UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

The genetic basis of lifespan in a host- pathogen model system

José Miguel Fernandes Bento

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Doutor Ivo Manuel Mimoso Vieira Chelo, FCUL

2023

Sumário

Um conceito que conecta intrinsecamente o envelhecimento e as interações hospedeiro-patogênico é a necessidade de adaptação do hospedeiro em resposta a influencias ambientais, como a presença de patógenos. Por exemplo, à medida que os indivíduos vão envelhecendo, o seu sistema imunitário sofre alterações que podem afetar a sua capacidade de defesa contra agentes patogénicos invasores. Outros fatores fisiológicos que afetam o progresso do envelhecimento serão, por exemplo, genéticos ou até mesmo estilos de vida e hábitos que os indivíduos tenham. Identificar a base genética da adaptação é fundamental para compreender e prever a evolução. Anteriormente, adaptámos populações de *C. elegans* para retardar a reprodução na presença de uma estirpe patogênica de *E. coli* (IAI1), que demonstrou antecipar a postura de ovos no verme. Isto resultou num ambiente muito adverso ao qual *C. elegans* teve que se adaptar através de mecanismos de desenvolvimento e/ou resposta imunológica. *Caenorhabditis elegans*, muitas vezes referido como *C. elegans*, é um verme microscópico que se tornou um organismo modelo bastante proeminente na pesquisa biológica. O seu tamanho pequeno, corpo transparente e base genética bem definida tornaram-no ideal para realizar vários estudos relacionados com a biologia do desenvolvimento, genética e neurobiologia. *C. elegans* desempenhou um papel imperativo na descoberta de processos biológicos fundamentais, e a sua simplicidade e facilidade de manipulação continuam a fazer dele um ótimo organismo para pesquisa, contribuindo para a nossa compreensão de fenómenos biológicos complexos.

Este estudo teve como objetivo encontrar genes de *C. elegans* responsáveis pela variação do tempo de vida na presença de uma estirpe patogênica de *E. coli* IAI1, e compará-los com um conjunto complementar de genes encontrados na presença de uma estirpe não-patogênica de *E. coli*. Uma comparação de ambos os conjuntos de genes permitir-nos-á inferir até que ponto a imunidade impacta as "vias de envelhecimento" neste sistema, bem como oferecer uma visão geral inicial das atuais trade-offs genéticas na evolução do envelhecimento na presença de patogênicos. Para isso, é fornecido um painel de mapeamento previamente desenvolvido, do qual serão escolhidas 18 linhas e, após exposição a ambas as bactérias, o tamanho do corpo e expectativa de vida serão medidos. As medidas do primeiro fenótipo obtiveram-se através da medição do comprimento do corpo de indivíduos em imagens obtidos com uma câmara instalada num estereoscópio, enquanto que as medidas do segundo fenótipo obtiveram-se através de ensaios de sobrevivência. Após a obtenção destes dados, primeiro foi realizada uma análise para analisar as variações da expectativa de vida e do tamanho do corpo entre os dois tipos de estirpes bacterianas. A esta seguiu-se uma análise que avaliou o impacto que a estirpe bacteriana e o tamanho do corpo têm sobre a expectativa de vida, assim como se existe alguma correlação e de que tipo entre os dois fenótipos, na bactéria patogênica e na não-patogênica. Finalmente, realizou-se o

mapeamento de associação utilizando o software PLINK, para descobrir os genes e as interações genéticas de interesse para a possível associação entre as variações destes fenótipos e a presença de uma bactéria patogênica. Este trabalho apresenta um novo protocolo que permite a transferência de nematodes adultos de forma muito rápida e menos trabalhosa. Neste protocolo, temos uma transferência dos indivíduos por meio de filtros que conseguem separar os adultos dos ovos e larvas, assim como um enorme número de indivíduos a serem transferidos por cada vez que se aplica. Isto permite reduzir o tempo gasto na transferência e, assim, aumentar o número de populações experimentais.

