RESEARCH ARTICLE

A spontaneous complex structural variant in rcan-1 increases exploratory behavior and laboratory fitness of Caenorhabditis elegans

Yuehui Zhao¹, Lijiang Long¹,², Jason Wang³, Shweta Biliya¹, Shannon C. Brady⁴, Daehan Lee⁵, Akinade Ojemakinde¹, Erik C. Andersen⁴, Fredrik O. Vannberg¹,⁵, Hang Lu⁵,⁶, Patrick T. McGrath¹,²,⁶,⁷*

¹ School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, United States of America, ² Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, Georgia, United States of America, ³ The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, United States of America, ⁴ Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, United States of America, ⁵ Parker H. Petit Institute of Bioengineering and BioScience, Georgia Institute of Technology, Atlanta, Georgia, United States of America, ⁶ School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States of America, ⁷ School of Physics, Georgia Institute of Technology, Atlanta, Georgia, United States of America

* patrick.mcgrath@biology.gatech.edu

Abstract

Over long evolutionary timescales, major changes to the copy number, function, and genomic organization of genes occur, however, our understanding of the individual mutational events responsible for these changes is lacking. In this report, we study the genetic basis of adaptation of two strains of C. elegans to laboratory food sources using competition experiments on a panel of 89 recombinant inbred lines (RIL). Unexpectedly, we identified a single RIL with higher relative fitness than either of the parental strains. This strain also displayed a novel behavioral phenotype, resulting in higher propensity to explore bacterial lawns. Using bulk-segregant analysis and short-read resequencing of this RIL, we mapped the change in exploration behavior to a spontaneous, complex rearrangement of the rcan-1 gene that occurred during construction of the RIL panel. We resolved this rearrangement into five unique tandem inversion/duplications using Oxford Nanopore long-read sequencing. rcan-1 encodes an ortholog to human RCAN1/DSCR1 calcipressin gene, which has been implicated as a causal gene for Down syndrome. The genomic rearrangement in rcan-1 creates two complete and two truncated versions of the rcan-1 coding region, with a variety of modified 5’ and 3’ non-coding regions. While most copy-number variations (CNVs) are thought to act by increasing expression of duplicated genes, these changes to rcan-1 ultimately result in the reduction of its whole-body expression due to changes in the upstream regions. By backcrossing this rearrangement into a common genetic background to create a near isogenic line (NIL), we demonstrate that both the competitive advantage and exploration behavioral changes are linked to this complex genetic variant. This NIL strain does not phenocopy a strain containing an rcan-1 loss-of-function allele, which suggests that the residual expression of rcan-1 is necessary for its fitness effects. Our results demonstrate how colonization of new environments, such as those encountered in the laboratory, can create...
evolutionary pressure to modify gene function. This evolutionary mismatch can be resolved by an unexpectedly complex genetic change that simultaneously duplicates and diversifies a gene into two uniquely regulated genes. Our work shows how complex rearrangements can act to modify gene expression in ways besides increased gene dosage.

Author summary

Evolution acts on genetic variants that modify phenotypes that increase the likelihood of staying alive and passing on these genetic changes to subsequent generations (i.e. fitness). There is general interest in understanding the types of genetic variants that can increase fitness in specific environments. One route that fitness can be increased is through changes in behavior, such as finding new food sources. Here, we identify a spontaneous genetic change that increases exploration behavior and fitness of animals in laboratory environments. Interestingly, this genetic change is not a simple genetic change that deletes or changes the sequence of a protein product, but rather a complex structural variant that simultaneously duplicates the \textit{rcan-1} gene and also modifies its expression in a number of tissues. Our work demonstrates how a complex structural change can duplicate a gene, modify the DNA control regions that determine its cellular sites of action, and confer a fitness advantage that could lead to its spread in a population.

Introduction

Structural variation, resulting in the removal, duplication, insertion, or rearrangement of large (> 50bp) genomic regions, makes up a significant component of natural genetic variation in many different species [1–6]. The largescale rearrangement of DNA can truncate genes, modify transcriptional regulatory regions, and/or increase gene dosage and expression. Consequently, structural variation can have profound, detrimental effects on phenotype, including a variety of human diseases [7–11]. However, structural variants are also thought to be important for adaptive evolution in natural populations [2, 12–14] and domesticated plants and animals [15], including a number of examples that link structural changes to putative adaptive phenotypic variation [16–20]. From an evolutionary perspective, these larger genomic changes are interesting for a number of reasons. A gene duplication creates a new genetic substrate for evolution to act on, and over long evolutionary timescale, can result in the creation of a paralogous gene [21, 22]. Inversion events can both change the chromatin state that a gene is found in and also suppress recombination events within the inverted region [23, 24]. Finally, structural changes can create incompatibilities between populations, contributing to speciation [14, 25].

For these reasons, it is desirable to understand how genomic rearrangements modify phenotype and spread through populations. However, determining the effect of naturally-occurring genomic rearrangement on phenotype and fitness is very difficult due to linkage of nearby mutations. Experimental evolution is a powerful approach to study adaptation in real time due to the lower rate of nucleotide diversity between the selected strains, aiding in the identification of causal mutations [26–38]. These studies typically utilize microorganisms with short generation times such as \textit{E. coli} or \textit{S. cerevisiae}, elucidating the molecular basis of adaptation and profiling genome dynamics in evolving population under diverse laboratory settings. By identifying and studying causal genetic variants, important insights into beneficial
mutations have been gained, such as their occurrence frequency, the complexity of their molecular basis, the role of contingency and genetic background into their effect, and their fitness effects in specific environments. A number of studies have demonstrated that genomic rearrangements can spread in these populations due to the actions of positive selection [28, 29, 39–44]. In some of these experiments, gene duplicates are thought to facilitate the metabolism or transport of a limiting nutrient due to increased protein product responsible for a rate-limiting step of metabolism.

While these experiments have led to fundamental advances in our understanding of evolution in real time, it is desirable to perform similar experiments in multicellular organisms, with specialized tissues and the ability to respond to their environment using a nervous system. However, long-term adaptation studies are still less advanced in multicellular animals. In our lab, we use the nematode *C. elegans* to study the connection between genotype and phenotype. Compared to other species, *C. elegans* has a high-rate of spontaneous structural mutations, as inferred by their presence in mutation accumulation lines and laboratory strains [41, 45–47]. In general, most of these structural changes are thought to be deleterious; they are purged in populations with higher effective populations sizes [46]. However, spread of copy number variants are also observed in animals carrying deleterious mutations, suggesting that positive selection also acts on copy number variants in certain contexts [41]. Structural changes are also common in wild strains of *C. elegans*, consistent with a role of structural variants being beneficial in certain natural environments [1, 48].

Here, we study two historical laboratory strains of *C. elegans*, called N2 and LSJ2. These two strains share the same hermaphrodite ancestor, which was isolated in 1951 from mushroom compost collected in Bristol, UK (Fig 1A). In 1958, descendants of this animal were split into two distinct lineages and cultured in different laboratory conditions. The N2 lineage grew on agar plates seeded with bacteria (standard conditions for a *C. elegans* genetics laboratory). After about two decades growing in this environment, this lineage was cryopreserved in Sydney Brenner’s lab and named as N2. After Sydney Brenner introduced *C. elegans* to the genetics research community, N2 became the standard reference strain used across the world [49, 50]. The second lineage was cultured in liquid, axenic media composed of liver and soy peptone extract as a food source for about fifty years before it was cryopreserved and named as LSJ2 [51]. In the time between their separation into two lineages and cryopreservation, approximately 300 mutations arose and fixed in either of the lineages [51]. Previous work has identified six causal mutations of these 300 that confer phenotypic change and competitive advantage in the conditions these mutations arose in [51–56].

In this study, we used recombinant inbred lines (RILs) created using the N2 and LSJ2 strains combined with quantitative trait loci mapping (QTL mapping) to non-biasedly identify any additional mutations between N2 and LSJ2 that confer competitive advantage in standard N2-like laboratory growth conditions. During these experiments, we identified a beneficial, spontaneous, and complex inversion/duplication mutation in the *rcan-1* gene that occurred during the construction of the RILs. This complex genomic rearrangement results in the partial duplication and inversion of five different regions, simultaneously duplicating the *rcan-1* coding region and modifying the upstream promoter regions. While the gene copy number of *rcan-1* is duplicated, the changes in upstream regions result in an overall decrease in *rcan-1* expression. Our work demonstrates how the initial mutational events that create gene duplicates can be complicated, result in unexpected changes in gene expression, and provide fitness increases that will result in its fixation in a population.
Results

A N2/LSJ2 recombinant inbred line (RIL<sub>hf</sub>) with increased competitive fitness and exploration behavior than either parental strain

Previously, we developed an assay to estimate the competitive fitness difference between two strains. Briefly, two strains are directly competed against each other in standard laboratory conditions.
growth conditions (i.e. a single agar plate seeded with the OP50 strain of *E. coli* bacteria). Initially, 10 L4 hermaphrodite larva from each strain are transferred to the first plate where they are allowed to eat and reproduce until their grandchildren reach the L1 stage. At this point, ~1000 L1 larva are transferred to a new plate to eat and reproduce until their progeny reach the L1 stage. Subsequently, each generation, ~1000 L1 larva are transferred to a new plate for a total of five to seven generations depending on the experiment. In these conditions, outcrossing is minimized due to the low spontaneous rate of males (confirmed by observing the populations before their transfer to a new plate). The proportion of each strain is then estimated every other generation by isolating genomic DNA from the mixed population and using digital PCR with detection by fluorescent hydrolysis probes targeted to a specific allele pair that distinguishes the two strains. In general, either a naturally-occurring genetic difference between the two strains or a CRISPR-edited silent mutation in the *dpy-10* gene is used (listed in Materials and Methods). Finally, relative fitness is estimated by fitting a haploid model to the measured allele frequencies. This assay is a more direct measure of competitive fitness in laboratory conditions than growth rate, fecundity, or other fitness-proximal traits that are often used in *C. elegans*.

