A powerful *C. elegans* resource for identifying the genetic determinants underlying complex traits

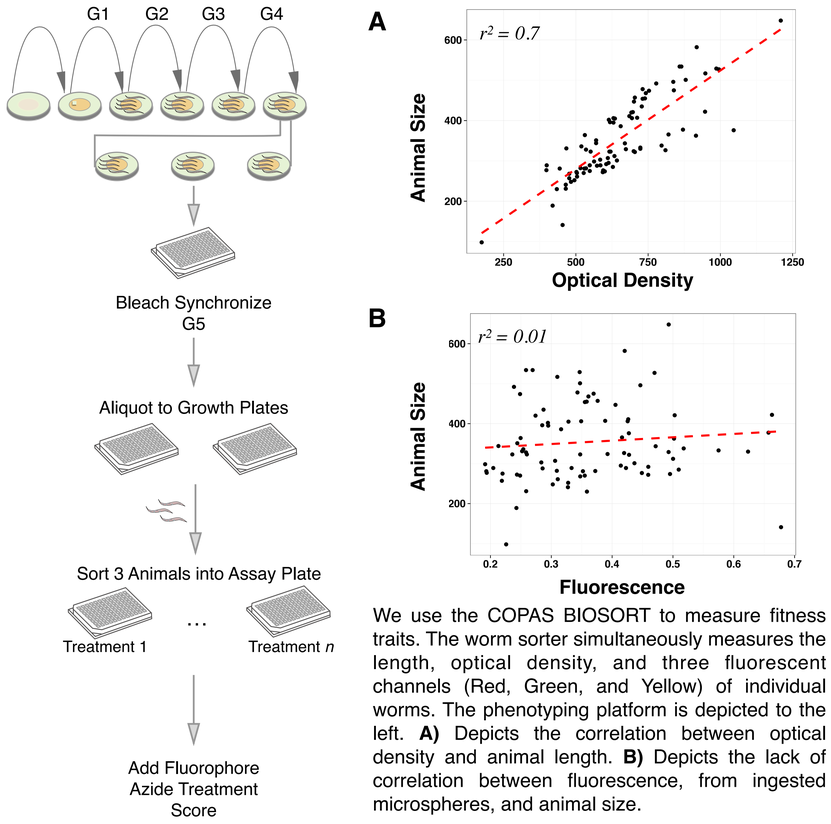
Stefan Zdraljevic[1,2], Sam K. Rosenberg[1], Robyn E. Tanny[1], Tyler C. Shimko[1], and Erik C. Andersen[1,2]

[1] Molecular Biosciences, Northwestern University, Evanston, IL, United States,
[2] Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL, United States

