12-17-2022

Sequencing an inserted concatemer in Caenorhabditis elegans strain hlh-29::gfp

Simon T. Anderson
James Madison University

Sequencing an inserted concatemer in *Caenorhabditis elegans* strain *hlh-29::gfp*

An Honors College Project Presented to

the Faculty of the Undergraduate

College of Science and Mathematics

James Madison University

By Simon Timothy Anderson

April 2022

Accepted by the faculty of the Department of Biology, James Madison University, in partial fulfillment of the requirements for the Honors College.

FACULTY COMMITTEE:

Project Advisor: Casonya M. Johnson, Ph.D.,
Affiliate Professor, Biology

Reader: Karen B. Barnard-Kubow, Ph.D.,
Genomics Research Director and Education Coordinator, Biology

Reader: Tim A. Bloss, Ph.D.,
Associate Professor, Biology

HONORS COLLEGE APPROVAL:

Fletcher Linder, Ph.D.,
Interim Dean, Honors College

PUBLIC PRESENTATION

This work is accepted for presentation, in part or in full, at the Honors College Symposium on April 22, 2022 and at the Biotech Symposium on April 29, 2022.

Table of Contents

List of Figures **3**

Acknowledgements **4**

Abstract **6**

Introduction **7**

Materials and Methods **11**

Worm propagation 11

DNA extraction 11

Library prep, sequencing, and basecalling 12

Mapping raw reads 12

Assembly and mapping of de novo contigs 12

BLAST identification of genes 12

Visualization 13

Results **14**

DNA extraction and sequencing prep 14

Sequencing output 14

Location and deleted genes 15

Size and contents of insertion 16

Assembled contigs 17

Discussion **21**

Future Work **23**

References **25**

List of Figures

Figures

1. Creation of the transgenic *hlh-29::gfp* strain 8
2. Wild type and roller phenotype *C. elegans* 9
3. Distribution of read lengths from MinION nanopore sequencing 14
4. Location of the *hlh-29::gfp* and *rol-6* concatemer in Chromosome III 15
5. Schematic of the Contig 37 and Contig 45 sections of the insertion 16
6. Overlaps between assembled contigs 19

Tables

1. Characteristics of assembled contigs 17

Acknowledgements

I would like to formally thank my faculty advisor, Dr. Casonya M. Johnson, for her continuous support of my work, both in and out of the laboratory, during her time at JMU and beyond. Her interest in, and unwavering encouragement of, my endeavors within the fields of biotechnology and health inequality have been instrumental in shaping my academic identity as a future health disparity researcher. It doesn't seem like that long ago that she was teaching me how to pick up worms without crushing them; two years later, I only now realize how much my technical skills have grown under her guidance and the guidance of the rest of my committee.

I would like to thank Dr. Karen B. Barnard-Kubow, without whom this project would never have proceeded past the propagation of worms. Her contributions to every step of this study, from DNA extraction to the writing of this very manuscript, have been extraordinary. She has taught me not only how to pursue creative and outside-the-box solutions, but also how to efficiently manage time and materials to meet the goals of my research.

I would also like to thank Dr. Tim A. Bloss, who not only has provided his time as a reader on my capstone committee, but who also fostered within me an even deeper understanding and appreciation of Mendelian and non-Mendelian genetics than I could have ever expected. In my soon-to-be nine semesters at JMU, Dr. Bloss's class on the Genetics of Cancer is easily one of the most fascinating and illuminating courses I have ever taken.

I would like to thank all the other members of the Johnson lab for their scientific camaraderie and intellectual stimulation. Even though our time in the Johnson lab has been ruled by the COVID pandemic, we have found ways to stay connected, and have learned from and supported each other where we can.

Finally, I would like to thank the JMU Department of Biology, JMU's Center for Genome and Metagenome Studies, and the National Science Foundation for their substantial financial and material support of Dr. Johnson's research into bHLH proteins, and of my work within that project.

Abstract

Hairy/Enhancer of Split (HES) proteins and their homologs play vital roles in the development of animals, and in the regulation of cancer and other diseases. HLH-29 is an REF-1 family protein and a *Caenorhabditis elegans* homolog of human HES proteins. A transgenic animal model has been developed to facilitate analyses of this protein, containing an inserted concatemer of the green fluorescent protein reporter gene, regulated by the *hlh-29* promoter, and a mutant *rol-6* allele. Though the strain has been used previously, the location and structure of the insertion has been unknown. Sequencing of the whole *hlh-29::gfp* genome with high-molecular weight Nanopore sequencing on the MinION platform has elucidated the structure. The concatemer was found to have been inserted into Chromosome III of Bristol N2 oocytes, deleting two genes with no immediately observable phenotypic abnormalities. The insertion was found to be at least 188.5 kb long, containing 26 copies of *hlh-29::gfp* and 11 copies of *rol-6*. While the *hlh-29::gfp* strain was confirmed to be a good tool for investigating expression of HLH-29, the insertion could not be assembled in its entirety and is likely much longer than 188 kb. Future use of this strain may seek to model dysregulation of HLH-29 in response to oxidative stress, which has implications for the alleviation of health disparities regarding various types of breast cancer.

Introduction

For over a century, biologists in various fields have used the nematode *Caenorhabditis elegans* to learn about eukaryote genetics, behavior, and structure.[1] In the age of biotechnology, we have the ability to make deliberate modifications to model organisms such as *C. elegans* quickly and efficiently. Instead of observing natural phenomena as they arise, we can convert what nature has given us into tools. However, before using a tool, we must first know all its details.