To determine if additional LSJ2/N2 fixed mutations can affect fitness in N2-like laboratory conditions, we used a previously described panel of 89 recombinant inbred lines (RILs) between the CX12311 and LSJ2 strains [51]. CX12311 is a near isogenic line that carries ancestral *npr-1* and *glb-5* alleles from the CB4856 Hawaiian wild strain introgressed into an N2 background (Fig 1B—henceforth referred to as N2’). Using N2’ as a parental strain eliminates the fitness effect of the derived alleles of N2 *npr-1* and *glb-5* [56]. Using the competition assay described above (Fig 1C), we measured the competitive advantage of each of the RIL strains against the N2’ strain. A bimodal distribution of relative fitness values was observed in the RIL strains, suggesting that a single genetic locus accounted for the majority of the variation in the RIL strains (Fig 1D and S1 Table). Using these measured evolutionary fitness values for QTL mapping, we identified a single significant QTL on the right arm of Chromosome II centered over the *nurf-1* gene (Fig 1E). We previously have shown that *nurf-1* contains two fixed mutations from both the N2 and LSJ2 lineages that each affect animal’s fitness in N2-like laboratory conditions [54, 57]. These results are consistent with our QTL analysis, suggesting that *nurf-1* plays an important role in adapting to laboratory conditions.

Interestingly, we found that one of the 89 RILs, CX12348 (henceforth called RIL$_{hf}$ for high fitness), had significantly higher fitness than either of the LSJ2 or N2’ parental strains, which we validated in an independent competition experiment (Fig 1D and Fig 2A). RIL$_{hf}$ contained a mixture of DNA from both the N2’ and LSJ2 parental strains, with N2’ DNA on the left arm of chromosome I, the entire chromosomes II, III, and V, and portions of the X chromosome (Fig 2B). The higher fitness of RIL$_{hf}$ strain could be caused by two possible reasons: 1) higher-order epistatic interaction between three or more of the 300 derived alleles or 2) a *de novo* beneficial mutation that occurred during construction of the RIL panel. We decided to focus on this unusual strain to determine the genetic basis of its higher fitness.

Wild strains of *C. elegans*, as well as the N2’ and LSJ2 parental strains, feed in groups on the borders of bacterial lawns, a strategy known as social behavior [52]. While growing the RIL$_{hf}$ strain on standard plates, we noticed that animals had a stronger propensity to explore the centers and regions outside of the bacterial lawns, causing an increased number of worms and tracks in the center and outside of the lawn (Fig 2B). It has been previously shown that increases in exploration behavior is caused by changes in the relative time *C. elegans* spends in roaming and dwelling states in the presence of O$_2$ and other chemical gradients created by the bacteria [58, 59]. To quantify this behavioral difference, we modified a previously described exploration assay [60] to measure long-term (16hrs) exploration behavior in the presence of
circular lawns (instead of uniform lawns used in standard assays). The RIL_{hf} strain explored a substantially larger fraction of the bacterial lawn than either of the parental strains (Fig 2C).

Mapping the causal mutation responsible for higher exploration behavior in RIL_{hf}

The change in exploration behavior is potentially an adaptive strategy for RIL_{hf} animals to increase their evolutionary fitness in the laboratory or it might be a pleiotropic effect of the underlying genetic basis of this fitness gain. Since this exploration trait is easier to assay than relative fitness, we first focused on mapping this phenotype using a bulk-segregant approach. We created two new small panels of 48 RILs between RIL_{hf} and either the N2 or LSJ2 parental strains and measured their exploration behavior (Fig 3A and 3B). An approximately equal number of RIL strains showed each parental phenotype, suggesting that this trait was controlled by a single locus. We grouped these strains into low or high exploration groups for each RIL panel and performed pooled genomic sequencing (~60x coverage) on the four groups that were created (Fig 3C). For each group, we estimated the allele frequency of each of the ~300 N2/LSJ2 genetic variants across the genome. In bulk-segregant analysis, the genetic loci which are not responsible for the exploration behavioral difference are expected to have approximately equal N2/LSJ2 allele proportion. The genetic loci that contribute to exploration behavioral difference are expected to show a larger difference pattern of N2/LSJ2 allele proportion. By analyzing the sequencing result, a large allelic imbalance between the pooled sequencing groups was observed in the center of chromosome III in the RIL_{hf} x LSJ2 panels (Fig 3D). This result is expected if a de novo mutation arose and fixed in the RIL_{hf} strain in the center of chromosome III.
chromosome III (which contains the N2 haplotype for the entire chromosome). In this scenario, a similar imbalance would occur for the de novo, causal variant in the RILhf x N2⁺ cross, however, because the two strains are largely identical on chromosome III, we could not observe it using the LSJ2/N2 SNVs. In addition to the center region of chromosome III, we also detected a large allele frequency difference on the center of chromosome V, suggesting that genetic variation on chromosome V also contributes to exploration behavior. However, the allelic imbalance on V was opposite as our expectation (i.e. the higher exploration group contained LSJ2 alleles on V while the RILhf strain contains N2 alleles on V). Since the chromosome III shows a stronger allelic imbalance signal than V (max = ~0.8 vs ~0.6) and goes in the expected direction, we focused on identifying the causal genetic variation in this region.

Fig 3. Exploration behavior differences of the RILhf strain maps to the center of chromosome III. (A) To map the changes in RILhf exploration behavior and fitness, we generated two panels of RILs (n = 48) between the RILhf and N2⁺ strains or the RILhf and LSJ2 strains. (B) Each RIL was measured for exploration behavior using the assay shown in Fig 2C. Color coding shows how strains were combined into low (green) or high (orange) groups for bulk-segregant analysis. Uncolored RILs were not included. The histograms on the right displayed the distribution of RIL’s exploratory fraction. (9 replicates were performed for each RIL). (C) Overview of bulk-segregant approach using pooled genomic DNA to calculate LSJ2/N2 allele frequency. (D) Allele frequency for LSJ2/N2 genetic differences was calculated for each population. A large allelic frequency difference was observed on chromosomes III and V in the LSJ2/RILhf.
However, it is possible that variation on chromosome V also contributes to exploration, although potentially in a manner unrelated to the RILhf phenotype.

**A de novo complex genomic rearrangement is identified in the rcan-1 gene**

To determine if the RILhf strain contains any *de novo* mutations in the center region of chromosome III, we sequenced genomic DNA isolated from the RILhf, N2*, and LSJ2 strains using Illumina short read sequencing. Although we did not identify any *de novo* SNVs or small indels on chromosome III in the RILhf strain, we did identify a large increase in coverage (2x – 8x) in the rcan-1 gene, which is an ortholog of human Down Syndrome gene RCAN1 ([Fig 4A](#)) [61]. This coverage increase was detected in the high exploration groups of both RIL panels, consistent with this genetic change causing increased exploration behavior ([Fig 4A](#)). The increased sequencing coverage suggests that the rcan-1 gene region has been amplified in the RILhf strain.

While a simple gene duplication event would cause an increase in coverage, we observed a non-uniform change in coverage across the affected region. We also identified a large number of chimeric or split reads (reads which partially align to two unique locations) that mapped to multiple locations within the rcan-1 locus ([S1 Fig](#)). The sequence of these chimeric reads within these groups were consistent with each other, and suggest that at least five new fusions between DNA sequence has occurred in the rcan-1 region of the RILhf strain. In other words, the rcan-1 genetic change consists of multiple inversion and/or duplication events. To resolve the precise mutation, we first attempted to amplify the entire affected region using PCR without success. As a complementary approach, we sequenced the RILhf strain using an Oxford Nanopore sequencing MinION, a long-read single molecule sequencing device with reported read lengths that could resolve the complex rearrangement [62]. By selecting reads that mapped to the rcan-1 region, we identified a single, ~34.5 kb long read that spanned the entire rcan-1 region ([Fig 4B](#) and [S1 Data](#)). This read resolved the large structural changes of this complex genomic rearrangement. By combining this long Nanopore long read with the DNA fusion events predicted by the Illumina short read sequencing, we resolved the complex rearrangement into five unique tandem inversions interspaced within the rcan-1 locus ([Fig 4C](#) and [S2–S4 Data](#)). This proposed rearrangement was consistent with other Oxford Nanopore reads that did not entirely span the rearrangement and resolved the coverage increase and chimeric reads from the Illumina short-read resequencing ([S2 Fig](#) and [S3 Fig](#)) as well as smaller PCR products that cover the new junctions ([S4 Fig](#) and [S2 Table](#)).

**rcan-1 complex genomic rearrangement is linked to changes in fitness and exploration behavior**

To determine if this rearrangement was responsible for the increases in exploration behavior and relative fitness of the RILhf strain, we created two near isogenic lines (NILs) by backcrossing the rcan-1 rearrangement from the RILhf strain into the N2* background ([Fig 5A](#)). Genomic DNA from these NILs was sequenced to confirm that LSJ2-derived DNA and RILhf-specific mutations besides the rearrangement were removed from both NILs ([S3 Table](#)). As expected, both of these NILs explored a higher fraction of the bacterial lawn ([Fig 5B](#)). Pairwise competition experiments between the NILs and the N2* strain also demonstrated that this rearrangement is associated with the increases in fitness ([Fig 5C](#)). Finally, we were interested in whether the rearrangement affected fitness-proximal traits such as body size, growth rate, or reproduction. We used a high-throughput COPAs worm sorter to demonstrate that the NIL animals were shorter than wild type controls ([Fig 5D](#)), indicating that at least one fitness-proximal trait (body length) was affected. However, we cannot say whether this difference in body
length is responsible for the change in fitness. These data support a causal role for the rcan-1 rearrangement for both the competitive fitness advantage and the exploration behavioral...
changes of the RILhf strain. However, we do not exclude a role for additional genetic mutations in regulating these phenotypes, such as genetic variation on chromosome V suggested by the bulk-segregant analysis of exploration behavior.

The rcan-1 rearrangement is predicted to cause a number of changes to the rcan-1 gene. First, it creates two full-length versions of the rcan-1 coding region (Fig 4C). However, the upstream region for each is modified by an inversion event in between the two coding regions.
The first and second copies of \textit{rcan-1} contain 857 and 1,725 bp of endogenous upstream sequence before the inversion event occurs. While the core promoter region is likely conserved in both of the full-length \textit{rcan-1} versions, enhancers and other regulatory regions are probably missing or perturbed in the rearranged region, which might cause decreased, increased, or ectopic expression of \textit{rcan-1} (e.g. in \textit{C. elegans}, at least 3kb of upstream DNA is typically used to estimate the expression pattern of a given gene). Indeed, analysis of previously published ChIP-seq data [63–65] identified 39 transcription factors that bind throughout the upstream region of \textit{rcan-1} (S5 Fig and S4 Table). Second, the second copy of the \textit{rcan-1} gene also contains a small inversion in the 3’ UTR region. This inversion could modify binding sites for small RNAs or other RNA-binding proteins that regulate the stability or translation of the mRNA product. Additionally, the 3’ end of the small inversion is fused to an upstream promoter region, consequently, the native transcriptional terminator is missing from the second full-length copy of \textit{rcan-1}. Finally, two truncated copies of the \textit{rcan-1} gene are also created, containing the first two exons of the gene. It is possible that truncated peptides with novel C-terminal fragments are produced from these copies and modify wildtype phenotype, although they lack the PxIxIT motif encoded by the last exon of \textit{rcan-1} that is required for RCAN-1 to bind with Calcineurin/TAX-6 [66]. It is difficult to predict \textit{a priori} which of these changes alone or in combination could cause the changes to exploration behavior and/or evolutionary fitness.