Após a realização da componente experimental, pode-se concluir a eficiência notável oferecida por este novo protocolo de transferência baseada na utilização de filtros, facilitando a seleção e transferência rápida de indivíduos adultos. No entanto, é crucial reconhecer que esta abordagem inovadora apresenta a sua quota-parte de desafios. Um desses desafios envolve o facto de que apesar da grande eficácia para separar adultos de larvas e ovos, ainda é possível encontrar uma mínima quantidade dessas larvas no filtrado final. No entanto, este método de transferência baseado em filtros melhorou significativamente a eficiência e a precisão das transferências de vermes, o que nos permitiu usar populações experimentais com maiores números, assim como mais condições experimentais por cada ensaio de sobrevivência. Os nossos primeiros resultados demonstraram uma diferença significativa na expectativa de vida entre OP50 e IAI1, sendo maior a expectativa de vida quando na presença da não-patogênica como se esperava. Quanto ao tamanho do corpo os resultados foram mais inesperados, apesar de não-significativos, pois notou-se que esse tamanho era maior na presença da bactéria patogénica. Além disso, os nossos resultados também apresentaram: que tanto o tipo de estirpe bacteriana de que os *C. elegans* se alimentam, como o tamanho destes nematodes, vão influenciar a expectativa de vida, mas igualmente que existe uma correlação negativa entre os fenótipos da expectativa de vida e tamanho corporal. As nossas descobertas ressaltam a complexidade das interações hospedeiro-patógeno, como elas impactam as principais características da história de vida em *C. elegans*, e como esses fenótipos se correlacionam. Também são já a primeira indicação de uma possível relação com o desenvolvimento. Por fim, a nossa análise de associação revelou uma diversidade de regiões genéticas associadas às variações da expectativa de vida e do tamanho corporal. Na verdade, apenas se conseguiu obter resultados significativos após se proceder a análise utilizando os fenótipos compostos da expectativa de vida e do tamanho do corpo, criados a partir da divisão dos valores desses fenótipos nas diferentes bactérias (os da IAI1 a dividir pelos da OP50, para ambos os fenótipos). Estabeleceu-se um intervalo para cada marcador significativo que o estudo de associação encontrou e anotou-se todos os genes que estavam nas proximidades do marcador, depois procedendo-se á investigação dos mesmos quanto ás suas funções e fenótipos aos quais estão ligados. Estas regiões

abrangiam uma ampla gama de genes, incluindo vários genes que apesar de não terem ligação direta aos fenótipos da expectativa de vida e do tamanho corporal, estavam associados a outros genes e vias genéticas que possuem essas ligações. Muitos destes genes associados estão envolvidos na modulação da expectativa de vida e na tolerância ao stresse, além de contribuírem para a regulação do desenvolvimento e do crescimento. Notavelmente, a sua presença tanto na expectativa de vida quanto no tamanho corporal, sugere a possibilidade de um atraso no desenvolvimento ou crescimento como uma hipótese para corroborar os resultados mencionados anteriormente relativos á correlação negativa entre os dois fenótipos, e por sua vez esses atrasos podem estar associados à presença de um patógeno.

Em conclusão, o nosso estudo em *C. elegans* fornece uma nova compreensão precisa e crítica sobre a arquitetura genética que rege ás variações na expectativa de vida e no tamanho do corpo quando exposto a uma estirpe de bactéria patogênica. Não só confirmou que existe uma base genética associada a estas variações e á presença de um patógeno, mas os genes encontrados indicam um cenário em que essa associação provém de alterações em características fisiológicas, como o desenvolvimento e o crescimento. Além disso, as regiões e vias genéticas identificadas oferecem caminhos promissores para futuras pesquisas. Exemplos podem incluir o teste para tentar avaliar o verdadeiro impacto destas regiões genómicas, ou genes de interesse, assim como avaliar a sua associação com a característica que está a ser investigada.

Palavras-chave: Envelhecimento; Interação hospedeiro-patógeno; Expectativa de vida; Tamanho do corpo; Análise de associação.

Abstract

A concept that intricately connects aging and host-pathogen interactions is the host's need for adaptation in response to dynamic environmental challenges, such as the presence of pathogens. Identifying the genetic basis of adaptation is critical to understanding and predicting evolution. We previously adapted *C. elegans* populations to delayed reproduction in the presence of a pathogenic *E. coli* strain (IAI1), shown to anticipate egg laying in the worm. This has resulted in a very adversarial environment to which *C. elegans* had to adapt via developmental and immunological response mechanisms. This study aimed to find *C. elegans* genes responsible for lifespan variation in the presence of the pathogenic *E. coli* IAI1 and compare them to a complementary set of genes found in the presence of a non-pathogenic *E. coli* strain. A previously developed mapping panel is provided, from which 18 inbred lines will be chosen and their body length and lifespan measured. First, an analysis for differences and correlations between these two phenotypes was performed, followed by association mapping to discover genes and gene interactions of interest. Initial analysis revealed a significant difference in lifespan between OP50 and IAI1 and a negative correlation between the lifespan and the body size phenotypes. Furthermore, our association analysis unveiled a diversity of genetic regions associated with lifespan and body size variations. These regions encompassed a broad array of genes involved in stress tolerance, development and growth regulation. In conclusion, our study in *C. elegans* provides critical insights into the genetic architecture governing lifespan and body size variations when exposed to a pathogenic bacteria strain. A comparison of both gene sets will allow us to infer the amount to which immunity impacts "aging pathways" in this system and offer an initial overview of current genetic trade-offs in the evolution of aging in the presence of pathogens.

