# THE PHENOTYPIC AND GENOMIC CONSEQUENCES OF TRANSPOSABLE ELEMENTS IN *C. ELEGANS* BERGERAC STRAINS

An Undergraduate Research Scholars Thesis

by

AUSTIN DAIGLE

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Faculty Research Advisor: Dr. Vaishali Katju

May 2021

Major: Genetics

# RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

I, Austin Daigle, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Research Faculty Advisor prior to the collection of any data used in this final thesis submission.

This project required approval from the Texas A&M University Research Compliance & Biosafety office. This IBC Permit approves the use of *Escherichia coli* K-12 strains (NIH Guidelines III-F-8), *Escherichia coli* non-K-12 strains (NIH Guidelines III-E), and *Caenorhabditis elegans* (NIH Guidelines III-D-4) in BSL 1 conditions within the Katju-Bergthorsson laboratories. All NIH guidelines were followed throughout the project.
TAMU IBC 2018-119 Approval Date: 11/19/2018 Expiration Date: 11/19/2021

# TABLE OF CONTENTS

Page

# ABSTRACT

The Phenotypic and Genomic Consequences of Transposable Elements in *C. elegans* Bergerac Strains

Austin Daigle
Department of Biochemistry and Biophysics
Texas A&M University

Research Faculty Advisor: Dr. Vaishali Katju
Department of Veterinary Integrative Biosciences
Texas A&M University

One of the earliest samples of *Caenorhabditis elegans* was isolated in Bergerac, France. Early estimates indicated that a Bergerac strain had a high copy-number of the transposable element *Tc1* and also displayed diminished fitness compared to lower *Tc1* bearing counterparts. In this study, to clarify the extent of phenotypic disruption caused by high TE copy number, four fitness traits (developmental rate, longevity, survivorship, and productivity) were analyzed in three Bergerac strains (RW7000, RW6999, and CB4851) and compared to a wildtype N2 control. All three Bergerac strains were shown to have significantly reduced fitness compared to the control for all traits measured with specific traits displaying significant differences between Bergerac strains. To understand the molecular basis for these differences, whole-genome sequencing was completed on each Bergerac strain. The *Tc1* copy-number for each strain was estimated using the McClintock meta-pipeline, and *Tc1* copy-number was shown to be negatively correlated with fitness values. A genomic analysis of the location of *Tc1* insertions and the genes disrupted by *Tc1* has revealed the probable cause of low fitness in these strains,

while also reinforcing the target site preferences of *Tc1*. This study sets the stage for a search for mutations associated with *Tc1* proliferation, a comparison of the relative amount of RNA transcripts in the Bergerac strains, and long-term experimental evolution at high population sizes, which will reveal the causes and consequences of *Tc1* proliferation and the genetic basis for adaptive, compensatory evolution.

# ACKNOWLEDGEMENTS


## Contributors

I would like to thank my faculty advisor, Dr. Vaishali Katju, and my collaborators, Dr. Thaddeus Deiss and Robert Melde, for their guidance, support, and suggestions throughout the course of this research. Thanks also go to my friends and colleagues and the faculty and staff in the Departments of Veterinary Integrative Biosciences and Genetics for making my time at Texas A&M University a great experience.

The fitness assays depicted in this thesis were conducted by Austin Daigle and Robert Melde, with the guidance of Dr. Katju. The bioinformatic analyses of whole-genome sequences depicted in this study were conducted in part by Dr. Thaddeus Deiss, a postdoc in Dr. Katju's laboratory, with the assistance of Austin Daigle. All other work conducted for the thesis was completed by the student independently.

## Funding Sources

This work was also made possible in part by start-up funds from the Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences at Texas A&M University. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Texas A&M University.

# 1. INTRODUCTION

## *1.1 Caenorhabditis elegans*

The nematode, *Caenorhabditis elegans*, is a model organism that has been studied for decades to gain a greater understanding of molecular biology and evolution. *C. elegans* are egg-laying microscopic nematodes, and their sizes range from 0.25 mm larvae to 1 mm adults (Fig. 1.1). The species has a short generation time, developing from egg to egg-laying adult within 3.5 days at 20°C. Despite its small size and short life-cycle, *C. elegans* displays a wide range of behaviors and has a variety of organs and tissues, including a nervous system, muscles, intestines, and reproductive organs, providing a more complex study system than other small model organisms like *E. coli* or yeast, without the time-consuming maintenance required by larger models like mice (Corsi *et al.* 2015). *C. elegans* is an androdioecious species, meaning it has a rare reproductive system with two sexes: hermaphrodites capable of self-fertilization, and males capable of fertilizing hermaphrodites. Males arise at a low frequency in normal conditions, making up less than 0.2% of the population; however, researchers can take advantage of these varied reproductive systems to cross strains of interest in genetic studies (Fatt and Dougherty 1963; Vertino *et al.* 2011; Corsi *et al.* 2015; Nigon and Félix 2017). The ease of manipulation, speed of reproduction, and visual transparency of *C. elegans* allows researchers to easily perform a variety of experiments to answer questions about molecular genetics, evolution, and developmental biology (Corsi *et al.* 2015).

*C. elegans* has become an even more advantageous model organism in the genomics era, being the first multicellular eukaryote to have its genome sequenced (The *C. elegans* Sequencing Consortium 1998). The ability to quickly and cheaply sequence the genomes of experimentally

evolved strains has recently allowed direct empirical estimates of the genome-wide mitochondrial and nuclear mutation rates, providing insight into the sources of variation in the evolutionary process (Konrad *et al.* 2017; 2018; 2019). By comparing and contrasting the genomes of divergent strains, whole-genome sequencing reveals the function of genes and the details of molecular evolution by linking phenotypes to genotypes (Cutter 2010).

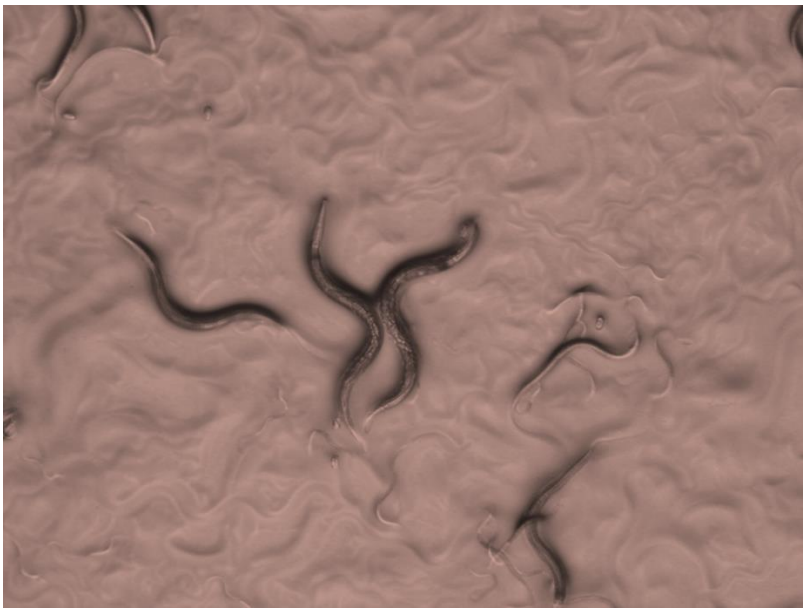

*Figure 1.1: C. elegans, living on a standard agar plate seeded with E. coli, in various stages of their life cycle.*

## 1.2 The Bergerac strains

Victor Nigon, one of the first scientists to study *C. elegans*, used a strain that he isolated from soil in Bergerac, France. As new strains continued to be isolated by other researchers, the Bergerac strain fell out of use because of its sensitivity to heat and the infertility of Bergerac males. However, this strain continued to be propagated independently in multiple laboratories for several decades while accumulating genetic differences. Eventually, *C. elegans* biologists developed techniques for long-term cryopreservation of nematode stock (Brenner 1974). Today,

several sublines of Bergerac with distinct phenotypes are available for scientific research, though some lineages have gone extinct or are missing (Nigon and Félix 2017).

Several studies have investigated the Bergerac nematodes to further elucidate some of their unique features. Fatt and Dougherty (1963) found that Bergerac strains harbor a recessive heat-sensitive mutation that renders them sterile above 23°C. Decades later, this mutation was found to be a point mutation in the *zyg-12* gene, which plays a role in attaching the centrosomes to the nucleus during cell division. *C. elegans* embryos with this mutation were found to be defective for this attachment at 25°C, leading to aneuploidy and early death (Malone *et al.* 2003). It has also been noted that Bergerac strains are less fit, produce less progeny, move with less coordination, and have a higher incidence of males relative to other *C. elegans* strains, despite the males being sterile (Hodgkin and Doniach 1997).

## 1.3 Transposable elements

### *1.3.1 What is a transposable element?*

It was generally thought that the cause of most of these unusual phenotypes in Bergerac strains was an unusually high transposable element copy-number (Hodgkin and Doniach 1997). Transposable elements (TEs) are small fragments of the genome that can occasionally excise themselves from the genome and move to new locations. TEs are broadly categorized into two categories: RNA transposons—which proliferate through reverse transcription into an RNA intermediate before reintegrating into different areas of the genome, and DNA transposons—which are directly excised from the genome as DNA by a double strand break before moving to a new location (Bessereau 2006).

TEs are viewed as parasitic and selfish because they rely on the host's cellular machinery to replicate, and can cause harmful phenotypes by inserting into genes and disrupting their functions (Orgel and Crick 1980; Muñoz-López and García-Pérez 2010). TEs are widespread in many organisms; though most TEs contain mutations that cause them to be inactive, they may be a major driver of evolutionary change (Lohe *et al.* 1995). For instance, in humans, where TEs are thought to make up 45% of the genome (Lander *et al.* 2001), TEs have been shown to be a major source of genetic variation, occasionally causing exon shuffling, deletions, inversions, and other rearrangements that can influence genome evolution due to misrepair of double strand breaks caused by transposon excision (Lohe *et al.* 1995; Prak and Kazazian 2000). Gaining a greater understanding of the variety of mechanisms that cause transposons to proliferate and decline within genomes would provide insight into the processes that shape genomes over time.

### *1.3.2 Transposable elements in C. elegans*

*C. elegans* has been a useful model for TE research since the discovery of the first *C. elegans* TE, a DNA transposon named *Tc1* (Liao *et al.* 1983). *Tc1* was studied in a Bergerac strain, and found to be significantly more active than in the standard laboratory strain N2, displaying site-specific insertion and excision from the muscle gene *unc-54* (Eide and Anderson 1985; 1988). While commonly used *C. elegans* strains like N2 have approximately 30 *Tc1* copies, Bergerac strains were estimated to possess about 300-550 copies using quantitative dot blot hybridization (Egilmez *et al.* 1995). It has been hypothesized that this change in *Tc1* copy-number occurred in the laboratory after Nigon isolated Bergerac, but the mechanism of increase has yet to be identified (Moerman and Waterston 1984; Egilmez *et al.* 1995). Since the discovery of *Tc1*, a wide variety of TEs have been defined in *C. elegans*. Overall, TEs comprise approximately 12% of the *C. elegans* genome (*C. elegans* Sequencing Consortium 1998). Unlike

humans, DNA transposons are much more common than RNA transposons in *C. elegans*, demonstrating the context-dependent evolutionary paths TEs can take across taxa (Bessereau 2006).

### 1.4 Transposable element regulation

In the N2 strain, *Tc1* activity was found to be common in somatic cells but silenced in the germ line, meaning that changes in TE location are not passed on to the next generation (Emmons and Yesner 1984). However, the Bergerac strain RW7000 was found to have active transposition of *Tc1* in the germline, indicating that TEs in this strain somehow overcame the regulatory mechanisms responsible for controlling TE expression in the germline (Eide and Anderson 1985).

The massive difference in *Tc1* activity and copy-number between Bergerac and other *C. elegans* strains could be due to a failure of the normal RNAi mechanism responsible for silencing TEs in the genome in the former (Bessereau 2006). Recent work has highlighted the importance of piRNAs (also called 21U-RNAs in *C. elegans*) and siRNAs in transposon silencing (Reed *et al.* 2020). These specialized small RNAs (smRNAs) are currently the subject of intense investigations due to advances in RNA sequencing technology (Bergthorsson *et al.* 2020; Weick and Miska 2014). However, the piRNA pathway's role in TE silencing has mainly been studied in *Drosophila melanogaster*, where TEs are targeted for silencing by small noncoding piRNAs derived from longer precursor molecules, which are produced by a family of proteins called Argonaute (AGO) proteins (Huang *et al.* 2017). *C. elegans* has homologous AGO proteins which may be used for silencing some TEs in the germline. However, there are several differences in the classes of AGO proteins in *C. elegans* compared to *D. melanogaster*, leading

some researchers to hypothesize that the nematode system for recognizing and silencing foreign DNA like TEs may incorporate other types of smRNA such as 22G RNAs, 26G RNAs, and siRNAs (Almeida *et al.* 2019). Studying the Bergerac strains could elucidate aspects of this pathway that are responsible for regulating *Tc1* in the *C. elegans* genome, contributing to a growing understanding of the evolution of TEs and the genomic defenses that control them.

## 1.5 Project components

This project aims to gain a better understanding of TE proliferation and regulation by studying the causes and consequences of TE activity in three Bergerac sublines (RW6999, RW7000, and CB4851). Four fitness traits (developmental rate, productivity, longevity, and survivorship) were analyzed using previously established assays (Katju *et al.* 2015; 2018; Dubie *et al.* 2020) to establish the phenotypic significance of high *Tc1* copy-number and activity in each strain. This project is the first to employ high-throughput Illumina whole-genome sequencing technology to sequence and analyze the entire genomes of three distinct Bergerac strains. While the genome of one Bergerac strain, CB4851, has previously been sequenced, it has not yet been analyzed in depth to study *Tc1* proliferation (Cook *et al.* 2016). By discovering the varying genomic disruptions and phenotypic consequences of *Tc1* proliferation in the Bergerac strains, this work sets the foundation for future experiments to understand how the *C. elegans* RNAi pathways may have been rendered defective or dysfunctional in these unique strains.

## 1.6 References

Almeida MV, Andrade-Navarro MA, Ketting RF. 2019. Function and evolution of nematode RNAi pathways. *Noncoding RNA*. 5(1):8.

Bergthorsson U, Sheeba CJ, Konrad A, Belicard T, Beltran T, Katju V, Sarkies P. 2020. Long-

term experimental evolution reveals purifying selection on piRNA-mediated control of transposable element expression. *BMC Biology*. 18(1):162.

Bessereau JL. 2006. Transposons in *C. elegans*. WormBook : the online review of *C. elegans* biology.1-13.

Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics*. 77(1):71-94.

*C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*. 282(5396):2012-2018.

Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga JE et al. 2016. The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics*. 204(1):371-383.

Corsi AK, Wightman B, Chalfie M. 2015. A transparent window into biology: A primer on *Caenorhabditis elegans*. *Genetics*. 200(2):387-407.

Cutter AD. 2010. Molecular evolution inferences from the *C. elegans* genome. *WormBook : the online review of C. elegans biology*.1-14.

Dubie JJ, Caraway AR, Stout MM, Katju V, Bergthorsson U. 2020. The conflict within: Origin, proliferation and persistence of a spontaneously arising selfish mitochondrial genome. *Philos. Trans. R. Soc. Lond. B Biol.* 375(1790):20190174.

Egilmez NK, Ebert RH, 2nd, Shmookler Reis RJ. 1995. Strain evolution in *Caenorhabditis elegans*: Transposable elements as markers of interstrain evolutionary history. *J. of Mol. Evol*. 40(4):372-381.

