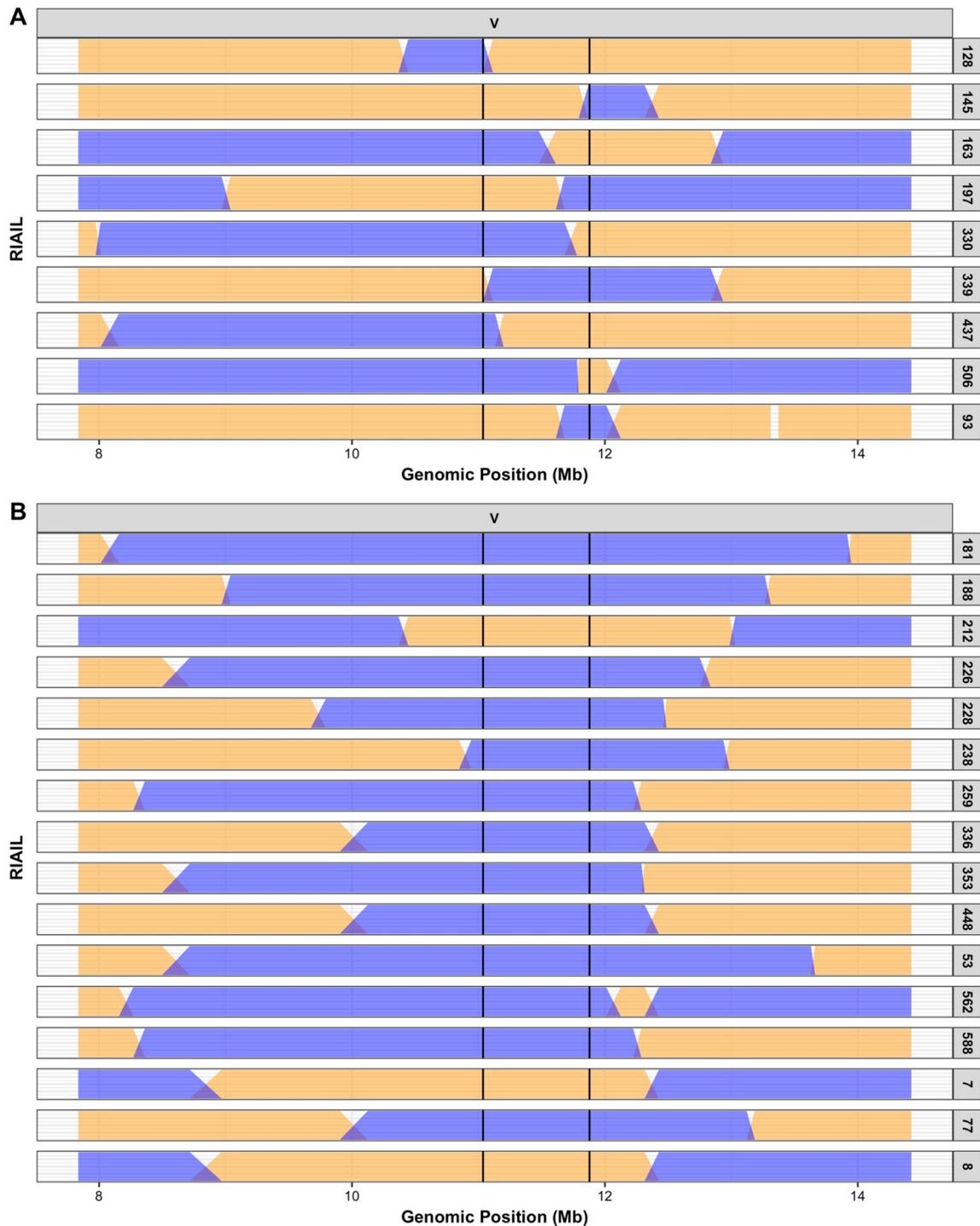
#### Appendix D Figure 6 Example *findN2fosmids* output

An example of the *findN2fosmids* output is shown. On the x-axis, genomic positions near the search interval are shown. The search interval is indicated by two vertical orange lines. Each fosmid that tiles across the interval is shown as a horizontal line, which indicates the length of the fosmid. Fosmid names are printed in text above each horizontal line. The y-axis is arbitrary.