Keywords: Aging; Host-pathogen interaction; Lifespan; Body Size; Association analysis.

Table of contents

Table of contents

Figure Index

I. Introduction

1. Aging and species interactions

Aging is the deterioration of an organism's structure and function throughout its lifetime; it is also defined as a complex biological process influenced by various factors, including genetics, lifestyle, and environmental conditions (Antell & Taczanowski, 1999; Rodríguez-Rodero et al., 2011). Genetics, for instance, can determine an individual's susceptibility to age-related diseases and the overall rate of aging (Rodríguez-Rodero et al., 2011). Lifestyle factors, such as diet, exercise, and stress management, also impact aging by affecting the body's health and leading to the early rise of age-related diseases (Jura & Kozak, 2016). And so, the process of aging, which involves a gradual decline in physiological functions and an increased susceptibility to infections, can also be influenced by other physiological factors like an organism's size (Collins et al., 2008). Longevity, the measure of how long an organism can live, is closely tied to aging, with some species exhibiting remarkable lifespans due to evolutionary adaptations.

Aging, longevity, and body size variation are intricately linked concepts that often intersect with the complex networks of host-microorganism interactions. These complex networks are created from the intimate functional interactions between microorganisms and their hosts (Radeke & Herman, 2021; R. Zhang & Hou, 2013).

Almost all organisms engage in host-microbe symbiosis, and these symbiotic bacteria may be advantageous to the host, detrimental to the host, or have no impact at all. For instance, we have the example of *Escherichia coli*, with benign strains that are a natural component of the gut flora and can benefit their hosts by preventing pathogenic bacteria from invading the intestine, for example. But other strains, such as *E. coli* O26, on the other hand, can infect their hosts and cause sickness (R. Zhang & Hou, 2013).

Additionally, host-microorganism interactions play a crucial role in an organism's health and longevity. The composition of an individual's microbiome, the community of microorganisms residing in and on the body, can significantly impact aging and overall well-being. Studies have shown that a balanced and diverse microbiome can support a longer and healthier life, whereas disruptions in this microbial equilibrium can contribute to age-related diseases and other health issues (Badal et al., 2020; Ghosh et al., 2022). Understanding the intricate interplay between aging and host-microorganism

interactions is a crucial area of research that holds promise for enhancing our understanding of longevity and the aging process and promoting healthier aging in diverse species, including humans.

2. Host-pathogen interactions

A pathogen, such as a virus, bacteria, fungus, or viroid, and its host can interact in various ways, typically referred to as a host-pathogen interaction. In order to live and infect a host, pathogens adapt and discover new strategies for these tasks, after which, if the host's immune system is unable to combat these infectious invaders, illnesses will begin to arise in the host's body (Sen et al., 2016). However, not all host-pathogen interactions result in disease and those that do typically have a complex development that results in this condition. Because of that, these interactions can be examined on various levels, with three typical examples: population, organismal, and molecular (Marsh & May, 2012; Sen et al., 2016).

Host-pathogen interactions can often lead to detrimental consequences for the host organism. When a pathogen, such as a virus, bacterium, or parasite, invades the host's body, it seeks to exploit the host's resources for its own survival and reproduction. This can trigger a cascade of immune responses, inflammation, and tissue damage as the host's defense mechanisms attempt to combat the invader. In many cases, these defensive measures can cause harm to the host's own cells and tissues, leading to symptoms of illness and potentially long-term health consequences. Additionally, the strategies employed by pathogens to evade the host's organism can further exacerbate the adverse effects on the host by exploiting the host's defense mechanisms (Alberts et al., 2002; Charles A Janeway et al., 2001). Therefore, these interactions end up having various negative effects that profoundly impact the host's overall health and well-being. In addition to the immediate impacts, host-pathogen interactions can adversely affect the host's growth and development mechanisms (Kong & Yang, 2023; Mordecai, 2011), potentially affecting the host's body size. One common consequence is the diversion of the host's resources and energy, leading to an allocation of these vital elements for the pathogen; this hinders the host's ability to grow, develop, and thrive, potentially resulting in stunted growth and delayed development (Cressler et al., 2014). Pathogens may also engage in nutrient competition (Schaible & Kaufmann, 2005), scavenging essential compounds required for the host's metabolic processes and thereby hindering growth.