Eide D, Anderson P. 1985. Transposition of *Tc1* in the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U S A*. 82(6):1756-1760.

Eide D, Anderson P. 1988. Insertion and excision of *Caenorhabditis elegans* transposable element *Tc1*. *Mol. Cell. Biol*. 8(2):737-746.

Emmons SW, Yesner L. 1984. High-frequency excision of transposable element *Tc1* in the nematode *Caenorhabditis elegans* is limited to somatic cells. Cell. 36(3):599-605.

Fatt HV, Dougherty EC. 1963. Genetic control of differential heat tolerance in two strains of the nematode *Caenorhabditis elegans*. *Science*. 141(3577):266-267.

Hodgkin J, Doniach T. 1997. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics*. 146(1):149-164.

Huang X, Fejes Tóth K, Aravin AA. 2017. piRNA biogenesis in Drosophila melanogaster. *Trends Genet.* 33(11):882-894.

Katju V, Packard LB, Bu L, Keightley PD, Bergthorsson U. 2015. Fitness decline in spontaneous mutation accumulation lines of *Caenorhabditis elegans* with varying effective population sizes. *Evolution*. 69(1):104-116.

Katju V, Packard LB, Keightley PD. 2018. Fitness decline under osmotic stress in *Caenorhabditis elegans* populations subjected to spontaneous mutation accumulation at varying population sizes. *Evolution*. 72(4):1000-1008.

Konrad A, Brady MJ, Bergthorsson U, Katju V. 2019. Mutational landscape of spontaneous base substitutions and small indels in experimental *Caenorhabditis elegans* populations of differing size. Genetics. 212(3):837-854.

Konrad A, Flibotte S, Taylor J, Waterston RH, Moerman DG, Bergthorsson U, Katju V. 2018. Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans Proc. Natl. Acad. Sci. U S A.* 115(28):7386-7391.

Konrad A, Thompson O, Waterston RH, Moerman DG, Keightley PD, Bergthorsson U, Katju V. 2017. Mitochondrial mutation rate, spectrum and heteroplasmy in caenorhabditis elegans spontaneous mutation accumulation lines of differing population size. *Mol. Biol. Evol*. 34(6):1319-1334.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409(6822):860-921.

Liao LW, Rosenzweig B, Hirsh D. 1983. Analysis of a transposable element in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U S A.* 80(12):3585-3589.

Lohe AR, Moriyama EN, Lidholm DA, Hartl DL. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol. Biol. Evol.*

12(1):62-72.

Malone CJ, Misner L, Le Bot N, Tsai MC, Campbell JM, Ahringer J, White JG. 2003. The *C. elegans* hook protein, zyg-12, mediates the essential attachment between the centrosome and nucleus. *Cell*. 115(7):825-836.

Moerman DG, Waterston RH. 1984. Spontaneous unstable *unc-22* IV mutations in *C. elegans* var. Bergerac. *Genetics*. 108(4):859-877.

Muñoz-López M, García-Pérez JL. 2010. DNA transposons: Nature and applications in genomics. *Curr Genomics*. 11(2):115-128.

Nigon VM, Félix M-A. 2017. History of research on *C. elegans* and other free-living nematodes as model organisms. *WormBook : the online review of C. elegans biology*. 2017:1-84.

Orgel LE, Crick FHC. 1980. Selfish DNA: The ultimate parasite. *Nature*. 284(5757):604-607.

Prak ET, Kazazian HH, Jr. 2000. Mobile elements and the human genome. *Nat. Rev. Genet.* 1(2):134-144.

Reed KJ, Svendsen JM, Brown KC, Montgomery BE, Marks TN, Vijayasarathy T, Parker DM, Nishimura EO, Updike DL, Montgomery TA. 2020. Widespread roles for piRNAs and wago-class sirnas in shaping the germline transcriptome of *Caenorhabditis elegans*. *Nucl. Acids Res*. 48(4):1811-1827.

Vertino A, Ayyadevara S, Thaden JJ, Shmookler Reis RJ. 2011. A narrow quantitative trait locus in *C. elegans* coordinately affects longevity, thermotolerance, and resistance to paraquat. *Front. Genet.* 2:63.

Vos JC, De Baere I, Plasterk RH. 1996. Transposase is the only nematode protein required for in vitro transposition of *Tc1*. *Genes Dev*. 10(6):755-761.

Weick EM, Miska EA. 2014. piRNAs: From biogenesis to function. *Development*. 141(18):3458-3471.

# 2. DIVERGENT PHENOTYPES IN BERGERAC STRAINS

## 2.1 Introduction

The Bergerac strains of *C. elegans*, descended from a single isolate from Victor Nigon's garden in Bergerac, France in 1944 (Nigon and Félix 2017), have long been observed to have divergent phenotypes compared to other *C. elegans* natural isolates. A few studies have sought to quantify these phenotypes in one or two Bergerac strains. However, in this chapter, four fitness-related traits are studied in three Bergerac strains, revealing distinct differences between strains and providing a comprehensive foundation for further comparative analysis and experimentation.

The first laboratory study of a genetic trait in *C. elegans* was conducted on the progenitor Bergerac strain, where the Bergerac strain was found to be sterile at ≥23°C, while the Bristol strain was unaffected. This trait was found to follow a Mendelian inheritance pattern when the two strains were crossed, indicating a simple recessive trait (Fatt and Dougherty 1963). Further studies of the temperature sensitivity of this strain showed that malformations such as reduced body length and zig-zag intestinal shape could be seen in earliest larval stage of *C. elegans*, with minor swelling being observed at temperatures as low as 18°C (Abdulkader and Brun 1980). Subsequent molecular studies linked the temperature sensitivity in Bergerac strains to a single amino acid change on the hook protein *zyg-12*, which is thought to be responsible for anchoring and moving centrosomes to the nucleus during cell division. When this mutation is present and temperatures are high, the centrosomes fail to attach to the nucleus, leading to aneuploidy and early death (Wood *et al.* 1980; Malone *et al.* 2003).

While the genetic basis for temperature sensitivity in Bergerac strains is well understood, many other divergent phenotypes have not been fully explained. Moerman and Waterston (1984)

explored a unique phenotype that spontaneously arose in a descendent of the Bergerac strain RW7000, which was characterized by slow movement and frequent twitching. However, they found that this mutation was unstable, occasionally disappearing or reappearing in each generation. When they outcrossed the same mutation into the standard laboratory strain N2, they found that the reversion rate of the Bergerac worms was at least 100-fold more than N2. Due to recent evidence for high transposable element content in Bergerac strains, they predicted that this unique mutator activity was caused by the frequent insertion and excision of the transposable element *Tc1* (Moerman and Waterston 1984). This evidence suggested that an unidentified mutation in the Bergerac strains was the source of other frequent, unstable mutations as TEs changed their position within the genome, while also occasionally increasing the TE copy-number (Hodgkin and Doniach 1997; Bessereau 2006).

Additional studies have highlighted more divergent phenotypes in Bergerac strains without clarifying their molecular origins. Even in the earliest studies of the progenitor Bergerac strain, rare males isolated from the population were unable to mate with hermaphrodites (Fatt and Dougherty 1963; Abdulkader and Brun 1980). In a comparative study of multiple *C. elegans* strains' copulatory plug formation after mating, the authors confirmed that the males of two Bergerac strains, RW7000 and CB4851, were unable to mate with females. The authors also noted that Bergerac individuals appeared unhealthy and had uncoordinated movements compared to the other strains studied (Hodgkin and Doniach 1997).

While some traits in Bergerac have been studied, mainly in the strain RW7000 (Fatt and Dougherty 1963; Shook and Johnson 1999; Vertino *et al.* 2011; Lee *et al.* 2016), a comprehensive study of multiple traits related to reproductive success—also called fitness—has yet to be conducted on multiple distinct Bergerac strains simultaneously. Combined with a

genomic analysis, this approach has the potential to reveal the variable effects of TE activity that occurred during their separate histories of laboratory propagation. A recent analysis of fitness traits in multiple *C. elegans* strains measured developmental time and brood size in the Bergerac strain RW7000, finding that RW700 developed ~58 hours after the L1 larval stage, while the strain N2 developed in ~50 hours. The reported average brood size for RW7000 was <50, in contrast to the N2 brood size, which was ~225 (Lee *et al.* 2016). While this study supports the hypothesis that the high *Tc1* copy-number in RW7000 led to a decrease in fitness, multiple unique Bergerac strains exist and have been noted to display strain specific differences in movement and health (Hodgkin and Doniach 1997). Herein, we quantified and compared four fitness-related traits (developmental rate, productivity, longevity, and survivorship) in order to discern extant phenotypic variation among the Bergerac strains RW6999, RW7000, and CB4851.

## 2.2 Methods

### *2.2.1 Strains used*

Throughout the history of laboratory propagation of Bergerac strains, and prior to the advent of cryopreservation techniques of *C. elegans* stocks, various sublineages diverged over the years as a result of evolution in the laboratories. According to an account by V.M. Nigon, who isolated the common ancestor of all Bergerac strains in 1944, the original Bergerac strain had males capable of crossing with hermaphrodites. A strain named BW28 (also known as Bergerac DO or Bergerac BE), was shared with Ellsworth Dougherty at Berkeley in 1948, and by 1960 males of this strain no longer possessed the ability to cross. However, this strain is now missing (Nigon and Félix 2017). Another pair of strains, Bergerac LY and Bergerac FR, were shared in 1977 and 1980, respectively. Bergerac FR was noted to have offspring numbers

comparable to the standard N2 strain, along with restored male reproduction, while the Bergerac LY sample from 1977 displayed the low fitness associated with Bergerac strains. This strain also showed evidence of high *Tc1* copy-number according to a Southern Blot analysis; however, the exact copy-number could not be determined (Liao *et al.* 1983). Though these strains are no longer publicly available, it is possible they are preserved in various laboratories (Nigon and Félix 2017).

In this study, three Bergerac strains are utilized to quantify the phenotypic variation, if any, that may have occurred during laboratory evolution and divergence. The first, RW7000 (also known as Bergerac BO), was given to David Hirsh by Nigon's student Jean-Louis Brun in 1983, and used in many of the original studies of TEs in *C. elegans*, as described above (Liao *et al.* 1983; Rosenzweig *et al.* 1983; Moerman and Waterston 1984; Mori *et al.* 1988;). The second, known as RW6999, is relatively understudied and reported to be a subclone of RW7000 on the *Caenorhabditis* Genetics Center (CGC) website (https://cgc.umn.edu/strain/RW6999). The final strain, CB4851, was shared with Sydney Brenner in 1969, and thus could have diverged from RW7000 for 14 years (Nigon and Félix 2017).

### *2.2.2 Fitness assays*

To explore the deleterious phonotypes exhibited by the Bergerac strains RW6999, RW7000, and CB4851 relative to the laboratory strain N2, four fitness-related traits (developmental rate, productivity, longevity, and survivorship) were maintained on Nematode Growth Medium (NGM) agar plates seeded with the *E. coli* strain OP50 at 20°C, the standard temperature for *C. elegans* culturing (Corsi *et al.* 2015). To complete the assays, frozen stocks for each of the chosen strains were thawed and individual worms were isolated onto NGM plates. In order to establish independent replicates, which enables within-strain replication, 15 and 20

worms were isolated for the N2 control and each Bergerac strain, respectively. After isolation, the worms were allowed to reproduce and individuals in the L4 larval stage were transferred with a worm-picker to a new plate, establishing five sub-replicates for each replicate ($n$ = 75 lines for N2 and $n$ = 100 lines for each Bergerac strain). To negate the possibility of maternal or grandmaternal effects from cryopreservation on the assays (Lynch 1985), each sub-replicate was transferred for one more generation. Third generation worms were isolated in the L1 larval stage for the assays. This hierarchical structure (strains, replicates, and sub-replicates) combined with the fact that hermaphrodites self-fertilize to produce each new generation, minimizes the possibility of genetic divergence within a sub-replicate, allowing for a measure of environmental variance by comparing sub-replicates.

The first three assays (development, productivity, and longevity) were conducted on a single worm isolated from each subline. To assay developmental rate, starting 36 hours after L1s are isolated, worms were checked every two hours to identify the time (hours) until the first egg reached the worm's uterus. When an egg was identified in the uterus, a worm was scored as having developed to adulthood. This initial measurement yielded the developmental time. Worms that died before reaching adulthood were not scored. The inverse of the developmental time yielded the developmental rate. To assay productivity, each worm that developed to adulthood was transferred to a new plate every 24 hours for eight days. After transferring the worm, the eggs from the previous plate were allowed to hatch for an additional 24 hours, then stored at 4°C for a minimum of one month to allow the progeny to die without producing offspring. Counts were conducted by staining each plate with a 0.075% water dilution of toluidine blue dye, which temporarily makes the progeny stand out white against a purple background to enable easier counting. After the eight days of productivity transfers, the worms

were left on a single agar plate seeded with the E. coli strain OP50 until death in order to score longevity in days. To score longevity, the worms transferred for the productivity assay were monitored each day for movement and pharyngeal pumping. When no movement was detected, the agar pad near the worm was gently tapped. If no response was detected, the tail of the worm was tapped. If the worm still did not respond, it was recorded as dead and days to mortality was calculated.

For the survivorship assay, 10 L1 siblings for each third-generation subline used for the other assays were isolated on the same day, onto a 60mm agar plate. For some sublines, less than 10 L1 individuals were isolated due to the low and delayed productivity of Bergerac worms. 36 hours after isolation, the plates were checked for worms that survived to adulthood, and each plate was scored using the fraction of worms that survived to adulthood (values ranged from 0 to 1). For plates with desiccated worms on the edge of the plate, worms were still scored as surviving if eggs were observed in the uterus, to provide a more conservative estimate of survivorship.

### *2.2.3 Statistical analyses*

A two-level nested analysis of variance (ANOVA) for unequal sample sizes was conducted for each fitness trait. This statistical test compares the variance among strains (N2, CB4851, RW6999, and RW7000), among lines (15 N2 and 20 Bergerac lines), and among sublines (5 replicates for each line) in order to distinguish the between strain variance from the line-specific and environmental variance (Sokal 1995). To conduct pairwise comparisons for each pair of strains for each trait, a Tukey-Kramer HSD test for unequal sample sizes was used, assuming a standard experimental error of 5%.

## 2.3 Results

### *2.3.1 Absolute fitness values*

The fitness assays were conducted as described in the methods section and analyzed using a two-level ANOVA. The mean fitness values for each trait measured are displayed in Figure 2.1. For the survivorship assay, N2 had a survival rate to adulthood of approximately 98%, while the strains CB4851, RW6999, and RW7000 had survivorship rates of approximately 91%, 89%, and 83%, respectively. In the longevity assay, N2 worms survived an average of 13.5 days, while the Bergerac strains survived from 8.54 to 10 days, on average. Bergerac strains produced a much smaller number of offspring compared to N2, with N2 worms producing ~308 offspring and the strains CB4851, RW6999, and RW7000 producing an average of 75, 107, and 43 offspring, respectively. As reported in a previous study (Shook and Johnson 1999), many Bergerac worms were observed to die by bagging, a phenotype characterized by larvae hatching within the hermaphrodite before egg-laying occurs. However, these worms were observed to lay some eggs before bagging, and were included in the analysis to provide a realistic estimate of mean productivity in laboratory conditions. The mean developmental times for N2, CB4851, RW6999, and RW7000 were 47.6, 60.0, 53.4, and 60.5 hours, respectively.