HLH-29 is a basic helix-loop-helix (bHLH) transcription factor that plays a role in early development.[2] It is one of the six members of the REF-1 family in *C. elegans*, which are homologous to Hairy/Enhancer of Split (HES) proteins in humans.[2,3] HES and REF-1 proteins regulate cardiovascular (in mammals) and neural development, as well as endoderm and mesoderm specification in primarily Notch-dependent, but also Notch-independent, pathways.[2–4] In *C. elegans*, HLH-29 regulates neurogenesis, fatty acid metabolism, the oxidative stress response, egg-laying, aging, and other processes.[3,5] The protein is most highly expressed in the embryonic and L1 stages; expression then drops to near zero and increases slowly until the worm is mature.[4] HLH-29 is expressed throughout the body in early development, but expression in mature worms is localized to the spermatheca, uterine and vulval structures, intestinal wall, and select neurons.[2,4,6] The *hlh-29* gene is located on the X-chromosome and shares a promoter with *hlh-28*, another REF-1 family protein.[4,6]

To aid in the study of *C. elegans* bHLH transcription factors, a novel strain of *C. elegans* was created through microinjection of linearized plasmids into oocytes, which were consequently ligated *in vivo* into an extrachromosomal concatemer[4,7–10] (Fig. 1). Gamma irradiation of microinjected embryos facilitated integration of the concatemer into an as of yet unknown location in the worms' genome[4,7–10] (Fig. 1).

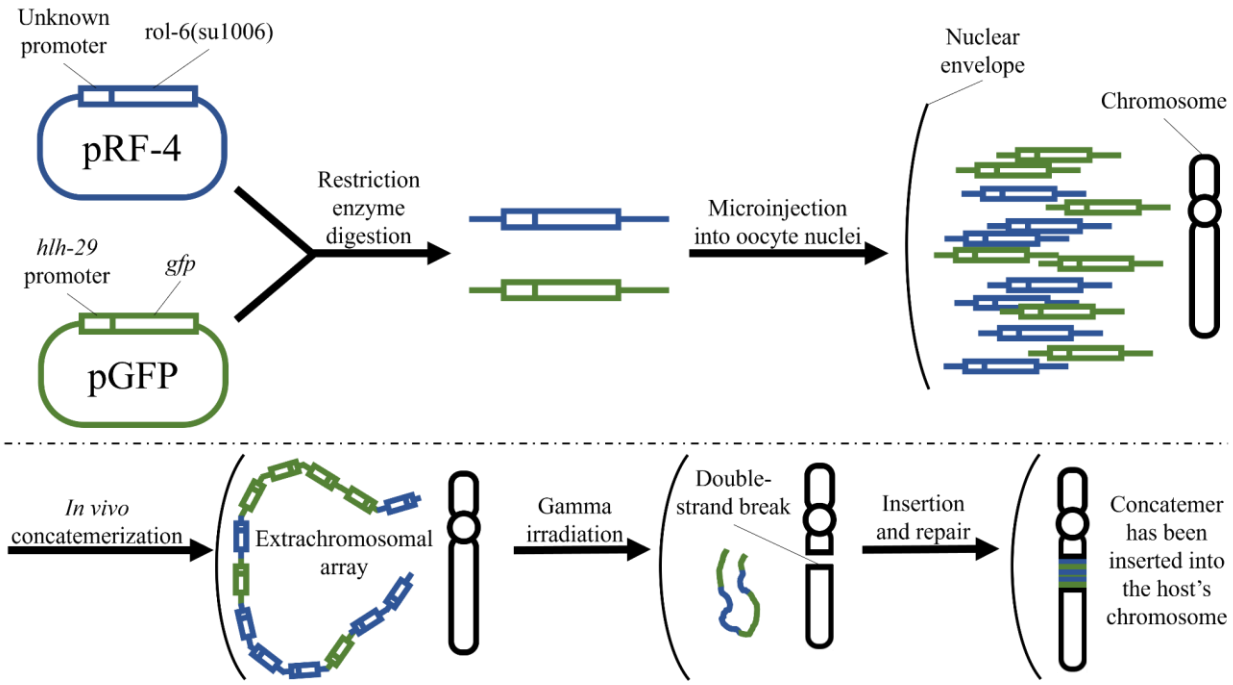

Figure 1. Creation of the transgenic *hlh-29::gfp* strain. The plasmids in the top left were obtained/created. The plasmids were linearized and microinjected into wild type (Bristol N2) maturing oocyte nuclei. Within these cells, foreign DNA can be concatemerized into an extrachromosomal array. Breakage of the host cell's chromosomes with ionizing radiation activates damage response pathways, which can result in non-homologous end-joining repair that includes the concatemer. The array becomes inserted into the host chromosome. Structures are not shown to scale.

The inserted construct contains the reporter gene green fluorescent protein (*gfp*), whose expression is controlled by the *hlh-29* promoter. The *hlh-29* promoter is responsible for the protein's pattern of expression, both temporal and spatial. The *hlh-29* promoter is a target for several transcription factors, including Notch, LIN-12 (which can redundantly supply some of the same signals as Notch), and MED-1 (one of the GATA factor family proteins, activated by SKN-1).[4] McMiller et al.[4] show that genes behind an *hlh-29* promoter will be expressed in the same

structures as HLH-29 itself, and that such proteins will not be expressed if Notch, LIN-12, and SKN-1 have been knocked out. Due to the similarities in HLH-29 and *hlh-29*::transgene expression patterns, the *hlh-29::gfp* strain should allow the expression of GFP - and thus HLH-29 itself - to be visualized through fluorescence, rather than through qPCR. The construct in the *hlh-29::gfp* strain is also known to contain the *rol-6* allele *su1006*, which encodes an abnormal collagen that causes the worms to roll in a right-handed helix.[9,11,12] *rol-6* is semi-dominantly expressed, therefore any worm showing bent rolling, as opposed to sinusoidal swimming (Fig. 2), must carry the abnormal *rol-6*.[9,11,12] Due to the method of mutagenesis, such worms will also carry the *hlh-29::gfp* insertion.