The efficient treatment of illnesses brought on by harmful microorganisms and the promotion of animal health by benign bacteria can be accomplished by understanding these relationships, existing

several methods for their investigation (R. Zhang & Hou, 2013). These molecular techniques are supported by animal studies, which look at the infection from the perspective of the entire organism. Thus, finding species to conduct extensive, ethical research on these connections has become more important than ever (Marsh & May, 2012). A remarkable advancement in this field was learning of many model organisms that have been used to understand the mechanisms behind these interactions, including *Arabidopsis thaliana*, *Drosophila melanogaster*, *Caenorhabditis elegans*, zebrafish, and mice, as they are simple, genetically tractable, and happen to be susceptible to several human pathogens (Marsh & May, 2012; R. Zhang & Hou, 2013). *C. elegans* stands out among them for having a variety of distinctive traits that make it advantageous as a model organism (R. Zhang & Hou, 2013).

3. *C. elegans* as a model organism

The nematode worm *Caenorhabditis elegans* has become a very used model organism, being employed for practically all facets of biology (Meneely et al., 2019). For example, in the experimental study at metabolic and genomic levels in vivo, when used to recapitulate the majority of human illnesses (S. Zhang et al., 2020). The fact that is simple to work with in the lab, has minimal nutritional and growth requirements, produces a significant number of offspring by self-fertilization within a short period, and has been extensively studied, all make *C. elegans* a very useful model organism (Meneely et al., 2019). This eukaryotic organism also possesses a genetic profile that has been fully sequenced (S. Zhang et al., 2020), being the first sequenced genome from a multicellular organism (Meneely et al., 2019), and it is a recognized genetic model organism for the investigation of aging.

These studies on aging in *C. elegans* have aided in uncovering molecular signals, transcriptional regulators, and epigenetic alterations linked to longevity, expanding our understanding of an organism's aging process (S. Zhang et al., 2020).

A renowned molecular scientist named Sydney Brenner made the decision in 1963 to shift his study focus from bacteria and phage to an animal (Goldstein, 2016; Meneely et al., 2019). Brenner and other scientists were supporters of genetic analysis, which stands for the use of biologically disruptive mutations as a tool to comprehend regularly occurring processes, and had utilized it for various purposes. Therefore, he carefully studied and tested different species in his lab before selecting to work with *C. elegans* as a model organism for research of development and behavior (Marsh & May, 2012; Meneely et al., 2019).

The nematode *C. elegans* is a non-lethal, global, free-living (non-parasitic) organism that consumes microorganisms (Meneely et al., 2019; Radeke & Herman, 2021; S. Zhang et al., 2020), and inhabits both compost piles and the soil (Huang & Kammenga, 2020; Marsh & May, 2012; Radeke & Herman, 2021). With very simple and affordable maintenance, these typically self-fertilizing

hermaphrodites (XX), and rather small proportion of occurring males (X0) (Marsh & May, 2012), possess a typical lifespan that lasts 18 to 20 days when maintained at 20°C. Adult individuals are about 1 mm in length (Meneely et al., 2019; S. Zhang et al., 2020).

Mitotic divisions start after fertilization when an eggshell develops around the egg. An egg will be deposited roughly after fertilization, and around 12 hours after that egg-laying, the worm will emerge with the embryogenesis finished (Meneely et al., 2019). *C. elegans* progress through the L1–L4 larval stages immediately after hatching, in which they develop and replace their cuticles (Meneely et al., 2019; S. Zhang et al., 2020). The L4 larvae undergo their last molt to become adult hermaphrodites, which then begin reproducing on their own 12 hours later (Meneely et al., 2019). But there is a second pathway, where L2 worms molt into a unique stage known as the dauer, which possess a more resistant cuticle. This developmentally arrested stage helps organisms endure stress conditions, and when the circumstances are better, the dauer larvae bypass the L3 stage and molt to create the L4 stage, resuming normal development (Meneely et al., 2019; S. Zhang et al., 2020).

C. elegans has long served as a valuable model organism for studying the genetic basis of lifespan and body size. Several key genes and gene pathways have been identified in *C. elegans* that play crucial roles in regulating these traits. The insulin/insulin-like growth factor (IGF-1) signaling pathway, conserved across species, is a central player in controlling both lifespan and body size. Mutations in genes such as daf-2, which encodes the insulin/IGF-1 receptor, can extend lifespan and reduce body size (Iser & Wolkow, 2007; Venz et al., 2021). Additionally, the daf-16 gene, which acts as the target of the daf-2 pathway, plays a pivotal role in promoting longevity while also influencing body size (Lin et al., 2001; Uno et al., 2021). Other pathways, such as the TOR (target of rapamycin) pathway and the sirtuin pathway, have also been implicated in regulating these traits (Uno & Nishida, 2016; Vellai et al., 2003). Through the intricate interplay of these genes and pathways, *C. elegans* provides valuable insights into the genetic mechanisms governing lifespan and body size, offering potential implications for understanding aging and growth regulation in higher organisms, including humans.