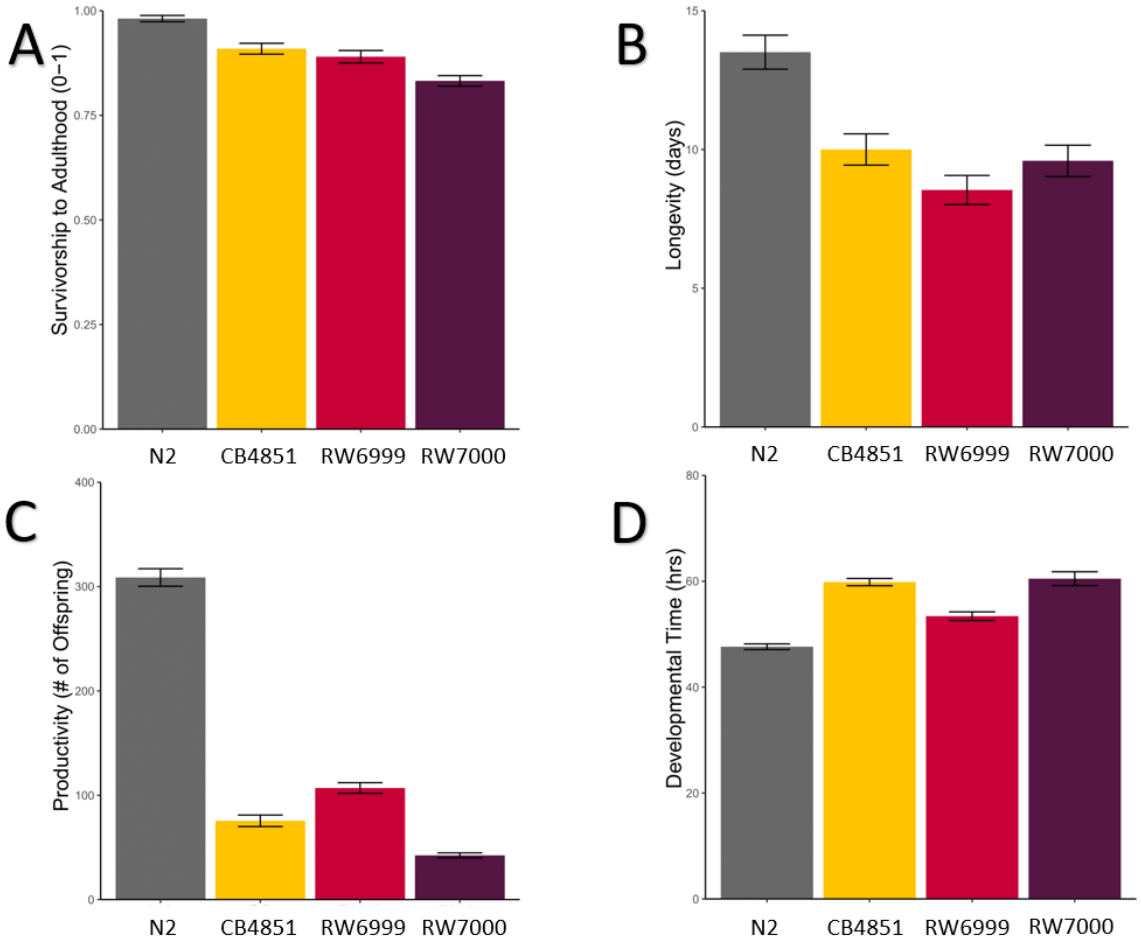


*Figure 2.1: Absolute fitness values for survivorship to adulthood (A), longevity (B), productivity (C), and developmental time (D) in three Bergerac strains and the N2 strain, a wildtype control, with error bars representing ± 1 standard error.*

### *2.3.2 Analysis of variance*

Twenty replicates were established for each Bergerac strain in the fitness assays, while the N2 control had 15 replicates. Five descendants of each replicate were chosen to establish sub-replicates during the assay. For some Bergerac replicates, the high mortality rate led to a smaller number of sub-replicates being established. Exact replicate numbers can be found in Table 2.1. A nested ANOVA revealed significant declines in each of the four fitness traits in all three Bergerac strains relative to the N2 control strain (developmental time: $F = 35.5$, $p < 2 \times 10^{-16}$;

longevity: $F = 12.73$, $p = 1.05 \times 10^{-7}$; productivity: $F = 409.34$, $p < 2 \times 10^{-16}$; survivorship: $F = 23.79$, $p = 1.02 \times 10^{-13}$). For the among-replicates comparison, all traits showed no significant differences between replicates except for survivorship ($0.01 \leq p \leq 0.05$). The final level of comparison tests for environmental variance within replicates. According to this test, no significant differences within replicates were detected, indicating that no environmental variation significantly affected the results any fitness assay.

*Table 2.1: Two-level nested ANOVA results for all assays*

| ***Survivorship*** | **Df** | **SS** | **MS** | **F** |
|---|---|---|---|---|
| Among Strains | 3 | 0.945 | 0.315 | 23.79*** |
| Among Replicates | 69 | 1.275 | 0.018 | 1.39* |
| Within Replicates | 274 | 3.629 | 0.013 | |
| Total | 346 | | | |
| ***Longevity*** | | | | |
| Among Strains | 3 | 1041 | 347 | 12.733*** |
| Among Replicates | 69 | 2035 | 29.5 | 1.082 |
| Within Replicates | 220 | 5995 | 27.3 | |
| Total | 292 | | | |
| ***Productivity*** | | | | |
| Among Groups | 3 | 3069207 | 1023069 | 324.49*** |
| Among Replicates | 69 | 217546 | 3153 | 1.08 |
| Within Replicates | 223 | 650800 | 2918 | |
| Total | 297 | | | |
| ***Development*** | | | | |
| Among Strains | 3 | 7804 | 2601.3 | 35.509*** |
| Among Replicates | 69 | 5348 | 77.5 | 1.058 |
| Within Replicates | 223 | 16337 | 73.3 | |
| Total | 295 | | | |

*2.3.2 Strain comparisons*

Next, to determine which strains significantly differ from each other, Tukey-Kramer HSD tests were used to compare strain pairs for each of the four fitness traits analyzed. The results of these comparisons are presented in Figure 2.2, and details regarding mean differences and exact *p*-values for the Tukey-Kramer HSD tests are shown in Tables 2.2-2.5. For all traits measured, each Bergerac strain showed significant differences compared to N2.

The fitness assays were also able to detect significant differences among Bergerac strain comparisons for all traits measured. With the exception of the CB4851 vs. RW6999 comparison, all other pair-wise comparisons showed significant differences for survivorship to adulthood (RW7000 < RW6999, $p = 4.70 \times 10^{-3}$; RW7000 < CB4851, $p = 1.90 \times 10^{-4}$). The three Bergerac strains were not significantly different from each other with respect to longevity, with all Bergerac worms surviving an average of 9 days after the L1 larval stage. Productivity was the most variable trait among the Bergerac strains, given that all pair-wise comparisons of strains yielded highly significant trait differences (RW7000 < CB4851, $p = 1.60 \times 10^{-3}$; CB4851 < RW6999, $p = 1.10 \times 10^{-3}$; RW7000 < RW6999, $p = 8.75 \times 10^{-13}$). The developmental rate assay also revealed significant differences among the Bergerac strains, with only CB4851 and RW7000 displaying a significant difference in developmental time (RW6999 < RW7000, $p = 1.87 \times 10^{-6}$; RW6999 < CB4851, $p = 5.94 \times 10^{-5}$).

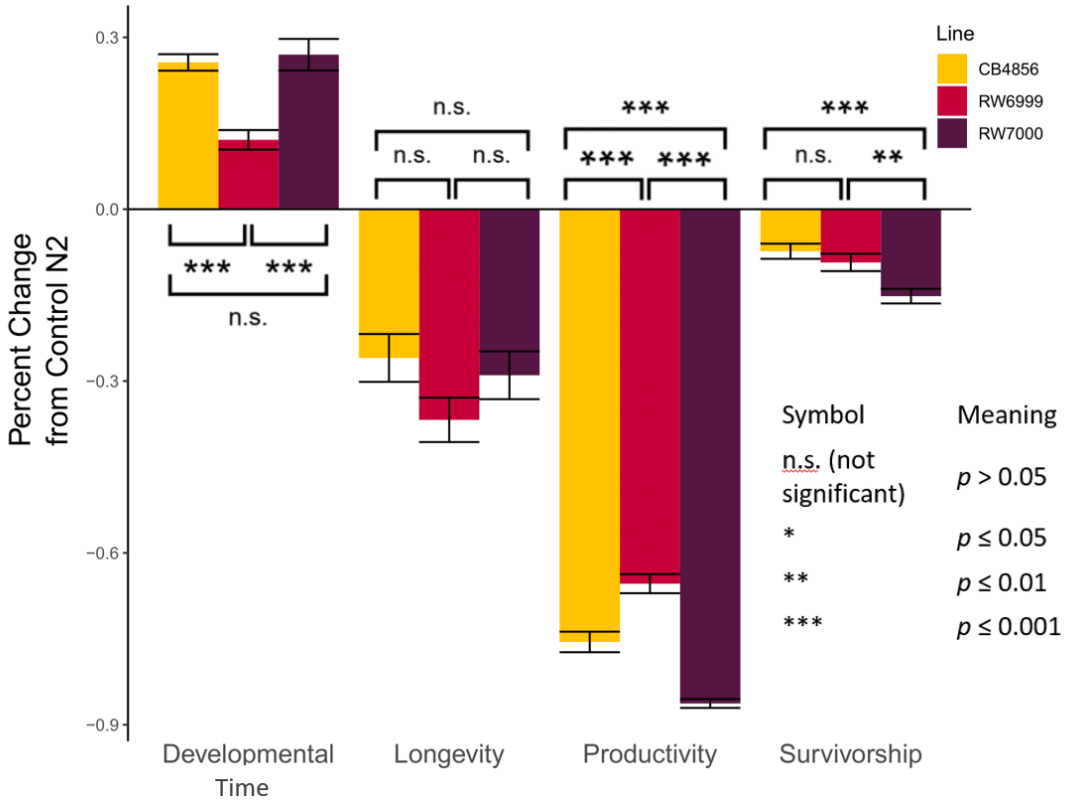


*Figure 2.2: Relative fitness for traits in Bergerac strains compared to the wildtype control, N2. Compared to the N2 strain, all strains showed highly significant differences in trait values ($p \leq 0.001$). The stars on brackets summarize p-values for Tukey-Kramer HSD comparisons between Bergerac strains. Exact p-values for the Tukey-Kramer HSD comparisons can be found in tables 2.2-2.5*

*Table 2.2: Tukey-Kramer HSD results for survivorship to adulthood*

| Strain pair | Difference in means (fraction surviving) | Adjusted *p*-value | Level of significance |
|---|---|---|---|
| N2-CB4851 | 0.072014245 | 0.000783512 | *** |
| RW6999-CB4851 | -0.019167573 | 0.68979707 | n.s. |
| RW7000-CB4851 | -0.076898036 | 0.000102266 | *** |
| RW6999-N2 | -0.091181818 | 2.62E-06 | *** |
| RW7000-N2 | -0.148912281 | 2.99E-13 | *** |
| RW7000-RW6999 | -0.057730463 | 0.003112075 | ** |

*Table 2.3: Tukey-Kramer HSD results for longevity*

| Strain pair | Difference in means (days) | Adjusted *p*-value | Level of significance |
|---|---|---|---|
| N2-CB4851 | 3.507042254 | 0.000740046 | *** |
| RW6999-CB4851 | -1.459770115 | 0.32715947 | n.s. |
| RW7000-CB4851 | -0.408450704 | 0.968806964 | n.s. |
| RW6999-N2 | -4.966812368 | 6.30E-08 | *** |
| RW7000-N2 | -3.915492958 | 7.39E-05 | *** |
| RW7000-RW6999 | 1.051319411 | 0.589877349 | n.s. |

*Table 2.4: Tukey-Kramer HSD results for productivity assay*

| Strain pair | Difference in means (hours) | Adjusted *p*-value | Level of significance |
|---|---|---|---|
| N2-CB4851 | 233.2421194 | 0.0010053 | *** |
| RW6999-CB4851 | 201.8332524 | 0.0016465 | *** |
| RW7000-CB4851 | 265.2577835 | 0.0018911 | *** |
| RW6999-N2 | -31.408867 | 0.0010053 | *** |
| RW7000-N2 | 32.01566416 | 0.0010053 | *** |
| RW7000-RW6999 | 63.42453116 | 0.0010053 | *** |

*Table 2.5: Tukey-Kramer HSD results for developmental time*

| Strain pair | Difference in means (hours) | Adjusted *p*-value | Level of significance |
|---|---|---|---|
| N2-CB4851 | -12.20490686 | 1.98E-13 | *** |
| RW6999-CB4851 | -6.436410827 | 5.79E-05 | *** |
| RW7000-CB4851 | 0.648132428 | 0.970999599 | n.s. |
| RW6999-N2 | 5.768496034 | 0.000212342 | *** |
| RW7000-N2 | 12.85303929 | 8.52E-14 | *** |
| RW7000-RW6999 | 7.084543255 | 1.90E-06 | *** |

## 2.4 Discussion

All Bergerac strains exhibited severe fitness decline when compared to the wildtype N2 strain for each of the four fitness traits, supporting previous observations and the hypothesis that the high *Tc1* copy-number in these strains is associated with decreased fitness (Hodgkin and Doniach 1997). Fitness values for the wildtype control N2 were comparable to previous literature. Average values for developmental time and longevity for the N2 control were all similar to a previous set of assays (Dubie *et al.* 2020). The average values for percent survivorship and productivity matched the results of several assays conducted using identical protocols (Katju *et al.* 2015; Dubie *et al.* 2020).

In addition to confirming that all Bergerac strains had low fitness, our study is the first to quantify significant differences between different Bergerac strains for several fitness traits. Overall, strain RW7000 (also known as Bergerac-BO), which was used in many transposon-tagging studies in the 1990s (Korswagen *et al.* 1996), had the lowest fitness or was tied for lowest fitness for each trait measured, including productivity, which was 58% lower than the next least-fit strain, CB4851. Interestingly, the strain RW6999, listed as an RW subclone of RW7000 on the CGC website, displayed the highest fitness among the Bergerac strains. Assuming that this strain is descended from RW7000, this suggests that the strain has evolved higher fitness during laboratory evolution. However, without any knowledge of the history of these strains before stock delivery to our laboratory, any conclusions would be premature.

Since *Tc1* insertions have been documented to be the most common form of *de novo* germline mutations in Bergerac strains (Collins *et al.* 1987), many of the observed differences in fitness in distinct Bergerac strains could be due to differing copy-numbers of *Tc1* or differing insertion sites. Although the exact histories of each strain before reaching our laboratory was

unknown at the time of this study, it was likely that these strains were isolated and cultured in various laboratories for years before cryopreservation. After several hundred generations of experimental evolution, the average fitness of *C. elegans* lines improved with high effective population size (Katju *et al.* 2015), demonstrating that different laboratory propagation methods could have influenced the evolution of the Bergerac strains. It is possible that propagation methods unique to each laboratory influenced the phenotypic divergence of the strains. The next section of this thesis will explore the genomes of the Bergerac strains using whole-genome sequencing to analyze the details of transposon insertions in the Bergerac genomes, clarifying the genetic causes of the phenotypic differences observed in this section.

## 2.5 References


Abdulkader N, Brun J. 1980. A temperature-sensitive mutant of *Caenorhabditis elegans* var. Bergerac affecting morphological and embryonic development. *Genetics*. 51(2):81-92.