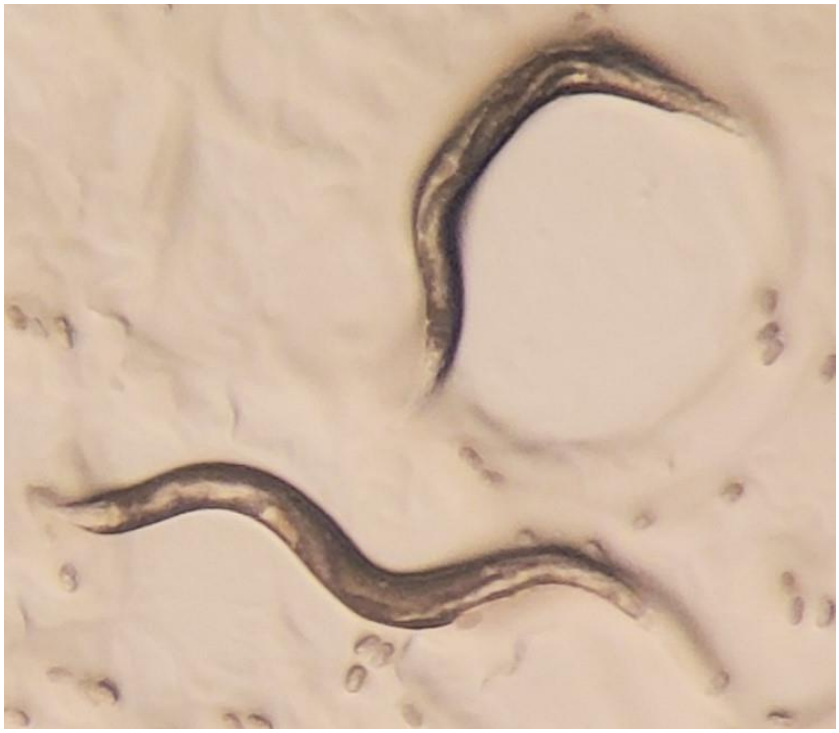

Figure 2. Wild type and roller phenotype *C. elegans*. Wild type (Bristol N2) worms swim in a sinusoidal fashion (bottom). The semi-dominantly expressed *rol-6* gene causes the top (*hlh-29::gfp*) worm to roll in a circle with an easily identifiable angular, bent shape.

While the *hlh-29::gfp* strain has been used in previous studies,[2,4] we still lack basic information about the insertion found in these worms. Location of the concatemer and copy number, orientation, and order of the genes contained within are unknown, and so were the focus of this study. Short-read (e.g. Illumina) sequencing would be insufficient to resolve this type of chromosomal modification due to its size and repetitive nature. We have opted to investigate the structure of the *hlh-29::gfp* insertion using third generation long-read sequencing on the Oxford Nanopore platform.

Materials and Methods

Worm propagation

The *hlh-29::gfp* strain was created by McMiller et al.[4] using methods previously described by McMiller and Johnson.[7] Worms were grown as previously described in Lewis and Fleming,[13] using OP50 *E. coli* on nematode growth medium (NGM) Lite agar plates. Worms were grown at 16°C.

The *hlh-29::gfp* strain was obtained as a population containing heterozygotes for the reporter gene *rol-6*, as identified through both presence and absence of the roller phenotype in the population. Single worms showing the roller phenotype were removed from the population and plated separately. The progeny of each worm was screened for non-rollers, indicating a lack of the inserted construct in that offspring and thus heterozygosity of the parent worm. Populations that contained only rollers were assumed to consist solely of individuals homozygous for insertion of the construct. Homozygous populations of *hlh-29::gfp* were used for all DNA extractions.

DNA extraction

Worms were washed off of six 6-cm plates using M9 buffer. The plates contained only homozygous roller worms and had no visible lawn of bacteria. The populations of worms on each plate were too numerous to count, but were in the order of hundreds. Extraction was completed using the NEB Monarch HMW DNA Extraction Kit. DNA quality and quantity was checked using a Nanodrop spectrophotometer and a Qubit 3.0 Fluorometer.

Library prep, sequencing, and basecalling

1 μg of DNA was used for library construction. DNA was cleaned using Ampure XP magnetic beads. Cleaned DNA was prepared for MinION sequencing using Nanopore's SQK-LSK110 ligation sequencing kit. 75 μL of prepped DNA solution (12 μL DNA, 37.5 μL sequencing buffer, 25.5 μL loading beads) was loaded onto a FLO-MIN106 flow cell. Sequencing ran for 41 hours. The binary FAST5 file output by the sequencer was fed into Nanopore's Guppy basecaller (version cpu_4.5.4_linux64). A minimum quality score of 7 was set.

Mapping raw reads

Basecalled reads (output as a FASTQ file) were mapped to the WBcel235 Bristol N2 reference genome using minimap2 (version 2.22). Mapping was run using the option "-ax map-ont".

Assembly and mapping of de novo contigs

Basecalled reads were also assembled into de novo contigs using NECAT (version necat_20200803_Linux-amd64). Parameters set included a genome size of 100 Mb and a minimum read length of 3 kb. The resulting polished contigs were mapped to the WBcel235 Bristol N2 reference genome using minimap2 (version 2.22). Mapping was again run using the option "-ax map-ont".