4. E. coli and the effects of the strain IAI1

Theodor Escherich described the rod-shaped, gram-negative bacterium known as *Escherichia coli* in 1885, initially isolated from infant stool (Braz et al., 2020; Pakbin et al., 2021). It is both an adaptable and facultative anaerobic bacterium from the Enterobacteriaceae family of bacteria (enteric bacteria) (Jnani & Ray, 2022; Pakbin et al., 2021; Robbens et al., 2014) that is simple to detect and

susceptible to random and spontaneous genetic mutation (Braz et al., 2020). This family is one of the most significant groups of bacteria in medicine (Robbens et al., 2014). The numerous *E. coli* genomes that have been sequenced show a wide range within the same bacterial species, demonstrated by the various sizes and genetic diversity across commensal and pathogenic (Braz et al., 2020), besides being the most prevalent facultative anaerobic bacteria of the human intestinal flora (Jaureguy et al., 2008), beginning to colonize the infants' gastrointestinal tracts a few hours after birth (Pakbin et al., 2021).

The non-pathogenic members of this family typically function as commensal inhabitants of the human gastrointestinal tract, as well as the gut of other animals, coexisting without any negative impacts and with a number of reciprocal advantages (Braz et al., 2020; Pakbin et al., 2021).

Escherichia coli is a versatile bacterium that plays a pivotal role in scientific research and various biotechnological applications. While some strains of *E. coli* are pathogenic and can cause illnesses in humans and animals, many strains, such as the OP50 strain, are non-pathogenic and have been extensively studied for their beneficial properties (de Sousa Figueiredo et al., 2021). OP50 is a laboratory-adapted strain of *E. coli* often used as a food source for the nematode *Caenorhabditis elegans*, a model organism in biology research (Arata et al., 2020). Moreover, non-pathogenic *E. coli* strains like OP50 are frequently employed in environmental studies to monitor the quality and safety of drinking and recreational water (Merkx-Jacques et al., 2013). *E. coli*, just like the non-pathogenic OP50 strain, showcases the remarkable diversity within the species and underscores its significance in both basic and applied scientific research. Various genera of the Enterobacteriaceae family are considered human intestinal pathogens, including Salmonella, Shigella, and Yersinia. Within this group of pathogenic strains, *E. coli* O157:H7 can cause a range of health issues when ingested. These bacteria produce toxins that lead to very severe symptoms (Jnani & Ray, 2022).

The known pathogenic *E. coli* strain IAI1, belonging to the phylogenetic group B1 (Touchon et al., 2009), is predominantly associated with food and animals, yet its presence in humans is relatively scarce (Aguirre-Sánchez et al., 2022). Together with phylogenetic group A, it contains the vast majority of commensal and intraintestinal pathogenic *E. coli* strains (Diard et al., 2007).

Additionally, while other reports and studies have presented measurements of body size of several *C. elegans* strains and other nematodes while feeding on the *E. coli* strain OP50 (Flemming et al., 2000; Woodruff et al., 2018); surprisingly, there is a dearth of information concerning body size measurements on IAI1 strain and differences in body length between the two bacteria. This intriguing knowledge gap underscores the need for further investigation into the specific interactions and effects of different bacterial strains on the body size of *C. elegans*, shedding light on the nuanced intricacies of this model organism's biology.

5. Main Objectives and Working Strategy

This project has two main objectives: firstly, is to pinpoint the *C. elegans* genes that cause lifespan variation in the presence of *E. coli* IAI1, and secondly, to compare those genes to a similar set of genes acquired when a non-pathogenic *E. coli*, the control, was present. With these goals in mind, the body length and lifespan of approximately 100 inbred lines are proposed to be measured, using a previously developed mapping panel. The next step will involve association mapping, which will be used to find important genes and important gene interactions.

Through the comparison of both gene sets, it will be possible to infer the degree to which immunity affects "aging pathways" in this system. It will also present a first overview of existing genetic trade-offs in the evolution of aging in the presence of pathogens.

II. Material and Methods

1. *Caenorhabditis elegans* and Bacterial Strains

In this research, we used a non-pathogenic *Escherichia coli* strain, *E. coli* OP50 (Brenner, 1974), as well as *E. coli* IAI1, which is known to have pathogenic effects in *C. elegans* (Diard et al., 2007; Picard et al., 1999). The *E. coli* IAI1 strain, which will be referred to in this work as the IAI1 strain, was generously donated by Ivan Matic. In contrast, the *E. coli* OP50, referred to in this work as the OP50 strain, was acquired from the Caenorhabditis Genetics Center (CGC).