Bessereau JL. 2006. Transposons in *C. elegans*. *WormBook : the online review of C. elegans biology*.1-13.

Collins J, Saari B, Anderson P. 1987. Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. *Nature*. 328(6132):726-728.

Dubie JJ, Caraway AR, Stout MM, Katju V, Bergthorsson U. 2020. The conflict within: Origin, proliferation and persistence of a spontaneously arising selfish mitochondrial genome. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 375(1790):20190174.

Fatt HV, Dougherty EC. 1963. Genetic control of differential heat tolerance in two strains of the nematode *Caenorhabditis elegans*. *Science*. 141(3577):266-267.

Hodgkin J, Doniach T. 1997. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics*. 146(1):149-164.

Katju V, Packard LB, Bu L, Keightley PD, Bergthorsson U. 2015. Fitness decline in spontaneous

mutation accumulation lines of *Caenorhabditis elegans* with varying effective population sizes. *Evolution*. 69(1):104-116.

Korswagen HC, Durbin RM, Smits MT, Plasterk RHA. 1996. Transposon *Tc1*-derived, sequence-tagged sites in *Caenorhabditis elegans*; as markers for gene mapping. *Proc. Natl. Acad. Sci. U S A*. 93(25):14680.

Lee Y, Hwang W, Jung J, Park S, Cabatbat JJT, Kim P-J, Lee S-JV. 2016. Inverse correlation between longevity and developmental rate among wild *C. elegans* strains. *Aging*. 8(5):986-999.

Liao LW, Rosenzweig B, Hirsh D. 1983. Analysis of a transposable element in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U S A*. 80(12):3585-3589.

Lynch M. 1985. Spontaneous mutations for life-history characters in an obligate parthenogen. *Evolution*. 39(4):804-818.

Malone CJ, Misner L, Le Bot N, Tsai MC, Campbell JM, Ahringer J, White JG. 2003. The *C. elegans* hook protein, zyg-12, mediates the essential attachment between the centrosome and nucleus. *Cell*. 115(7):825-836.

Moerman DG, Waterston RH. 1984. Spontaneous unstable *unc-22* IV mutations in *C. elegans* var. Bergerac. *Genetics*. 108(4):859-877.

Mori I, Moerman DG, Waterston RH. 1988. Analysis of a mutator activity necessary for germline transposition and excision of *Tc1* transposable elements in *Caenorhabditis elegans*. *Genetics*. 120(2):397-407.

Rosenzweig B, Liao LW, Hirsh D. 1983. Sequence of the *C. elegans* transposable element *Tc1*. *Nucl. Acids Res*. 11(12):4201-4209.

Shook DR, Johnson TE. 1999. Quantitative trait loci affecting survival and fertility-related traits in *Caenorhabditis elegans* show genotype-environment interactions, pleiotropy and epistasis. *Genetics*. 153(3):1233-1243.

Sokal RR. 1995. The principles and practice of statistics in biological research. *Biometry*. 451-554.

Strain: RW6999, Genotype: C. elegans wild isolate. - Caenorhabditis Genetics Center (CGC) - College of Biological Sciences. 2021. cgc.umn.edu. [accessed 2021 Apr 12]. https://cgc.umn.edu/strain/RW6999

Vertino A, Ayyadevara S, Thaden JJ, Shmookler Reis RJ. 2011. A narrow quantitative trait locus in *C. elegans* coordinately affects longevity, thermotolerance, and resistance to paraquat. *Front. Genet*. 2:63.

Wood WB, Hecht R, Carr S, Vanderslice R, Wolf N, Hirsh D. 1980. Parental effects and phenotypic characterization of mutations that affect early development in *Caenorhabditis elegans*. *Dev. Biol*. 74(2):446-469.

# 3. GENOMIC DISRUPTION BY TRANSPOSABLE ELEMENTS

## 3.1 Introduction

The field of genomics, the study of the entire set of genetic instructions in an organism, has seen a period of rapid growth in the past twenty years. While sequencing all the exons of a gene was a difficult task at the turn of the century, now, entire genomes can be quickly sequenced at an affordable price (Fig. 3.1). Access to entire genomes allows scientists to ask more nuanced questions about parts of the genome that were neglected in the pre-genomics era, including non-coding RNA, epigenetics, *cis*-regulatory elements, and TEs (McGuire *et al.* 2020). The study of complex phenotypes using genomics has shown that the genetics of even a simple phenotype, like height, are governed by multiple genes and regulatory elements spread throughout genomes. This shows that, until the function and structure of all features of the genome are described, a full understanding of life will remain elusive (Boyle *et al.* 2017). The study of TEs has benefitted from the genomics revolution, as the movement of TEs throughout the genome made locating and sequencing TEs difficult before the advent of whole-genome sequencing. Now, genomics shows how TE movement can cause genomic variability by disrupting genes, moving or deleting DNA sequences, and even causing chromosomal rearrangements. These genetic changes occasionally lead to adaptive phenotypes, but host organisms have also evolved complex regulatory systems to keep TEs under control (Klein and O'Neill 2018; Payer and Burns 2019). In this section, whole-genome sequencing is used to gain a better understanding of the *Tc1* proliferation in the genomes of the Bergerac strains, setting the foundation for future studies that will establish a deeper understanding of TE regulation.

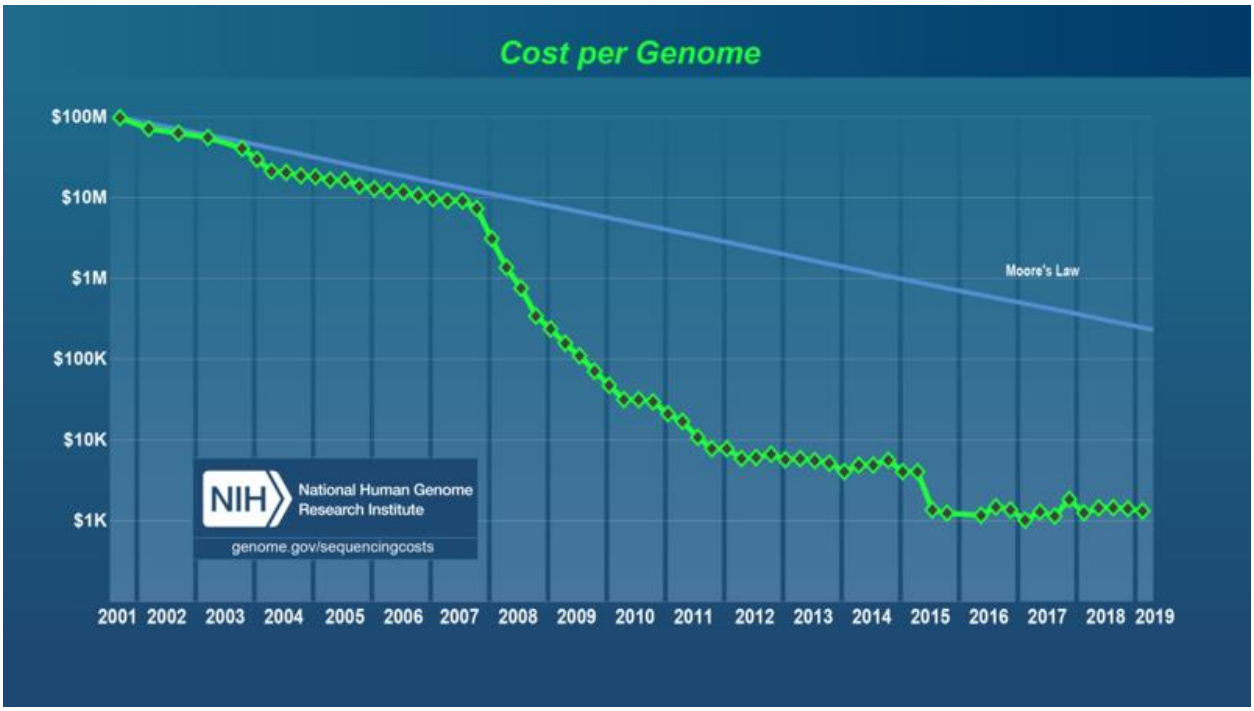


*Fig. 3.1 The cost of sequencing a human genome has dropped dramatically, allowing researchers studying a variety of organisms to answer questions about entire genomes rather than a small fraction of genes. This image is made available under the Creative Commons CC0 1.0 Universal Public Domain Dedication. Source: https://www.genome.gov/sequencingcosts/*

### *3.1.1 Previous estimates of Tc1 copy-number*

Before the era of genomics, various molecular techniques were used to study the genetics of Bergerac strains. Due to a unique mutator activity observed in specific Bergerac subcultures in which the descendants of mutated worms occasionally reverted to wildtype in a non-Mendelian fashion, researchers speculated and later confirmed that the transposable element *Tc1* was highly active in Bergerac strains. As mentioned earlier, early experiments estimated that *Tc1* activity was at least 100-fold higher by studying reversion phenotypes. However, at the same time researchers were working to estimate the copy-number of *Tc1* in the Bergerac strains (Moerman and Waterston 1984). A southern blot analysis, which used molecular probes to detect the sequence for *Tc1*, estimated that there were 31 copies of *Tc1* in the wildtype N2 wildtype

laboratory strain and a minimum 300 copies of *Tc1* in a strain called Bergerac-LY (Liao *et al.* 1983). This Southern blot technique had poor resolution above 100 copies, so this analysis was unable to estimate an exact *Tc1* copy-number. A decade later, a quantitative dot blot analysis, in which a strain's DNA sample is probed for *Tc1*, followed by an analysis of the light intensity of the resulting band, estimated that the strain RW7000 had 419-527 copies of *Tc1* (Egilmez *et al.* 1995).

### *3.1.2 Early attempts to locate locus of increased transposition*

Researchers previously attempted to identify the source of dramatic *Tc1* proliferation in the Bergerac strains, but the exact cause has not been identified. By crossing wildtype N2 worms with RW7000 worms, a large region of the Bergerac chromosome I, named *mut-4*, was found to be associated with increased *Tc1* activity, which was measured by the reversion of *unc-22* mutations in offspring of mutant worms. Further crossing revealed that this locus of increased *Tc1* activity occasionally moved to new locations, indicating that the source of *Tc1* proliferation in Bergerac worms could be a transposable element (Mori *et al.* 1988). Despite this promising initial evidence, this hypothesis has never been confirmed.

### *3.1.3 Genomic analysis of Bergerac strains*

While useful in revealing the drastic increase in *Tc1* copy-number in one Bergerac strains, these pre-genomic estimates were incapable of detecting the locations of all *Tc1* elements in the genome. Recently studies have completed whole genome sequencing on many *C. elegans* wild isolates, and one study sequenced the genome of the Bergerac strain CB4851 (Cook *et al.* 2016). The *Tc1* copy-number of this Bergerac strain was estimated to be 406 in a study of *Tc1* elements in 208 wild isolates, though no further analysis was completed using this genome (Laricchia *et al.* 2017). This section details a genomic analysis of the strains CB4851, RW7000,

and RW6999, revealing copy-number estimates and the genomic locations of *Tc1*. In addition, the phenotypic data for each strain are compared to *Tc1* copy-number, and the Bergerac genomes are compared to the N2 reference genome to identify potential sources of increased *Tc1* proliferation.

## 3.2 Methods

### *3.2.1 DNA extraction and Illumina sequencing*

Genomic DNA was extracted from Bergerac and N2 control lines as previously described (Konrad *et al.* 2018) with libraries prepared using the Nextera DNAflex library kit (Illumina, San Diego, CA). Libraries were sequenced on the Illumina Novaseq6000 platform (2 ×150bp) at the North Texas Genome Center at the University of Texas at Arlington.

### *3.2.2 Tc1 copy-number estimates*

*Tc1* copy-number in each genome was determined using the McClintock meta-pipeline, which combines many TE-detection algorithms to identify reference and non-reference TE insertions in each genome. Because each TE-detection algorithm has its own strengths and weaknesses, using multiple algorithms provides a range of estimates that will capture more TEs in the genome. For this analysis, the McClintock v0.2.1 was used. A standard consensus sequence of a 1610 bp *Tc1* element was used as an input, and two TE callers, relocaTE and Retroseq, were selected from McClintock to estimate *Tc1* copy-number. To determine the relationship between TE load and fitness, the correlation between *Tc1* copy-number and each fitness trait analyzed previously was determined using Pearson and Spearman statistical tests on R.

### *3.2.3 Genes with exons disrupted by Tc1*

Next, genes with exons disrupted by *Tc1* according to the TE caller relocaTE were identified. A BED file generated by the McClintock pipeline provided the location of *Tc1* elements in each strain's genome, and these locations were compared to exon regions in an annotated *C. elegans* reference genome (PRJNA13758.WS279). The tool SimpleMine (https://wormbase.org/tools/mine/simplemine.cgi) from the *C. elegans* database WormBase was used to extract the Public Name, Genetic Map Position, and RNAi phenotypes for each gene with an exon disrupted by *Tc1*. A word cloud was generated on https://monkeylearn.com/word-cloud/, to visualize the top 20 RNAi phenotypes for disrupted genes.

### *3.2.4 Tc1 landing site sequence analysis*

Next, the sequence context of each *Tc1* insertion predicted by relocaTE was analyzed in each Bergerac strain. The FASTQ reads generated by Illumina whole-genome sequencing were aligned to the *C. elegans* reference genome (PRJNA13758.WS279) with the Burrows-Wheeler Aligner (BWA), version 0.7.12-r1039. Using the BED file generated by the McClintock pipeline, sequences ± 25 bp from each insertion site were extracted from the BWA alignments. These sequences were aligned using ClustalW and a consensus sequence was generated for each strain with IUPAC nucleotide code for each strain by analyzing distribution of nucleotides ± 6bp from the center of the insertion.

## 3.3 Results

### *3.3.1 Tc1 copy-number estimates*

Whole-genome sequencing was completed on strains N2, CB4851, RW6999, and RW7000, to obtain information about the causes and consequences of *Tc1* proliferation in the

Bergerac strains. First, the copy number of *Tc1* was estimated using McClintock, a meta-pipeline that allows for multiple TE callers to be used in parallel (Nelson *et al.* 2017). Two main techniques are used to design TE calling algorithms for Illumina whole-genome sequencing data—read-pair methods and split-read methods. As there is no consensus on the most accurate method for calling TEs, a read-pair method (RetroSeq) and a split-read method (relocaTE) were used to obtain two estimates of *Tc1* copy-number within each genome (Keane *et al.* 2013; Robb *et al.* 2013). Figure 3.2 displays the *Tc1* copy-number estimates, confirming that the Bergerac strains contain extremely high *Tc1* copy-numbers, with the three strains containing 16 to 27 times more copies of *Tc1* in their genomes compared to N2. RelocaTE and RetroSeq also produced fairly consistent copy-number estimates for each Bergerac strain, with the largest discrepancy of 4.51% found in CB4851. In the wildtype N2 strain, RetroSeq estimated only one copy of *Tc1*, while RelocaTE predicted 28 *Tc1* elements. This result was expected, as RetroSeq only calls non-reference insertions within sequenced genomes. Since most of the ~30 *Tc1* copies in N2 are annotated in the reference genome used for this analysis, a low copy-number estimate of non-reference *Tc1* insertions from RetroSeq was expected (Bessereau 2006). RW7000 contains the most genomic *Tc1* insertions, while CB4851 contains the least insertions among the Bergerac strains.

