Dear Professor Rigden,

Thank you for handling our manuscript. We have addressed all of your and the reviewers' comments. Please find the reviews and comments in bold below and our responses in normal face font.

Editor Comments to Author:

Dear Dr. Anderson,

Thank you for giving us the opportunity to consider your manuscript.

The referees' comments were very positive and, in principle, we would be pleased to publish it. However, before we can accept the manuscript we ask that you consider the points raised by the referees, which you will find below in this email. Please either answer them in a revised version or explain why they are not relevant. The revised manuscript may be subject to another round of reviews by the same and/or additional referees.

The only substantial issue concerns Fig 2 and the simulated data. I agree the paper would benefit from a case study telling a story with real data, instead (or in addition if you like).

We agree as well and have changed Figure 2 to real data. We also addressed the concerns raised by each reviewer in the section below.

Referee Comments to Author:

Referee: 1

Comments for the Author
Crombie et al. report an update of the C. elegans Natural Diversity Resource (CeNDR) that includes two additional Caenorhabditis species. This new release (now called CaeNDR) includes over 1,000 strains of C. briggsae and over 600 strains of C. tropicalis, in addition to the previous C. elegans resources. The CeNDR database has been an invaluable resource for the C. elegans evolution community by making whole-genome variation data easily accessible to researchers. An expansion of this collection to include additional species only makes it more useful as it will now facilitate studies of evolutionary divergence. It will also accelerate quantitative genetic studies more suited to non-elegans species, such as temperature-related traits in the phylogeographical clades of C. briggsae. The CaeNDR website not only includes easily-accessible data across various workflow steps (.bam,.vcf,.csv, etc.), but also has a number of useful tools to explore the data, including an intuitive genome browser. As for the paper itself, it is a clear description of the updated resource. All of the links work, although I did not explore tools that required a user account. My review is short not because I have spent little time reading this manuscript. Rather, it is because I see no major issues with it. I hope it will be published soon.

Thank you for your positive feedback!

Figure 2. This really is a minor comment, as I understand this is the output of an automated GWAS report on simulated data. Yet, the values of the y-axis on the blow-up on Figure 2C are much higher than on Figure 3A. Was the y-axis clipped on Figure 2A, excluding the sites with the lowest p-values?

The fine mapping is a separate association test using all variants across the interval rather than only the variants above a minor allele frequency of 5% and not in high linkage disequilibrium with each other. That is why the plot is different. In the fine mapping plot, we excluded the highest p-values to emphasize the peak and linkage disequilibrium across the interval. We now include text in the figure 2 legend to address this comment.

NEW: “(**C**) A fine mapping plot is shown to visualize the significance of all bi-allelic SNVs in the QTL interval on the right of chromosome V. The x-axis is the physical position in the genome, and the y-axis is the -$\log_{10}$ *p*-value obtained from a statistical test of association using all SNVs in the interval. The inclusion of all SNVs alters the range of *p*-values relative to GWAS manhattan plot. We truncated the axes to focus on the center of the QTL.”

Referee: 2

Comments for the Author
In this manuscript, Crombie et al. present a significant update of the original C. elegans Natural Diversity Resource (CeNDR, 2016). The new platform, named Caenorhabditis Natural Diversity Resource (CaeNDR), further expands the work established in C. elegans and now includes two additional selfing species: C. briggsae and C. tropicalis. Compared to C. elegans, the amount of genetic research performed so far in C. briggsae and C. tropicalis is negligible. However, when it comes to answering fundamental questions about how natural genetic diversity shapes trait variation, it is truly critical to do so in an evolutionary framework. And this is precisely what the CaeNDR platform will empower researchers to do.

The three goals of CaeNDR are to:
1) Serve as a centralized repository for wild Caenorhabditis selfing strains (including submissions from both academic and citizen scientists).
2) Generate and annotate natural genetic variants across these strains.
3) Facilitate the use of this data and strains for GWAS and quantitative genetic studies.

These three goals have been carried out to extremely high standards. Navigating the CaeNDR website is exceptionally easy and intuitive. All the sequencing data, from aligned BAM files to variant effect predictions, are easily accessible and downloadable. The sheer amount of sequencing data is substantial, yet everything works seamlessly. It is evident that the authors have put a lot of thought and effort into ensuring smooth operation. What I like most about this platform is that it truly facilitates the inherently interactive process of identifying the genetic variants that underlie phenotypic changes. CaeNDR offers great tools and functionalities that impact and streamline every decision a researcher must make and every challenge they may face. From relatively minor tasks like finding suitable genetic markers and calculating heritability to the most challenging ones, such as leveraging SNP effect predictions, this platform provides comprehensive support. It is clear that this platform will be heavily used and cited. Lastly, unlike many other databases that are published year after year, this is clearly not something that would stop working after a few months.