Recombinant inbred lines (RILs) derived from a polymorphic lab-adapted population; the A6140 population, were employed in testing possible variations of survival and body size in the presence of the bacterial strains mentioned above. Noble et al. (2017) first described the A6140 population, presenting 507 genome-sequenced RILs, with Noble et al. (2021) adding to the number of already sequenced RILs. These lines were created through 13–16 generations of inbreeding by selfing of single individuals, from a population derived by 140–190 generations of experimental evolution using the A0 population. This hybrid population was itself descended through parallel intercrossing among 16 wild isolates, or founders.

In this work, it was possible to analyze 18 of these RILs, which were provided by the laboratory of Henrique Teotónio (ENS, Paris, France).

2. Strain maintenance

Using isolated colonies from a streak plate of LB agar, *E. coli* OP50 was grown in LB liquid medium through overnight incubation at 37°C. From the grown culture, 0.1 ml was used to seed 92 mm plates of solid Nematode Growth Medium-lite (Chelo, 2014; Stiernagle, 2006), previously prepared with 28 ml of NGM-lite agar. After using glass beads to spread the bacteria, plates were incubated overnight at 37°C to enable the development of a bacterial lawn (James et al., 2018). Seeded plates were kept at 4°C after incubation until used for maintenance or survival assays, which occurred within three weeks.

In accordance with established procedures, *C. elegans* hermaphrodites were maintained on the NGM-lite agar plates seeded with *E. coli* OP50, at 20 °C (Chelo, 2014; Stiernagle, 2006). Synchronization of populations was achieved through the "bleaching protocol". In that process, at day four of the life cycle, eggs and adult worms were washed off the plates with M9 buffer into 15 ml tubes

and then treated with a bleaching solution. The fundamental principle behind this technique is that, while embryos are protected from bleach by the eggshell, worms are vulnerable to this chemical, killing adults and larvae but permitting the isolation of unhatched embryos. In the following steps, isolated eggs were pelleted, thoroughly washed using M9 buffer, and left in 4 ml of M9 buffer under constant shaking with 2.5 mg/ml of tetracycline. First-stage larvae (L1) eclosed overnight, after which the development will be halted because of a lack of food, synchronizing this new population of L1 individuals. Finally, the worms in the L1 stage were seeded into NGM-lite plates, carrying a confluent lawn of *E. coli* OP50, becoming day one of the life cycle (Porta-de-la-Riva et al., 2012). Each plate was seeded with approximately 1000 larvae and then placed at 20°C and 80% (RH) for 72 hours until the next bleach.

3. Body Size Measurement

Body length was measured for each inbred line, in the presence of the two bacterial strains, 48 hours after L1 seed. Stock populations frozen at −80 ºC were thawed and maintained for two generations in the conditions described above and, in the third generation, L1 seed was performed on NGM-lite plates with each bacterial strain and the larvae were then left to mature for 48 hours. Worms were washed off the plates with M9 buffer into 15 ml tubes, and the number of live *C. elegans* was estimated for the setup of each survival assay, with individuals from two plates being used to create each estimate. The inbred lines from the A6140 population were grouped and tested in separate experimental blocks, with each block including the N2 strain feeding on OP50 strain and IAI1 strain, as a common reference and two or three inbred lines on both bacteria.

A CCD camera (DFK 23UX174), connected to a dissecting stereomicroscope (Nikon SMZ 18), was used to capture images of individuals present in droplets of 5 µl prepared on glass slides, all with the same settings, and the images were saved as tiff files in RGB format.

4. Survival assays

The impact of the two bacteria on the survival of each inbred line was monitored. After two generations, L1 larvae were seeded on 4 NGM-lite plates (1000 individuals/plate) with a lawn of one of the two bacteria (2 OP50 and 2 IAI1) and left to incubate, at 20 °C, until day 3 of the assay (48 hours later). On day 3, L4 hermaphrodites were washed, their numbers estimated, and seeded onto two new NGM-lite plates (250 individuals/plate), one with OP50 strain and the other with IAI1 strain. Each experimental block assessed the N2 line, which served as a common reference between them, and three new inbred lines. Also, each block was composed of eight plates, with each plate containing one of the two bacterial strains and individuals from one of the inbred lines, either N2 or one of the three. During

the reproductive period (approximately 6 days/10 days counting from the day of the setup), worms were transferred daily to fresh plates to separate offspring from the individuals whose survival was being measure. The assay continued until all the individuals were found dead or considered to be missing, with the sets of plates incubating at 20 °C throughout the entire assay. In certain cases, when there was evidence of plate contamination, the assay was terminated and all remaining individuals were deemed missing.