**The rearrangement of \textit{rcan-1} decreases its expression but is not a loss-of-function allele**

To gain insights into the transcriptional changes caused by the rearrangement, we used RNA-seq to compare the genome-wide expression differences between the two NIL strains and the N2* strain. Interestingly, the gene with the largest change in expression was \textit{rcan-1}, indicating that the rearrangement decreased transcription of the \textit{rcan-1} gene by about 75% (Fig 6A, S6 Fig, and S5 Table). In other contexts, gene duplications can modify phenotype by increasing gene dosage and expression; for the \textit{rcan-1} rearrangement, this is not the case.

Because the transcriptional profiling only reports whole-body changes in total \textit{rcan-1} expression, we created fusions of fluorescent proteins to the modified upstream regions of the different \textit{rcan-1} versions. We cloned the entire region between the two full-length versions of \textit{rcan-1} in both directions to create a P_{rcan-1-R1}::mCherry construct (reporting expression of the first full-length version of \textit{rcan-1} in the complex rearrangement) and a P_{rcan-1-R2}::mCherry construct (reporting expression of the second full-length version of \textit{rcan-1} in the complex rearrangement). As a control, we also cloned the first 5,085 bp of the upstream region from N2 and fused it to both GFP and mCherry (P_{rcan-1-WT}::GFP or P_{rcan-1-WT}::mCherry). We then simultaneously co-injected P_{rcan-1-WT}::GFP with P_{rcan-1-R1}::mCherry, P_{rcan-1-R1}::mCherry, or P_{rcan-1-R2}::mCherry (Fig 6B). Using a microfluidic device combined with confocal microscopy, we imaged whole-body expression from both green and red channels (Fig 6C). As expected from a previous publication, we observed wild-type expression of \textit{rcan-1} in a variety of tissues, including neurons, pharyngeal cells, and hypodermal cells [61]. We first measured how the modified upstream regions affected whole-body expression by measuring the total amounts of GFP and mCherry signals from ~30 animals for each promoter construct (Fig 6B). Both of the constructs from the complex rearrangement drove less mCherry expression than the wild-type construct to different extents, with the first upstream rearrangement more affected. The effect of the different constructs on mCherry fluorescence levels was also tissue-specific, which we measured by quantifying the appropriate anatomical regions. For example, the head fluorescence was significantly more affected then the body fluorescence in both constructs (Fig 6B).
Fig 6. The *rcan-1* rearrangement allele decreases expression of *rcan-1*. (A) A volcano plot of expression differences between the *rcan-1* NIL1 and N2* strains. RNA was isolated from synchronized, L4 animals. The gene with the largest and most significant expression decrease was *rcan-1*. Red: $p < 0.01$, log$_2$(Fold Change) $> 1$. Cyan: $p < 0.01$, log$_2$(Fold Change) $< -1$. (The list of differential expressed genes with significance are available in S5 Table). (B) Co-injection of wild-type *rcan-1* promoters driving GFP with wild-type *rcan-1* or rearranged *rcan-1* upstream promoter regions driving mCherry created from the RIL$_{AI}$ strain. Each dot represents the ratio of total GFP expression divided by total mCherry expression from a single animal. Fluorescence was also segmented into head or body expression and compared separately (**: $p < 0.001$, one-way ANOVA tests followed by Tukey's honest significant difference test. The top comparison group refers to the head fluorescence signal of the two truncated *rcan-1* promoters compared to head of wild-type *rcan-1*). (C) Representative fluorescence images of (B). The white arrow in the middle panel indicates the neurons that retain high level of expression of mCherry driven by $P_{rcan-1-R1}$. The white arrow in the bottom panel indicates the expression of mCherry driven by $P_{rcan-1-R2}$ in the head region. Scale bar is 100μm. (A: anterior; P: posterior). (D) Representative fluorescence images of the head showing a pair of interneurons with less affected expression. (These are different animals from C). The white arrow in the middle panel indicates the neurons that retain high level of expression of mCherry driven by $P_{rcan-1-R1}$. Scale bar is 50μm. (A: anterior; P: posterior).
Further, while the transcriptional reporter Prcan-1-R1 shows decreased fluorescence in the pharynx, fluorescence in two neurons are mostly unaffected. The cell bodies of these neurons are found in the retrovesicular ganglion and send a single process to the nerve cord. We tentatively identified these neurons as RIF or RIG. The reporter Prcan-1-R2 universally decreased the expression in the head (Fig 6D). Combined with the whole-body RNA-seq data, we suggest that rcan-1 expression is largely reduced in the RILhf strain because of changes to the upstream regions of the new versions of rcan-1. Potentially, there are cell-type specific changes in transcription, however, extrachromosomal arrays are composed of dozens to hundreds of copies of the promoter region that might not reflect the expression of the genomic promoter.

The above experiments suggest that the rcan-1 rearrangement could be beneficial because of a global reduction of rcan-1 transcription. However, an alternative hypothesis is simply that the rearrangement is beneficial because loss of rcan-1 activity is beneficial in laboratory conditions and the remaining residual expression is unrelated to the fitness of the animals. To test the second hypothesis, we used CRISPR-enabled genomic editing to delete rcan-1 in the N2+ strain (S7 Fig). This knockout strain showed an intermediate phenotype between the N2+ and the rcan-1 NIL strains in the modified exploration behavior (Fig 7A). When we competed this strain against the two rcan-1 NILs, we found that the rcan-1 rearrangement was substantially more fit than the rcan-1 deletion (Fig 7B). We also competed the strain containing the rcan-1 deletion against wild-type N2+ and found no significant difference in fitness (Fig 7C). These data are consistent with the rearrangement producing residual or ectopic expression of rcan-1 that is necessary for the fitness gains. We attempted to rescue the RILhf exploration phenotype using a transgene created from a PCR product amplified from the wildtype rcan-1 region, however, this construct was unable to rescue the exploration behavior (S8 Fig). There are a number of putative explanations for this result. Potentially, the transgene does not fully recapitulate wildtype expression of rcan-1, lacking upstream elements required for expression in cells necessary for the changes in the exploration behavior. Alternatively, additional genetic variants in RILhf also promote exploration behavior independent of the rcan-1 rearrangement.

While the changes in exploration behavior are genetically linked to the changes in fitness on laboratory plates, it is unknown whether the changes in exploration behavior are required for the fitness gains. To test this, we used plates seeded with a uniform bacteria lawn (UBL) across the entire plate. These plates lack the lawn borders that create O$_2$ gradients and suppress aggregation behavior of N2+ animals [56]. We competed RILhf against N2+ on UBL plates and found that RILhf still showed a fitness advantage (Fig 7D), indicating that the behavioral change is not solely responsible for the fitness gains. We also tested whether the RILhf strain consumed more food than the N2+ strain. We had previously found that a strain containing derived, beneficial alleles of npr-1 and glb-5 consumed more food on plates in an equal amount of time. However, the food consumption of the RILhf strain was statistically indistinguishable than the N2+ strain (S9 Fig).

Finally, in order to explore the role between gene dosage of rcan-1 and exploration behavior, we assayed heterozygotes between the RILhf NIL and rcan-1 deletion strains (Fig 7E). These experiments suggest that there is a strong relationship between rcan-1 dosage and exploration behavior, as the heterozygotes for each of these crosses were intermediate to the parental strains.

**Discussion**

In this study, we identified an outlier RIL with higher relative fitness than either parental strain. This RIL also displayed a new behavioral phenotype not seen in either parental strain, resulting in increased exploration activity on laboratory agar plates. By mapping this trait, we
identified a tandem set of inversion/duplications in the rcan-1 gene that seemed to influence both the exploration behavior and relative fitness of this RIL in standard laboratory conditions. A complex genomic rearrangement affecting phenotype have been found in C. elegans before [67], however, this was created in response to a chemical mutagen. The genomic rearrangement of rcan-1 was unexpectedly complex and provides insight into how gene duplication and rearrangement can occur in microevolutionary timescales.
Gene duplicates are thought to be a primary source of genetic material for the generation of evolutionary novelty, however, it is unclear how duplicates can arise and then navigate an evolutionary trajectory from redundancy to a state where both copies are maintained by natural selection as paralogs [68]. Two major issues in understanding how new gene copies evolve are understanding how gene duplicates initially spread through a population and the evolutionary forces responsible for functional differences in the two copies. Our work here suggests how both can occur due to a single mutational event. While some models of gene duplication have focused on the role of masking deleterious mutations or the role of genetic drift and purifying selection in spreading gene duplicates in a population [69, 70], our results suggest that positive selection can also be involved. After isolation from the wild, evolutionary mismatch between *C. elegans* and its laboratory environment resulted in N2 being at a point away from an adaptive peak. One route to increase its fitness was by changing *rcan-1* activity, which was accomplished by a complex genetic change that creates two duplicated copies of the *rcan-1* coding region. This complex genomic rearrangement was created naturally during the short RIL
construction period (10 generations). We propose this genomic rearrangement occurred as a single genomic instability event, potentially caused by the replication stress or mis-annealing during Okazaki fragment processing in DNA replication. The complex rearrangement might be a unique repair result induced by an initial error that activated a DNA replication checkpoint and the DNA repair machinery [71–73]. Although the RIL
strain shows a fitness advantage, the breeding pedigree of the RIL panel is designed to minimize fitness effects. We do not propose that the RIL
was selected for by positive selection despite its high fitness. However, our work demonstrates how the origin of new gene copies can provide a fitness advantage in new environments, where large functional changes to specific genes can be advantageous.

