

## **CaeNDR, the *Caenorhabditis* Natural Diversity Resource**

Timothy A. Crombie<sup>1,2</sup>, Ryan McKeown<sup>2,3</sup>, Nicolas D. Moya<sup>4,5</sup>, Kathryn S. Evans<sup>2</sup>, Samuel J. Widmayer<sup>2</sup>, Vincent LaGrassa<sup>6</sup>, Natalie Roman<sup>6</sup>, Orzu Tursunova<sup>6</sup>, Gaotian Zhang<sup>2</sup>, Sophia B. Gibson<sup>2</sup>, Claire M. Buchanan<sup>2</sup>, Nicole M. Roberto<sup>2</sup>, Rodolfo Vieira<sup>6</sup>, Robyn E. Tanny<sup>4</sup>, and Erik C. Andersen<sup>4\*</sup>

1. Department of Biomedical and Chemical Engineering and Sciences, Florida Institute of Technology, Melbourne, FL, USA
2. Department of Molecular Biosciences, Northwestern University, Evanston, IL USA
3. Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL USA
4. Department of Biology, Johns Hopkins University, Baltimore, MD, USA
5. Cell, Molecular, Developmental biology, and Biophysics Graduate Program, Johns Hopkins University, Baltimore, MD, USA
6. Northwestern University Information Technology, Media and Technology Innovation, Northwestern University, Evanston, IL USA

\* Corresponding author

### **Erik C. Andersen**

Professor of Biology  
Johns Hopkins University  
Department of Biology, 383 UTL  
3400 N. Charles St.  
Baltimore, MD 21218  
Email: erik.andersen@gmail.com

**Journal:** [Nucleic Acids Research](#)

**Running title:** The *Caenorhabditis* Natural Diversity Resource

**Keywords:** *Caenorhabditis*, natural diversity, wild isolates, genome-wide association, functional variation

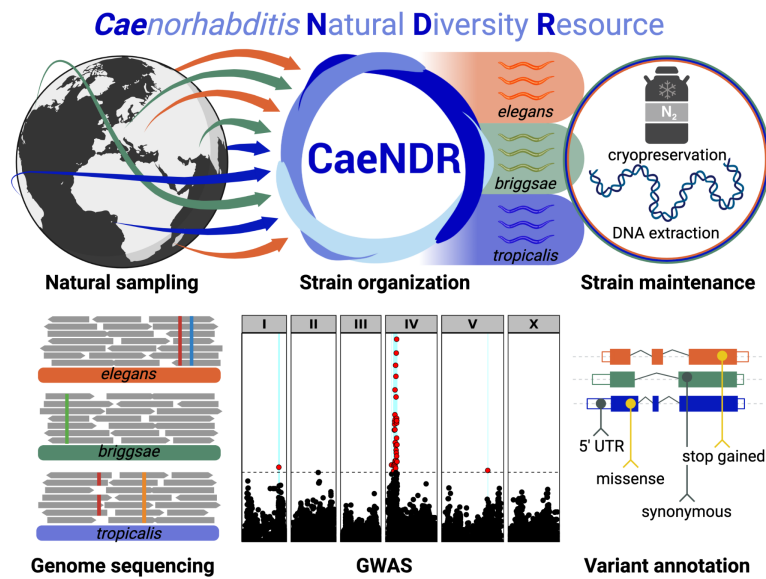
### **Key points:**

1. CaeNDR serves as a centralized repository for all wild selfing *Caenorhabditis* strains.
2. CaeNDR provides genome-wide sequence and variant data for all selfing *Caenorhabditis* strains.
3. CaeNDR enables comparative quantitative genetics across selfing *Caenorhabditis* species.

### **Emails and ORCIDs:**

Timothy A. Crombie, tim.crombie@gmail.com, ORCID 0000-0002-5645-4154  
Ryan McKeown, ryanmckeown2021@u.northwestern.edu, 0000-0003-3570-2494  
Nicolas D. Moya, nicolasmoya2024@u.northwestern.edu, 0000-0002-6817-1784  
Kathryn S. Evans, katiasevans9@gmail.com, 0000-0002-1388-8155  
Sophia B. Gibson, sophiabg@uw.edu, 0000-0001-9839-9045  
Sam Widmayer, samuel.widmayer@jax.org, 0000-0002-1200-4768  
Gaotian Zhang, gaotian52@gmail.com, 0000-0001-6468-1341  
Robyn E. Tanny, rtanny1@jhu.edu, 0000-0002-0611-3909  
Claire M. Buchanan, claire.buchanan.920@gmail.com, 0000-0003-3725-3469  
Nicole M. Roberto, nmroberto@sbcglobal.net, ORCID 0000-0002-8339-0627  
Vincent LaGrassa, vincent.lagrassa@northwestern.edu  
Orzu Tursunova, orzu.tursunova@northwestern.edu  
Natalie Roman, natalie.roman@northwestern.edu  
Rodolfo Viera, rodolfo@northwestern.edu  
Erik Andersen, erik.andersen@gmail.com, ORCID 0000-0003-0229-9651

## Graphical Abstract



## Abstract

Studies of model organisms have provided important insights into how natural genetic differences shape trait variation. These discoveries are driven by the growing availability of genomes and the expansive experimental toolkits afforded to researchers using these species. For example, *Caenorhabditis elegans* is increasingly being used to identify and measure the effects of natural genetic variants on traits using quantitative genetics. Since 2016, the *C. elegans* Natural Diversity Resource (CeNDR) has facilitated many of these studies by providing an archive of wild strains, genome-wide sequence and variant data for each strain, and a genome-wide association (GWA) mapping portal for the *C. elegans* community. Here, we present an updated platform, the *Caenorhabditis* Natural Diversity Resource (CaeNDR), that enables quantitative genetics and genomics studies across the three selfing *Caenorhabditis* species: *C. elegans*, *C. briggsae*, and *C. tropicalis*. The CaeNDR platform hosts several databases that are continually updated by the addition of new strains, whole-genome sequence data, and annotated variants. Additionally, CaeNDR provides new interactive tools to explore natural variation and enable GWA mappings. All CaeNDR data and tools are accessible through a freely available web portal located at [caendr.org](http://caendr.org).

## Introduction

*Caenorhabditis elegans* is regarded as a powerful model because of its experimental tractability, which has facilitated major contributions to the fields of developmental, cellular, and neurobiology. *C. elegans* is a small (~1 mm), free-living nematode that has a near-global distribution and is most often found in association with decomposing plant material (1). Several attributes make *C. elegans* an exceptional animal model. First, the species has a short generation time of approximately four days, wherein an individual produces hundreds of offspring (2). Its genome is highly compact (~100 Mb), extensively annotated, and amenable to genome editing (3, 4). The body is transparent, which enables direct observation of developmental processes and fluorescent reporter constructs (5). Additionally, because *C. elegans* is a facultative outcrosser (males are rare and self-fertile hermaphrodites are the norm), genotypes can be recombined easily using controlled crosses and homogenized by repeated selfing. Moreover, strains can be cryopreserved indefinitely (6), which prevents the accumulation of novel mutations and ensures the fidelity of genomes over time. Finally, the other selfing *Caenorhabditis* species, *C. briggsae* and *C. tropicalis*, share the advantageous attributes of *C. elegans* and enable comparative studies within the genus.