The transfer of *C. elegans* was accomplished through the use of a newly devised approach developed in the host lab, with its main advantages being the transfer of a large number of individuals and the separation of adults from eggs, L1 and L2 larvae. Upon an initial count of the dead worms, M9 buffer was used to wash the plates, after which it was poured and filtered through a 40 μm filter (a pore diameter previously tested in Laranjeira (2016)) fitted on a 50 ml Falcon. The suspension passes slowly through the net, retaining the adults, and then to eliminate any lingering juvenile stages, the content of the filter is washed with M9. Adults were collected through centrifugation (1800 rpm; 1 min.; 20 ºC) and a glass pipet and then transferred to a fresh Petri plate.

When counting, worms that did not move in reaction to physical stress induced by gentle poking, and did not exhibit any pharyngeal pumping activity, were classified as dead, as well as those that displayed an egg-laying deficiency (bagging) or perished due to vulva rupture. On the other hand, individuals that vanished most likely because they crawled off the plate and dried, or disintegrated after dying or even perished during transfer, were classified as censored data. Counting of the dead before the passage through the filters and the counting of the living after were done using a stereoscope with a (20x-40x magnification). These daily observations were used to calculate the probability of survival.

5. Data analysis

R (*R: The R Project for Statistical Computing*, n.d.) was used to carry out the statistical analysis. To maintain track of the individuals that were alive and dead for each plate, raw data from the survival assay was recorded in MS Excel 2019.

Image processing and body length measurement were done using the ImageJ ver. 1.53t software. The tiff pictures were converted to 8-bit greyscale images and scaled using a previously determined calibration line (60× magnification). Finally, measurement of the worms' body length was done by manually tracing a line through the mid-body section. Body size mean estimates were obtained from at least 50 individuals. The measurements from the different lines were divided into blocks and taken at separate time points.

Analysis of the survival data was performed with the Cox regression (proportional hazards

analysis; (Cox, 1972)), used to assess variations in survivorship. The N2 line in OP50 strain was used to calculate the reference risk. The *Surv* and *coxph* functions from the R *survival* package (Therneau, 2023) were both used to create the following model: Surv(Time,Event)~pop*bact, where "Time" is the moment at which an individual was discovered dead or missing ("Event"), the second being considered censored data, “bact” refers to the two bacterial strains used, and “pop” represents the recombinant inbred lines that were investigated. Survival data for day 4 refers to dead or missing individuals between 72 and 96 hours after the L1 seed, between 96 and 120 hours for day 5, between 120 and 144 hours for day 6, and so on. The Kaplan-Meier estimation (Kaplan & Meier, 1958) was performed in the data to estimate the mean life span and create survival curves with the aid of R packages, with an example of these curves represented in Supplementary Figure 1. The proportion of surviving individuals plotted versus time was used to generate these survival curves. Using the results of N2 acquired from each block, the mean lifetime estimates based on the Kaplan-Meier model were corrected, with these values being applied later in the genetic analysis via the PLINK software.

5.1 Comparison analysis of phenotypes

Further analyses consisted of studying possible correlations between the body size and lifespan phenotypes on both OP50 and IAI1 strains.

Using the *lm*() function in the R software, a multiple linear regression was used to test if body size and bacterial strain significantly predicted variance in lifespan. More importantly, an ANOVA (analysis of variance) was carried out to investigate possible statistically significant differences between both phenotypes and between both bacteria, in which the threshold for statistical significance was $p < 0.05$.Correlation analysis between body size and lifespan for OP50 strain and IAI1 strain, and between OP50 strain and IAI1 strain for body size and lifespan, were performed by employing the *cor.test*() function in the R software environment, utilizing Pearson's product moment correlation coefficient. Finally, a two-sample t-test was used to compare the body size phenotype in OP50 strain and IAI1 strain and then repeated for lifespan.

5.2 PLINK and Association analysis

PLINK, a free, open-source whole-genome association analysis toolbox, was created to address the computational and analytical problems that result from whole-genome association studies (WGAS) (Purcell et al., 2007). This is done by carrying out several fundamental, large-scale analyses quickly and effectively of primarily genotype/phenotype data, not supporting any processes that come before this analysis (*PLINK: Whole Genome Data Analysis Toolset*, n.d.). Data administration,

summary statistics, population stratification, association analysis, and identity-by-descent estimates are just five of the primary functional domains of PLINK (Purcell et al., 2007).