Our work also suggests how functional variation between two gene copies, a second major issue for understanding the evolution of paralogous genes, can arise in a single mutational step. The rearrangement of *rcan-1* causes large-scale changes to upstream noncoding regions, 3’UTR regions, and the creation of two truncated versions of *rcan-1* coding sequence. For short evolutionary timescales, this type of genetic variant could potentially access changes to gene function that would be difficult for a single SNV, insertion-deletion, or tandem duplication to cause. For example, besides changing the upstream promoter region that determines the exact levels and tissues the *rcan-1* gene is expressed in, the complex rearrangement is also predicted to create an *rcan-1* mRNA with a modified 3’U TR, potentially modifying its translational regulation or mRNA stability. Due to the differences in promoter region and 3’ UTR, it is possible that the two copies of *rcan-1* are not functionally redundant because they may have different expression levels or tissue-specific expression. It will be interesting to determine the precise amounts of protein that are produced by each copy and whether deleting each of these copies of *rcan-1* has a negative effect on fitness.

Our data indicates that the rearrangement reduces expression of *rcan-1*. Further, analysis of heterozygotes suggests that exploration behavior is sensitive to gene dosage of *rcan-1*. Our working model is that changes in expression of *rcan-1* is responsible for the changes to exploration behavior. Exploration is controlled by a distributed neural circuit [60, 74, 75]. Modifying *rcan-1* activity in these neurons could be responsible for the behavioral changes.

*rcan-1* encodes an ortholog of the human *RCAN1* gene [66], which encodes a calcipressin family protein that inhibits the calcineurin A protein phosphatase [76]. In humans, *RCAN1* plays an important role in human health; it has been proposed to be a key contributor to Down Syndrome phenotypes in patients with trisomy 21 [76, 77] and chronic overexpression of *RCAN1* in mice results in phenotypes related to Alzheimer’s disease [78]. In *C. elegans*, *rcan-1* is required for memory of temperature exposure through a *tax-6*/calcineurin-family and *crh-1*/CREB-dependent pathway [66]. Thermotaxis, however, is not predicted to be
important for laboratory fitness, and it is likely that the rcan-1 rearrangement regulates other unknown aspects of C. elegans biology on which selection can act. Unlike the standard N2 strain, which is potentially more fit in laboratory environments due to its ability to consume more food than the N2\(^*\) strain, a strain containing the rcan-1 rearrangement showed no difference in food consumption compared to the N2\(^*\) strain. However, we found that animals that carry the rcan-1 rearrangement were shorter than the N2\(^*\) strain. rcan-1 was previously shown to regulate body size using loss-of-function mutations [79]. It should be interesting to determine the exact phenotypes that are responsible for the gains in fitness in laboratory conditions. While an increasing number of causal genetic variants that modify phenotype and fitness are being identified, few examples demonstrating the exact phenotypes responsible for fitness changes have been worked out.

It will be interesting to study the continued evolution of a strain carrying this rearrangement, as it is unlikely that this strain has reached its adaptive peak in a single mutational step. Will additional beneficial mutations act through rcan-1? One possibility is that cis-regulatory mutations could fine tune the expression of each copy of rcan-1 in causal tissues. These mutations could act to further diversify the function of each copy of rcan-1. Alternatively, one of the duplicated rcan-1 copies could be subsequently lost, as seen in experimental evolution of poxviruses [40].

As long-read sequencing technology improves, the ability to identify complex structural variants similar to the one that we described here will increase. It will be interesting to see how often these types of variants survive the actions of purifying and positive selection to become common in natural populations of C. elegans and other animals.

**Materials and methods**

**C. elegans growth conditions**

Animals were cultivated on standard nematode growth medium (NGM) plates containing 2\% agar seeded with 200 \(\mu\)L of an overnight culture of the E. coli strain OP50 [50]. Ambient temperature was controlled using an incubator set at 20˚C. Strains were grown for at least three generations without starvation before any experiments were conducted.

**Strains**

The following strains were used in this study. For each figure, a list of strains used is included in S1 Table.

Near isogenic lines (NILs):

- CX12311 (N2\(^*\))—kyIR1(V, CB4856\(>\)N2), qgIR1(X, CB4856\(>\)N2)
- PTM413 (rcan-1 NIL 1) kahIR16(III, CX12348\(>\)N2), kylIR1(V, CB4856\(>\)N2), qgIR1(X, CB4856\(>\)N2)
- PTM414 (rcan-1 NIL 2), kahIR17(III, CX12348\(>\)N2), kylIR1(V, CB4856\(>\)N2), qgIR1(X, CB4856\(>\)N2)

Recombinant inbred lines (RILs):

- CX12311—LSJ2 RILs: CX12312-19, CX12321-27, CX12346-52, CX12354-60, CX12362-66, CX12368-75, CX12381-88, CX12414-37, CX12495-99, CX12501-08, CX12510, CX12361
- CX12311—CX12348 (RIL\(_{100}\)) RILs: PTM378-397, PTM421-434, PTM494-503
- LSJ2—CX12348 (RIL\(_{100}\)) RILs: PTM435-478

CRISPR-generated knockout and barcoded strains:

- PTM505: dpy-10 (kah83) II, rcan-1(kah183) III, kylIR1(V, CB4856\(>\)N2), qgIR1(X, CB4856\(>\)N2)
- PTM288: kylIR1(V, CB4856\(>\)N2), qgIR1(X, CB4856\(>\)N2) dpy-10(kah83)II;
Extrachromosomal array strains:
PTM553 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx169[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::WP::mCherry 25ng/µL] Isolate 1.
PTM544 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx170[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::WP::mCherry 25ng/µL] Isolate 2.
PTM555 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx171[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::WP::mCherry 25ng/µL] Isolate 3.
PTM556 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx172[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::R2::mCherry 25ng/µL] Isolate 1.
PTM557 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx173[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::R2::mCherry 25ng/µL] Isolate 2.
PTM559 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx174[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::R2::mCherry 25ng/µL] Isolate 3.
PTM560 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx175[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::R1::mCherry 25ng/µL] Isolate 1.
PTM561 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx176[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::R1::mCherry 25ng/µL] Isolate 2.
PTM562 kyIR1(V, CB4856>N2), aqIR1(X, CB4856>N2), kahEx177[P_rcan-1::WP::GFP 25ng/µL; P_rcan-1::R1::mCherry 25ng/µL] Isolate 3.
PTM563 CX12348 kahEx185[50ng/ul Prcan-1; 45ng/ul pSM; 5ng/ul pCFJ90] PTM564 CX12348 kahEx185[50ng/ul Prcan-1; 45ng/ul pSM; 5ng/ul pCFJ90] PTM565 CX12348 kahEx185[50ng/ul Prcan-1; 45ng/ul pSM; 5ng/ul pCFJ90] PTM566 CX12348 kahEx185[50ng/ul Prcan-1; 45ng/ul pSM; 5ng/ul pCFJ90] PTM567 CX12348 kahEx185[50ng/ul Prcan-1; 45ng/ul pSM; 5ng/ul pCFJ90]

Strain construction
To create the CX12311-CX12348 and LSJ2-CX12348 RILs, CX12311 males or LSJ2 males were crossed to CX12348 hermaphrodites. 96 F2 progeny (48 from CX12311 x CX12348; 48 from LSJ2 x CX12348) were cloned to individual plates and allowed to self-fertilize for 10 generations to create the inbred lines. One RIL line was lost from the LSJ2xCX12348 cross, creating 47 RILs.

To create the rcan-1 NILs (PTM413 and PTM414), CX12348 animals were backcrossed to CX12311 for 10 generations. Two completely independent sets of crosses were used to create two independent lines. Primers used to identify male animals containing the rearrangement were: 5’—ggacaataactctatgattaccaccca -3’ and 5’—gctgacacagcaatggtgacca -3’.

To create the rcan-1 deletion strain (PTM505), two sgRNAs targeting the 5’ region of rcan-1 and two sgRNAs targeting the 3’ end of rcan-1 were created: sgRNA1: 5’-atttggaagataacttcttt---TGG-3’; sgRNA2: 5’-catgactgacatgtc-----TGG-3’; sgRNA3: 5’-ctgtggatctttcatgacca-TGG-3’; sgRNA4: 5’-tcctagcaagaagccc-TGG-3’. CoCRISPR [80] was used to simultaneously edit the dpy-10 and rcan-1 genes using the following injection mix: 50ng/µL Pefl-3::Cas9, 10ng/µL dpy-10 sgRNA, 25ng/µL of each of the four rcan-1 sgRNAs, and 500nM dpy-10(c664) repair oligonucleotide. This mix was injected into CX12311 animals and Dpy or Rol animals were singled and genotyped using PCR. An animal with the deleted sequence 5’-catgacatgacca-3’ was identified.

To create the GFP/mCherry extrachromosomal lines (PTM553-PTM561), four constructs were created. P_rcan-1::WP::mCherry was created by inserting the rcan-1 promoter from CX12311 genomic DNA using primers 5’-ctgGGCCGGGCGCGGCATATCTG-3’ and 5’-ttGGGCGGGGCGCGCGTCGCACTTTG-3’ cloning into the pSM-GFP backbone with restriction enzyme sites 5’-FseI and 3’-Ascl. To create P_rcan-1::WP::mCherry, the rcan-1 promoter was amplified from CX12311 genomic DNA using primers 5’-tcggttcaaatctctgagacca-3’ and 5’-ttttgtgatataaaaaatttcagcaacca-3’ and a pCFJ90-mCherry backbone was
amplified using primers 5’- attttttctataagtttaacaaca aaaaAcaagtttgtacaaaaaagcaggct-3 and 5’- ccagtattgatgaagcgtgctgttctgaat-3’. The two fragments were assembled using HI-FI assembly (NEB E5520S). To construct the Prcan-1-R1::mCherry and Prcan-1-R2::mCherry plasmids, 5’- tttttgttgttaacttatagaaaaat ttcagca-3’ and 5’- gaaacgaaacaaggtgggt cc-3’ or 5’- tttttgttgttaacttatagaaaaat ttcagca-3’ and 5’- agcggacccaccttgtttc-3’ were used to amplify the rearranged promoters from CX12348 genomic DNA. These PCR products were cloned into a pCFJ90-mCherry backbone using HI-FI assembly. Concentrations of each plasmid are indicated for each strain in the strain description.