BLAST identification of genes

The construct in *hlh-29::gfp* was known to contain the *gfp(S65C)* coding sequence (JX171292.1, bases 2779-3648) immediately preceded by the 1000 bases upstream of the

chromosomal *hlh-29* coding sequence (that is, the *hlh-29* promoter; bases 17,549,640 to 17,548,640 on the X chromosome of WBcel235). The *gfp(S65C)* gene is terminated by the *unc-54* (myosin-4) terminator (JX171292.1, bases 3785-4519). The construct was also known to contain *rol-6(su1006)* (WBVar00248869) derived from the pRF-4 cloning vector.[14,15] BLAST (ncbi-blast-2.12.0+) analyses of contigs sought to identify and count the number of each of the above elements within the insertion.

BLAST analyses were conducted on all contigs to identify those containing *gfp(S65C)*, as this gene does not occur naturally in *C. elegans* and so can only have been introduced in the insertion. BLAST analyses were run on *gfp*-containing contigs to identify the locations of *hlh-29* promoters, *unc-54* 3' UTRs (terminator), and copies of *rol-6(su1006)*.

Visualization

Raw reads, assembled contigs, and their mapping to the reference genome were visualized using the Broad Institute's Integrative Genomics Viewer (IGV). The structure of each contig (the location and orientation of *rol-6* and *gfp* elements) was visualized in Desmos graphing calculator using data obtained from BLAST analyses.

Results

DNA extraction and sequencing prep

DNA extracted using the NEB Monarch kit was found to have a concentration of 116 ng/µL. The product of library prep was found to have a concentration of 15.8 ng/µL prior to addition of sequencing buffer and loading beads. We used 12 µL of this solution for sequencing; a total of 189.6 ng was added to the flow cell.

Sequencing output

582,770 reads were produced by sequencing. Read length ranged from 1 bp to 210,269 bp, with a mean read length of 8,731 bp and a median read length of 4,864 bp (Fig. 3). There were over 200 reads longer than 100 kb (Fig. 3 inset).

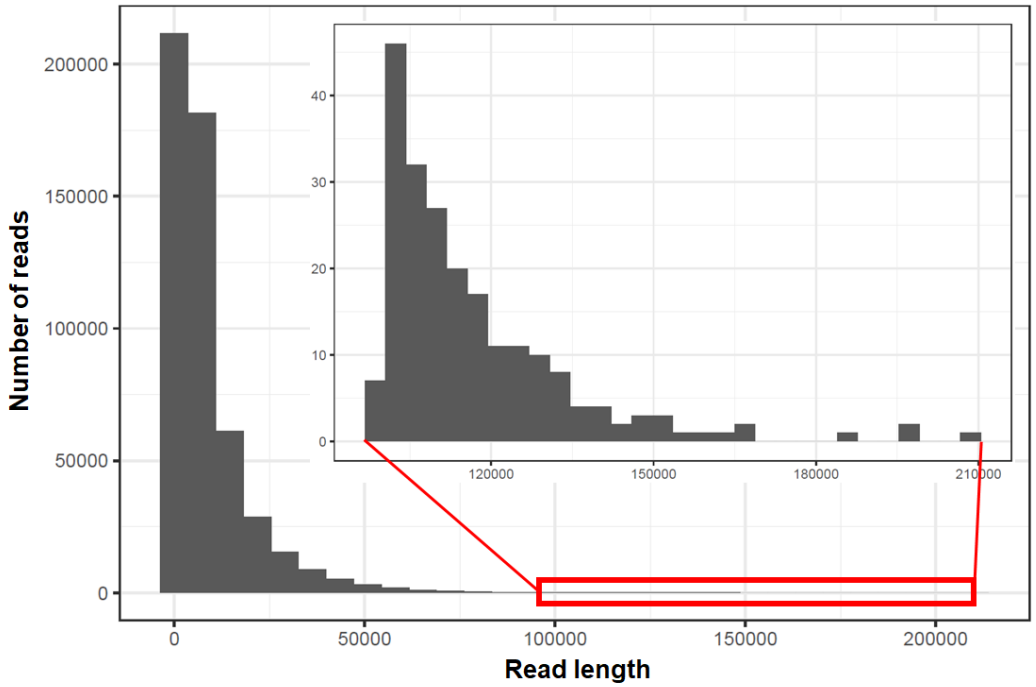

Figure 3. Distribution of read lengths from MinION nanopore sequencing. The inset shows all reads longer than 100 kb, up to 210 kb. The reads shown in the graph represent a total of 4.57

Gb. 75% of reads were longer than 2.5 kb and 25% of reads were longer than 10.6 kb. Over 200 reads were longer than 100 kb, and 12 were longer than 150 kb.

Location and deleted genes

The *hlh-29::gfp* and *rol-6* concatemer was found to have been inserted into Chromosome III between bases 1,756,475 and 1,759,577 in WBcel235. The insertion event resulted in the deletion of 3.1 kb from the chromosome (Fig. 4). The 3.1 kb region contained two genes: Y46E12A.2.1 and Y46E12A.5.1. The genes' deletion is non-lethal and results in no immediately obvious phenotypic effects.