Comparative studies have informed our understanding of several aspects of *Caenorhabditis* biology, including the evolution of self-fertilization (7, 8), the regulation of cell fates involved in vulva development (9–11), the thermotolerance and thermoplasticity of germline development (12), the developmental determinants of sperm size (13), the regulation of dauer formation (14), pheromone signaling and the biosynthesis of pheromones (15), the mutational process (16, 17), the abundance and distribution of selfish genetic elements such as transposable elements and toxin-antidote elements (18–21), and the resistance to anthelmintics (22). These investigations have also revealed the changes in overall genome structure, gene structure, and protein-coding gene content that are associated with the transition to selfing (23–26), and the strikingly high rates of intrachromosomal rearrangements in selfing species relative to their outcrossing relatives (27–30).

Despite this rich history of comparative studies, many of the insights from this work are based on just a small number of strains, or even a single strain, within each species. The most common strains used in comparative studies are N2 for *C. elegans*, AF16 for *C. briggsae*, and JU1373 for *C. tropicalis*. Although the

extensive focus on these particular strains has helped the community standardize experimental procedures and infer biology from the synthesis of independent studies, it also limits our understanding of natural diversity within these species. To help address this limitation, researchers and citizen scientists have isolated many wild *Caenorhabditis* strains from around the globe (31–43). These wild strains represent an enormous reservoir of genetic diversity that can be leveraged to understand the genetic basis of evolutionary processes and trait variation using genome-wide association studies (GWAS). The GWAS approach correlates genotypic variation with phenotypic differences across a set of strains to identify regions of the genome that underlie differences in a trait of interest. In turn, the identified regions, known as quantitative trait loci (QTL), can be dissected to reveal the precise identity of the genetic variants influencing a trait.

In recent years, the genetic and experimental tractability of *C. elegans* has helped it become a leading model for the discovery of genes and variants that underlie quantitative trait variation (44–46). In 2016, we launched the *C. elegans* Natural Diversity Resource (CeNDR) to serve as a public repository for wild *C. elegans* strains, their genome sequences, and variant data (47). Since CeNDR was launched, the platform has facilitated approximately 150 studies of natural variation, including the investigation of global population structure (39, 41, 42), the evolutionary origins of hyper-divergent regions within the genome (41), and the discovery of variants that influence ecologically and biomedically relevant traits using quantitative genetic mappings (35, 37, 40, 48–54). These discoveries highlight the utility of CeNDR as a centralized repository of genetic variation for the *C. elegans* species.

Here, we present the *Caenorhabditis* Natural Diversity Resource (CaeNDR), which enables comparative studies of natural diversity by including 1,680 *C. briggsae* and 681 *C. tropicalis* wild strains. The new CaeNDR platform provides three major functions. First, CaeNDR serves as a centralized hub for the collection, maintenance, and distribution of selfing *Caenorhabditis* strains isolated from nature. Second, CaeNDR leverages recently improved reference genomes for *C. briggsae* and *C. tropicalis* (21, 29, 55) to provide whole-genome sequence and variant data for strains in the archive. Third, CaeNDR offers web-based tools to help researchers design, perform, and analyze GWAS using an augmented version of the NemaScan GWAS pipeline (56). In the following sections, we describe how the major functions of CaeNDR are

implemented and explain how the online tools support these functions. The CaenDR platform can be found at [caendr.org](https://caendr.org).

## Implementation

### *A centralized repository for wild strains*

CaenDR provides the *Caenorhabditis* research community with a consolidated collection of wild strains that is continuously updated and rigorously maintained to ensure strain fidelity. We designed the new CaenDR platform to streamline the submission and acquisition of wild strains. Two options are available for community members that wish to submit strains to CaenDR. Research groups can contribute strains by completing the Submit a Strain web form or the batch strain submission form, then mailing their strains to CaenDR. Alternatively, citizen scientists can contribute wild strains by following the collection and processing guidelines at [caendr.org/get-involved/citizen-scientists](https://caendr.org/get-involved/citizen-scientists). We aim to preserve the ecological data associated with the wild strains, so we request that contributors include the time, location, photos, and descriptions of the landscape and substrate for each strain submission (57). These data, along with attributions for the individuals that sampled and isolated the strains, are added to the CaenDR database and can be browsed with an interactive map on the website home page. The complete strain dataset is available for download at [caendr.org/data/data-release](https://caendr.org/data/data-release).

Researchers can procure wild strains from CaenDR individually, or in bulk, using the Request Strains page ([caendr.org/request-strains/](https://caendr.org/request-strains/)). Additionally, CaenDR offers pre-selected sets of strains to facilitate GWAS in each species. These sets include a Divergent Set and several Mapping Sets. The Divergent Set contains 12 strains that are chosen to represent the genetic diversity across a particular species and can be used to assess any heritable trait variation in that subset of strains. If genetic differences among the divergent strains are shown to contribute strongly to trait variation, a GWAS using trait data from mapping sets will likely identify QTL for that trait (46). Each Mapping Set contains 48 strains. Because the statistical power to detect QTL with GWAS increases with the number of strains that are measured, we recommend using at least 96 strains (56).

CaenDR maintains and regularly updates whole-genome sequence and variant data for strains in the archive. Because selfing *Caenorhabditis* species reproduce predominantly by self-fertilization in the wild, nematodes isolated from the same substrate are often nearly identical, with only a small number of differences attributable to recent, lineage-specific mutations (33, 38, 39, 42, 58, 59). Therefore, we examine the rates of shared variation among all strains and group strains that are nearly identical into genome-wide haplotypes called isotypes (Supplemental Figure 1). The aligned sequence and variant data for every wild strain that we have sequenced are freely downloadable for each data release. Because strains within isotypes share nearly all variation, their inclusion in a GWAS can lead to false association signals (60). Therefore, the majority of CaenDR tools, including GWAS, are focused on exploring genetic variation among isotypes. We achieve this focus by choosing a single representative strain for each isotype group, called the isotype reference strain. These strains are selected by choosing the strain from the isotype group with the most sequence coverage. For each species, the single-nucleotide variants (SNVs) and short insertions and deletions (indels) for all isotype reference strains can be downloaded as Variant Call Format (VCF) files (61). CaenDR also provides additional genomic datasets relevant to the study of natural variation. For example, each of the selfing *Caenorhabditis* species carry punctuated regions of extreme sequence divergence within their genomes that are called hyper-divergent regions (21, 41, 62). Because these hyper-divergent regions can influence the performance of GWAS and complicate QTL dissection, the physical positions of hyper-divergent regions for all isotype reference strains within *C. elegans* are available for download as Browser Extensible Data (BED) files. In the near future, we will add these datasets for *C. briggsae* and *C. tropicalis*. Importantly, the hyper-divergent regions are identified by extreme sequence divergence relative to the reference genomes for each species, *i.e.*, N2 for *C. elegans*, QX1410 for *C. briggsae*, and NIC58 for *C. tropicalis*. Therefore, the reference strains will not contain hyper-divergent regions because of reference bias (21, 29). The software and methods used to produce all the genomic datasets are available at the Data Release page ([caendr.org/data/data-release](https://caendr.org/data/data-release)).