**Competition experiment**

Competition experiments were performed as described previously [56]. In the standard assays, 9 cm NGM plates were seeded with 300 μL of an overnight E. coli OP50 culture and incubated at room temperature for three days. In the competition experiment using uniform bacteria lawn (UBL), the 9cm UBL plates were made by pouring overnight E. coli OP50 culture onto the NGM plate to cover the whole plate. Excess culture on the plate was removed by pouring off and the plates were left at 20˚C overnight for forming uniform bacteria lawn. In each competition experiment, ten L4 larvae from each strain were picked onto a single plate and cultured for five days. Animals were transferred to identically prepared NGM plates and subsequently transferred every four days. Depending on the experiment, five or seven total transfers were performed. For each transfer, animals were washed off the plates using M9 buffer and collected into 1.5 mL centrifuge tubes. The animals were then mixed by inversion and allowed to stand for approximately one minute to settle adult animals. 50 μL of the supernatant containing approximately 1000–2000 L1-L2 animals were seeded onto fresh plates. The remaining animals were concentrated and used for genomic DNA isolation. Genomic DNA was collected every odd generation using a Zymo DNA isolation kit (D4071). To quantify the relative proportion of each strain, a digital PCR assay was performed with custom TaqMan fluorescent-quenching probes (Applied Biosciences). Genomic DNA was digested with SacI or EcoRI for 30 min at 37 °C. The digested products were purified using a Zymo DNA cleanup kit (D4064) and diluted to approximately 1–2 ng/μL. Seven TaqMan probes were designed using ABI software that targeted WBVar00051876, WBVar00601322, WBVar00167214, WBVar00601493, WBVar00601538, dpy-10 (kah82), or tbc-10(kah185) (S6 Table). Digital PCR assays were performed using a Biorad QX200 digital PCR machine with standard probe absolute quantification protocol. The relative allele proportion was calculated for each DNA sample using the count number of the droplet with fluorescence signal (Eq 1). To calculate the relative fitness of the two strains using three or four measurements of relative allele proportion, we used linear regression to fit this data to a one-locus generic selection model (Eqs 2 and 3), assuming one generation per transfer.

\[
P(A)_t = \frac{\text{No. Allele } A}{\text{No. Allele } A + \text{No. Allele } a}
\]

\[
P(A)_t = \frac{P(A)_0 W_{AA}^t}{P(A)_0 W_{AA}^t + (1 - P(A)_0) W_{aa}^t}
\]

\[
\log\left(\frac{P(A)_t}{1 - P(A)_t}\right) = \left(\log\left(\frac{W_{aa}}{W_{AA}}\right)\right) t
\]

The relative fitness value and Taqman assay information for each competition experiment are included in S1 Table.
Exploration behavioral assay

The exploration assays from Flavell et al. were modified to study exploration in the presence of circular lawns [60]. 35 mm Petri dishes were seeded with 150 μL OP50 E. coli Bacteria for 24 h before the start of the assay. Individual L4 hermaphrodites were placed in the center of the plate and cultivated in 20˚C for 16 hours. The plates were placed on a grid that has 100 squares that cover the whole bacteria lawn. To calculate the exploration fraction, the number of full or partial squares that contained animal’s tracks out of bacteria lawn border was quantified. The number of full or partial squares that contain the bacteria lawn was also counted (about 94–96 grids). The exploration fraction was calculated (Eq 4).

\[
\text{Exploration fraction} = \frac{\text{No. grids contained tracks}}{\text{No. grids contained bacteria lawn}}
\] (Eq 4)

Heterozygous exploration assay

Heterozygous F1 was created by mating PTM288 males with the other strain of interest at L4 stage for one day. Fertilized hermaphrodites were then single into individual plates. After two days, L4 hermaphrodites were picked from plates where a lot of males were present, indicating successful mating. After the assay, these animals are individually lysed and genotyped at dpy-10(kah83)II site to confirm they are heterozygous. The genotyping primers: Forward primer: 5’–gtcagatgatctaccggtgtgtcac —3’, reverse primer: 5’–gtctctcctgttgctccgtcttcac– 3’.

rcan-1 rescue assay

A PCR fragment that covers 4.5kb upstream to 0.7kb downstream of rcan-1 was cloned using NEB Phusion Q5 PCR system (Forward primer: 5’–gtctctagaagctctgcattctc– 3’, reverse primer: 5’–gtctctcagcttgctccgtcttcac– 3’). The PCR product was purified and injected at 50ng/μL with 5ng/μL f pCFJ90 and 45ng/μL pSM. The exploration behavior fraction of the animals expressing mCherry was quantified using standard exploration behavior assay method.

Bulk-segregant analysis of exploration behavior

The exploration behavioral assays were performed on 48 CX12311-CX12348 RILs and 47 LSJ2-CX12348 RILs. In the CX12311/CX12348 RILs, 28 RILs with median exploration fraction less than 0.575 were assigned to the low exploration group and the 20 RILs with median exploration fraction greater than or equal to 0.575 were assigned to the high exploration group. In the LSJ2/CX12348 RILs group, the 17 RILs with median exploration fraction less than 0.620 were assigned to the low exploration group, the 20 RILs with median exploration behavior greater than or equal to 0.870 were assigned to high exploration group, and the rest of the RILs were excluded from further analysis. Genomic DNA from each RIL (100 ng) was isolated and pooled into the four described groups for whole-genome resequencing.

Whole-genome sequencing

Genomic DNA was isolated using Qiagen Gentra Puregene Kit (158667) following the supplementary protocol for nematodes. The genomic DNA was further purified using Zymo Quick-DNA kit (D4068). DNA libraries were prepared using an Illumina Nextera DNA kit (FC-121-1030) with indexes (FC-121-1011). The prepared libraries were sequenced at 35 bp or 150 bp paired-read using an Illumina NextSeq 500. The reads were aligned to reference genome using BWA-aligner v0.7.17 [81]. BAM files were deduplicated and processed using SAMtools v1.9 [82] and Picard[83] (http://broadinstitute.github.io/picard/). SNVs were called by Freebayes and annotated by SnpEff [84, 85]. Custom Python scripts using the pysam library.
(https://github.com/pysam-developers/pysam) were used to identify regions of the genome with a large number of clipped and chimeric reads. Reads depths were visualized using IGV [86]. The sequencing reads were uploaded to the SRA under BioProject PRJNA526525.

**Oxford Nanopore long-read sequencing**

Genomic DNA of CX12348 was isolated from animals grown on 89 cm NGM plates using Qiagen Gentra Puregene Kit (158667) following the supplementary protocol for nematodes. The genomic DNA was concentrated and purified using Zymo Quick-DNA kit (D4068). Size-selection to collect DNA fragments from 10 kbp–50 kbp was carried out using a Blue-pippin. The sequencing library was prepared using 1D ligation kit (SQK-LSK108) following the standard protocol. DNA was repaired using the NEBNext FFPE Repair Mix (M6630). After DNA repair, end preparation was performed and the adapter was ligated. 600 ng prepared library was loaded in the Nanopore R9 flow cell in MinION sequencer. The standard 48 hours sequencing protocol was performed and approximately 5 Gb of sequencing data was generated. To resolve the structure of rcan-1 complex rearrangement, the FASTQ files were aligned to reference genome using BWA aligner. Reads that covered the rcan-1 gene region and contained a gap in alignment were fetched using pysam (https://github.com/pysam-developers/pysam). These reads were then mapped to rcan-1 using BLAST and visualized with matplotlib (https://matplotlib.org) to show the rearrangement events. The structure of the complex rearrangement was verified by using BWA and IGV to map the Illumina short reads or FlexiDot [87] to map and visualize the Oxford Nanopore reads. The sequencing reads were uploaded to the SRA under BioProject PRJNA526525.

**RNA-seq and transcriptome analysis**

CX12311, PTM413, and PTM414 were synchronized using alkaline-bleach to isolate embryos, which were washed with M9 buffer and placed on a tube roller overnight. Approximately 400 hatched L1 animals were placed on NGM agar plates for each strain and incubated at 20°C for 48 hours. The ~L4 stage animals were washed off for standard RNA isolation using Trizol. Four replicates for each strain were performed on different days. The RNA libraries were prepared using the NEB Next Ultra II Directional RNA Library Prep Kit (E7760S) following its standard protocol. The libraries were sequenced by Illumina NextSeq 500. The reads were aligned by HISAT2 using default parameters for pair-end sequencing. Transcript abundance was calculated using HTseq and then used as inputs for the SARTools [88, 89]. edgeR v3.16.5 was used for normalization and differential analysis[55][90]. The analysis result was shown in a volcano plot. CX12311 was treated as the wild type. The genes show significant differential expression in the volcano plot are under thresholds | log2(fold) | > 1 and FDR adjusted p-value < 0.01. Sequencing reads were uploaded to the SRA under BioProject PRJNA526525.

**Imaging**

The detailed steps of microfluidic device fabrication were previously reported [91]. For each experiment, about 100–150 animals were suspended in 1 mL of S Basal and delivered into using a syringe. Animals were immobilized using 1 mL of tetramisole hydrochloride (200 mM) (Sigma-Aldrich cas. 5086-74-8) in S Basal. Imaging were acquired on a spinning disk confocal microscope (PerkinElmer UltraVIEW VoX) with a Hamamatsu FLASH 4 sCMOS camera. Images of the animals were quantified using ImageJ. A region-of-interest (ROI) was drawn around the entire worm, and the mean intensity of the GFP and mCherry images were calculated across the ROI. Relative fluorescence intensity was calculated as (Mean Intensity of mCherry)/(Mean Intensity of GFP).
Food consumption assay

The experimental method was described previously [56]. In brief, The 24-well plates were prepared by pipetting 0.75mL NGM agar contain 25 μM FUDR and 1x Antibiotic-Antimycotic (ThermoFisher 15240062) to each well. Each well was seeded with 20μL of freshly cultured OD600 of 4.0 (CFU ~ 3.2×10^9/mL) E. coli OP50-GFP(pFPV25.1). The plates were dried in a fume hood and dried with air flow for 1.5hr. The fluorescence signal of OP50-GFP was quantified by area scanning protocol using BioTek Synergy H4 multimode plate reader. The synchronized L4 animals were placed in the wells in the first five columns and the last column is used as control column. Each well was placed with 10 animals, and the plate was incubated in a 20˚C incubator for 18 hours and the fluorescence signal was quantified again as the ending time point. The relative food consumption amount was calculated using the equations reported previously [56].