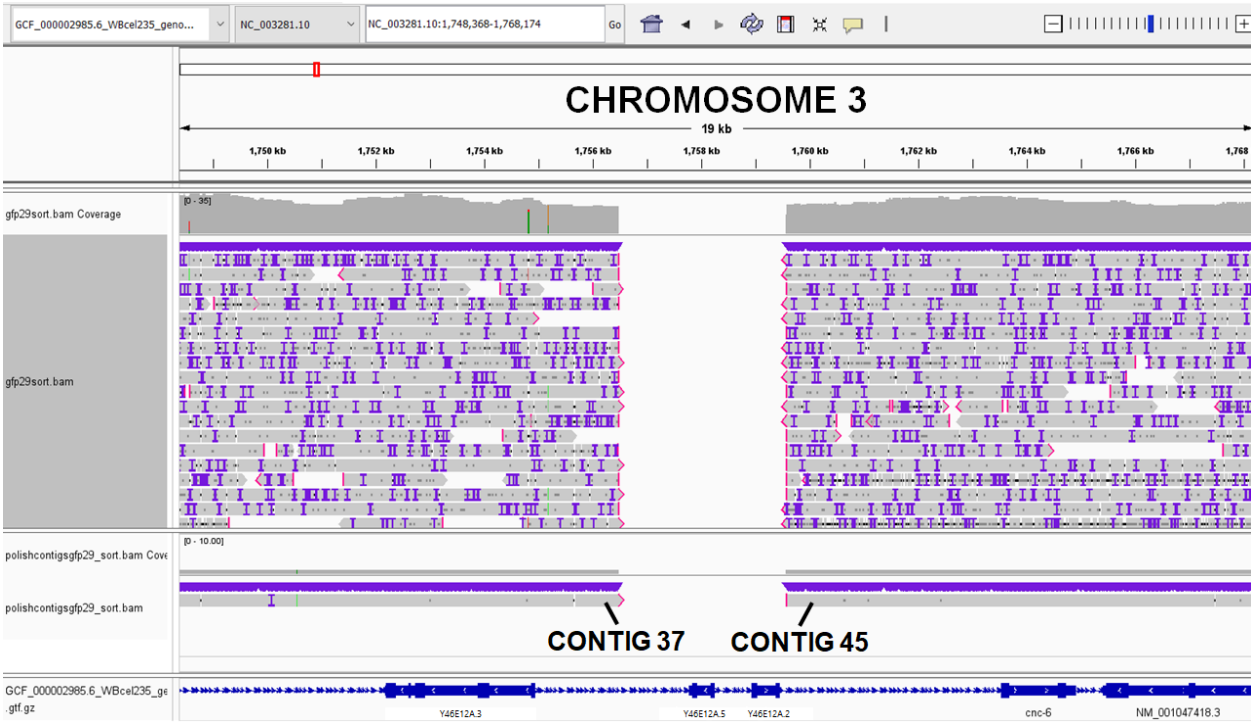

Figure 4. Location of the *hlh-29::gfp* and *rol-6* concatemer in Chromosome III. This IGV screenshot shows the location of the insertion in the *hlh-29::gfp* strain. The insertion of the *hlh-29::gfp* and *rol-6* concatemer deleted Y46E12A.2.1 and Y46E12A.5.1, shown in the bottom track. Contig 37 extends 136 kb into the insertion from the upstream end and Contig 45 extends 53 kb into

the insertion from the downstream end. The first 463 kb of Contig 37 and the last 332 kb of Contig 45 map to the reference genome.

Size and contents of insertion

135,928 bp of the upstream end of the insertion was assembled (Contig 37) and 52,651 bp of the downstream end of the insertion was assembled (Contig 45) from the MinION reads. Contig 37 has 20 copies of the *hlh-29::gfp(S65C)::unc-54* transgene that are likely to be transcribed, while Contig 45 has six copies. Contigs 37 and 45 contain eight and three copies of *rol-6*, respectively. Summing the regions of Contigs 37 and 45 that do not map to the reference genome gives a minimum length for the insertion of 188,579 bp, containing 26 copies of *hlh-29::gfp* and 11 copies of *rol-6* (Fig. 5).

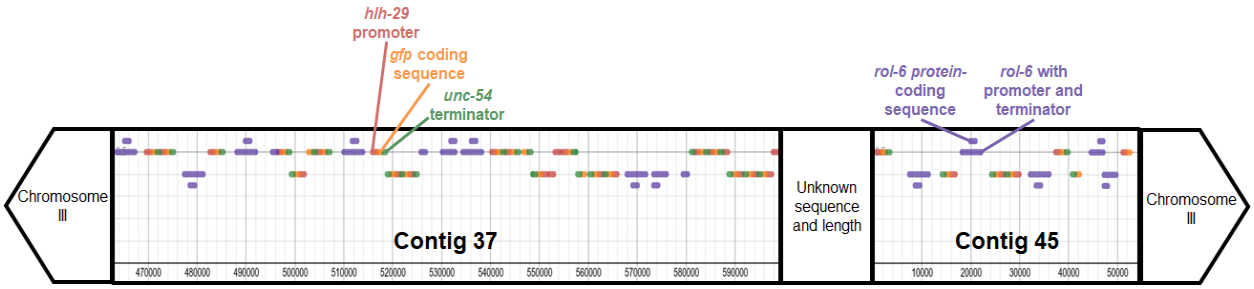

Figure 5. Schematic of the Contig 37 and Contig 45 sections of the insertion. These portions of the insertion shown comprise 188.5 kb, containing 26 copies of *hlh-29:gfp* and 11 copies of *rol-6* that are likely to be transcribed (here "likely to be transcribed" means that most of the promoter region and the majority of the protein coding region is present). An unknown amount of DNA is present between Contigs 37 and 45. The top row of features represents the forward strand of Chromosome III, the bottom row represents the reverse strand. The coordinates below the contig features refer to the contigs themselves, not coordinates on Chromosome III.

No single read of DNA was found to go all the way through the insertion. Furthermore, the reads and contigs generated cannot be assembled in such a way as to link the 3' end of Contig 37 to the 5' end of Contig 45. The full structure of the *hlh-29::gfp* insertion could not be determined from the DNA extracted and analyzed in this study.

Assembled contigs

26 contigs containing *gfp(S65C)* were assembled using the basecalled MinION reads. The length, *hlh-29::gfp(S65C)::unc-54* copy number, and *rol-6* copy number of each of these contigs is shown in Table 1. If all of these contigs were mostly independent (overlapped negligibly), we would expect the full length of the insertion to be approximately 2.2 Mb, and to contain over 500 copies of *hlh-29::gfp* and just less than 40 copies of *rol-6*.