We appreciate the supportive feedback!

I have only a few minor comments that, in my opinion, should not preclude the publication of this manuscript:

1. In figure 2C, there appears to be no mention of the significance of the blue dashed vertical line.

We have added a description to the Figure 2 legend text.

NEW: “The vertical dashed line (cyan) represents the physical position of the most significantly associated SNV identified on the right of chromosome V in the GWAS.”

2. Figure 2C legend mentions that "the axes are identical to the Manhattan plot above." Specifically, the Y-axis of Fig. 2A and Fig. 2C is -log10(p), but it is not clear if we are referring to the same "p-value". Could this please be clarified?

As mentioned in response to reviewer #1, the association test is different between the two figures. We edited the figure description to clarify this point.

NEW: “(**C**) A fine mapping plot is shown to visualize the significance of all bi-allelic SNVs in the QTL interval on the right of chromosome V. The x-axis is the physical position in the genome, and the y-axis is the -log$_{10}$ *p*-value obtained from a statistical test of association using all SNVs in the interval. The inclusion of all SNVs alters the range of *p*-values relative to GWAS manhattan plot. We truncated the axes to focus on the center of the QTL.”

More broadly speaking, I didn't quite follow the choice of using simulated data for Figure 2. On the one hand, the main purpose of the figure is to highlight the graphical outputs generated by running the GWAS mapping tool, so the specifics of the trait may not be relevant. On the other hand, this figure shows what appears to be an overall "not very successful GWAS" with only a marginally significant QTL. So, not surprisingly, plots like the one in Fig 2B don't really show any visually striking data, and Fig 2C, while technically an interesting way to plot variant effects, doesn't really showcase the practical utility of this kind of plot. I would personally find it more impactful to showcase one or two published examples of successful GWAS studies down to the gene level. Alternatively, and my apologies if I missed this, it would be nice to have such "case study" datasets easily accessible on the website to facilitate the work of non-experts in quantitative genetics and also for teaching purposes.

We agree and have changed the manuscript to use data from a real study published previously on the abamectin response trait (Evans *et al.* 2021). We updated Figure 2 and added text describing how the mapping report can be used to identify genes and natural variants that underlie trait differences and act as a mini case-study to help non-experts see the utility of the report plots.

NEW: The paragraph describing the Genetic Mapping tool is updated (new text in red):
The Genetic Mapping tool (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) 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. To illustrate the utility of the mapping report, we used the Genetic Mapping tool to process data from a previously published GWAS of abamectin responses (67). Abamectin is a widely used agricultural pesticide and the threat of pervasive nematode resistance motivates the search for natural resistance alleles. The Manhattan plot taken from the abamectin response mapping report shows the physical positions of the three QTL identified across the genome (Figure 2A). To visualize the strength of these associations, a phenotype-by-genotype plot for the most significant markers in each QTL is included in the report (Figure 2B). 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. Fine mapping tests the association between genotype and phenotype at every SNV in a QTL, not just the marker loci, and can help users identify the variants that are most likely to underlie trait differences at a QTL (Figure 2C). In the abamectin example, the gene *gcl-1* is labeled because natural variation in this gene was previously shown to confer abamectin resistance (52, 68). Even without this knowledge, the low *p*-value and predicted high effect variant in *glc-1* (arrow, red bar, Figure 2C) would suggest that natural variation in *glc-1* could underlie differences in abamectin responses. Importantly, only two other genes with high-effect variants that have lower *p*-values than *glc-1* were found. This example demonstrates how the list of candidate genes in a region is greatly reduced by cross-referencing the fine-mapping *p*-values and predicted variant impacts provided in the mapping report.

NEW: Figure 2 and its updated legend are included below for convenience (new text in red).