The Genome Browser tool ([caendr.org/tools/genome-browser/](https://caendr.org/tools/genome-browser/)) enables users to visualize the genomic datasets from any of the three species. For example, a user can query a gene or region of interest to visualize genetic variation across all isotype reference strains and inspect the aligned sequencing reads supporting

variant calls at that locus. The browser tool uses an embedded JavaScript implementation of the Integrative Genomics Viewer (IGV) (63), which can be configured to show a number of genomic datasets as tracks. By default, the browser displays gene and transcript isoform tracks (Figure 1A-B). Two optional tracks display hyper-divergent regions, either as a species-wide summary, or for particular isotype reference strains (Figure 1C-D). The optional VCF track displays high-quality SNVs and indels identified in each isotype reference strain in blue overlaid onto the species-wide variant set in gray (Figure 1E). The optional BAM track shows coverage statistics and paired-end read alignments that give context for how the variants within an isotype reference strain are called (Figure 1F).



### Figure 1 - The CaenDR genome browser

The user specifies the Species in the setup panel to the left. The optional Gene Search field will direct the browser to that gene in the genome (A) and display transcripts in the track below (B). The Divergent Regions Summary track shows regions of the genome that are classified as hyper-divergent relative to the reference genome in any isotype reference strain within the species (C). The Divergent Regions track displays regions that are hyper-divergent relative to the reference genome for individual isotype reference strains, each track feature can be clicked to reveal the isotype reference strain the feature belongs to (D). The Isotype field in the setup window on the left is optional and allows users to add VCF and BAM tracks to the browser. The VCF track shows species-wide SNV and short indels in gray, with the variants specific to the isotype reference strain in blue (E). The BAM track shows read level data from which the variants were called. The top of the track shows read depth and the bottom shows the paired-end reads (F).

To investigate a region further, researchers can use the Variant Annotation tool ([caendr.org/tools/variant-annotation](https://caendr.org/tools/variant-annotation)), which shows how variants within coding sequences are distributed across the population and supplies predictions for how the variants might impact protein function. The annotation database includes the physical positions of the variants, gene names, names of strains sharing the variants, predicted transcript consequences and protein impacts, among other annotations. Users can filter the database by strain, gene, genetic interval, and functional impact class, then browse the filtered database or download it as a Comma-Separated Values (CSV) file for further processing. Importantly, CaeNDR now uses the BCFtools/csq program to predict functional consequences and impacts for variants by considering nearby, *cis*-variants jointly rather than singly when classifying variant consequences (64). To assign variant impacts within transcripts, we group consequences that include changes to amino acids, start and stop codon positions, or splice variants as “HIGH” and all others as “LOW” impact. Beyond this binary impact score, the variant annotation database also includes BLOSUM and Grantham scores, which can help evaluate the potential impact of amino acid substitutions in coding sequences with more nuance (65, 66). These scores measure how conserved or radical a particular amino acid substitution is from an evolutionary perspective. The BLOSUM scale ranges from -4 to +11 and the Grantham scale ranges from 5 to 215. Low BLOSUM and high Grantham scores indicate radical amino acid substitutions that are more likely to impact protein function.

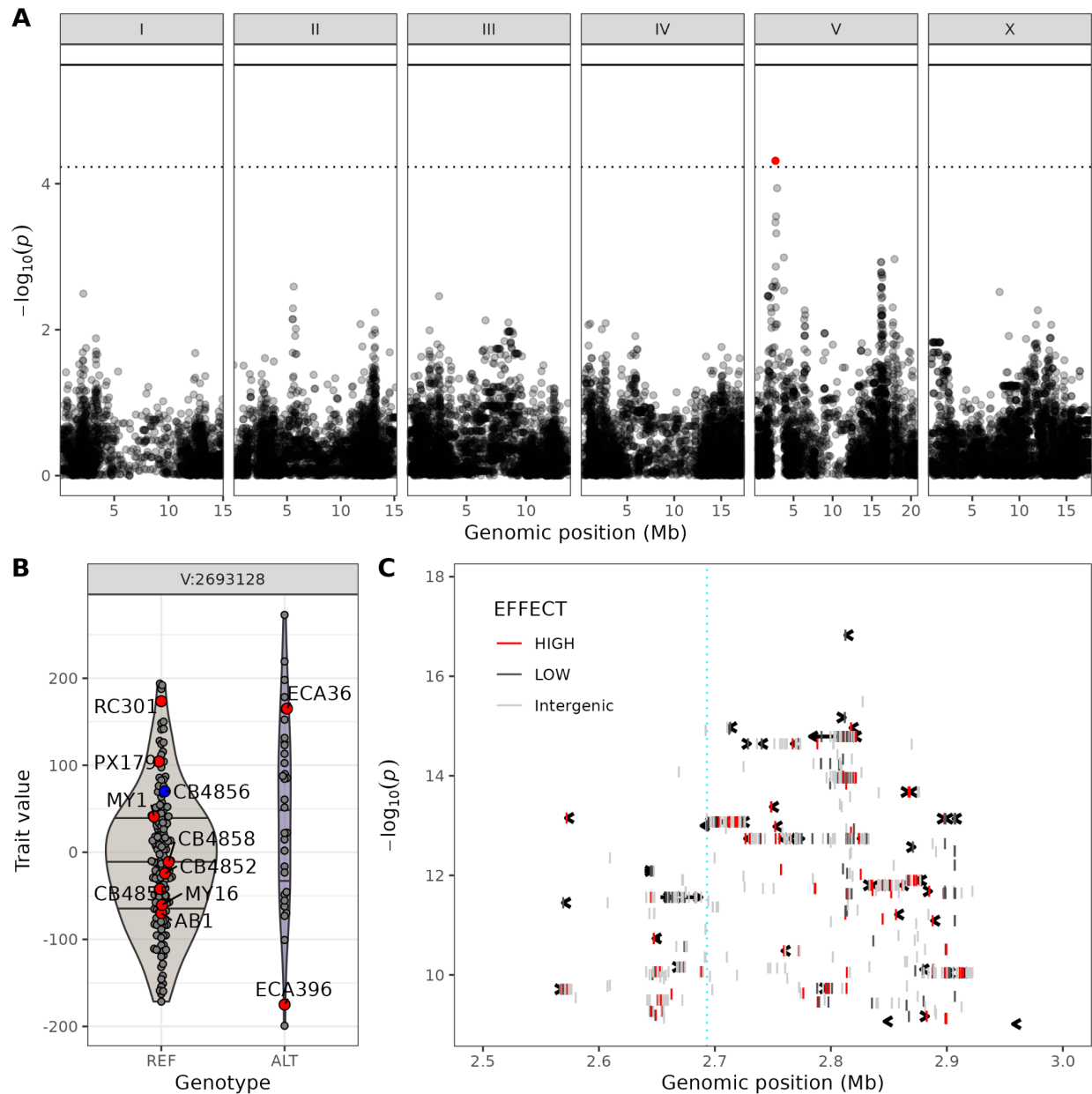
#### *A Caenorhabditis GWAS toolbox*

CaeNDR hosts several web-based tools to help researchers design, perform, and analyze GWAS. A typical GWAS begins with a pilot study to measure a trait of interest across a small number of divergent strains in independent replicates. Researchers then calculate the proportion of trait variance explained by genetic differences among those strains, which is called broad-sense heritability ( $H^2$ ). In the context of GWAS, the  $H^2$  of a trait is useful for determining whether a particular GWAS design will be sufficiently powered to identify QTL. For example, based on simulations in *C. elegans*, 500 strains should be measured in a GWAS to achieve approximately 70% power for detecting a single QTL that underlies a trait with a  $H^2$  of 0.2 (56). The Heritability Calculator tool ([caendr.org/tools/heritability-calculator](https://caendr.org/tools/heritability-calculator)) enables users to quickly estimate  $H^2$  by submitting data from a pilot experiment as a Tab-Separated Values (TSV) file. The tool reports a  $H^2$  estimate along with a plot