High-throughput growth rate analysis

The high-throughput growth rate and brood size assays were performed as described previously [92]. In short, approximately 25 bleach-synchronized embryos were aliquoted into each well of 96-well plates, and fed 5 mg/mL HB101 bacterial lysate on the following day [93]. After 48 hours of growing at 20˚C, 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 with 50 μL of K medium with HB101 lysate (10 mg/mL) and Kanamycin (50 μM). Animals were grown for 96 hours at 20˚C and were then treated with sodium azide (50 mM in M9). Animal number (n) and animal length (time of flight, TOF) were measured by the BIOSORT. For each well, animal growth was measured as the median length of the population, and brood size was measured as the number of progeny per sorted animal. The experiments were replicated in two independent assays, and the linear model with the formula (phenotype ~ assay) was applied to normalize the differences among assays [94].

Statistical test

The raw data are included in S1 Table. To assess statistical significance, we performed one-way ANOVA tests followed by Tukey’s honest significant difference test to correct for multiple comparisons or the Wilcoxon-Mann-Whitney nonparametric test for pairwise comparisons. NS: not significant; *: p < 0.05; **: p < 0.01; ***: p < 0.001.

QTL mapping

The average of the log_2(w) of each N2'/LSJ2 RIL was used as phenotype with 192 previously genotyped SNPs. R/qtl was used to perform a one-dimensional scan using marker regression on the 192 markers. The genome-wide error rate (p = 0.05) was determined by 1000 permutations test[95].

List of key resources and reagents

The key resources and reagents used in this study are listed in S6 Table.

Supporting information

S1 Fig. Illumina reads mapped to the rcan-1 locus. (A) IGV plot of illumina sequencing short reads align to rcan-1 genomic locations. (B) Chimeric reads align to rcan-1 genomic locations. Reads are from the resequencing of the N2’(CX12311), LSJ2, and RIL_hf (CX12348) strain. Besides an increase in coverage at the rcan-1 locus, a large number of chimeric reads
(i.e. reads that partially map to two locations) were found in the RIL$_{hf}$ strain. (Reads with grey color indicates they are normal reads (Pair orientations: LR); Reads with cyan color imply inversion (Pair orientations: LL); Reads with blue color imply inversion (Pair orientations: RR); Reads with green color imply duplication or translocation (Pair orientations: RL). Reads with red color have larger than expected inferred sizes.)

**S2 Fig. Dot plot of the nanopore sequencing reads align to proposed rcan-1 rearrangement.** 10 nanopore sequencing reads that overlapped the rcan-1 structural variant were used to generate a dot plot with proposed rcan-1 rearrangement.

**S3 Fig. Illumina short sequencing reads aligned to the proposed rcan-1 structural variant.** Top: All reads aligned to the rcan-1 rearrangement. Bottom: Chimeric reads aligned to the rcan-1 rearrangement. The uniform coverage and lack of chimeric reads is consistent with the proposed structure of the rearrangement. (Reads with grey color indicates they are normal reads (Pair orientations: LR); Reads with cyan color imply inversion (Pair orientations: LL); Reads with red color have larger than expected inferred sizes. Reads with empty color have low mapping quality.)

**S4 Fig. The PCR products include the rearranged regions.** Red arrows are the PCR products that include the rearranged regions. The detail information of the primers, the expected length and observed length in agarose gel of each PCR product is listed in S2 Table.

**S5 Fig. Transcription factor binding regions at rcan-1 5'-UTR.** The green bars represent the transcription factor binding region. The red bars represent the two truncated promoter regions that drive full length of rcan-1 gene body in the complex rearrangement. The blue bar represents the highly occupied target region (‘HOT’). The figure is generated from Wormbase J-browser by adding the feature of transcription factor binding regions. The information of the transcription factors is listed in S4 Table.

**S6 Fig. Volcano plot of rcan-1 NIL2 gene expression vs. N2*.** Red dots indicate genes with increased expression in rcan-1 NIL2 vs. N2* (p < 0.01, log2(Fold Change) > 1). Cyan dots indicate genes with decreased expression in rcan-1 NIL2 vs. N2* (p < 0.01, log2(Fold Change) < -1). The list of differential expressed genes with significance are available in S5 Table.

**S7 Fig. Strategy for creating a knockout allele of rcan-1 using CRISPR/Cas9.** The position of two pairs of sgRNAs that target the 5’ and 3’ end of the rcan-1 coding region. The resulting deletion allele is shown as a blue box.

**S8 Fig. Exploration fraction of rcan-1 rescue lines.** The RIL$_{hf}$ animals were co-injected with 50ng/µL $P_{rcan-1(4.5Kbp)}$:rcan-1 PCR product, 5ng/µL pCFJ90, and 45ng/µL pSM. The exploration fraction of the animals that express mCherry were measured.

**S9 Fig. Food consumption assay of RIL$_{hf}$ and rcan-1 NILs.** Relative food consumption of indicated strains. Each dot indicates one experimental replicate.
S1 Data. *rcan-1_NanoporeReads.txt*. This file contains the sequence of the Oxford Nanopore reads (Fig 4B and S2 Fig) that overlap the structural variant in fasta format.

S2 Data. *rcan-1_RearrangementSequence.txt*. This file contains sequence information of the proposed *rcan-1* structural variant in fasta format.

S3 Data. *rcan-1_RearrangementSequence.txt*. This file contains annotated gene and junction information for the structural variant in Genbank format.

S4 Data. *rcan-1_RearrangementSequence.dna*. This file annotated gene and junction information for the structural variant in SnapGene format. It contains the primer information for study the structural variant. This file can be viewed by SnapGene software or SnapGene Viewer software (SnapGene Viewer is a free software).

S1 Table. Raw data. This table includes the raw experimental data of Figs 1–7 and S8 Fig and S9 Fig.

S2 Table. Rearranged junction sequences. This table includes the junction sequences for the *rcan-1* structural variant. The primer’s information and the information of each PCR product’s size are also included.

S3 Table. NIL resequencing. This table includes all genetic variants identified in the *rcan-1* near isogenic lines (NILs).

S4 Table. TF binding regions in 5 UTR. This table summarizes the transcription factor binding information at *rcan-1* 5’ upstream region from Wormbase.

S5 Table. NIL_RNA-Seq. This table includes all gene expression data for *rcan-1* NILs.

S6 Table. Sequence information of TaqMan probes and summary of resources and reagents. This table lists sequence information for the TaqMan fluorescent quenching probes used for competition experiments. This table also includes the information of key resources and reagents used in this study.

Acknowledgments

We thank the *Caenorhabditis* Genetics Center for strains, Todd Streelman, Levi Morran, Chao Jiang, Wei Zhang, Will Ratcliff, Annalise Paaby, and members of the Streelman and McGrath lab for discussions, and WormBase.

Author Contributions

Conceptualization: Yuehui Zhao, Patrick T. McGrath.
Data curation: Yuehui Zhao, Lijiang Long, Jason Wan, Shannon C. Brady, Daehan Lee, Akinade Ojemakinde, Fredrik O. Vannberg, Hang Lu.

Formal analysis: Yuehui Zhao, Lijiang Long, Shweta Biliya, Erik C. Andersen, Patrick T. McGrath.

Funding acquisition: Erik C. Andersen, Patrick T. McGrath.

Investigation: Lijiang Long, Shweta Biliya, Patrick T. McGrath.

Methodology: Lijiang Long, Jason Wan, Fredrik O. Vannberg, Hang Lu.

Project administration: Patrick T. McGrath.

Resources: Fredrik O. Vannberg.

Supervision: Erik C. Andersen, Hang Lu.

Validation: Yuehui Zhao, Patrick T. McGrath.

Visualization: Yuehui Zhao, Lijiang Long, Jason Wan, Shannon C. Brady, Daehan Lee, Patrick T. McGrath.

Writing – original draft: Yuehui Zhao, Patrick T. McGrath.

Writing – review & editing: Yuehui Zhao, Erik C. Andersen, Patrick T. McGrath.

References

1. Maydan JS, Lorch A, Edgley ML, Flibotte S, Moerman DG. Copy number variation in the genomes of twelve natural isolates of Caenorhabditis elegans. BMC Genomics. 2010; 11:62. Epub 2010/01/27. https://doi.org/10.1186/1471-2164-11-62 PMID: 20103950; PubMed Central PMCID: PMC2822765.

2. Katju V, Bergthorsson U. Copy-number changes in evolution: rates, fitness effects and adaptive significance. Front Genet. 2013; 4:273. Epub 2013/12/26. https://doi.org/10.3389/fgene.2013.00273 PMID: 24368910; PubMed Central PMCID: PMC3857721.

3. Sudmant PH, Rauutsch 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(7571):75–81. Epub 2015/10/04. https://doi.org/10.1038/nature15394 PMID: 26432246; PubMed Central PMCID: PMC4617611.

4. Long E, Evans C, Chaston J, Udall JA. Genomic Structural Variations Within Five Continental Populations of Drosophila melanogaster. G3 (Bethesda). 2018; 8(10):3247–53. Epub 2018/08/17. https://doi.org/10.1534/g3.118.200631 PMID: 30111620; PubMed Central PMCID: PMC6169376.

5. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, et al. Characterizing the Major Structural Variant Alleles of the Human Genome. Cell. 2019; 176(3):663–75 e19. Epub 2019/01/22. https://doi.org/10.1016/j.cell.2018.12.019 PMID: 30661756; PubMed Central PMCID: PMC6438697.

6. Fuentes RR, Chebotarow D, Duitama J, Smith S, De la Hoz JF, Mohiyuddin M, et al. Structural variants in 3000 rice genomes. Genome Res. 2019; 29(5):870–80. Epub 2019/04/18. https://doi.org/10.1101/gr.241240.118 PMID: 30992303; PubMed Central PMCID: PMC6499320.

7. Lupski JR. Genomic rearrangements and sporadic disease. Nat Genet. 2007. 39(7 Suppl):S43–7. Epub 2007/09/05. https://doi.org/10.1038/ng2084 PMID: 17597781.

8. Chen JM, Cooper DN, Ferec C, Kehrer-Sawatzki H, Patrinos GP. Genomic rearrangements in inherited disease and cancer. Semin Cancer Biol. 2010; 20(4):222–33. Epub 2010/06/15. https://doi.org/10.1016/j.semcancer.2010.05.007 PMID: 20541013.