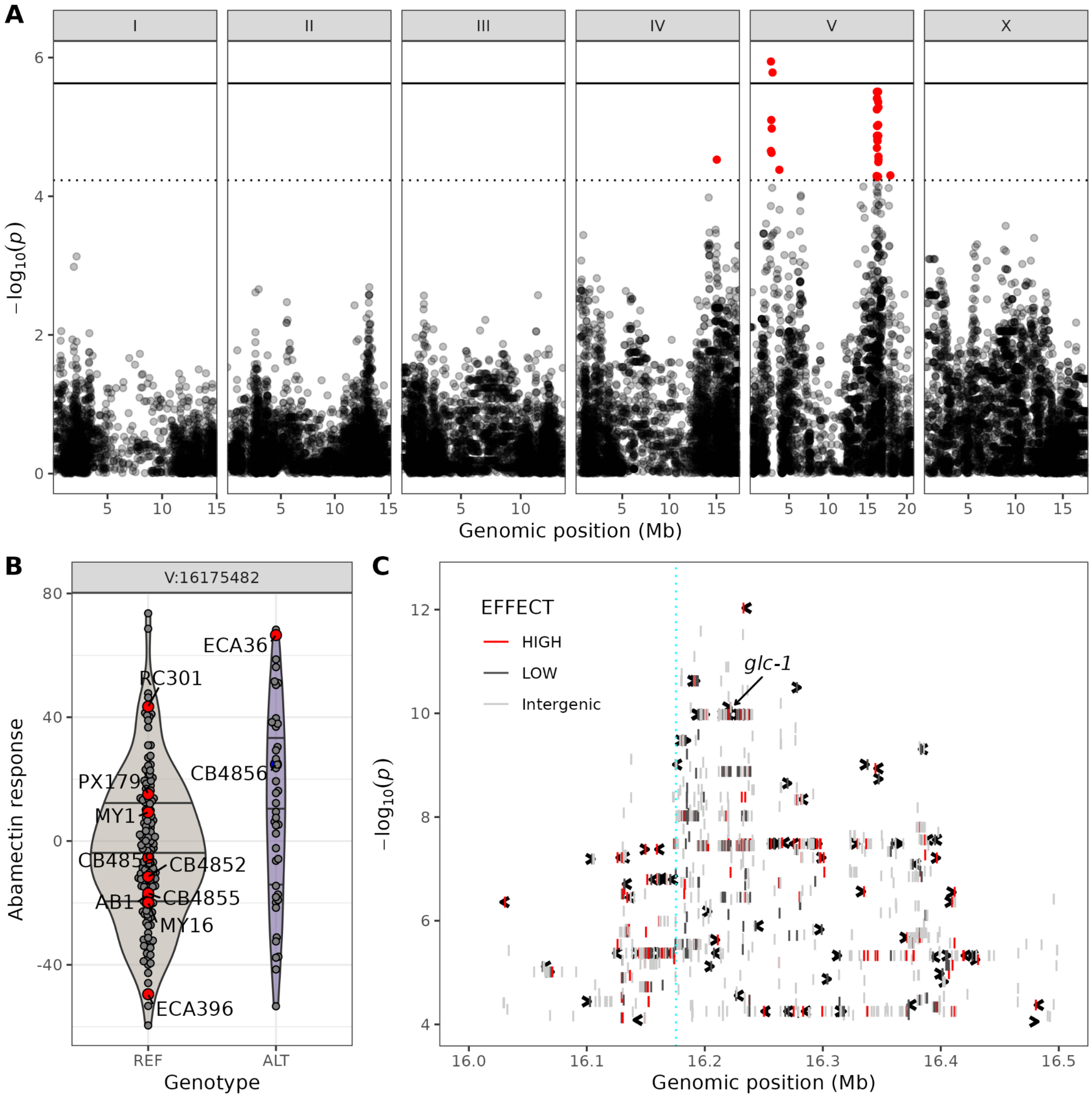

Figure 2 - Elements of the abamectin response GWAS mapping report from CaeNDR.

(**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 -log10 *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 the right of chromosome V (16.175 Mb). The plot shows the trait values for the strains with the reference (REF) allele compared to the strains with the alternative (ALT) allele for that marker. The abamectin response values shown here are mean centered; the strains with larger values develop better in the presence of abamectin. The horizontal lines indicate the 75th percentile, median, and 25th percentile of the data for each genotype. The trait values for some strains are highlighted and labeled because these strains are commonly used to measure dose responses by the community (69–72). 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 the right of chromosome V. The x-axis is the physical position in the genome, and the y-axis is the -log10 *p*-value obtained from a statistical test of association using all SNVs in the interval. The inclusion of all SNVs alters the range of *p*-values relative to GWAS Manhattan plot. We truncated the axes to focus on the center of the QTL. Each SNV is 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 -log10 *p*-value of all variants in the gene. The vertical dashed line (cyan) represents the physical position of the most significantly associated SNV identified on the right of chromosome V in the GWAS. The gene gcl-1 is labeled in the plot because natural variation in this gene was previously shown to confer abamectin resistance (52, 68).

Referee: 3

Comments for the Author
This paper does a very nice job of providing an overview of the expansion of the C. elegans natural variation resource, which now includes nearly 2,000 strains from two additional species. This alone merits new publication about the database and resource in NAR. In addition, however, the website and database hosts additional analytical tools that are not included in the original publication.

This is naturally a fairly unusual publication form, as it is highly descriptive. From a reviewer's point of view, I found the descriptions appropriate and comprehensive, without being pedantic. I have no specific comments or criticisms on the text itself, which seem very publishable as is to my eyes.

Thank you for your feedback!