to help visualize the effect of genetic differences among the strains on the trait. Additionally, the narrow-sense heritability ( $h^2$ ) is reported, which measures the portion of trait variance explained by additive genetic variance among the strains and is more representative of the heritability that is relevant to GWAS. The heritability report helps users plan the scale of their GWAS and procure the required strains using the Request Strains page. After measuring a trait of interest in a GWAS strain set, mappings can be conducted within CaenDR's GWAS mapping portal.

The Genetic Mapping tool ([caendr.org/tools/genetic-mapping](https://caendr.org/tools/genetic-mapping)) allows users to perform a GWAS by uploading a TSV file of trait measurements for isotype reference strains. The tool runs a containerized version ([hub.docker.com/r/andersenlab/nemascan-nxf](https://hub.docker.com/r/andersenlab/nemascan-nxf)) of the NemaScan mapping pipeline (56) to process the trait data and test for correlations between genotype and phenotype at over 50,000 marker loci (bi-allelic SNVs) across the *C. elegans*, *C. briggsae*, or *C. tropicalis* genomes. The tool produces a detailed mapping report that can be downloaded as an HTML file. The report contains plots and interactive data tables to help users analyze their QTL (Figure 2). Because long-range linkage disequilibrium (LD) is strong in selfing *Caenorhabditis* species, NemaScan collapses nearby markers with significant correlations into a single QTL (56). Consequently, most QTL identified by NemaScan will contain hundreds of variant sites among the strains in the mapping that could contribute to differences in the trait. Because GWAS is a correlative approach and many variant sites might exist within QTL, each QTL must be dissected to validate whether specific variants actually cause differences in the trait. Several methods are used to dissect QTL for *Caenorhabditis* (46)



**Figure 2 - Elements of the GWAS mapping report from CaenDR.**

The phenotype data used to generate these CaenDR report figures were simulated using the NemaScan simulation profile. **(A)** A Manhattan plot is shown to visualize the significance values for all markers used in the statistical test of association between genotype and phenotype. The y-axis is the negative base 10 log of the  $p$ -value obtained from the statistical test of association. The x-axis is the genomic position in millions of base pairs and is faceted by chromosome. Markers (SNVs) with a  $-\log_{10} p$ -value greater than the Bonferroni-corrected significance threshold (solid line) or the genome-wide eigendecomposition significance threshold (dotted line) are significantly correlated with the phenotype and are colored red. Significant markers denote QTL and indicate that genetic variation linked with the marker could cause differences in the phenotype. **(B)** Violin plots are shown for the most significant marker in the QTL on chromosome V. The plot shows the simulated trait values for the strains with the reference (REF) allele compared to the strains with the alternative (ALT) allele for that marker. The horizontal lines indicate the 75th percentile, median, and 25th percentile of the data. The trait values for some strains are highlighted and labeled because these strains are commonly used to measure dose responses by the community (67–70). A difference in the median trait between the two genotypes is expected for significant markers, and the genotypes of strains at this genomic

position can help users plan follow-up experiments to validate the effect of QTL on the trait. (C) A fine-mapping plot is shown to visualize the significance of all bi-allelic SNVs in the QTL interval on chromosome V. The axes are identical to the Manhattan plot above, with SNVs represented by a vertical line and colored by the predicted variant impact on genes: red = HIGH, black = LOW, gray = INTERGENIC. Genes are represented by black arrows showing the direction of the gene and are positioned on the y-axis based on the maximum  $-\log_{10} p$ -value of all variants in the gene.

A common method used to narrow and validate QTL is the creation and measurement of near-isogenic lines (NILs) (71). A NIL is a strain constructed using genetic crosses that carries a small region from one strain (the donor) introgressed into the genome of another strain (the recurrent parent) (46). Oftentimes, trait differences between the NIL and the recurrent parent are caused by an allelic difference within the introgressed region, validating the effect of the QTL (46). Many quantitative mapping studies in *Caenorhabditis* have successfully leveraged NIL panels to validate and narrow QTL intervals so they contain a small number of potentially causal genes, which were then explicitly tested using CRISPR-Cas9 genome-editing techniques (45). The CaeNDR platform simplifies the construction of NIL panels by providing the Pairwise Indel Finder tool ([caendr.org/tools/pairwise-indel-finder](https://caendr.org/tools/pairwise-indel-finder)) to help users accurately track the introgression of the donor region into the recurrent parent using small indels unique to either genetic background. To use the tool, researchers enter a species, a pair of strains, and a specific genomic region. The tool outputs indel variants between 50 and 500 bp that are unique to each strain along with PCR primers that researchers can use to screen progeny for the desired introgression. Additionally, the IGV view supplied by the tool contains a hyper-divergent region track to help avoid indels in hyper-divergent regions because these calls are less reliable and primer searches are more error-prone.

## Applications

### *Functional studies of natural variation*

Traditional studies of gene function often use loss-of-function or gain-of-function alleles created by mutagenesis to investigate how these changes influence traits (72). However, these mutant alleles might perturb gene function to such an extreme that they prevent researchers from observing traits of interest. To address this challenge, researchers can query genes of interest with the Variant Annotation tool ([caendr.org/tools/variant-annotation](https://caendr.org/tools/variant-annotation)) to see all of the natural variants, along with any predicted

coding-sequence impacts, in a particular genomic region. Because natural variation has vetted these alleles, they are less likely to be highly deleterious and more likely to be ecologically and physiologically relevant. Researchers can then evaluate the influence of these natural variants on traits by introducing them into a genetic background of choice, e.g., the *C. elegans* laboratory-adapted strain N2, using outcrossing or genome editing (46).

### *Population genomics and adaptive variation*

The CaeNDR platform gives researchers the genomic data required to explore how patterns of genetic variation within the three selfing *Caenorhabditis* species have been shaped and maintained by evolutionary processes. For example, an early study of genetic diversity in *C. elegans* identified chromosome-scale selective sweeps that purged genetic diversity in large portions of the global population (32). In agreement with evolutionary theory, the high rate of self-fertilization in *C. elegans* (33, 38, 58, 73) likely exacerbated the effects of selective sweeps (32). However, recent analyses of more geographically representative samples revealed additional structure within the species and relatively high genetic diversity within strains across the Pacific Rim and Hawaiian Islands (39, 41, 42). Comparatively, *C. briggsae* populations are structured into distinct phylogeographic groups and harbor greater genetic diversity than *C. elegans*; the two primary phylogeographic groups, Temperate and Tropical, are divided along latitudinal lines (62, 74–77), which is consistent with differences in thermal tolerance and fitness traits measured in wild strains (12, 78). Among the three species, *C. tropicalis* harbors the lowest levels of genetic diversity (79) and appears to be restricted to tropical regions with strong continental population structure and local heterogeneity (21).