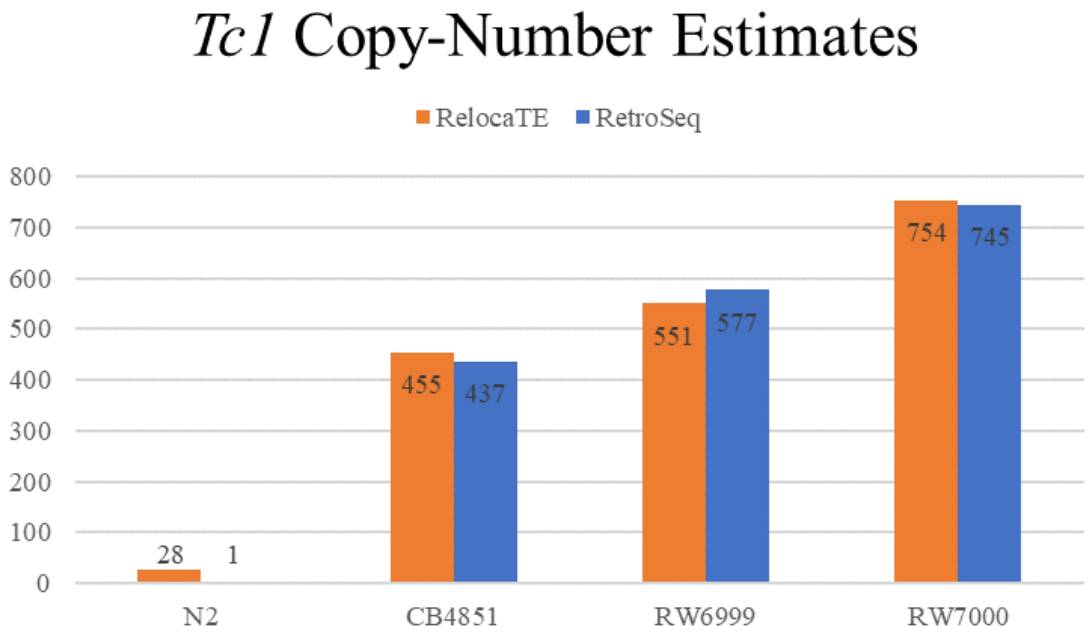


*Figure 3.2: Tc1 copy-number estimates based on Illumina whole-genome sequencing data, using two TE calling algorithms. The Bergerac strains studied contain 16 to 27 times more copies of Tc1 than the N2 wildtype control, and sizable copy-number differences between Bergerac strains are evident.*

### *3.3.2 Tc1 copy-number relationship to fitness traits*

Next, the relationship between the previously studied fitness traits and the genomic copy-number of *Tc1* was studied using Pearson and Spearman correlations. *Tc1* copy-number was found to be correlated with decreased fitness for each trait studied, supporting the hypothesis that strains with higher *Tc1* copy-number suffer phenotypic consequences due to genomic disruptions by *Tc1*. *p*-values for the Spearman and Pearson tests, shown in Tables 3.1 and 3.2, all were less than .001 and thus were highly significant. A visualization of the correlations of fitness traits and *Tc1* copy-number, as predicted by relocaTE, is displayed in Figure 3.3. As expected, all correlations are negative with the exception of developmental time. Despite the clear relationship between increased *Tc1* copy-number and decreased fitness, the line RW6999 is an exception, as it displays increased productivity and decreased developmental time relative to the other Bergerac strains, indicating increased fitness (Fig. 2.1 and Fig. 3.3).

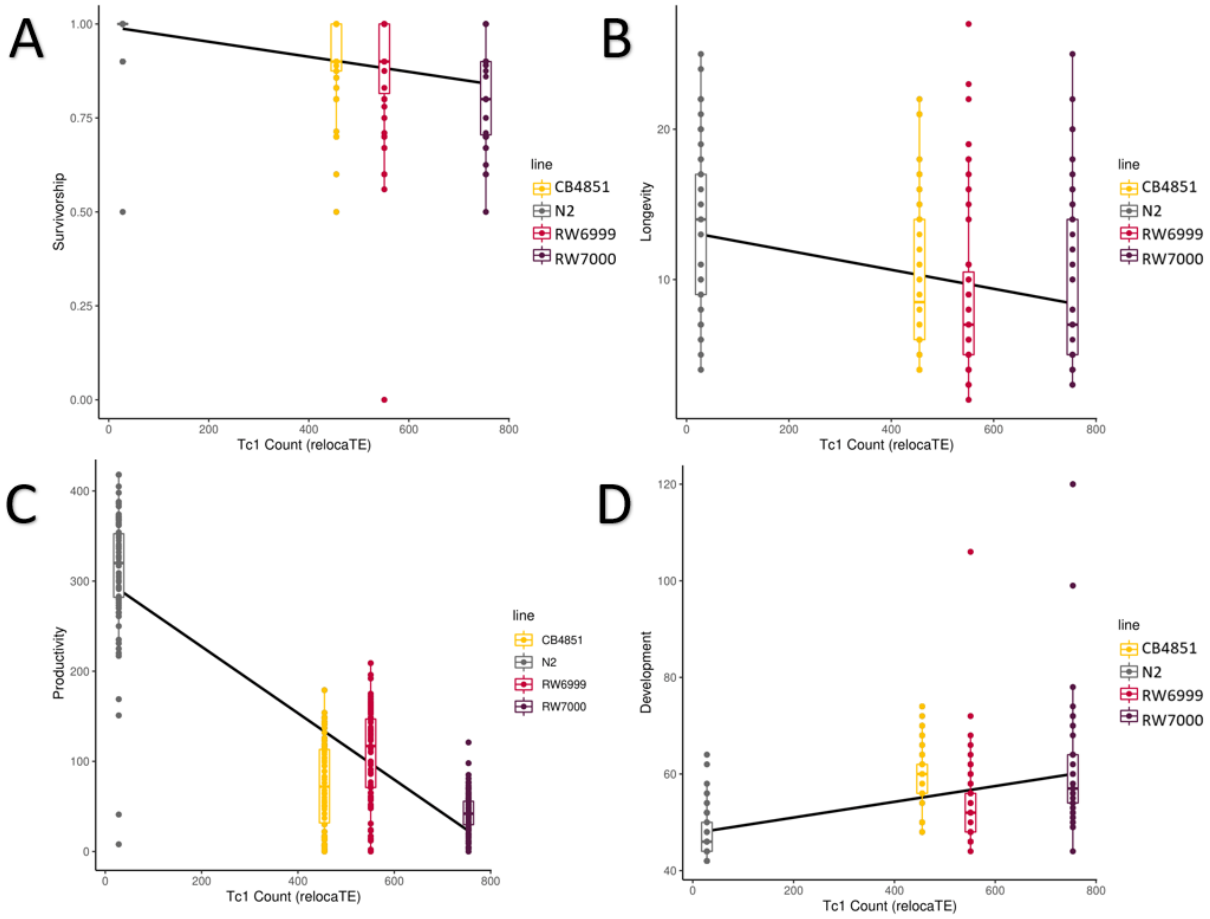


*Figure 3.3: Correlation of four fitness traits (survivorship (**A**), longevity (**B**), productivity (**C**), and development (**D**)) to Tc1 copy number in the N2 wildtype control and the Bergerac strains CB4851, RW6999, and RW7000. Box-and-whisker plots display the range of fitness values, and the x-axis represents Tc1 copy number estimated by relocaTE. All correlations were determined to be significant by Pearson and Spearman statistical tests.*

*Table 3.1: Pearson correlation coefficient data for comparison of fitness assay data to Tc1 copy-number*

| **Assay** | **Correlation Coefficient** | ***t*** | **df** | ***p*-value** |
|---|---|---|---|---|
| Survivorship | -0.397 | -8.0452 | 345 | 1.40E-14 |
| Longevity | -0.297 | -5.3129 | 291 | 2.14E-07 |
| Productivity | -0.852 | -27.888 | 294 | 2.2E-16 |
| Developmental Time | 0.430 | 8.1625 | 294 | 9.71E-15 |

*Table 3.2: Spearman correlation coefficient data for comparison of fitness assay data to Tc1 copy-number*

| **Assay** | **Correlation Coefficient** | ***S*** | **p-value** |
|---|---|---|---|
| Survivorship | -0.465 | 10198868 | 5.56E-20 |
| Longevity | -0.283 | 5376574 | 8.85E-07 |
| Productivity | -0.658 | 7166758 | 2.2e-16 |
| Developmental Time | 0.404 | 2575574 | 4.68E-13 |

Next, the locations of *Tc1* insertions provided by relocaTE were analyzed. A list of genes with exons disrupted by *Tc1* was generated using annotations in the *C. elegans* reference genome. The public gene name, genetic map position, and RNAi phenotypes of all genes with disrupted exons are displayed in the Appendix, in tables A.1-A.3. In total CB4851, RW7000, and RW6999 had 110, 224, and 165 genes predicted to have *Tc1* insertions in exons, respectively, matching the ratios seen when comparing their total *Tc1* copy-numbers. RNAi is a commonly

used method to knock down gene expression in *C. elegans* where double-stranded RNA is introduced to the organism, causing the mRNA matching the dsRNA sequence to be repressed (Han 2018). Since *Tc1* insertions in exons are expected to knock down or eliminate expression of a gene, RNAi phenotypes provide an estimate of the phenotypic consequences of *Tc1* insertions. A word cloud summarizing the top twenty RNAi phenotypes in the Appendix tables A.1-A.3 was created (Fig. 3.4). Many of the phenotypes in this word cloud complement the results of the fitness assays in the Bergerac strains, including reduced brood size (decreased productivity), slow growth (longer developmental time), embryonic lethality (survivorship), and shortened life span (longevity). Many other phenotypes reveal the promise of future assays and experiments on the Bergerac strains, including the predicted RNAi phenotypes involving transgene expression and locomotion.


*Figure 3.4: Word cloud displaying the 20 most common RNAi phenotypes for genes with disrupted exons in the Bergerac strains. Although Tc1 insertions may not completely knock out gene activity like RNAi, similar phenotypic consequences are expected if gene expression is reduced. All relevant genes and RNAi phenotypes are located in the Appendix.*

### *3.3.3 Analysis of Tc1 landing sites*

The analysis of base composition (± 6) bp surrounding all relocaTE-detected *Tc1* insertions was conducted as described in the methods section, with each strain being analyzed independently to identify a shared motif surrounding *Tc1* landing sites. In IUPAC nucleotide code, the sequence WWAYRTAYRTWW was found to be supported in each strain. As expected, positions ± 1 were found to be TA in all *Tc1* insertions. Also, positions ± 4 were a pair of A and T in ~80% of detected *Tc1* insertions.

## 3.4 Discussion

This study represents the first comprehensive analysis of *Tc1* insertions in the Bergerac strains using whole-genome sequencing data. While previous methods used to study the genetics of these unique strains demonstrated that *Tc1* activity and copy-number were high in certain Bergerac strains, whole-genome sequencing provides a more precise picture of the unique features spread throughout each genome. Two different TE callers detected ~750 copies of *Tc1* in the genome of RW7000, which is at least 125 copies more than the previous estimate using quantitative dot blot hybridization (Egilmez *et al.* 1995). This could reflect the shortcomings of the previous method or differences in the laboratory strains of RW7000, as the culturing history of each laboratory strain is unclear and *Tc1* proliferation is still active in RW7000 (Moerman and Waterston 1984). The results of the TE callers are also comparable to a previous genomic estimate of *Tc1* copy-number in the strain CB4851; however, the estimates reported here (437 and 455), are ~30 to 50 copies higher than the previous estimate of 406 copies (Laricchia *et al.* 2017). As this study used the results of a different sequencing run and different TE callers, rerunning the analysis on both genomes could reveal the cause of the discrepancy. The accuracy

of current TE calling algorithms using Illumina PE sequencing data is not perfect, so efforts are underway to improve and synthesize the outputs of different TE callers (Vendrell-Mir *et al.* 2019). Whole-genome sequencing reads from this project will remain publicly available for future analyses as the algorithms improve.

The output from the TE caller relocaTE was chosen for further analyses of *Tc1* copy-number, location, and distribution in the Bergerac genomes. This TE caller uses split-read methods, which were found to be more positionally accurate in an analysis of TE caller performance in detecting synthetic TE insertions in yeast genomes. While relocaTE did produce many positionally inaccurate predictions in this study, this caller produced more correct TE predictions within 100 bp of the true synthetic insertion than other split-read methods (Nelson *et al.* 2017).

A strong correlation was observed when comparing the results of the Bergerac fitness assays to predicted *Tc1* copy-number in each strain, supporting the general scientific consensus that uncontrolled TE proliferation leads to fitness decline (Hodgkin and Barnes 1991; Bessereau 2006). By analyzing *C. elegans* strains with a spectrum of *Tc1* copy-numbers, this relationship was demonstrated even more clearly than before. In addition to improved copy-number estimation, whole-genome sequencing also allowed all *Tc1* elements in each genome to be located. *Tc1* insertions within exons are known to be capable of gene knockouts (Moerman and Waterston 1984). However, some insertions are spliced out of mRNAs before translation, which can lead to silent TE insertions or TE insertions that partially knock down gene transcription (Kiff *et al.* 1988; Rushforth and Anderson 1996). Studies of the relative mRNA transcript levels for these genes compared to the N2 control may provide promising insight into the spectrum of effects caused by TE insertions.

RNA interference (RNAi) is a biological process that causes sequence-specific gene silencing when double-stranded RNA is introduced into an organism. RNAi has been used for targeted silencing of many *C. elegans* genes, and the divergent phenotypes of worms with silenced genes have been studied to elucidate the functions of these genes (Han 2018). Since *Tc1* insertions into exons can cause gene knockouts (Moerman and Waterston 1984), RNAi phenotypes of genes with *Tc1* insertions in exons could be used to estimate the phenotypic effects of these insertions. The top 20 RNAi phenotypes of genes with *Tc1* insertions in the Bergerac genomes are displayed as a word cloud in Figure 3.3, showcasing a variety of predicted phenotypes and supporting observations from the fitness assays. Additional predicted phenotypic effects, including changes in transgene expression and locomotion, could be the subject of future investigations.

An obvious exception to the correlation of *Tc1* copy-number and fitness traits is the strain RW6999, which displayed the highest fitness values of the Bergerac strains, despite having more *Tc1* copies than the strain CB4851. This strain, identified as a subclone of the strain RW7000 on the CGC website (https://cgc.umn.edu/strain/RW6999), was not studied in the past like other Bergerac strains, so its history is unknown. However, as mentioned previously, it appears possible that this strain has recovered fitness, as it is a descendant of RW7000, which has lower fitness and higher *Tc1* copy-number. *C. elegans* strains with low fitness have been shown to evolve higher fitness within several hundred generations in laboratory conditions, so it is possible that different laboratory propagation techniques allowed this strain to adapt to its TE load (Farslow *et al.* 2015). Further investigation of the RW6999 genome may reveal the mechanistic basis for adaptation to TE invasions, which frequently occur in wild populations of many organisms (Weick and Miska 2014).

The *Tc1* landing sites of detected insertions were analyzed to identify a sequence motif surrounding the *Tc1* insertions. This reproduces previous work on *Tc1* insertions in different *C. elegans* strains, confirming that *Tc1* movement throughout the genome is limited by target site preferences (Préclin *et al.* 2003). Future research that takes these target site preferences into account could uncover even more information about *Tc1* insertions in the Bergerac strains, including an analysis of the relative distribution of *Tc1* in various genomic features and chromosomal locations. This initial genomic work on the Bergerac strains, combined with the phenotypic analysis presented earlier, sets the stage for future searches for the mechanistic basis of *Tc1* proliferation, and provides an invaluable resource for any future research using these strains.