Table 1. Characteristics of assembled contigs. It is evident that many more copies of *gfp* - perhaps an order of magnitude more - are present versus *rol-6* in the inserted concatemer. Contig 67 is known to be contained within Contig 37 (Fig. 6a), and so its length and contents were not included in the calculation of the total length and copy numbers of the insertion.

Contig	Length (kb)	*gfp* copies	*rol-6* copies
37	135.9	20	8
45	52.6	6	3
48	266.0	72	1
50	226.7	47	6
53	182.2	34	3
54	174.2	37	7
55	170.8	44	2
56	157.2	34	2
57	151.7	44	1
59	125.4	38	1
61	97.6	22	1

64	82.3	21	1
66	60.7	19	0
67	56.0	8	2
68	52.5	12	1
70	49.2	13	0
71	48.2	12	0
72	39.7	10	0
75	33.5	7	1
78	26.2	8	0
80	23.5	7	0
81	23.2	6	0
82	23.1	6	0
84	17.4	5	0
87	14.5	4	0
89	4.9	1	0
Contig	**Length (kb)**	***gfp* copies**	***rol-6* copies**
Total	2192.5	529	38

However, these 26 contigs are likely not all independent of each other. It was found that Contig 67 is contained entirely within Contig 37 (Fig. 6a). Furthermore, it was found that Contigs 57 and 59 overlap by approximately 37.5 kb and that Contigs 64 and 70 overlap by approximately 21 kb (Fig. 6b). The ends of Contigs 37 and 45 were not found to overlap with any of the 24 other contigs to such a significant degree.

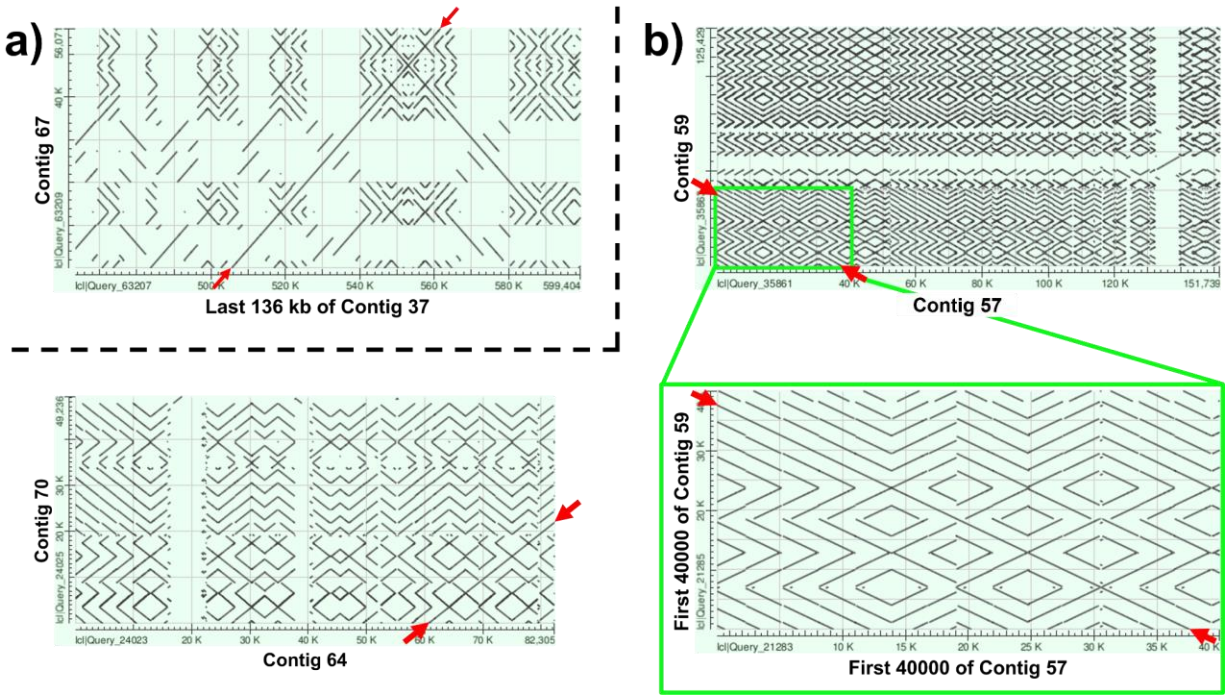

Figure 6. Overlaps between assembled contigs. Despite being assembled by NECAT into separate entities, some contigs were found to have significant overlaps. Contig 67 was found to be entirely contained within Contig 37, and two other pairs of contigs were found to have overlapping ends. Long, unbroken lines in BLAST dotplots (bookended by red arrows) that terminate at two edges of the graph were taken to be evidence of overlaps between contigs.

Panel a: Contig 67 (56 kb) was found to be contained within Contig 37, one of the contigs of the insertion that connects to Chromosome III.

Panel b: The 3' end of Contig 64 overlaps the 21 kb at the 5' end of Contig 70. The two contigs are connected in tandem; however, not all contigs produced by NECAT are tandem. This is illustrated by Contigs 57 and 59 overlapping by 37.5 kb at what would be their 5' ends. Thus, what NECAT declared to be the 5' end of one of these contigs is actually its 3' end.

While not able to be connected to Contig 37 or 45 (and thus not connected to Chromosome III), the largest *gfp*-containing contig assembled was Contig 48. Contig 48 is 266 kb long and contains 72 copies of *gfp*, but only a single copy of *rol-6*. Addition of Contig 48 to the estimate

produced by summing just Contigs 37 and 45 indicates that the insertion is likely more than 454.5 kb long and contains at least 98 copies of *hlh-29::gfp* and 12 copies of *rol-6*.