Abstract

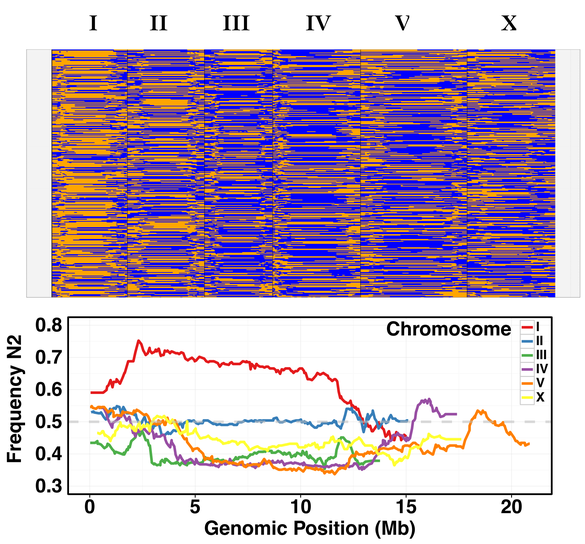
Identifying the genetic determinants of complex traits depends on our ability to gather accurate phenotype data and to correlate these data to genetic differences present in a population. Furthermore, a large population of individuals is required to identify genomic regions that have subtle contributions to overall phenotypic differences. To uncover genomic loci that control complex traits, we have developed a high-throughput and high-accuracy phenotyping pipeline and generated a panel of 1,200 recombinant inbred *Caenorhabditis elegans* lines. The recombinant inbred collection was derived from a modified version of the standard laboratory strain N2 and the highly diverged wild strain CB4856. To abrogate the confounding pleiotropic effects of the laboratory-adapted N2 *npr-1* allele, we replaced this allele with the CB4856 allele. Additionally, the N2 *peel-1* gene was disrupted to eliminate the incompatibility locus. Our phenotyping pipeline relies on the COPAS BIOSORT large particle nematode sorter to accurately quantify various fitness related traits in a population of individuals, including animal size, fecundity, optical density, and pharyngeal pumping rate. Using this pipeline, we are able to phenotype 96 recombinant lines exposed to 24 different conditions in one week. Altogether, we exposed 359 recombinant inbred lines to 70 different conditions. This massive effort led to the identification of 550 unique quantitative trait loci (QTL) that explain phenotypic variation in response to 46 environmental perturbations, including chemotherapeutics, anthelmintics, heavy metals, pesticides, and different food sources and temperatures. We observe both complex and simple genetic architectures in response to different perturbations. For example, variation in response to the anthelmintic abamectin maps to six QTL that together account for 90% of the phenotypic variance that can be attributed to genetic factors. By contrast, a single QTL on chromosome V explains 55% of the phenotypic variation in response to the chemotherapeutic bleomycin. The confidence interval surrounding this peak contains five candidate genes, one of which is known to be involved in the DNA damage response. Though we have identified approximately 10 QTL with strikingly large effects, we expect to increase the number of QTL substantially using the full recombinant inbred line collection. This approach will expedite the discovery of causal genes contributing to phenotypic variation in response to therapeutics that can increase the likelihood that one day we will have individualized medicine.

High-throughput Fitness Quantification



We use the COPAS BIOSORT to measure fitness traits. The worm sorter simultaneously measures the length, optical density, and three fluorescent channels (Red, Green, and Yellow) of individual worms. The phenotyping platform is depicted to the left. **A)** Depicts the correlation between optical density and animal length. **B)** Depicts the lack of correlation between fluorescence, from ingested microspheres, and animal size.