These *Caenorhabditis* population genomic studies are useful for several reasons. First, they provide a lens through which researchers can observe how functional variation is shaped and maintained by the effects of mutation, selection, recombination, genetic drift, and demography. Beyond contextualizing the adaptive qualities of functional variation, population genomic studies quantify the genetic relatedness among individuals and characterize population structure, which can influence GWAS performance (80). For instance, in *C. elegans*, the optimal GWAS mapping panel is dependent on trade-offs between the amount of natural variation surveyed, the statistical power to detect QTL, and the false discovery rate (56). Therefore, as

CaeNDR continues to grow, updated population genomic studies will improve the ability to dissect complex traits using quantitative genetics approaches.

### *Identifying genotype-phenotype correlations*

The primary goal of CaeNDR is to facilitate the identification of genes and genetic variants that drive phenotypic differences in natural populations. CaeNDR's GWAS mapping tool outputs a detailed report to help users achieve this goal. The report includes Manhattan plots that show the statistical significance of correlations between genetic markers and the variation in the trait of interest across the genome (Figure 2A). Additionally, plots are included to assess the strength of the correlation between genotype and phenotype at the most significant markers in the mapping (Figure 2B). Because most selfing *Caenorhabditis* species have long-range LD, even among chromosomes (62, 81, 82), the identification of multiple QTL on a single chromosome or even on separate chromosomes could be caused solely by LD. For this reason, the mapping reports show the LD between the most significant markers from each QTL. Another source of spurious associations is population structure among the strains in the mapping population. To help control for false positive associations in the presence of population structure, the NemaScan pipeline uses the resource-efficient linear mixed model approach called fastGWA, which is implemented in the GCTA program (60, 83–85). However, strong population structure might still cause spurious associations in GWAS, so we include measures of genomic inflation for users to check for systemic biases in their results. Additionally, users are provided with fine-mapping plots (Figure 2C) and interactive data tables that are useful for narrowing QTL to specific genes and variants that can underlie trait differences. For example, variants that (1) fall within genes, (2) are strongly associated with trait differences, and (3) are predicted to have high impacts on gene function are good candidates to be tested for their influence on the trait.

### **Usage and impact**

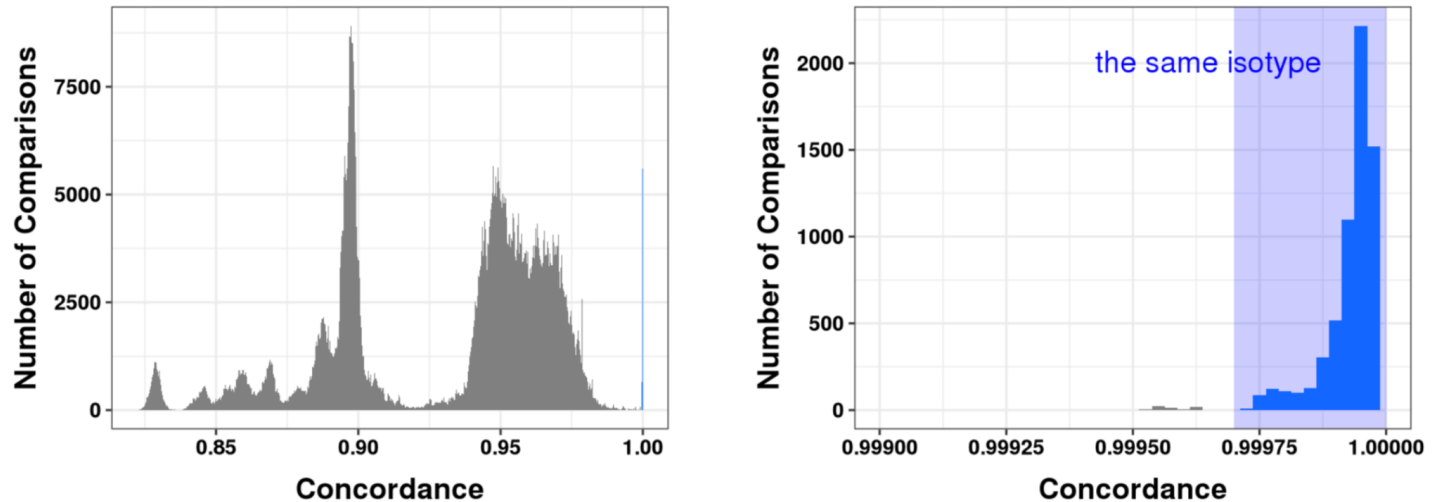
The CeNDR platform (now CaeNDR) has made a significant scientific impact since its inception seven years ago by supporting nearly 150 publications exploring natural variation. Moreover, the current engagement with CaeNDR is robust, with an average of 490 new visitors per month (average calculated from April 2023 to July 2023) and over 800 registered users in total (July 2023). Registration is free and requires only a valid email

address. After registering, users can procure stains, run all tools on the site, and easily retrieve data and reports from their personalized results pages.

### **Future Directions**

CaeNDR is updated regularly to incorporate newly isolated wild strains and their genomes. Furthermore, the variant calling pipelines used by CaeNDR will be expanded to incorporate all major classes of structural variants, including deletions, duplications, insertions, inversions, and translocations. We also plan to release gene expression data for isotype reference strains, which will enable mediation analyses (86). Additionally, we will continue to release new tools to help mine and explore natural variation across the three species. For example, we are currently developing a phenotype database to store and release trait data that users make public. These data will help the community explore correlations among traits and perform comparative studies. We are also preparing a tool to expedite CRISPR-Cas9 genome editing in wild strains. The tool will output strain specific guide RNA sequences to target edits at user supplied loci, which will simplify the process of testing natural variants for functional effects on *Caenorhabditis* traits.

## Supplemental Materials



### Supplemental Figure 1 - Strain pairwise concordance histogram for *C. elegans*

Concordance analysis is used to group strains that are genetically similar into isotypes. Concordance values for every pair of strains are calculated as the number of shared variant sites divided by the total number of unique variants called within the pair. A strain pair is grouped into the same isotype if the concordance score is greater than 0.9997. The distribution of pairwise concordance values above 0.8 are shown on the left. The plot on the right shows the pairwise concordance values above 0.999. For both histograms the concordance values above 0.9997 are in blue.

## **Data Availability Statement**

The data underlying this article are available in the GitHub repository [NAR\_CaeNDR\_manuscript], at [https://github.com/AndersenLab/NAR\\_CaeNDR\\_manuscript](https://github.com/AndersenLab/NAR_CaeNDR_manuscript).