9. Martin CL, Kirkpatrick BE, Ledbetter DH. Copy number variants, aneuploidies, and human disease. Clin Perinatol. 2015; 42(2):227–42, vii. Epub 2015/06/05. https://doi.org/10.1016/j.clp.2015.03.001 PMID: 26042902; PubMed Central PMCID: PMC4459515.

10. Rice AM, McLysaght A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. Nat Commun. 2017; 8:14366. Epub 2017/02/09. https://doi.org/10.1038/ncomms14366 PMID: 28176757; PubMed Central PMCID: PMC5309798.

11. Hieronymus H, Murali R, Tin A, Yadav K, Abida W, Moller H, et al. Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. Elife. 2018; 7. Epub 2018/09/05. https://doi.org/10.7554/elife.37294 PubMed Central PMCID: PMC6145837. PMID: 30176746
12. Chain FJ, Feulner PG. Ecological and evolutionary implications of genomic structural variations. Front Genet. 2014; 5:326. Epub 2014/10/04. https://doi.org/10.3389/fgen.2014.00326 PMID: 25278961; PubMed Central PMCID: PMC4165313.

13. Fan S, Meyer A. Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. Front Genet. 2014; 5:163. Epub 2014/06/12. https://doi.org/10.3389/fgene.2014.00163 PMID: 24917883; PubMed Central PMCID: PMC4042683.

14. Wellenreuther M, Merot C, Berdan E, Bernatchez L. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. Mol Ecol. 2019; 28(6):1203–9. Epub 2019/03/06. https://doi.org/10.1111/mec.15066 PMID: 30834648.

15. Lye ZN, Purugganan MD. Copy Number Variation in Domestication. Trends Plant Sci. 2019; 24(4):352–65. Epub 2019/02/13. https://doi.org/10.1016/j.tplants.2019.01.003 PMID: 30745056.

16. Dorshorst B, Molin AM, Rubin CJ, Johansson AM, Stromstedt L, Pham MH, et al. A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet. 2011; 7(12):e1002412. Epub 2012/01/05. https://doi.org/10.1371/journal.pgen.1002412 PMID: 22216010; PubMed Central PMCID: PMC3245302.

17. Cook DE, Lee TG, Guo X, Melfo S, Wang K, Bayless AM, et al. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science. 2012; 338(6111):1206–9. Epub 2012/10/16. https://doi.org/10.1126/science.1228746 PMID: 23065905.

18. Durkin K, Coppieters W, Drogemuller C, Ahariz N, Cambisano N, Druet T, et al. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature. 2012; 482(7383):81–4. Epub 2012/02/03. https://doi.org/10.1038/nature10757 PMID: 22297974.

19. Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, et al. Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat Genet. 2015; 47(8):944–8. Epub 2015/07/07. https://doi.org/10.1038/ng.3346 PMID: 26147619.

20. Yassin A, Delaney EK, Reddiex AJ, Seher TD, Bastide H, Appleton NC, et al. The pdm3 Locus Is a Hotspot for Recurrent Evolution of Female-Limited Color Dimorphism in Drosophila. Curr Biol. 2016; 26(18):2412–22. Epub 2016/08/23. https://doi.org/10.1016/j.cub.2016.07.016 PMID: 27546577; PubMed Central PMCID: PMC5450831.

21. Ohno S. Evolution by gene duplication. Berlin, New York,: Springer-Verlag; 1970. xv, 160 p. p.

22. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet. 2004; 38:615–43. Epub 2004/12/01. https://doi.org/10.1146/annurev.genet.38.072902.092831 PMID: 15568988.

23. Kunte K, Zhang W, Tenger-Trojan A, Palmer DH, Martin A, Reed RD, et al. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. Curr Biol. 2016; 26(3):344–50. Epub 2016/10/01. https://doi.org/10.1016/j.cub.2015.11.069 PMID: 26804558; PubMed Central PMCID: PMC4747794.

24. Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, et al. Chromosomal Rearrangements Shape Adaptation and Speciation: Case studies in Drosophila. Curr Biol. 2019; 29(6):1203–9. Epub 2019/06/06. https://doi.org/10.1017/pnas.0803151105 PMID: 19836166.

25. Blount ZD, Borland CZ, Lenski RE. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proc Natl Acad Sci U S A. 2008; 105(23):7899–906. Epub 2008/06/06. https://doi.org/10.1073/pnas.0803151105 PMID: 18524956; PubMed Central PMCID: PMC2430337.

26. Brown CJ, Todd KM, Rosenzweig RF. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol. 1998; 15(8):931–42. Epub 1998/08/27. https://doi.org/10.1093/oxfordjournals.molbev.a026009 PMID: 9718721.

27. Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, et al. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2002; 99(25):16144–9. Epub 2002/11/26. https://doi.org/10.1073/pnas.242624799 PMID: 12446845; PubMed Central PMCID: PMC3138579.

28. Brown CJ, Todd KM, Rosenzweig RF. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol. 1998; 15(8):931–42. Epub 1998/08/27. https://doi.org/10.1093/oxfordjournals.molbev.a026009 PMID: 9718721.

29. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet. 2003; 4(6):457–69. Epub 2003/05/31. https://doi.org/10.1038/nrg1088 PMID: 12776215.
Structural variation influences exploratory behavior and fitness in C. elegans

31. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, et al. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 2008; 4(12): e1000303. Epub 2008/12/17. https://doi.org/10.1371/journal.pgen.1000303 PMID: 19079573; PubMed Central PMCID: PMC2586090.

32. Kao KC, Sherlock G. Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. Nat Genet. 2008; 40(12): 1499–504. Epub 2008/11/26. https://doi.org/10.1038/ng.280 PMID: 19029969; PubMed Central PMCID: PMC2596280.

33. Kasahara T, Abe K, Mekada K, Yoshiki A, Kato T. Genetic variation of melatonin productivity in laboratory mice under domestication. Proc Natl Acad Sci U S A. 2010; 107(14): 6412–7. Epub 2010/03/24. https://doi.org/10.1073/pnas.0914399107 PMID: 20308563; PubMed Central PMCID: PMC2851971.

34. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature. 2015; 519(7542):181. https://doi.org/10.1038/nature14279 PMID: 25731169

35. Orozco-terWengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlotterer C. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol Ecol. 2012; 21(20):4931–41. Epub 2012/06/26. https://doi.org/10.1111/j.1365-294X.2012.05673.x PMID: 22726122; PubMed Central PMCID: PMC3533796.

36. Ratcliff WC, Denison RF, Borrello M, Travisano M. Experimental evolution of multicellularity. Proc Natl Acad Sci U S A. 2012; 109(5):1595–600. Epub 2012/02/07. https://doi.org/10.1073/pnas.1115321109 PMID: 223077146.

37. Rose MR. Artificial Selection on a Fitness-Component in Drosophila melanogaster. Evolution. 1984; 38(3):516–26. Epub 1984/05/01. https://doi.org/10.1111/j.1558-5646.1984.tb00317.x PMID: 2855975.

38. Stanley CE Jr., Kalathinal RJ. Genomic signatures of domestication on neurogenetic genes in Drosophila melanogaster. BMC Evol Biol. 2016; 16:6. Epub 2016/01/06. https://doi.org/10.1186/s12862-015-0580-1 PMID: 26728183; PubMed Central PMCID: PMC4700609.

39. Castagnone-Sereno P, Mulet K, Danchin EGJ, Koutsovoulos GD, Kauralick M, Da Rocha M, et al. Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode Meloidogyne incognita. Mol Ecol. 2019; 28(10):2559–72. Epub 2019/04/10. https://doi.org/10.1111/mec.15095 PMID: 30964953.

40. Elde NC, Child SJ, Eckbush MT, Kitzman JO, Rogers KS, Shendure J, et al. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell. 2012; 150(4):831–41. Epub 2012/08/21. https://doi.org/10.1016/j.cell.2012.05.049 PMID: 22901812; PubMed Central PMCID: PMC3499626.

41. Farslow JC, Lipinski KJ, Packard LB, Edgley ML, Taylor J, Filbotte S, et al. Rapid Increase in frequency of gene copy-number variants during experimental evolution in Caenorhabditis elegans. BMC Genomics. 2015; 16:1044. Epub 2015/12/10. https://doi.org/10.1186/s12864-015-2253-2 PMID: 26645535; PubMed Central PMCID: PMC4673709.

42. Lauer S, Aveclila G, Spealman P, Sethia G, Brandt N, Levy SF, et al. Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. PLoS Biol. 2018; 16(12):e3000069. Epub 2018/12/19. https://doi.org/10.1371/journal.pbio.3000069 PMID: 30562346; PubMed Central PMCID: PMC6298651.

43. Lauer S, Gresham D. An evolving view of copy number variants. Curr Genet. 2019. Epub 2019/05/12. https://doi.org/10.1007/s00294-019-00980-0 PMID: 31076843.

44. Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, et al. Development of a comprehensive genotype-to-phenotype map of adaptation-driving mutations in yeast. Cell. 2016; 166(6):1585–96. e22. https://doi.org/10.1016/j.cell.2016.08.002 PMID: 27594428

45. Vergara IA, Mah AK, Huang JC, Tariao-Graovac M, Johnsen RC, Baillie DL, et al. Polymorphic segmental duplication in the nematode Caenorhabditis elegans. BMC Genomics. 2009; 10:329. Epub 2009/07/23. https://doi.org/10.1186/1471-2164-10-329 PMID: 19622155; PubMed Central PMCID: PMC2728738.

46. Konrad A, Filbotte S, Taylor J, Waterston RH, Moyerman DG, Berghorthson U, et al. Mutational and transcriptional landscape of spontaneous gene duplications and deletions in Caenorhabditis elegans. Proc Natl Acad Sci U S A. 2018; 115(28):7386–91. Epub 2018/06/27. https://doi.org/10.1073/pnas.1801930115 PMID: 29941601; PubMed Central PMCID: PMC6048555.

47. Yoshimura J, Ichikawa K, Shoura MJ, Artiles KL, Gabdank I, Wahba L, et al. Recompleting the Caenorhabditis elegans genome. Genome Res. 2019; 29(6):1009–22. Epub 2019/05/28. https://doi.org/10.1101/gr.244830.118 PMID: 31123080; PubMed Central PMCID: PMC6581061.