Discussion

HES proteins and their homologs play vital roles in the development of animals and in the regulation of cancer and other diseases.[2–5,16–19] As such, it is a worthwhile task to develop a transgenic animal model to facilitate analyses of these proteins. Such a model has been created in *C. elegans* through gamma-mediated insertion of an extrachromosomal array.[4,7] However, the location and structure of this insertion was previously unknown. While NECAT was not able to fully resolve the structure of the insertion, likely due to its highly repetitive nature and the lack of a read that spans the entire construct, we have identified its location in the *C. elegans* genome to be Chromosome III. We have identified several copies of the *hlh-29::gfp* transgene and *rol-6*, and provided estimates of the insertion's length and content. The *hlh-29::gfp* strain contains an insertion at least 188.5 kb long, with at least 26 copies of the *gfp* transgene and 11 copies of *rol-6*, though the insertion could be as long as two million bases and contain as many as 500 copies of *hlh-29::gfp* and 40 copies of *rol-6*. A more complete assembly of this insertion would require further extraction of high molecular weight DNA and additional MinION sequencing runs, with the goal of obtaining reads that span the entire length of the insertion.

An unintended consequence of the method of insertion was the deletion of a 3.1-kb stretch of Chromosome III containing two genes: Y46E12A.2.1 and Y46E12A.5.1. Y46E12A.2.1 has been found to be transcriptionally active throughout the organism's life[20,21] and may be a target of HLH-1 activity.[22,23] Y46E12A.5.1 is transcribed in the mesoderm and ectoderm of late-stage embryos.[24,25] Neither of the two genes are known to produce a protein.[22,26] Based on observations made during this study, deletion of the two genes is non-lethal and results in no immediately obvious phenotypic effects. Thus, further use of the *hlh-29::gfp* strain is unlikely to be confounded by the effects of knocking out Y46E12A.2.1 and Y46E12A.5.1. Taken with the data presented

above regarding the presence of multiple copies of the *hlh-29::gfp* transgene within the insertion, our findings indicate that this strain of *C. elegans* is a good model for investigating HLH-29 expression.

Future work

We have demonstrated that the transgenic *hlh-29::gfp* strain of *C. elegans* is a good model for investigating expression of HLH-29, an REF-1 family protein. HES-1, a human HES homolog of nematode REF-1 family proteins, is responsible for preventing premature differentiation of cells via downregulation of proliferating cell nuclear antigen (PCNA, PCN-1 in *C. elegans*).[16,17] Downregulation of HES-1 has been associated with more aggressive cancers with a heightened ability to proliferate.[16] This is particularly true of hormone-dependent breast cancers (HDBCs), in which HES-1 is a target of estradiol signaling.[16]

Ström et al.[16] have shown that supplementation of exogenous HES-1 to colon cancers and HDBCs has the effect of decreasing PCNA, which inhibits proliferation of the tumor. A study by Xu et al.[18] indicates that HES-1 will be upregulated by oxidative stress in the human trabecular meshwork cells of glaucomatous eyes, contributing to pathogenesis. McMiller et al.[4] show where we can expect HLH-29 to be naturally expressed over the lifespan of *C. elegans*. Our contribution to this literature shows that the *hlh-29::gfp* strain will be useful for modeling HLH-29 expression in *C. elegans*. In future projects, we hope to, in effect, utilize the sum of this literature to attempt to show that oxidative stress acts on the *hlh-29* promoter in *hlh-29::gfp* to increase GFP expression in all cell types.

This project would have implications for the mitigation of racial disparities in health. It has been found that, despite having roughly equal rates of breast cancer as non-Hispanic white women, African-American women are 40% more likely to die from the disease than white women, have the highest breast cancer mortality of any racial group, and have an increasing rate of incidence of breast cancer, despite an overall decreasing mortality rate.[27,28] Moreover, African-American women are more likely to have HDBCs, which are more deadly than hormone-

independent tumors.[29,30] If we can provide evidence through this future project that oxidative stress is capable of upregulating HLH-29, HES proteins, or a gene with an *hlh-29* promoter, then we have also laid the groundwork for potential clinical applications. One such application could be the induction of localized oxidative stress in HDBC tumors in an attempt to increase endogenous HES-1 and stall the cancer's proliferation, which would provide other therapies with an opportunity to shrink and kill the tumor.

Our HES protein animal model may also have implications for triple-negative breast cancer (TNBC). Upregulation of HES-1 in these cancers promotes proliferation;[19] this is contrary to the behavior of HDBC. Regarding the study proposed above, should we find that oxidative stress decreases HLH-29 expression in the animal model, instead of increasing it, we may then be able to treat the worms as a model for TNBC, not HDBC.

References

1. Nigon VM, Félix M-A. History of research on C. elegans and other free-living nematodes as model organisms. *WormBook*. Published online 2018. Accessed March 31, 2021. https://www.ncbi.nlm.nih.gov/books/NBK453431/

2. White A, Fearon A, Johnson CM. HLH-29 regulates ovulation in C. elegans by targeting genes in the inositol triphosphate signaling pathway. *Biol Open*. 2012;1(3):261-268. doi:10.1242/bio.2012046

3. Quach TK, Chou HT, Wang K, Milledge GZ, Johnson CM. Genome-Wide Microarrray Analysis Reveals Roles for the REF-1 Family Member HLH-29 in Ferritin Synthesis and Peroxide Stress Response. *PLoS One*. 2013;8(3):e59719. doi:10.1371/journal.pone.0059719

4. McMiller TL, Sims D, Lee T, Williams T, Johnson CM. Molecular characterization of the Caenorhabditis elegans REF-1 family member, hlh-29/hlh-28. *Biochim Biophys Acta - Gene Struct Expr*. 2007;1769:5-19. doi:10.1016/j.bbaexp.2006.12.001