## **Acknowledgements**

We would like to thank members of the Andersen laboratory for their helpful comments on this manuscript. We also thank the CaeNDR scientific advisory committee: Bhagwati Gupta, Yen-Ping Hsueh, Matthew Rockman, Ann Rougvie, Jessica Sowa, and Paul Sternberg; and the major CaeNDR collaborators (and funding for nematode collections): Michael Allion, Christian Braendle, Marie-Anne Félix (Centre National de la Recherche Scientifique), Matthew Rockman (GM141906) Hinrich Schulenberg (Collaborative Research Center 1182 on the Origin and Function of Metaorganisms, project A1), Mark Sterken, and John Wang (National Science and Technology Council, Taiwan (NSTC 111-2311-B-001-036-MY3) and Academia Sinica).

## **Funding**

This work was supported by the National Science Foundation Capacity: Biological Collections grant (2224885) awarded to R.E.T. and E.C.A.



## References

1. Frézal,L. and Félix,M.-A. (2015) *C. elegans* outside the Petri dish. *Elife*, **4**.
2. Zhang,G., Mostad,J.D. and Andersen,E.C. (2021) Natural variation in fecundity is correlated with species-wide levels of divergence in *Caenorhabditis elegans*. *G3* , **11**.
3. *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, **282**, 2012–2018.
4. Dickinson,D.J. and Goldstein,B. (2016) CRISPR-Based Methods for *Caenorhabditis elegans* Genome Engineering. *Genetics*, **202**, 885–901.
5. Chalfie,M., Tu,Y., Euskirchen,G., Ward,W.W. and Prasher,D.C. (1994) Green fluorescent protein as a marker for gene expression. *Science*, **263**, 802–805.
6. Jarret,R.L. and McCluskey,K. (2019) *The Biological Resources of Model Organisms* CRC Press.
7. Wei,Q., Zhao,Y., Guo,Y., Stomel,J., Stires,R. and Ellis,R.E. (2014) Co-option of alternate sperm activation programs in the evolution of self-fertile nematodes. *Nat. Commun.*, **5**, 5888.
8. Zhao,Y., Tan,C.-H., Krauchunas,A., Scharf,A., Dietrich,N., Warnhoff,K., Yuan,Z., Druzhinina,M., Gu,S.G., Miao,L., *et al.* (2018) The zinc transporter ZIPT-7.1 regulates sperm activation in nematodes. *PLoS Biol.*, **16**, e2005069.
9. Sulston,J.E. and Horvitz,H.R. (1977) Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.*, **56**, 110–156.
10. Kirouac,M. and Sternberg,P.W. (2003) cis-Regulatory control of three cell fate-specific genes in vulval organogenesis of *Caenorhabditis elegans* and *C. briggsae*. *Dev. Biol.*, **257**, 85–103.
11. Kiontke,K., Barrière,A., Kolotuev,I., Podbilewicz,B., Sommer,R., Fitch,D.H.A. and Félix,M.-A. (2007) Trends, stasis, and drift in the evolution of nematode vulva development. *Curr. Biol.*, **17**, 1925–1937.
12. Poulet,N., Vielle,A., Gimond,C., Ferrari,C. and Braendle,C. (2015) Evolutionarily divergent thermal sensitivity of germline development and fertility in hermaphroditic *Caenorhabditis* nematodes. *Evol. Dev.*, **17**, 380–397.
13. Vielle,A., Callemeyn-Torre,N., Gimond,C., Poulet,N., Gray,J.C., Cutter,A.D. and Braendle,C. (2016) Convergent evolution of sperm gigantism and the developmental origins of sperm size variability in *Caenorhabditis* nematodes. *Evolution*, **70**, 2485–2503.
14. Inoue,T., Ailion,M., Poon,S., Kim,H.K., Thomas,J.H. and Sternberg,P.W. (2007) Genetic analysis of dauer formation in *Caenorhabditis briggsae*. *Genetics*, **177**, 809–818.
15. Cohen,S.M., Wrobel,C.J.J., Prakash,S.J., Schroeder,F.C. and Sternberg,P.W. (2022) Formation and function of dauer ascarosides in the nematodes *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *G3* , **12**.
16. Baer,C.F., Shaw,F., Steding,C., Baumgartner,M., Hawkins,A., Houppert,A., Mason,N., Reed,M., Simonelic,K., Woodard,W., *et al.* (2005) Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 5785–5790.
17. Ostrow,D., Phillips,N., Avalos,A., Blanton,D., Boggs,A., Keller,T., Levy,L., Rosenbloom,J. and Baer,C.F. (2007) Mutational bias for body size in rhabditid nematodes. *Genetics*, **176**, 1653–1661.

18. Laricchia,K.M., Zdraljevic,S., Cook,D.E. and Andersen,E.C. (2017) Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol. Biol. Evol.*, 10.1093/molbev/msx155.
19. Dolgin,E.S., Charlesworth,B. and Cutter,A.D. (2008) Population frequencies of transposable elements in selfing and outcrossing *Caenorhabditis* nematodes. *Genet. Res.* , **90**, 317–329.
20. Ben-David,E., Pliota,P., Widen,S.A., Koreshova,A., Lemus-Vergara,T., Verpukhovskiy,P., Mandali,S., Braendle,C., Burga,A. and Kruglyak,L. (2021) Ubiquitous Selfish Toxin-Antidote Elements in *Caenorhabditis* Species. *Curr. Biol.*, **31**, 990–1001.e5.
21. Noble,L.M., Yuen,J., Stevens,L., Moya,N., Persaud,R., Moscatelli,M., Jackson,J.L., Zhang,G., Chitrakar,R., Baugh,L.R., *et al.* (2021) Selfing is the safest sex for *Caenorhabditis tropicalis*. *Elife*, **10**.
22. Zamanian,M., Cook,D.E., Zdraljevic,S., Brady,S.C., Lee,D., Lee,J. and Andersen,E.C. (2018) Discovery of genomic intervals that underlie nematode responses to benzimidazoles. *PLoS Negl. Trop. Dis.*, **12**, e0006368.
23. Thomas,C.G., Li,R., Smith,H.E., Woodruff,G.C., Oliver,B. and Haag,E.S. (2012) Simplification and desexualization of gene expression in self-fertile nematodes. *Curr. Biol.*, **22**, 2167–2172.
24. Fierst,J.L., Willis,J.H., Thomas,C.G., Wang,W., Reynolds,R.M., Ahearne,T.E., Cutter,A.D. and Phillips,P.C. (2015) Reproductive Mode and the Evolution of Genome Size and Structure in *Caenorhabditis* Nematodes. *PLoS Genet.*, **11**, e1005323.
25. Kanzaki,N., Tsai,I.J., Tanaka,R., Hunt,V.L., Liu,D., Tsuyama,K., Maeda,Y., Namai,S., Kumagai,R., Tracey,A., *et al.* (2018) Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nat. Commun.*, **9**, 3216.
26. Yin,D., Schwarz,E.M., Thomas,C.G., Felde,R.L., Korf,I.F., Cutter,A.D., Schartner,C.M., Ralston,E.J., Meyer,B.J. and Haag,E.S. (2018) Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *Science*, **359**, 55–61.
27. Stevens,L., Félix,M.-A., Beltran,T., Braendle,C., Caurcel,C., Fausett,S., Fitch,D., Frézal,L., Gosse,C., Kaur,T., *et al.* (2019) Comparative genomics of 10 new *Caenorhabditis* species. *Evolution Letters*, **3**, 217–236.
28. Stein,L.D., Bao,Z., Blasiar,D., Blumenthal,T., Brent,M.R., Chen,N., Chinwalla,A., Clarke,L., Clee,C., Coghlan,A., *et al.* (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.*, **1**, E45.
29. Stevens,L., Moya,N.D., Tanny,R.E., Gibson,S.B., Tracey,A., Na,H., Chitrakar,R., Dekker,J., Walhout,A.J.M., Baugh,L.R., *et al.* (2022) Chromosome-Level Reference Genomes for Two Strains of *Caenorhabditis briggsae*: An Improved Platform for Comparative Genomics. *Genome Biol. Evol.*, **14**.
30. Hillier,L.W., Miller,R.D., Baird,S.E., Chinwalla,A., Fulton,L.A., Koboldt,D.C. and Waterston,R.H. (2007) Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. *PLoS Biol.*, **5**, e167.
31. Kiontke,K., Félix,M.-A., Ailion,M., Rockman,M.V., Braendle,C., Pénigault,J.-B. and Fitch,D.H.A. (2011) A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol. Biol.*, **11**, 339.
32. Andersen,E.C., Gerke,J.P., Shapiro,J.A., Crissman,J.R., Ghosh,R., Bloom,J.S., Félix,M.-A. and Kruglyak,L. (2012) Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.*, **44**, 285–290.