## 3.5 References


Bessereau JL. 2006. Transposons in *C. elegans*. *WormBook : the online review of C. elegans biology*.1-13.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: From polygenic to omnigenic. *Cell*. 169(7):1177-1186.

Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga JE et al. 2016. The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics*. 204(1):371-383.

Corsi AK, Wightman B, Chalfie M. 2015. A transparent window into biology: A primer on *Caenorhabditis elegans*. *Genetics*. 200(2):387-407.

Egilmez NK, Ebert RH, 2nd, Shmookler Reis RJ. 1995. Strain evolution in *Caenorhabditis elegans*: Transposable elements as markers of interstrain evolutionary history. *J. Mol. Evol*. 40(4):372-381.

Farslow JC, Lipinski KJ, Packard LB, Edgley ML, Taylor J, Flibotte S, Moerman DG, Katju V, Bergthorsson U. 2015. Rapid increase in frequency of gene copy-number variants during experimental evolution in *Caenorhabditis elegans*. *BMC G*. 16:1044.

Free Word Cloud Generator – Results. 2021. Monkeylearn.com. [accessed 2021 Apr 12]. https://monkeylearn.com/word-cloud/

Han H. 2018. RNA interference to knock down gene expression. *Methods Mol Biol*. 1706:293-302.

Hodgkin J, Barnes TM. 1991. More is not better: Brood size and population growth in a self-fertilizing nematode. *Proc. Biol. Sci.* 246(1315):19-24.

Hodgkin J, Doniach T. 1997. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics*. 146(1):149-164.

Keane TM, Wong K, Adams DJ. 2013. Retroseq: Transposable element discovery from next-generation sequencing data. *Bioinformatics*. 29(3):389-390.

Konrad A, Flibotte S, Taylor J, Waterston RH, Moerman DG, Bergthorsson U, Katju V. 2018. Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans. Proc. Natl. Acad. Sci. U S A.* 115(28):7386-7391.

Kiff JE, Moerman DG, Schriefer LA, Waterston RH. 1988. Transposon-induced deletions in *unc-22* of *C. elegans* associated with almost normal gene activity. *Nature*. 331(6157):631-633.

Klein SJ, O'Neill RJ. 2018. Transposable elements: Genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res*. 26(1-2):5-23.

Laricchia KM, Zdraljevic S, Cook DE, Andersen EC. 2017. Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol. Biol. Evol*. 34(9):2187-2202.

Liao LW, Rosenzweig B, Hirsh D. 1983. Analysis of a transposable element in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U S A*. 80(12):3585-3589.

McGuire AL, Gabriel S, Tishkoff SA, Wonkam A, Chakravarti A, Furlong EEM, Treutlein B, Meissner A, Chang HY, López-Bigas N et al. 2020. The road ahead in genetics and genomics. *Nat. Rev. Genet*. 21(10):581-596.

Moerman DG, Waterston RH. 1984. Spontaneous unstable *unc-22* IV mutations in *C. elegans*

var. Bergerac. *Genetics*. 108(4):859-877.

Mori I, Moerman DG, Waterston RH. 1988. Analysis of a mutator activity necessary for germline transposition and excision of *Tc1* transposable elements in *Caenorhabditis elegans*. *Genetics*. 120(2):397-407.

Nelson MG, Linheiro RS, Bergman CM. 2017. Mcclintock: An integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. *G3 - Genes Genom. Genet*. 7(8):2763.

Nigon VM, Félix M-A. 2017. History of research on *C. elegans* and other free-living nematodes as model organisms. *WormBook : the online review of C. elegans biology*. 2017:1-84.

Payer LM, Burns KH. 2019. Transposable elements in human genetic disease. *Nat. Rev. Genet*. 20(12):760-772.

Préclin V, Martin E, Ségalat L. 2003. Target sequences of *Tc1*, *Tc3* and *Tc5* transposons of *Caenorhabditis elegans*. *Genet Res*. 82(2):85-88.

Robb SMC, Lu L, Valencia E, Burnette JM, III, Okumoto Y, Wessler SR, Stajich JE. 2013. The use of relocaTE and unassembled short reads to produce high-resolution snapshots of transposable element generated diversity in rice. *G3 - Genes Genom. Genet*. 3(6):949-957.

Rushforth AM, Anderson P. 1996. Splicing removes the *Caenorhabditis elegans* transposon *Tc1* from most mutant pre-mrnas. *Mol. Cell. Biol*. 16(1):422-429.

Strain: RW6999, Genotype: C. elegans wild isolate. - Caenorhabditis Genetics Center (CGC) - College of Biological Sciences. 2021. cgc.umn.edu. [accessed 2021 Apr 12]. https://cgc.umn.edu/strain/RW6999

Vendrell-Mir P, Barteri F, Merenciano M, González J, Casacuberta JM, Castanera R. 2019. A benchmark of transposon insertion detection tools using real data. *Mobile DNA*. 10(1):53.

Weick EM, Miska EA. 2014. piRNAs: From biogenesis to function. *Development*. 141(18):3458-3471.

Wormbase.org. [accessed 2021 Apr 12]. https://wormbase.org/tools/mine/simplemine.cgi

# 4. CONCLUSION

Rapid advances in genomics have deepened our understanding of molecular evolution by providing a clearer view of complex differences across genomes. The model organism *C. elegans* is particularly useful in studying the evolution of diverse molecular pathways in multicellular eukaryotes due to its wide range of phenotypes and rapid life cycle. In this work, three unique *C. elegans* strains known to have high activity of the *Tc1* transposable element were examined with multiple traditional *C. elegans* fitness assays and whole-genome sequencing to provide a more comprehensive understanding of the genomic and phenotypic effects of *Tc1* proliferation.

The origin of *Tc1* proliferation in the Bergerac strains remains a mystery. This work represents the first steps to solve this mystery using genomics-era techniques, which will eventually provide more information about TE regulation in *C. elegans* and multicellular eukaryotes in general. A variety of fitness traits in multiple Bergerac strains were fully examined, showing that, despite a likely common origin of *Tc1* proliferation and fitness decline, a range of phenotypes developed in different laboratory backgrounds. The genomic analysis shows these varying phenotypes are mostly explained by varying accumulation of *Tc1* in the genome, with one exception. A list of genes predicted to have exons disrupted by *Tc1* was generated, supporting previous hypotheses that increased TE transposition harms organisms by causing TEs to insert into genes.

Several experiments are planned to continue research on the Bergerac strains. First, a comprehensive scan of each genome will reveal potential genetic variants contributing to enhanced *Tc1* proliferation. An RNA-Seq experiment will follow upon the conclusion of the

genomic analysis. RNA-Seq is a technique for analyzing the entire complement of gene transcripts produced by an organism, thereby allowing an estimation of gene expression. While some gene sequences may not be mutated directly, mutations in different regulatory genes can affect gene expression, leading to the unique phenotypes observed. In addition to a deeper understanding of the Bergerac phenotypes, this work will reveal the relative abundance of smRNA transcripts, showing if the Bergerac RNAi pathways are functioning normally. Finally, long-term experimental evolution with the Bergerac strains is underway. The maintenance of *C. elegans* at large population size enables evolution under strong natural selection, which may lead to the fixation of new compensatory mutations that serve to suppress the expression of transposable elements, resulting in fitness recovery of these low fitness strains. This technique has been observed to cause mutant *C. elegans* lines to recover ancestral fitness levels. If an adaptive response occurs in any of the lines in this experiment, genomic sequencing could reveal the molecular mechanisms that evolve to overcome TE invasions.

# APPENDIX: GENES WITH *TC1* INSERTIONS IN EXONS

*Table A.1: Names, genetic map positions, and RNAi phenotypes for genes with exons disrupted by Tc1 insertions in the genome of CB4851, as predicted by the TE caller relocaTE.*

| Public Name | Genetic Map Position | RNAi Phenotype Observed |
|---|---|---|
| B0198.3 | X 6.998073 | organism development variant |
| B0310.1 | X -19.701050 | fat content reduced |
| B0554.7 | V -19.995560 | fat content reduced |
| kin-4 | IV 4.980146 | aging variant, bag of worms, embryonic lethal, protruding vulva, sterile progeny |
| tiam-1 | I -1.307499 | N.A. |
| linc-56 | X -12.878100 | N.A. |
| npr-4 | X 7.175480 | fewer egg laying events during active |
| inx-1 | X -1.949599 | N.A. |
| pals-6 | I 17.223000 | N.A. |
| C25F6.7 | X -5.309794 | N.A. |
| C25H3.11 | II -0.970698 | cytoplasmic processing body variant, slow growth, sterile progeny, transgene subcellular localization variant |
| linc-156 | I -0.420670 | N.A. |
| C31B8.1 | V -12.840720 | N.A. |
| gadr-2 | IV 4.279294 | N.A. |
| C33D3.5 | X 2.219612 | N.A. |
| oac-8 | IV 4.640526 | N.A. |
| C45G9.10 | III -2.204094 | organism development variant, slow growth |
| pqn-22 | IV 2.939407 | embryonic lethal |
| C49A9.3 | IV 3.105703 | N.A. |
| cka-2 | X -7.798743 | reduced brood size, transgene expression increased |
| ets-9 | X -7.800850 | transgene expression reduced |
| D1007.15 | I -1.052969 | N.A. |
| srd-59 | II 0.499292 | dauer lifespan extended |
| F02C12.1 | X 11.884010 | N.A. |
| F07C6.8 | IV 6.345075 | N.A. |
| adt-2 | X -1.478892 | annulae morphology variant, avoids bacterial lawn, body length variant, clear, cuticle morphology variant, dumpy, engulfment variant, gene expression level reduced, larval arrest, late larval arrest, locomotion variant, molt defect, reactive oxygen species homeostasis variant, shortened life span, slow growth, small, thin |
| oac-14 | X 1.821830 | cadmium hypersensitive, slow growth |
| F11C3.1 | I 27.967630 | growth variant |

| | | |
|---|---|---|
| mec-2 | X -4.705517 | N.A. |
| F16B12.1 | X 16.416130 | N.A. |
| pab-2 | X 12.688190 | embryonic lethal, reduced brood size, sterile |
| gcy-20 | V 9.885293 | N.A. |
| F22F7.7 | V -15.789820 | body wall muscle sarcomere morphology variant, mitochondria alignment variant, protein expression reduced |
| fkh-5 | III -2.425467 | organism development variant |
| rrn-3.1 | I 29.999500 | N.A. |
| F32B4.10 | I 9.047656 | N.A. |
| ssp-32 | IV 4.441184 | slow growth, sluggish |
| F32H2.8 | I 3.350936 | N.A. |
| jkk-1 | X -5.403745 | pathogen susceptibility increased, transgene expression reduced, transgene induced cosuppression variant |
| tat-5 | I 3.707366 | accumulated cell corpses, accumulated germline cell corpses, apoptosis variant, cell suppression contacts abnormal early emb, cell ebb contacts abnormal in four cell embryo, cell membrane morphology variant, cell membrane organization biogenesis variant, development phenotype, embryonic lethal, embryonic morphology defective early emb, gastrulation variant, germ cell compartment morphology variant, germ cell compartment size variant, germ cell partition morphology variant, germline nuclear positioning variant, gonad morphology variant, lethal, lysosome morphology variant, maternal sterile, membrane trafficking variant, morphology phenotype, oocyte morphology variant, oocyte septum formation variant, plasma membrane leaflet composition variant, protruding vulva, rachis wide, seam cell morphology variant, slow growth, spindle orientation defective early emb, sterile, sterile progeny, transgene subcellular localization variant |
| gst-19 | II 18.821430 | cadmium hypersensitive, slow growth |
| oac-22 | IV -1.953758 | N.A. |
| cyld-1 | III 1.156097 | genotoxic chemical induced apoptosis variant, transgene expression increased |
| F40H3.3 | II -0.413437 | N.A. |
| oac-26 | I 13.087920 | N.A. |
| cnc-8 | X 1.950310 | N.A. |
| F42E8.1 | V 4.859907 | N.A. |
| fbxb-5 | II -14.525630 | N.A. |
| F49B2.7 | I 24.369030 | N.A. |
| F53A10.2 | II -15.561830 | N.A. |
| F55G1.15 | IV 3.339690 | N.A. |
| F56H6.7 | I 13.112030 | N.A. |
| F58D5.8 | I 13.022130 | N.A. |
| F59B10.6 | II 3.052886 | N.A. |
| syd-2 | X 2.238295 | aldicarb resistant, mitochondria alignment variant, protein expression reduced |

| | | |
|---|---|---|
| emc-1 | II -15.434200 | development phenotype, levamisole resistant, locomotor coordination variant, pattern of transgene expression variant, receptor mediated endocytosis defective, slow growth, transgene expression increased |
| K02A4.11 | X 8.766142 | N.A. |
| K02A4.13 | X 8.766339 | N.A. |
| K02F6.4 | II -12.502580 | N.A. |
| K04H8.3 | I 24.099319 | N.A. |
| fbp-1 | I -8.302657 | N.A. |
| col-49 | I -4.476120 | transgene expression reduced |
| linc-151 | IV 5.301830 | N.A. |
| C07A4.3 | X 1.869302 | N.A. |
| C31E10.1 | X 15.918010 | N.A. |
| fbxb-8 | I 24.358980 | N.A. |
| R02E4.1 | X -8.464085 | N.A. |
| T16G12.1 | III 1.724937 | embryonic lethal, pronuclear migration defective early emb |
| T22D1.1 | IV 3.250104 | embryonic lethal |
| Y8A9A.2 | II -6.211108 | N.A. |
| str-158 | IV 1.389915 | fat content reduced |
| srh-302 | V 13.305090 | N.A. |
| hum-8 | IV -25.996889 | embryonic lethal, germ cell morphology variant, reduced brood size |
| ZK856.5 | V 2.315676 | N.A. |
| C28D4.11 | IV 4.397570 | N.A. |
| C45B11.7 | V 2.919090 | N.A. |
| F09F3.8 | V 5.701095 | N.A. |
| F35E2.11 | I 9.609425 | N.A. |
| F49A5.12 | V 11.241040 | N.A. |
| T12B5.9 | III -25.066370 | N.A. |
| srz-84 | IV 3.789728 | N.A. |
| Y71A12B.18 | I 22.473009 | N.A. |
| 21ur-59 | IV 11.499300 | N.A. |
| 21ur-10003 | IV 11.499300 | N.A. |
| R05H11.1 | III -0.944825 | fat content reduced |
| R11F4.1 | II -5.628728 | N.A. |
| cest-19 | X -1.910930 | N.A. |
| R173.5 | X -1.919596 | N.A. |
| R173.9 | X -1.919596 | N.A. |
| T07A9.12 | IV -26.055479 | N.A. |
| T10A3.5 | X -1.781509 | N.A. |
| T10H9.9 | V 0.132287 | N.A. |
| T12A7.2 | IV 5.278917 | N.A. |
| srg-38 | V 2.962244 | N.A. |
| T22F3.14 | V -8.581398 | N.A. |
| trak-1 | I 0.789029 | transgene expression reduced |

| | | |
|---|---|---|
| W01H2.9 | X -7.896780 | N.A. |
| ant-1.2 | I 1.300410 | N.A. |
| dct-15 | IV 4.421897 | N.A. |
| lgc-45 | II -6.241304 | N.A. |
| Y40C5A.4 | IV 3.331152 | N.A. |
| Y45F10C.4 | IV 10.549580 | N.A. |
| Y51H4A.25 | IV 15.293360 | N.A. |
| Y70D2A.4 | X 21.567440 | N.A. |
| Y73F8A.1168 | IV 14.126320 | N.A. |
| ZK250.13 | II -14.116210 | N.A. |
| ZK470.14 | X -8.109752 | N.A. |
| ZK484.11 | I 0.778769 | N.A. |
| ZK688.5 | III -0.510266 | N.A. |