5. Q19917 | BHLH domain-containing protein hlh-29, Caenorhabditis elegans. UniProtKB. Published 2020. Accessed April 2, 2021. https://www.uniprot.org/uniprot/Q19917

6. hlh-29 (WBGene00001968). WormBase. Published 2013. Accessed April 2, 2021. http://www.wormbase.org/db/get?name=WBGene00001968;class=Gene

7. McMiller TL, Johnson CM. Molecular characterization of HLH-17, a C. elegans bHLH protein required for normal larval development. *Gene*. 2005;356:1-10. doi:10.1016/j.gene.2005.05.003

8. Hobert O. PCR Fusion-Based Approach to Create Reporter Gene Constructs for Expression Analysis in Transgenic C. elegans. *Biotechniques*. 2002;32(4):728-730.

9. Mello CC, Kramer JM, Stinchcomb D, Ambros V. Efficient gene transfer in C. elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J*. 1991;10(12):3959-3970.

10. Fire A. Integrative transformation of Caenorhabditis elegans. *EMBO J*. 1986;5(10):2673-2680. doi:10.1002/j.1460-2075.1986.tb04550.x

11. Kramer JM, French RP, Park EC, Johnson JJ. The Caenorhabditis elegans rol-6 gene, which interacts with the sqt-1 collagen gene to determine organismal morphology, encodes a collagen. *Mol Cell Biol*. 1990;10(5):2081-2089. doi:10.1128/mcb.10.5.2081

12. Kim TH, Kim YJ, Cho JW, Shim J. A novel zinc-carboxypeptidase SURO-1 regulates cuticle formation and body morphogenesis in Caenorhabditis elegans. *FEBS Lett*. 2010;585(1):121-127. doi:10.1016/j.febslet.2010.11.020

13. Lewis JA, Fleming JT. Basic culture methods. In: Epstein HF, Shakes DC, eds. *Caenorhabditis Elegans, Modern Biological Analysis of an Organism*. Academic Press; 1995:3-29.

14. su1006 (variation). WormBase. Published 2017. Accessed March 31, 2022. www.wormbase.org/db/get?name=WBVar00248869;class=variation

15. pRF4 (clone). WormBase. Published 2001. Accessed March 31, 2022. www.wormbase.org/db/get?name=pRF4;class=clone

16. Ström A, Arai N, Leers J, Gustafsson JÅ. The Hairy and Enhancer of Split homologue-1 (HES-1) mediates the proliferative effect of 17β-estradiol on breast cancer cell lines. *Oncogene*. 2000;19:5951-5953. doi:10.1038/sj.onc.1203990

17. pcn-1 (WBGene00003955). WormBase. Published 2019. Accessed April 9, 2021. http://www.wormbase.org/db/get?name=WBGene00003955;class=Gene

18. Xu L, Zhang Y, Guo R, et al. HES1 promotes extracellular matrix protein expression and inhibits proliferation and migration in human trabecular meshwork cells under oxidative stress. *Oncotarget*. 2017;8(13):21818-21833. doi:10.18632/oncotarget.15631

19. Li X, Cao Y, Li M, Jin F. Upregulation of HES1 promotes cell proliferation and invasion in breast cancer as a prognosis marker and therapy target via the AKT pathway and EMT process. *J Cancer*. 2018;9(4):757-766. doi:10.7150/jca.22319

20. Levin M, Hashimshony T, Wagner F, Yanai I. WBGene00021592 Expression. Developmental Cell. Published 2012. Accessed March 31, 2022. https://wormbase.org/img-static/pictures/WBPerson4037/WBGene00021592.jpg

21. Levin M, Hashimshony T, Wagner F, Yanai I. Developmental Milestones Punctuate Gene Expression in the Caenorhabditis Embryo. *Dev Cell*. 2012;22(5):1101-1108. doi:10.1016/j.devcel.2012.04.004

22. Y46E12A.2.1 (transcript). WormBase. Accessed March 31, 2022. www.wormbase.org/db/get?name=Y46E12A.2.1;class=Transcript

23. Fox RM, Watson JD, Von Stetina SE, et al. The embryonic muscle transcriptome of Caenorhabditis elegans. *Genome Biol*. 2007;8:R188. doi:10.1186/gb-2007-8-9-r188

24. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I. WBGene00194738 (Y46E12A.5) expression in embryos. Nature. Published 2015. Accessed March 31, 2022. https://wormbase.org/img-static/pictures/WBPerson4037/Hashimshony_WBGene00194738.jpg

25. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature*. 2015;519(7542):219-222. doi:10.1038/nature13996

26. Y46E12A.5.1 (transcript). WormBase. Accessed March 31, 2022. www.wormbase.org/db/get?name=Y46E12A.5.1;class=Transcript

27. U.S. Cancer Statistics Working Group. United States Cancer Statistics Data Visualization Tool. Centers for Disease Control. Published 2019. Accessed April 9, 2021. https://gis.cdc.gov/Cancer/USCS/DataViz.html

28. Breast Cancer Prevention Partners. *African American Women & Breast Cancer*.; 2018.

29. Tao L, Gomez SL, Keegan THM, Kurian AW, Clarke CA. Breast cancer mortality in African-American and non-hispanic white women by molecular subtype and stage at diagnosis: A population-based study. *Cancer Epidemiol Biomarkers Prev*. 2015;24(7):1039-1045. doi:10.1158/1055-9965.epi-15-0243

30. American Cancer Society. *Cancer Facts and Figures for African Americans 2019-2021*.; 2019. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-facts-and-figures-for-african-americans/cancer-facts-and-figures-for-african-americans-2019-2021.pdf