33. Félix,M.-A. and Duveau,F. (2012) Population dynamics and habitat sharing of natural populations of *Caenorhabditis elegans* and *C. briggsae*. *BMC Biol.*, **10**, 59.
34. Barrière,A. and Félix,M.-A. (2014) Isolation of *C. elegans* and related nematodes. *WormBook*, 10.1895/wormbook.1.115.2.
35. Cook,D.E., Zdraljevic,S., Tanny,R.E., Seo,B., Riccardi,D.D., Noble,L.M., Rockman,M.V., Alkema,M.J., Braendle,C., Kammenga,J.E., *et al.* (2016) The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics*, 10.1534/genetics.116.191148.
36. Ferrari,C., Salle,R., Callemeyn-Torre,N., Jovelín,R., Cutter,A.D. and Braendle,C. (2017) Ephemeral-habitat colonization and neotropical species richness of *Caenorhabditis* nematodes. *BMC Ecol.*, **17**, 43.
37. Hahnel,S.R., Zdraljevic,S., Rodríguez,B.C., Zhao,Y., McGrath,P.T. and Andersen,E.C. (2018) Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog.*, **14**, e1007226.
38. Richaud,A., Zhang,G., Lee,D., Lee,J. and Félix,M.-A. (2018) The Local Coexistence Pattern of Selfing Genotypes in *Caenorhabditis elegans* Natural Metapopulations. *Genetics*, **208**, 807–821.
39. Crombie,T.A., Zdraljevic,S., Cook,D.E., Tanny,R.E., Brady,S.C., Wang,Y., Evans,K.S., Hahnel,S., Lee,D., Rodríguez,B.C., *et al.* (2019) Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. *Elife*, **8**, e50465.
40. Lee,D., Zdraljevic,S., Cook,D.E., Frézal,L., Hsu,J.-C., Sterken,M.G., Riksen,J.A.G., Wang,J., Kammenga,J.E., Braendle,C., *et al.* (2019) Selection and gene flow shape niche-associated variation in pheromone response. *Nat Ecol Evol*, **3**, 1455–1463.
41. Lee,D., Zdraljevic,S., Stevens,L., Wang,Y., Tanny,R.E., Crombie,T.A., Cook,D.E., Webster,A.K., Chirakar,R., Baugh,L.R., *et al.* (2021) Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nat Ecol Evol*, 10.1038/s41559-021-01435-x.
42. Crombie,T.A., Battlay,P., Tanny,R.E., Evans,K.S., Buchanan,C.M., Cook,D.E., Dilks,C.M., Stinson,L.A., Zdraljevic,S., Zhang,G., *et al.* (2022) Local adaptation and spatiotemporal patterns of genetic diversity revealed by repeated sampling of *Caenorhabditis elegans* across the Hawaiian Islands. *Mol. Ecol.*, **31**, 2327–2347.
43. Sloat,S.A., Noble,L.M., Paaby,A.B., Bernstein,M., Chang,A., Kaur,T., Yuen,J., Tintori,S.C., Jackson,J.L., Martel,A., *et al.* (2022) *Caenorhabditis* nematodes colonize ephemeral resource patches in neotropical forests. *Ecol. Evol.*, **12**, e9124.
44. Gaertner,B.E. and Phillips,P.C. (2010) *Caenorhabditis elegans* as a platform for molecular quantitative genetics and the systems biology of natural variation. *Genet. Res.*, **92**, 331–348.
45. Evans,K.S., van Wijk,M.H., McGrath,P.T., Andersen,E.C. and Sterken,M.G. (2021) From QTL to gene: *C. elegans* facilitates discoveries of the genetic mechanisms underlying natural variation. *Trends Genet.*, 10.1016/j.tig.2021.06.005.
46. Andersen,E.C. and Rockman,M.V. (2022) Natural genetic variation as a tool for discovery in *Caenorhabditis* nematodes. *Genetics*, **220**.
47. Cook,D.E., Zdraljevic,S., Roberts,J.P. and Andersen,E.C. (2017) CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.*, **45**, D650–D657.
48. Zdraljevic,S., Strand,C., Seidel,H.S., Cook,D.E., Doench,J.G. and Andersen,E.C. (2017) Natural variation in a single amino acid substitution underlies physiological responses to topoisomerase II poisons. *PLoS*