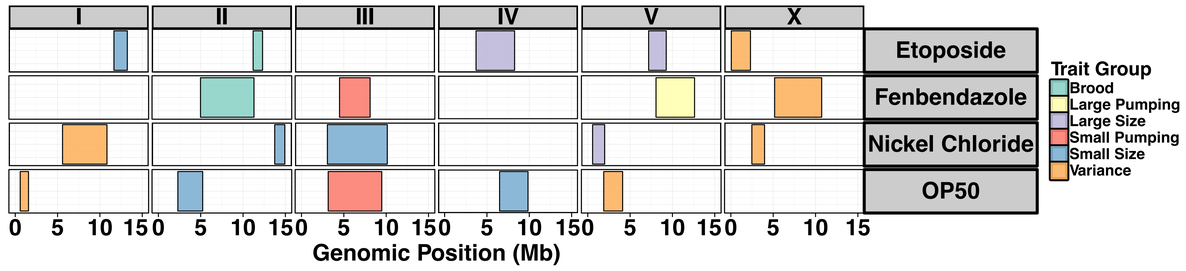
Recombinant Inbred Collection



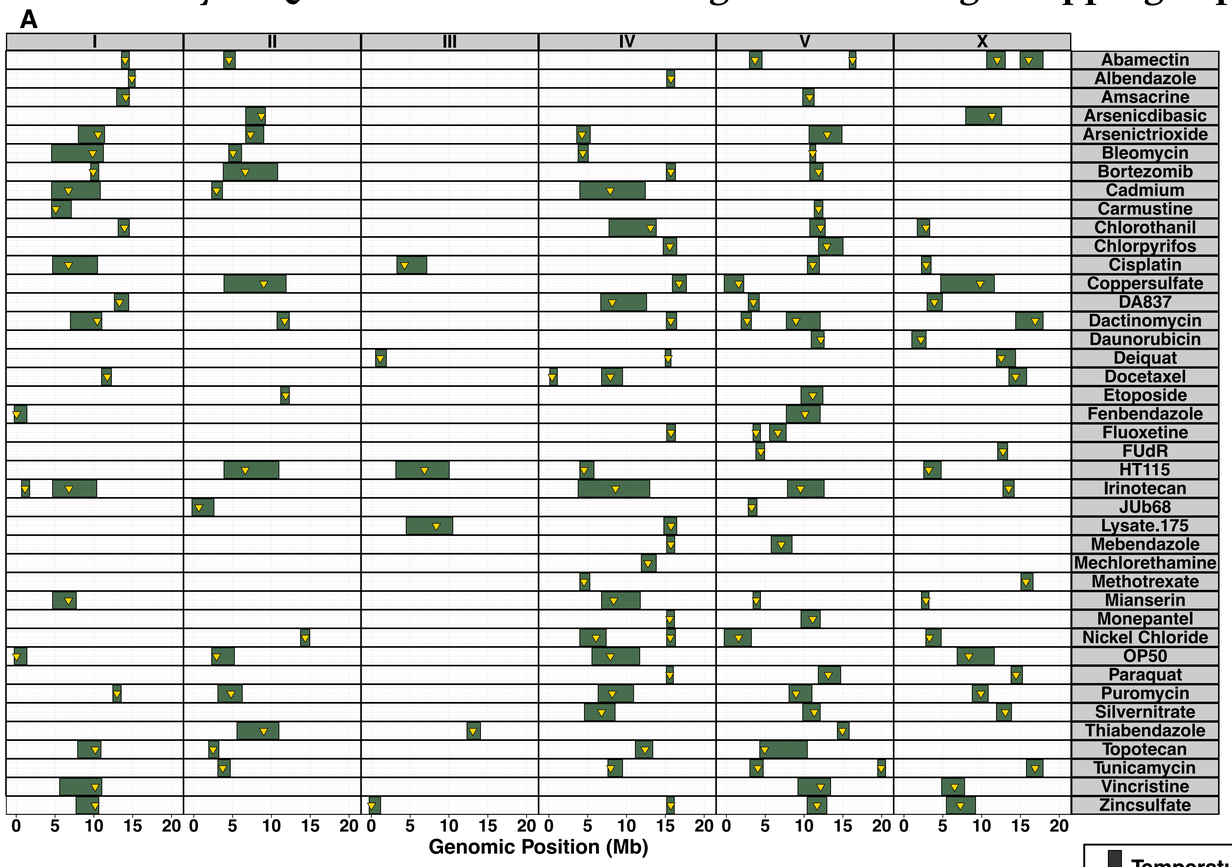
Our collection of 359 recombinant inbred advanced intercross lines (RIAILs) generated between N2 and CB4856. Each horizontal line in **A** corresponds to one RIAIL (N2, orange, CB4856, blue). These lines were genotyped at 1454 markers across the genome. Figure **B** shows the fraction of recombinant lines with the N2 genotype at every sequenced position in the genome. We are currently in the process of acquiring whole-genome sequence data for these lines and an additional 650 recombinant lines we generated last year.

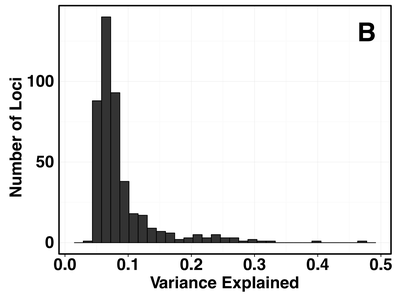
Distribution Parameters as Quantitative Fitness Traits

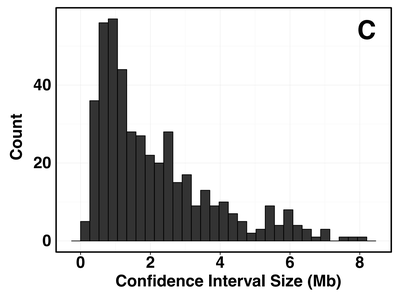
We transform raw distribution data from one strain into a variety of quantitative traits, including distribution parameters quantities, variance, median. These different traits map to distinct regions of the genome as indicated in the figure to the right. Each rectangle corresponds to a trait confidence interval (1.5 LOD drop) and colors correspond to trait class. We scored 41 drugs and four representative examples are shown to the right.