*Table A.2: Names, genetic map positions, and RNAi phenotypes for genes with exons disrupted by Tc1 insertions in the genome of RW7000, as predicted by the TE caller RelocaTE.*

| Public Name | Genetic Map Position | RNAi Phenotype Observed |
|---|---|---|
| B0205.4 | I 5.076996 | N.A. |
| B0207.5 | I 0.481204 | N.A. |
| sra-33 | II -4.041706 | N.A. |
| B0410.3 | X -12.680680 | N.A. |
| unc-30 | IV 8.219272 | locomotion variant, shrinker, transgene expression increased |
| rpm-1 | V 1.620292 | apoptosis reduced, germ cell hypersensitive ionizing radiation, protein aggregation variant, transgene expression increased, transgene expression reduced, transgene subcellular localization variant |
| C01C4.3 | X -9.730689 | N.A. |
| srh-22 | V -3.702482 | N.A. |
| C02H6.3 | V 0.747965 | N.A. |
| srd-15 | V -0.776262 | N.A. |
| C04F1.1 | I 0.415726 | N.A. |
| C05D12.1 | II 3.465246 | N.A. |
| C09D4.2 | I 0.036484 | N.A. |
| str-143 | V 1.006824 | N.A. |
| str-182 | V 2.394336 | N.A. |
| C13C4.4 | V 3.660373 | N.A. |
| spin-1 | V 3.662700 | fat content reduced, peptide uptake by intestinal cell decreased |
| C15B12.4 | X -2.895033 | N.A. |
| gar-1 | X -2.893147 | frequency body bend reduced, locomotion variant |
| srz-25 | V 12.918070 | N.A. |

| | | |
|---|---|---|
| cab-1 | X 7.008985 | aldicarb resistant, avoids bacterial lawn, Bacillus thuringiensis toxin hypersensitive, embryonic lethal, larval arrest, pore forming toxin hypersensitive, reduced brood size, transgene expression reduced |
| C25D7.19 | V 7.303129 | N.A. |
| wdfy-3 | IV 6.130780 | N.A. |
| linc-156 | I -0.420670 | N.A. |
| tbc-7 | X -6.172492 | aging variant, aldicarb resistant, body wall muscle morphology variant, locomotion reduced, shortened life span |
| ugt-21 | IV 4.283442 | fat content increased |
| C33D12.11 | X -12.421870 | N.A. |
| sru-27 | V -18.813431 | N.A. |
| C39E9.7 | IV 7.806450 | N.A. |
| C43D7.8 | V 21.061069 | N.A. |
| 21ur-919 | IV 2.945240 | N.A. |
| fhod-1 | I -0.421363 | body wall cell development variant, body wall muscle development variant |
| deg-1 | X -1.279535 | N.A. |
| unc-130 | II 3.404041 | body wall muscle sarcomere morphology variant, embryonic lethal, organism development variant |
| C50F4.16 | V 2.046092 | N.A. |
| ptr-5 | X 24.087980 | body vacuole, intestinal vacuole, locomotion variant, molt defect, small |
| srh-25 | V 4.211771 | N.A. |
| D2023.6 | V 3.394153 | N.A. |
| srt-12 | V -6.237041 | N.A. |
| F01G10.5 | IV 4.562107 | N.A. |
| flr-1 | X 12.576590 | avoids bacterial lawn, clear, embryonic lethal, extended life span, gene expression level high, increased pathogen accumulation, pale, reduced brood size, slow growth, sterile, transgene expression increased |
| F07C6.8 | IV 6.345075 | N.A. |
| adt-2 | X -1.478892 | annulae morphology variant, avoids bacterial lawn, body length variant, clear, cuticle morphology variant, dumpy, engulfment variant, gene expression level reduced, larval arrest, late larval arrest, locomotion variant, molt defect, reactive oxygen species homeostasis variant, shortened life span, slow growth, small, thin |
| fbxb-111 | II -11.223560 | N.A. |
| F08G5.3 | IV 5.915981 | N.A. |
| clec-54 | V 6.420877 | N.A. |
| F09F9.3 | X -8.348946 | N.A. |
| F09F9.5 | X -8.359201 | N.A. |
| F10D7.5 | X 24.089161 | embryonic lethal, maternal sterile, sick |
| F11A5.4 | V 9.836763 | N.A. |
| F11C3.1 | I 27.967630 | growth variant |
| 21ur-6044 | IV 12.263880 | N.A. |
| F14D2.19 | II -6.671661 | N.A. |
| mec-2 | X -4.705517 | N.A. |

| | | |
|---|---|---|
| ubql-1 | I 1.653790 | frequency body bend reduced, protein expression increased, protein ubiquitination variant, shortened life span, transgene expression increased |
| lys-10 | IV 3.708310 | N.A. |
| F18A12.2 | II -6.309660 | N.A. |
| gpa-13 | V 4.619757 | N.A. |
| srx-98 | II -6.231755 | N.A. |
| F23H12.21 | V 4.007497 | N.A. |
| F25C8.6 | V 25.587049 | N.A. |
| acr-16 | V 1.462229 | dauer lifespan extended, locomotion variant, transgene expression increased |
| tli-1 | I 5.054717 | N.A. |
| slc-28.2 | V -11.545600 | N.A. |
| F28C1.3 | V 4.285339 | N.A. |
| rrn-3.1 | I 29.999500 | N.A. |
| F31E8.17 | II 0.115929 | N.A. |
| F31E8.20 | II 0.115929 | N.A. |
| srx-21 | V -19.966000 | N.A. |
| jkk-1 | X -5.403745 | pathogen susceptibility increased, transgene expression reduced, transgene induced cosuppression variant |
| gst-38 | V 9.133283 | cadmium hypersensitive, chemical hypersensitive, dauer lifespan extended, organism electrophilic stress hypersensitive, slow growth |
| srw-56 | V 9.296890 | N.A. |
| oac-23 | IV -1.997559 | dauer lifespan extended |
| F41C6.14 | X -2.092150 | N.A. |
| vet-6 | I 17.347500 | N.A. |
| F47F2.1 | X -9.193692 | body wall muscle myosin organization defective |
| F48A11.4 | II -15.785650 | N.A. |
| F49C5.12 | II 10.320890 | N.A. |
| F49C12.5 | IV 4.155943 | N.A. |
| F53B3.3 | X -12.686970 | N.A. |
| F53F4.22 | V 5.520665 | N.A. |
| F53F4.24 | V 5.520695 | N.A. |
| srbc-48 | V 8.964798 | N.A. |
| attf-3 | III -27.063520 | cortical dynamics defective early emb, embryonic lethal, exploded through vulva, larval arrest, locomotion variant, long, pattern of transgene expression variant, receptor mediated endocytosis defective, slow growth, sterile, sterile progeny |
| F54C4.9 | III -27.055780 | N.A. |
| srt-34 | V -13.029860 | N.A. |
| F54E7.6 | III -1.462378 | N.A. |
| F55B11.4 | IV 11.928930 | fat content reduced |
| F56D5.9 | IV 4.239560 | fat content increased |
| oac-34 | I 7.950776 | N.A. |
| nhr-192 | V 2.244982 | N.A. |
| F57B1.5 | V 4.983706 | N.A. |

| | | |
|---|---|---|
| F58D5.8 | I 13.022130 | N.A. |
| str-87 | V 1.900382 | N.A. |
| emc-1 | II -15.434200 | development phenotype, levamisole resistant, locomotor coordination variant, pattern of transgene expression variant, receptor mediated endocytosis defective, slow growth, transgene expression increased |
| H24K24.2 | V -19.933371 | N.A. |
| nhr-97 | IV 4.400859 | linker cell migration variant |
| cdr-7 | V 4.139004 | N.A. |
| K02E10.5 | X -15.099810 | fat content increased, transgene expression increased |
| K02E10.10 | X -14.874850 | N.A. |
| K02F6.4 | II -12.502580 | N.A. |
| K04H8.3 | I 24.099319 | N.A. |
| vit-6 | IV 3.700448 | cell homeostasis metabolism variant, lipid metabolism variant, nicotine hypersensitive, organism pathogen response variant, pathogen susceptibility increased, transgene subcellular localization variant |
| nac-3 | III 1.946985 | life span phenotype |
| K10G4.5 | V 12.853470 | N.A. |
| srw-111 | I 13.058530 | shortened life span |
| M57.4 | IV -3.247869 | N.A. |
| C01G10.14 | V 7.397334 | N.A. |
| mdf-1 | V 2.046342 | anaphase bridging, cell cycle variant, dauer lifespan extended, embryonic arrest, embryonic lethal, endomitotic oocytes, germline proliferation variant, gonad development variant, high incidence male progeny, larval arrest, locomotion variant, masculinization of germline, organism development variant, organism starvation response variant, protruding vulva, reduced brood size, sterile F1, transgene subcellular localization variant, tumorous germline |
| srt-9 | V -12.791650 | N.A. |
| D1007.18 | I -1.042692 | body wall muscle myosin organization defective, cytoplasmic processing body variant, protein expression reduced, sterile, transgene subcellular localization variant |
| F07G11.2 | V 0.649514 | N.A. |
| srw-57 | V 9.296940 | N.A. |
| F49B2.4 | I 24.385719 | N.A. |
| cutl-1 | V 5.255365 | N.A. |
| abt-3 | IV 6.706862 | N.A. |
| emb-9 | III 0.421546 | adult lethal, cell membrane organization biogenesis variant, cytoplasmic processing body variant, developmental delay postembryonic, distal tip cell migration variant, embryonic lethal, gonad morphology variant, larval arrest, locomotion variant, maternal sterile, mRNA surveillance defective, nonsense mRNA accumulation, oocyte morphology variant, oocyte septum formation variant, pattern of transgene expression variant, pharyngeal morphology variant, receptor mediated endocytosis defective, slow growth, sterile, transgene expression increased, transgene subcellular localization variant |
| str-224 | V -19.858130 | N.A. |

| | | |
|---|---|---|
| srw-136 | V -12.796800 | N.A. |
| R02C2.1 | V -20.007820 | aldicarb resistant |
| str-178 | V 4.614000 | N.A. |
| R193.1 | X -18.483761 | N.A. |
| col-147 | V 2.058280 | embryonic lethal |
| ocr-3 | X 22.261299 | N.A. |
| sav-1 | X -15.884500 | N.A. |
| gem-4 | IV 5.272073 | gonad development variant |
| T12A7.6 | IV 5.274561 | N.A. |
| T15B7.14 | V 0.313354 | N.A. |
| T22D1.1 | IV 3.250104 | embryonic lethal |
| T28C6.7 | IV 3.980457 | N.A. |
| W05H9.2 | X -3.219023 | N.A. |
| twk-33 | V 13.057120 | N.A. |
| W07G4.2 | V 4.823104 | N.A. |
| gst-34 | II 24.420321 | N.A. |
| gei-18 | IV 10.988270 | locomotion variant |
| Y43B11AR.1 | IV 3.288180 | N.A. |
| Y50D4B.4 | V -19.853300 | N.A. |
| Y53C10A.10 | I 13.021760 | N.A. |
| srt-24 | IV -23.591591 | N.A. |
| srh-302 | V 13.305090 | N.A. |
| srg-67 | V -1.881114 | N.A. |
| ZC506.1 | X 1.731564 | N.A. |
| ZK287.4 | V 2.067648 | N.A. |
| ZK856.5 | V 2.315676 | N.A. |
| C01G10.19 | V 7.396645 | N.A. |
| F22E5.19 | II -12.212680 | N.A. |
| F29C4.3 | IV -26.980511 | N.A. |
| F33E2.8 | I 13.642840 | N.A. |
| F36A4.9 | IV 0.002292 | N.A. |
| srw-131 | V -12.797300 | N.A. |
| srh-107 | V 7.912514 | N.A. |
| srz-84 | IV 3.789728 | N.A. |
| Y105C5B.30 | IV 14.545170 | N.A. |
| R05H11.1 | III -0.944825 | fat content reduced |
| R07B1.5 | X 1.730264 | N.A. |
| R09E10.13 | IV 4.572156 | N.A. |
| R11F4.1 | II -5.628728 | N.A. |
| npr-24 | X 24.147699 | N.A. |
| cest-19 | X -1.910930 | N.A. |
| R173.5 | X -1.919596 | N.A. |
| R173.9 | X -1.919596 | N.A. |

| | | |
|---|---|---|
| T01B11.1 | IV 3.806196 | N.A. |
| npr-25 | V 3.075950 | amplitude of sinusoidal movement decreased, frequency body bend reduced, locomotion variant, sluggish |
| srj-49 | V -12.936250 | N.A. |
| T04G9.7 | X -19.501150 | embryonic lethal |
| T05A6.8 | II 0.568487 | N.A. |
| nhr-102 | V 9.050699 | N.A. |
| sri-7 | V 8.037158 | N.A. |
| fbxa-104 | V 8.078202 | N.A. |
| T07F10.3 | V 4.719372 | N.A. |
| ztf-4 | I 1.539785 | cell proliferation increased, excess intestinal cells |
| T10G3.4 | V 5.349897 | N.A. |
| srt-55 | III 1.829700 | N.A. |
| gck-1 | V 1.870474 | accumulated germline cell corpses, antibody staining variant, apoptosis increased, apoptosis variant, cell membrane organization biogenesis variant, chromosome condensation variant, chromosome segregation variant, cleavage furrow termination defective early emb, cortical dynamics defective early emb, diakinesis progression during oogenesis variant, diplotene region organization variant, fewer germ cells, germ cell compartment expansion variant, germ cell compartment large, germ cell compartment morphology variant, germ cell compartment multinucleate, germ cell compartment nuclei number variant, germ cell compartment size variant, germ cell compartment small, germ cell morphology variant, gonad vesiculated, maternal sterile, meiosis variant, meiotic progression prophase variant, mitosis variant, nuclear appearance variant, nuclei small, oocyte accumulation, oocyte morphology variant, oocytes small, pachytene region organization variant, proximal germ cell proliferation variant, rachis morphology variant, reduced brood size, reproductive system morphology variant, sterile, sterile F1, transgenerational loss of fertility |
| nhr-219 | V 1.882494 | N.A. |
| cyp-29A2 | V 2.972350 | egg laying defective, fat associated body size decreased, fat content reduced, lethal |
| srh-252 | V 12.834750 | N.A. |
| T19D12.15 | II -0.045708 | N.A. |
| T21C9.6 | V 2.603103 | transgene induced cosuppression variant |
| T22F3.14 | V -8.581398 | N.A. |
| str-19 | V 8.706384 | N.A. |
| T24H7.8 | II -0.395747 | larval arrest, larval lethal, lethal, protruding vulva, sick, small, sterile |
| T25B6.4 | X 0.534987 | N.A. |
| bbs-8 | V 0.146771 | dauer lifespan extended |
| fbxb-115 | V 20.713751 | N.A. |
| T27B7.9 | V -14.892180 | N.A. |
| hlh-30 | IV -1.013807 | autophagy variant, bacterially unswollen, developmental delay, fat content reduced, mRNA levels increased, mRNA levels reduced, reduced brood size, sterile, transgene expression reduced |