48. Kim C, Kim J, Kim S, Cook DE, Evans KS, Andersen EC, et al. Long-read sequencing reveals intra-species tolerance of substantial structural variations and new subtelomere formation in C. elegans.
49. Sterken MG, Snoek LB, Kammenga JE, Andersen EC. The laboratory domestication of Caenorhabditis elegans. Trends Genet. 2015; 31(5):224–31. Epub 2015/03/26. https://doi.org/10.1016/j.tig.2015.02.009 PMID: 25804345; PubMed Central PMCID: PMC4417040.

50. Brenner S. The genetics of Caenorhabditis elegans. Genetics. 1974; 77(1):71–94. Epub 1974/05/01. PMID: 466476; PubMed Central PMCID: PMC1213120.

51. McGrath PT, Xu Y, Alion M, Garrison JL, Butcher RA, Bargmann CI. Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. Nature. 2011; 477(7364):321–5. Epub 2011/08/19. https://doi.org/10.1038/nature10378 PMID: 21849976; PubMed Central PMCID: PMC3257054.

52. 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(5):679–89. Epub 1998/09/19. https://doi.org/10.1016/s0092-8674(00)81609-8 PMID: 9741632.

53. Duveau F, Felix MA. Role of pleiotropy in the evolution of a cryptic developmental variation in Caenorhabditis elegans. PLoS Biol. 2012; 10(1):e1001230. Epub 2012/01/12. https://doi.org/10.1371/journal.pbio.1001230 PMID: 22235190; PubMed Central PMCID: PMC4417040.

54. 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. 2016; 12(7):e1006219. Epub 2016/07/29. https://doi.org/10.1371/journal.pgen.1006219 PMID: 27467070; PubMed Central PMCID: PMC4965130.

55. 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(5):692–9. Epub 2009/03/17. https://doi.org/10.1016/j.neuron.2009.02.012 PMID: 19285466; PubMed Central PMCID: PMC2772867.

56. Zhao Y, Long L, Xu W, Campbell RF, Large EE, Greene JS, et al. Changes to social feeding behaviors are not sufficient for fitness gains of the Caenorhabditis elegans N2 reference strain. Elife. 2018; 7. Epub 2018/10/18. https://doi.org/10.7554/eLife.38675 PMID: 30328811; PubMed Central PMCID: PMC5905201.

57. Xu W, Long L, Zhao Y, Stevens L, Felipe I, Munoz J, et al. Evolution of Yin and Yang isoforms of a chromatin remodeling subunit precedes the creation of two genes. eLife. 2019; 8:e48119. https://doi.org/10.7554/eLife.48119 PMID: 31498079

58. de Bono M, Tobin DM, Davis MW, Avery L, Bargmann CI. Social feeding in Caenorhabditis elegans is induced by neurons that detect aversive stimuli. Nature. 2002; 419(6910):899–903. Epub 2002/11/01. https://doi.org/10.1038/nature01169 PMID: 12410303; PubMed Central PMCID: PMC3955269.

59. Gray JM, Karow BS, Lu H, Chang AJ, Chang JS, Ellis RE, et al. Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. Nature. 2004; 430(6997):317–22. Epub 2004/06/29. https://doi.org/10.1038/nature02714 PMID: 15220933.

60. Flavell SW, Pokala N, Macosko EZ, Albrecht DR, Larsch J, Bargmann CI. Serotonin and the neuropeptide PDF initiate and extend opposing behavioral states in C. elegans. Cell. 2013; 154(5):1023–35. https://doi.org/10.1016/j.cell.2013.08.001 PMID: 23972393

61. Lee JI, Dhakal BK, Lee J, Bandyopadhyay J, Jeong SY, Eom SH, et al. The Caenorhabditis elegans homolog of Down syndrome critical region 1, RCN-1, inhibits multiple functions of the phosphatase calcineurin. Journal of molecular biology. 2003; 328(1):147–56. https://doi.org/10.1016/s0022-2836(03)00237-7 PMID: 12684004

62. Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP. MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome Res. 2018; 28(2):66–74. Epub 2017/12/24. https://doi.org/10.1101/gr.221164.117 PMID: 29273626; PubMed Central PMCID: PMC5793790.

63. Jänes J, Dong Y, Schoof M, Serizay J, Appert A, Cerrato C, et al. Chromatin accessibility dynamics across C. elegans development and ageing. Elife. 2018; 7:e37344. https://doi.org/10.7554/eLife.37344 PMID: 30362940

64. Araya CM, Kawli T, Kundaje A, Jiang L, Wu B, Vafeados D, et al. Regulatory analysis of the C. elegans genome with spatiotemporal resolution. Nature. 2014; 512(7515):400. https://doi.org/10.1038/ nature13497 PMID: 25164749

65. Gerstein MB, Lu ZJ, Yan Nostrand EL, Cheng C, Arshinoff BI, Liu T, et al. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. 2010; 330(6012):1775–87. https://doi.org/10.1126/science.1196914 PMID: 21177976
66. Li W, Bell HW, Ahnn J, Lee SK. Regulator of Calcineurin (RCAN-1) Regulates Thermotaxis Behavior in Caenorhabditis elegans. J Mol Biol. 2015; 427(22):3457–68. Epub 2015/08/02. https://doi.org/10.1016/j.jmb.2015.07.017 PMID: 26232604.

67. Itani OA, Filbette S, Dumas KJ, Moerman DG, Hu PJ. Chromoanansythetic genomic rearrangement identified in a N-ethyl-N-nitrosourea (ENU) mutagenesis screen in Caenorhabditis elegans. G3: Genomes, Genomics, Genetics. 2016; (6)(2):351–6.

68. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010; 11(2):97–108. Epub 2010/01/07. https://doi.org/10.1038/nrg2689 PMID: 20051986.

69. Clark AG. Invasion and maintenance of a gene duplication. Proc Natl Acad Sci U S A. 1994; 91 (8):2950–4. Epub 1994/04/12. https://doi.org/10.1073/pnas.91.8.2950 PMID: 8159686; PubMed Central PMCID: PMC43492.

70. Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. Genetics. 2000; 154(1):459–73. Epub 2000/01/11. PMID: 10629003; PubMed Central PMCID: PMC1460895.

71. Neelsen KJ, Lopes M. Replication fork reversal in eukaryotes: from dead end to dynamic response. Nature Reviews Molecular Cell Biology. 2015; 16(4):207. https://doi.org/10.1038/nrm4035 PMID: 25714681.

72. Polleys EJ, House NC, Freundreich CH. Role of recombination and replication fork restart in repeat instability. DNA repair. 2017; 56:156–65. https://doi.org/10.1016/j.dnarep.2017.06.018 PMID: 28641941.

73. Kugelberg E, Kofoed E, Andersen DI, Lu Y, Mellor J, Roth FP, et al. The tandem inversion duplication in Salmonella enterica: selection drives unstable precursors to final mutation types. Genetics. 2010; 185(1):65–80. https://doi.org/10.1534/genetics.110.114074 PMID: 20215473.

74. Pradhan S, Quilez S, Homer K, Hendricks M. Environmental programming of adult foraging behavior in C. elegans. Current Biology. 2019; 29(17):2867–79. e4. https://doi.org/10.1016/j.cub.2019.07.045 PMID: 31422888.

75. Arron JR, Winslow MM, Polleri A, Chang CP, Wu H, Gao X, et al. NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. Nature. 2006; 441(7093):595–600. Epub 2006/03/24. https://doi.org/10.1038/nature04678 PMID: 16554754.

76. Li W, Choi T-W, Ahnn J, Lee S-K. Allele-Specific Phenotype Suggests a Possible Stimulatory Activity of RCAN-1 on Calcineurin in Caenorhabditis elegans. Molecules and cells. 2016; 39(11):827. https://doi.org/10.14348/molcells.2016.0222 PMID: 27871170.

77. Arribere JA, Bell RT, Fu BX, Artiles KL, Hartman PS, Fire AZ. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. Genet. 2014; 198(3):837–46. Epub 2014/08/28. https://doi.org/10.1534/genetics.114.169730 PMID: 25161212; PubMed Central PMCID: PMC4224473.

78. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012; 6(2):80–92. https://doi.org/10.4161/fly.19695 PMID: 22728672.

81. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168.

82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943.

83. http://broadinstitute.github.io/picard/index.html.

84. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907. 2012.
86. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature biotechnology. 2011; 29(1):24. https://doi.org/10.1038/nbt.1754 PMID: 21221095

87. Seibt KM, Schmidt T, Heitkam T. FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. Bioinformatics. 2018; 34(20):3575–7. Epub 2018/05/16. https://doi.org/10.1093/bioinformatics/bty395 PMID: 29762645.

88. Varet H, Brillet-Guéguen L, Coppée J-Y, Dillies M-A. SARTools: a DESeq2-and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. PLoS One. 2016; 11(6):e0157022. https://doi.org/10.1371/journal.pone.0157022 PMID: 27280887.

89. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31(2):166–9. https://doi.org/10.1093/bioinformatics/bту638 PMID: 25260700

90. Chen Y, Lun AT, Smyth GK. Differential expression analysis of complex RNA-seq experiments using edgeR. Statistical analysis of next generation sequencing data: Springer; 2014. p. 51–74.

91. Lee H, Kim SA, Coakley S, Mugno P, Hammarlund M, Hilliard MA, et al. A multi-channel device for high-density target-selective stimulation and long-term monitoring of cells and subcellular features in C. elegans. Lab on a Chip. 2014; 14(23):4513–22. https://doi.org/10.1039/c4lc00789a PMID: 25257026

92. 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(4):1509–25. https://doi.org/10.1534/genetics.118.301311 PMID: 30341085

93. Garcia-Gonzalez AP, Ritter AD, Shrestha S, Andersen EC, Yilmaz LS, Walhout AJM. Bacterial Metabolism Affects the C. elegans Response to Cancer Chemotherapeutics. Cell. 2017; 169(3):431 –41 e8. Epub 2017/04/22. https://doi.org/10.1016/j.cell.2017.03.046 PMID: 28431244; PubMed Central PMCID: PMC5484065.

94. Shimko TC, Andersen EC. COPASutils: an R package for reading, processing, and visualizing data from COPAS large-particle flow cytometers. PLoS One. 2014; 9(10):e111090. Epub 2014/10/21. https://doi.org/10.1371/journal.pone.0111090 PMID: 25329171; PubMed Central PMCID: PMC4203834.

95. Broman KW, Sen S. A Guide to QTL Mapping with R/qtl: Springer; 2009.