Expression levels for many genes have been measured previously for the N2xCB4856 recombinant lines using a microarray [158]. These expression data were used in a linkage mapping to identify regions of the genome that affect expression levels of each gene. These

expression QTL (eQTL) can be searched using the *checkeQTLintervals* function. The function finds eQTL peaks that overlap with a region of interest. The function outputs a data frame of quantitative information about detected eQTL in the interval as well as a data frame that contains qualitative data about the probes that mapped to those eQTL. A positive effect size means that the strains with the CB4856 allele at the QTL peak marker have higher expression of the probe than strains with the N2 allele at the QTL peak marker, and *vice versa*.

Users may wish to identify recombinant strains that can be used to generate near-isogenic lines (NILs) to isolate a QTL in a consistent genetic background. Usually, ideal recombinant lines for making NILs do not contain genetic recombination events within the region of interest, unless the user is attempting to generate NILs that tile across a region. The *FindRIALsforNILs* function can be used to identify recombinant lines that could be used for NIL construction. This function outputs a list of four elements, a data frame of all recombinant lines with a continuous genotype across the region, a data frame of all recombinant strains with breaks within the interval, a plot of all recombinant lines that have breakpoints inside and nearby the outside of the region, and a plot of all recombinant lines with a continuous genotype across the region. An example of the plots output from the *is* function is shown in **Figure 10-7**.



#### Appendix D Figure 7 Example *findRIALsforNILs* output

An example of the *findRIALsforNILs* output is shown. On the x-axis, positions across chromosome V are shown. Vertical black lines indicate the search region. **A.** Each RIAL with genetic recombination within the search interval is shown on the y-axis. **B.** Each RIAL with a constant genotype through the search interval is shown on the y-axis.

## Contributions

This package was constructed by Tyler Shimko. I spent a significant amount of time learning the functionalities of the package, fixing issues, and expanding the functionality of the package. Here, I describe my contributions to the *linkagemapping* package.

### Fixing bugs in package

In August of 2015, we were working on a collaboration with Patrick McGrath using the N2xLSJ2 recombinant lines. We noticed that our mapping results were not matching up, and he was not using the *linkagemapping* package. I compared Patrick's cross object with our N2xLSJ2cross object and noticed that the *merge\_pheno* function was not properly matching strains to their phenotypes. This mismatch problem was resolved when I reconstructed the N2xLSJ2cross object after ordering the strains in alphanumeric order before running the *read.cross* function.

In May of 2018, I noticed that setting the GWER or FDR threshold for QTL significance was not working. In particular, all iterations beyond the first iteration were being calculated with FDR, regardless of which method was selected in the *fsearch* argument. The *threshold* flag was being overwritten to the first-iteration threshold. I fixed this issue by changing the flag from "*threshold*" to "*thresh*" so the argument was not overwritten internally. Katie and I then reran all of the mappings to correct for this thresholding error.

## Expanding functionality

During my rotation in the Andersen Lab, I developed a fosmid search. I incorporated this functionality into the package, as described above. I also developed the eQTL search functionality of the package.

In April 2017, I worked with Dr. Daniel E. Cook to establish an expanded marker set for the N2xCB4856cross object. We had whole-genome sequence data from the RIAILs and we wanted to add more markers into the cross object to improve mapping resolution. We had extensively sequenced the N2 and CB4856 parental strains and had identified 195,655 high-quality genotypes that differed between the two strains. We analyzed the sequences of the RIAILs at these sites and identified 13,003 SNPs that were reliably called in the RIAIL set and that were informative of genetic recombination events in the panel. Dan compiled the genotypes for the 598 RIAILs in the N2xCB4856cross object as well as 548 newly constructed recombinant lines across the 13,003 SNPs. I then created the N2xCB4856cross\_full object by following the procedure described in the *qtl* package [259].

In 2018, Kathryn Evans and I were working on the QTL hotspot paper, when we realized that our method for defining confidence intervals differed from the technique used in the *qtl* package [259]. The *qtl* package defined a confidence interval as including all loci on the same chromosome as a QTL that have a LOD score above a 1.5-LOD drop from the QTL peak. Our method was to walk marker-by-marker to the left of the QTL until the LOD score reached below the 1.5-LOD drop and to repeat the process to the right of the QTL peak. The valleys most proximal to the QTL peak marker defined the confidence interval, even if other markers climbed back up above that

threshold. We thought that both types of confidence intervals would be informative to *linkagemapping* package users, so we incorporated this option for defining confidence intervals into the package.

## Acknowledgements

The *linkagemapping* package was initially constructed by Tyler Shimko. Erik Andersen supervised the construction and development of the package. Dan Cook helped create the `N2xCB4856cross_full` object, and Stefan Zdraljevic fixed several bugs in the package. Katie Evans supported the development of the proximal and chromosomal confidence interval thresholds.