| | | |
|---|---|---|
| srw-77 | V 10.794480 | N.A. |
| W07G4.7 | V 4.821764 | N.A. |
| lgc-45 | II -6.241304 | N.A. |
| Y26E6A.2 | X 15.147980 | N.A. |
| Y37A1B.17 | IV 10.949970 | aldicarb resistant, locomotion variant |
| nhr-235 | II 10.645140 | N.A. |
| cyd-1 | II 13.290650 | distal tip cell migration variant, extended life span, gonad arm morphology variant, gonad development variant, larval arrest, larval lethal, late larval lethal, locomotion variant, male gonad development variant, male somatic gonad development variant, pattern of transgene expression variant, receptor mediated endocytosis defective, somatic gonad development variant, sterile |
| Y45G12C.1 | V -14.679450 | N.A. |
| srw-99 | V -17.781820 | extended life span |
| Y50D4C.14 | V -19.930321 | N.A. |
| Y50D7A.8 | III -26.895100 | N.A. |
| gst-32 | II 24.384230 | dauer lifespan extended |
| Y55F3C.9 | IV -23.602791 | N.A. |
| srh-134 | V 13.449110 | N.A. |
| Y73A3A.1 | I -18.222040 | N.A. |
| Y73B6BL.278 | IV 3.194489 | N.A. |
| Y73F8A.1168 | IV 14.126320 | N.A. |
| Y105C5B.19 | IV 14.516190 | N.A. |
| Y105E8A.55 | I 25.801189 | N.A. |
| ZC13.2 | X -19.460239 | N.A. |
| mam-1 | X -19.458900 | larval lethal, molt defect |
| ZC449.4 | X -6.196723 | N.A. |
| ZC449.5 | X -6.196686 | N.A. |
| srbc-44 | V 4.704627 | N.A. |
| nhr-253 | V -19.999001 | locomotion variant |
| adm-4 | X -1.206455 | cell fate specification variant, maternal sterile, multiple anchor cells, organism development variant, sick |
| ZK180.13 | IV 0.451227 | N.A. |
| ZK180.19 | IV 0.451227 | N.A. |
| ZK250.13 | II -14.116210 | N.A. |
| ZK381.37 | IV 3.275232 | N.A. |
| ZK484.11 | I 0.778769 | N.A. |
| ZK822.5 | IV 5.434181 | body wall muscle myosin organization defective |
| wrt-7 | V 7.823690 | alae secretion variant, body vacuole, intestinal vacuole, locomotion variant, multiple alae, multivulva, small, transgene expression increased |
| ZK1225.5 | I 17.276890 | N.A. |
| | | |
| twk-10 | V 0.074709 | dauer lifespan extended |
| twk-24 | V 3.965970 | N.A. |

*Table A.3: Names, genetic map positions, and RNAi phenotypes for genes with exons disrupted by Tc1 insertions in the genome of RW6999, as predicted by the TE caller relocaTE.*

| Public Name | Genetic Map Position | RNAi Phenotype Observed |
|---|---|---|
| B0205.4 | I 5.076996 | N.A. |
| B0410.3 | X -12.680680 | N.A. |
| unc-30 | IV 8.219272 | locomotion variant, shrinker, transgene expression increased |
| srh-22 | V -3.702482 | N.A. |
| C02H6.3 | V 0.747965 | N.A. |
| C04C3.6 | IV -4.417310 | N.A. |
| C04F1.1 | I 0.415726 | N.A. |
| C09D4.2 | I 0.036484 | N.A. |
| str-143 | V 1.006824 | N.A. |
| str-182 | V 2.394336 | N.A. |
| gar-1 | X -2.893147 | frequency body bend reduced, locomotion variant |
| cft-1 | V -1.582813 | N.A. |
| srz-25 | V 12.918070 | N.A. |
| srx-24 | V -13.190050 | N.A. |
| C25D7.19 | V 7.303129 | N.A. |
| linc-156 | I -0.420670 | N.A. |
| tbc-7 | X -6.172492 | aging variant, aldicarb resistant, body wall muscle morphology variant, locomotion reduced, shortened life span |
| ugt-21 | IV 4.283442 | fat content increased |
| C33D12.11 | X -12.421870 | N.A. |
| C39E9.7 | IV 7.806450 | N.A. |
| C43D7.8 | V 21.061069 | N.A. |
| deg-1 | X -1.279535 | N.A. |
| unc-130 | II 3.404041 | body wall muscle sarcomere morphology variant, embryonic lethal, organism development variant |
| cka-2 | X -7.798743 | reduced brood size, transgene expression increased |
| ptr-5 | X 24.087980 | body vacuole, intestinal vacuole, locomotion variant, molt defect, small |
| srh-25 | V 4.211771 | N.A. |
| F07C6.8 | IV 6.345075 | N.A. |
| adt-2 | X -1.478892 | annulae morphology variant, avoids bacterial lawn, body length variant, clear, cuticle morphology variant, dumpy, engulfment variant, gene expression level reduced, larval arrest, late larval arrest, locomotion variant, molt defect, reactive oxygen species homeostasis variant, shortened life span, slow growth, small, thin |
| F08G5.3 | IV 5.915981 | N.A. |
| clec-54 | V 6.420877 | N.A. |
| F11A5.4 | V 9.836763 | N.A. |
| F11C3.1 | I 27.967630 | growth variant |
| F14D2.19 | II -6.671661 | N.A. |

| | | |
|---|---|---|
| ubql-1 | I 1.653790 | frequency body bend reduced, protein expression increased, protein ubiquitination variant, shortened life span, transgene expression increased |
| lys-10 | IV 3.708310 | N.A. |
| gpa-13 | V 4.619757 | N.A. |
| srx-98 | II -6.231755 | N.A. |
| F21D9.4 | V 20.816690 | N.A. |
| slc-28.2 | V -11.545600 | N.A. |
| F28C1.3 | V 4.285339 | N.A. |
| rrn-3.1 | I 29.999500 | N.A. |
| srx-21 | V -19.966000 | N.A. |
| jkk-1 | X -5.403745 | pathogen susceptibility increased, transgene expression reduced, transgene induced cosuppression variant |
| gst-38 | V 9.133283 | cadmium hypersensitive, chemical hypersensitive, dauer lifespan extended, organism electrophilic stress hypersensitive, slow growth |
| srw-56 | V 9.296890 | N.A. |
| oac-23 | IV -1.997559 | dauer lifespan extended |
| vet-6 | I 17.347500 | N.A. |
| F49C12.5 | IV 4.155943 | N.A. |
| F53B3.3 | X -12.686970 | N.A. |
| srt-34 | V -13.029860 | N.A. |
| F54E7.6 | III -1.462378 | N.A. |
| F55B11.4 | IV 11.928930 | fat content reduced |
| F56D5.9 | IV 4.239560 | fat content increased |
| nhr-192 | V 2.244982 | N.A. |
| F57B1.5 | V 4.983706 | N.A. |
| F58D5.8 | I 13.022130 | N.A. |
| str-87 | V 1.900382 | N.A. |
| emc-1 | II -15.434200 | development phenotype, levamisole resistant, locomotor coordination variant, pattern of transgene expression variant, receptor mediated endocytosis defective, slow growth, transgene expression increased |
| H24K24.2 | V -19.933371 | N.A. |
| nhr-97 | IV 4.400859 | linker cell migration variant |
| cdr-7 | V 4.139004 | N.A. |
| K02E10.10 | X -14.874850 | N.A. |
| K02F6.4 | II -12.502580 | N.A. |
| K04H8.3 | I 24.099319 | N.A. |
| K10G4.5 | V 12.853470 | N.A. |
| srw-111 | I 13.058530 | shortened life span |
| M57.4 | IV -3.247869 | N.A. |
| C01G10.14 | V 7.397334 | N.A. |
| srx-25 | V -13.196710 | N.A. |
| C33D9.8 | IV 3.961655 | N.A. |

| | | |
|---|---|---|
| D1007.18 | I -1.042692 | body wall muscle myosin organization defective, cytoplasmic processing body variant, protein expression reduced, sterile, transgene subcellular localization variant |
| F07G11.2 | V 0.649514 | N.A. |
| srw-57 | V 9.296940 | N.A. |
| F36H5.8 | II -14.500710 | N.A. |
| F49B2.4 | I 24.385719 | N.A. |
| cutl-1 | V 5.255365 | N.A. |
| abt-3 | IV 6.706862 | N.A. |
| emb-9 | III 0.421546 | adult lethal, cell membrane organization biogenesis variant, cytoplasmic processing body variant, developmental delay postembryonic, distal tip cell migration variant, embryonic lethal, gonad morphology variant, larval arrest, locomotion variant, maternal sterile, mRNA surveillance defective, nonsense mRNA accumulation, oocyte morphology variant, oocyte septum formation variant, pattern of transgene expression variant, pharyngeal morphology variant, receptor mediated endocytosis defective, slow growth, sterile, transgene expression increased, transgene subcellular localization variant |
| str-224 | V -19.858130 | N.A. |
| R02C2.1 | V -20.007820 | aldicarb resistant |
| str-178 | V 4.614000 | N.A. |
| R193.1 | X -18.483761 | N.A. |
| col-147 | V 2.058280 | embryonic lethal |
| ocr-3 | X 22.261299 | N.A. |
| sav-1 | X -15.884500 | N.A. |
| srab-19 | V 2.201741 | N.A. |
| gem-4 | IV 5.272073 | gonad development variant |
| T12A7.6 | IV 5.274561 | N.A. |
| T15B7.14 | V 0.313354 | N.A. |
| T22D1.1 | IV 3.250104 | embryonic lethal |
| T28C6.7 | IV 3.980457 | N.A. |
| gst-34 | II 24.420321 | N.A. |
| gei-18 | IV 10.988270 | locomotion variant |
| Y50D4B.4 | V -19.853300 | N.A. |
| Y53C10A.10 | I 13.021760 | N.A. |
| srt-24 | IV -23.591591 | N.A. |
| srh-302 | V 13.305090 | N.A. |
| srg-67 | V -1.881114 | N.A. |
| ZC506.1 | X 1.731564 | N.A. |
| ZK287.4 | V 2.067648 | N.A. |
| ZK856.5 | V 2.315676 | N.A. |
| C01G10.19 | V 7.396645 | N.A. |
| F22E5.19 | II -12.212680 | N.A. |
| F29C4.3 | IV -26.980511 | N.A. |
| F36A4.9 | IV 0.002292 | N.A. |

| | | |
|---|---|---|
| F36H5.15 | II -14.499190 | N.A. |
| srh-107 | V 7.912514 | N.A. |
| srz-84 | IV 3.789728 | N.A. |
| R05H11.1 | III -0.944825 | fat content reduced |
| R07B1.5 | X 1.730264 | N.A. |
| R11F4.1 | II -5.628728 | N.A. |
| cest-19 | X -1.910930 | N.A. |
| R173.5 | X -1.919596 | N.A. |
| R173.9 | X -1.919596 | N.A. |
| npr-25 | V 3.075950 | amplitude of sinusoidal movement decreased, frequency body bend reduced, locomotion variant, sluggish |
| T04G9.7 | X -19.501150 | embryonic lethal |
| T05H4.4 | V -0.166720 | cell secretion variant, lysosome-related organelle morphology variant, pattern of transgene expression variant, RAB-11 recycling endosome localization variant, RAB-11 recycling endosome morphology variant, slow growth, transgene subcellular localization variant |
| nhr-102 | V 9.050699 | N.A. |
| sri-7 | V 8.037158 | N.A. |
| fbxa-104 | V 8.078202 | N.A. |
| ztf-4 | I 1.539785 | cell proliferation increased, excess intestinal cells |
| gck-1 | V 1.870474 | accumulated germline cell corpses, antibody staining variant, apoptosis increased, apoptosis variant, cell membrane organization biogenesis variant, chromosome condensation variant, chromosome segregation variant, cleavage furrow termination defective early emb, cortical dynamics defective early emb, diakinesis progression during oogenesis variant, diplotene region organization variant, fewer germ cells, germ cell compartment expansion variant, germ cell compartment large, germ cell compartment morphology variant, germ cell compartment multinucleate, germ cell compartment nuclei number variant, germ cell compartment size variant, germ cell compartment small, germ cell morphology variant, gonad vesiculated, maternal sterile, meiosis variant, meiotic progression prophase variant, mitosis variant, nuclear appearance variant, nuclei small, oocyte accumulation, oocyte morphology variant, oocytes small, pachytene region organization variant, proximal germ cell proliferation variant, rachis morphology variant, reduced brood size, reproductive system morphology variant, sterile, sterile F1, transgenerational loss of fertility |
| cyp-29A2 | V 2.972350 | egg laying defective, fat associated body size decreased, fat content reduced, lethal |
| T21C9.6 | V 2.603103 | transgene induced cosuppression variant |
| T22F3.14 | V -8.581398 | N.A. |
| str-19 | V 8.706384 | N.A. |
| rocf-1 | V 8.242580 | N.A. |
| T24H7.8 | II -0.395747 | larval arrest, larval lethal, lethal, protruding vulva, sick, small, sterile |
| T25B6.4 | X 0.534987 | N.A. |
| bbs-8 | V 0.146771 | dauer lifespan extended |

| | | |
|---|---|---|
| fbxb-115 | V 20.713751 | N.A. |
| hlh-30 | IV -1.013807 | autophagy variant, bacterially unswollen, developmental delay, fat content reduced, mRNA levels increased, mRNA levels reduced, reduced brood size, sterile, transgene expression reduced |
| srw-77 | V 10.794480 | N.A. |
| lgc-45 | II -6.241304 | N.A. |
| Y22D7AR.7 | III -19.584240 | N.A. |
| Y37A1B.17 | IV 10.949970 | aldicarb resistant, locomotion variant |
| nhr-235 | II 10.645140 | N.A. |
| cyd-1 | II 13.290650 | distal tip cell migration variant, extended life span, gonad arm morphology variant, gonad development variant, larval arrest, larval lethal, late larval lethal, locomotion variant, male gonad development variant, male somatic gonad development variant, pattern of transgene expression variant, receptor mediated endocytosis defective, somatic gonad development variant, sterile |
| Y43C5A.2 | IV 4.564934 | N.A. |
| Y45G12C.1 | V -14.679450 | N.A. |
| Y50D4C.14 | V -19.930321 | N.A. |
| gst-32 | II 24.384230 | dauer lifespan extended |
| Y55F3C.9 | IV -23.602791 | N.A. |
| srh-134 | V 13.449110 | N.A. |
| Y73B6BL.278 | IV 3.194489 | N.A. |
| Y73F8A.1168 | IV 14.126320 | N.A. |
| Y105C5B.19 | IV 14.516190 | N.A. |
| Y105E8A.55 | I 25.801189 | N.A. |
| ZC13.2 | X -19.460239 | N.A. |
| mam-1 | X -19.458900 | larval lethal, molt defect |
| ZC449.4 | X -6.196723 | N.A. |
| nhr-253 | V -19.999001 | locomotion variant |
| adm-4 | X -1.206455 | cell fate specification variant, maternal sterile, multiple anchor cells, organism development variant, sick |
| ZK180.13 | IV 0.451227 | N.A. |
| ZK180.19 | IV 0.451227 | N.A. |
| ZK250.13 | II -14.116210 | N.A. |
| ZK381.37 | IV 3.275232 | N.A. |
| ZK484.11 | I 0.778769 | N.A. |
| ZK1225.5 | I 17.276890 | N.A. |
| | | |
| twk-10 | V 0.074709 | dauer lifespan extended |
| twk-24 | V 3.965970 | N.A. |
| twk-13 | V 3.999862 | linker cell migration variant |
| twk-25 | IV 4.990770 | N.A. |