Summary of QTL identified in our Large-scale Linkage Mapping Experiment

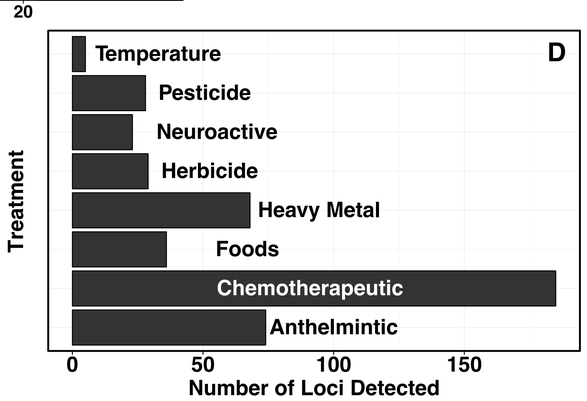


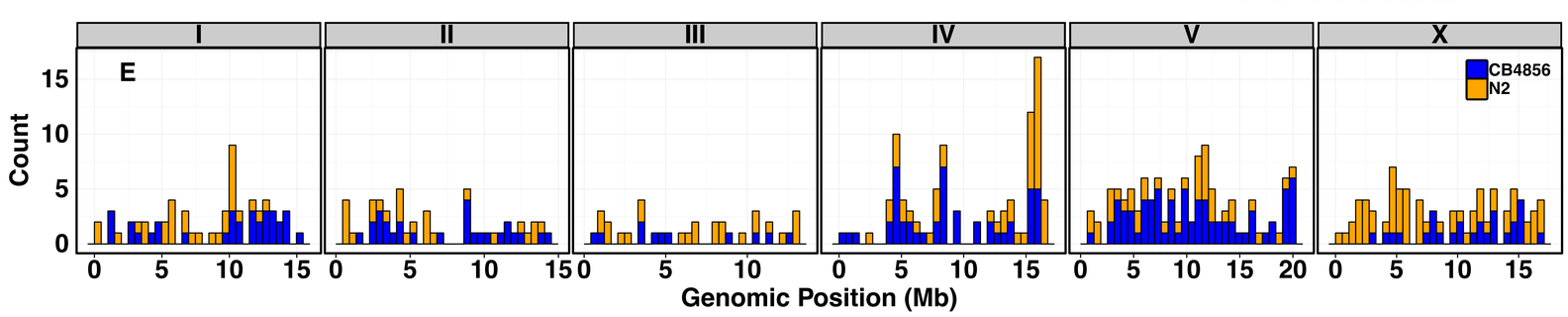




A) Summary of all identified QTL from linkage mapping. Green rectangles correspond to 1.5 LOD drop confidence intervals and yellow triangles represent the marker with the highest significance value in the confidence interval. The trait with the highest combined phenotypic variance explained is plotted for each experimental condition indicated on the right panels. B) Distribution of the variance explained for every detected QTL. C) Distribution of confidence interval sizes for every detected QTL. D) Number of QTL detected for each eight treatment classes. E) Histogram showing the distribution of all detected QTL across the genome. Colors represent the direction of effect, with orange indicating N2 having a relatively higher phenotypic values.

The focus of our lab is currently on narrowing these confidence intervals by generating nearly isogenic lines and other congenial approaches. The ultimate goal of this work is to identify the genetic variants in these intervals that are contributing to phenotypic variation in the population and determine the altered molecular mechanisms associated with the variants.





Funding

IBiS Travel Award

CMBD Training Grant