49. Lee,D., Yang,H., Kim,J., Brady,S., Zdraljevic,S., Zamanian,M., Kim,H., Paik,Y.-K., Kruglyak,L., Andersen,E.C., *et al.* (2017) The genetic basis of natural variation in a phoretic behavior. *Nat. Commun.*, **8**, 273.
50. Zdraljevic,S., Fox,B.W., Strand,C., Panda,O., Tenjo,F.J., Brady,S.C., Crombie,T.A., Doench,J.G., Schroeder,F.C. and Andersen,E.C. (2019) Natural variation in *C. elegans* arsenic toxicity is explained by differences in branched chain amino acid metabolism. *Elife*, **8**.
51. Brady,S.C., Zdraljevic,S., Bisaga,K.W., Tanny,R.E., Cook,D.E., Lee,D., Wang,Y. and Andersen,E.C. (2019) A Novel Gene Underlies Bleomycin-Response Variation in *Caenorhabditis elegans*. *Genetics*, 10.1534/genetics.119.302286.
52. Burga,A., Ben-David,E., Lemus Vergara,T., Boocock,J. and Kruglyak,L. (2019) Fast genetic mapping of complex traits in *C. elegans* using millions of individuals in bulk. *Nat. Commun.*, **10**, 2680.
53. Evans,K.S. and Andersen,E.C. (2020) The Gene *scb-1* Underlies Variation in *Caenorhabditis elegans* Chemotherapeutic Responses. *G3: Genes, Genomes, Genetics*, **10**, 2353–2364.
54. Evans,K.S., Zdraljevic,S., Stevens,L., Collins,K., Tanny,R.E. and Andersen,E.C. (2020) Natural variation in the sequestosome-related gene, *sqst-5*, underlies zinc homeostasis in *Caenorhabditis elegans*. *PLoS Genet.*, **16**, e1008986.
55. Moya,N.D., Stevens,L., Miller,I.R., Sokol,C.E., Galindo,J.L., Bardas,A.D., Koh,E.S.H., Rozenich,J., Yeo,C., Xu,M., *et al.* (2023) Novel and improved *Caenorhabditis briggsae* gene models generated by community curation. *bioRxiv*, 10.1101/2023.05.16.541014.
56. Widmayer,S.J., Evans,K.S., Zdraljevic,S. and Andersen,E.C. (2022) Evaluating the power and limitations of genome-wide association studies in *Caenorhabditis elegans*. *G3*, **12**.
57. Di Bernardo,M., Crombie,T.A., Cook,D.E. and Andersen,E.C. (2021) easyFulcrum: An R package to process and analyze ecological sampling data generated using the Fulcrum mobile application. *PLoS One*, **16**, e0254293.
58. Barrière,A. and Félix,M.-A. (2005) High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Curr. Biol.*, **15**, 1176–1184.
59. Petersen,C., Dirksen,P., Prah,S., Strathmann,E.A. and Schulenburg,H. (2014) The prevalence of *Caenorhabditis elegans* across 1.5 years in selected North German locations: the importance of substrate type, abiotic parameters, and *Caenorhabditis* competitors. *BMC Ecol.*, **14**, 4.
60. Yu,J., Pressoir,G., Briggs,W.H., Vroh Bi,I., Yamasaki,M., Doebley,J.F., McMullen,M.D., Gaut,B.S., Nielsen,D.M., Holland,J.B., *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.*, **38**, 203–208.
61. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T., *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
62. Thomas,C.G., Wang,W., Jovelin,R., Ghosh,R., Lomasko,T., Trinh,Q., Kruglyak,L., Stein,L.D. and Cutter,A.D. (2015) Full-genome evolutionary histories of selfing, splitting, and selection in *Caenorhabditis*. *Genome Res.*, **25**, 667–678.
63. Robinson,J.T., Thorvaldsdottir,H., Turner,D. and Mesirov,J.P. (2023) igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*, **39**.
64. Danecek,P. and McCarthy,S.A. (2017) BCFtools/csq: haplotype-aware variant consequences.

65. Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 10915–10919.
66. Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. *Science*, **185**, 862–864.
67. Evans, K.S., Brady, S.C., Bloom, J.S., Tanny, R.E., Cook, D.E., Giuliani, S.E., Hippleheuser, S.W., Zamanian, M. and Andersen, E.C. (2018) Shared Genomic Regions Underlie Natural Variation in Diverse Toxin Responses. *Genetics*, **210**, 1509–1525.
68. Wit, J., Rodriguez, B.C. and Andersen, E.C. (2021) Natural variation in *Caenorhabditis elegans* responses to the anthelmintic emodepside. *Int. J. Parasitol. Drugs Drug Resist.*, **16**, 1–8.
69. Widmayer, S.J., Crombie, T.A., Nyaanga, J.N., Evans, K.S. and Andersen, E.C. (2022) *C. elegans* toxicant responses vary among genetically diverse individuals. *Toxicology*, **479**, 153292.
70. Shaver, A.O., Wit, J., Dilks, C.M., Crombie, T.A., Li, H., Aroian, R.V. and Andersen, E.C. (2023) Variation in anthelmintic responses are driven by genetic differences among diverse *C. elegans* wild strains. *PLoS Pathog.*, **19**, e1011285.
71. Eshed, Y. and Zamir, D. (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics*, **141**, 1147–1162.
72. Brenner, S. (1974) The genetics of *Caenorhabditis elegans*. *Genetics*, **77**, 71–94.
73. Sivasundar, A. and Hey, J. (2003) Population Genetics of *Caenorhabditis elegans*: The Paradox of Low Polymorphism in a Widespread Species. *Genetics*, **163**, 147–157.
74. Cutter, A.D., Félix, M.-A., Barrière, A. and Charlesworth, D. (2006) Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of *Caenorhabditis briggsae*. *Genetics*, **173**, 2021–2031.
75. Dolgin, E.S., Félix, M.-A. and Cutter, A.D. (2008) Hakuna Nematoda: genetic and phenotypic diversity in African isolates of *Caenorhabditis elegans* and *C. briggsae*. *Heredity*, **100**, 304–315.
76. Raboin, M.J., Timko, A.F., Howe, D.K., Félix, M.-A. and Denver, D.R. (2010) Evolution of *Caenorhabditis* mitochondrial genome pseudogenes and *Caenorhabditis briggsae* natural isolates. *Mol. Biol. Evol.*, **27**, 1087–1096.
77. Félix, M.-A., Jovelín, R., Ferrari, C., Han, S., Cho, Y.R., Andersen, E.C., Cutter, A.D. and Braendle, C. (2013) Species richness, distribution and genetic diversity of *Caenorhabditis* nematodes in a remote tropical rainforest. *BMC Evol. Biol.*, **13**, 10.
78. Prasad, A. and Croydon-Sugarman, M.J.F. (2011) TEMPERATURE-DEPENDENT FECUNDITY ASSOCIATES WITH LATITUDE IN CAENORHABDITIS BRIGGSAE. *Journal of Organic ....*
79. Gimond, C., Jovelín, R., Han, S., Ferrari, C., Cutter, A.D. and Braendle, C. (2013) Outbreeding depression with low genetic variation in selfing *Caenorhabditis* nematodes. *Evolution*, **67**, 3087–3101.
80. Lander, E.S. and Schork, N.J. (1994) Genetic dissection of complex traits. *Science*, **265**, 2037–2048.
81. Cutter, A.D. (2006) Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. *Genetics*, **172**, 171–184.
82. Ross, J.A., Koboldt, D.C., Staisch, J.E., Chamberlin, H.M., Gupta, B.P., Miller, R.D., Baird, S.E. and Haag, E.S. (2011) *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and

the evolution of recombination. *PLoS Genet.*, **7**, e1002174.

83. Zhang,Z., Ersoz,E., Lai,C.-Q., Todhunter,R.J., Tiwari,H.K., Gore,M.A., Bradbury,P.J., Yu,J., Arnett,D.K., Ordovas,J.M., *et al.* (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.*, **42**, 355–360.
84. Yang,J., Lee,S.H., Goddard,M.E. and Visscher,P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
85. Jiang,L., Zheng,Z., Qi,T., Kemper,K.E., Wray,N.R., Visscher,P.M. and Yang,J. (2019) A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.*, **51**, 1749–1755.
86. Zhang,G., Roberto,N.M., Lee,D., Hahnel,S.R. and Andersen,E.C. (2022) The impact of species-wide gene expression variation on *Caenorhabditis elegans* complex traits. *Nat. Commun.*, **13**, 3462.