The initial input for the PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/; (Purcell et al., 2007)) to do the quantitative trait association analysis was composed of three parts of data. The first corresponded to the quantitative trait phenotype data, which included the corrected estimates for body size and mean lifespan in both OP50 and IAI1 strains. The second part was the genotype data, which originated from ("CeMEEv2_RIL_geno.csv"), consisting of the recombinant inbred lines genotype matrix for segregating diallelic single nucleotide variants (SNVs) coded [0,0.5,1] relative to the N2 reference genome. The third part contained information regarding the name and location of each SNP, including chromosome, physical position, reference and alternate alleles, and genetic position, all of which were derived from ("CeMEEv2_RIL_snps_ws245.csv").

To limit the impact of significant linkage disequilibrium (LD), an LD ($\mathbf{r^2}$) threshold of 0.1 for the SNP pairings was established in a sliding window of 50 SNPs that shifts 10 SNPs ahead with each pruning before starting the analysis. Thus, a pruned subset of 8,573 SNPs, in approximate linkage equilibrium with one another, was singled out using PLINK. The data, now containing only the information in regards to the pruned SNPs, was passed through a series of command lines of the PLINK software to perform the quantitative trait association analysis, followed by an adjustment for multiple testing. The final outputs encompassed a table with one row per single SNP association result, which included the p-values, and a pre-sorted list of the single SNP association results containing a range of adjusted p-values.

This section begins with loading the table with the non-adjusted association results, and the table containing the adjusted values in the case of the composite phenotypes, followed by the selection of the information of interest and the construction of new tables with them. The new non-adjusted table was runned through a series of R commands, and afterward, the new adjusted table and another series of R commands were employed to highlight the significant markers (FDR_BH<0.05). These significant SNPs were investigated in the tool JBrowse of Wormbase (*WormBase : Nematode Information Resource*, n.d.). From the *RAINBOWR* package (Hamazaki & Iwata, 2020) in the R software, the *manhattan*() function was employed in generating Manhattan plots utilizing the table with the non-adjusted values.

This mapping process was repeated for both the body size and mean lifespan phenotypes, on OP50 strain and IAI1 strain, as well as for the lifespan and body size composite phenotypes. The composite phenotypes were generated from the division of the IAI1 strain values for the OP50 strain values for both mean lifespan and body length. Their creation came from two reasons: primarily so that any significant results discovered stem from the specific response to the pathogenic bacteria rather than

being unrelated to this factor, and secondly, so that any slight difference that may be present in one or both bacteria will become more evident in the composite phenotype. This approach can pinpoint the unique regions of the host's reaction to the pathogen.

III. Results

1. Body Size and Lifespan Variation

The two bacteria strains each had a seemingly distinctive effect on the *C. elegans* A6140 population's body length and survival (Figure 3.1). Supplementary Table 1 contains a list of the RILs utilized, as well as information on mean lifespan and mean body size.

Through the use of the ImageJ ver. 1.53t software, the body sizes of 18 different *C. elegans* RILs were measured on both types of bacteria, OP50 strain and IAI1 strain. Though there was a slight difference in body size between OP50 strain (M = 944.15, SD = 64.13) and IAI1 strain (M = 987.03, SD = 66.78), with the estimated mean value for IAI1 strain being higher (Figure 3.1A), that difference was not statistically significant (t (34) = -1.910, p = 0.065). In this study, L61 has the highest body size on OP50 food (1040,5 μm), while L16 was the highest on IAI1 food (1125,4 μm), as seen in Supplementary Table 1. L16 was also the second largest on OP50 food (1034.1 μm), whilst L27 was the second largest on IAI1 food (1096.6 μm). Inversely, the smaller body sizes belonged to L62 (804.0 μm) and L5 (881.7 μm) on, respectively, OP50 strain and IAI1 strain.

In order to investigate differences in survival, *C. elegans* from 18 different RILs were grown on two *E. coli* bacterial strains, a control non-pathogenic (OP50 strain) and a pathogenic (IAI1 strain), at 20ºC. According to Figure 3.1B, *C. elegans* cultured on IAI1 strain exhibited a lower average mean lifespan when compared to individuals fed on the non-pathogenic control OP50 strain, demonstrating that the presence of a different bacteria had a statistically significant effect (t (29) = 6.257, p = 7.545e-07) on worm lifespan between OP50 strain (M = 12.84; SD = 1.16) and IAI1 strain (M = 10.74, SD = 0.76). One of the RILs was the only one to deviate from the pathogenic bacteria's general tendency to shorten lifespan: for the L7 genotype, lifespan in the presence of *E. coli* IAI1 was slightly higher than in the presence of *E. coli* OP50 (Supplementary Table 1).

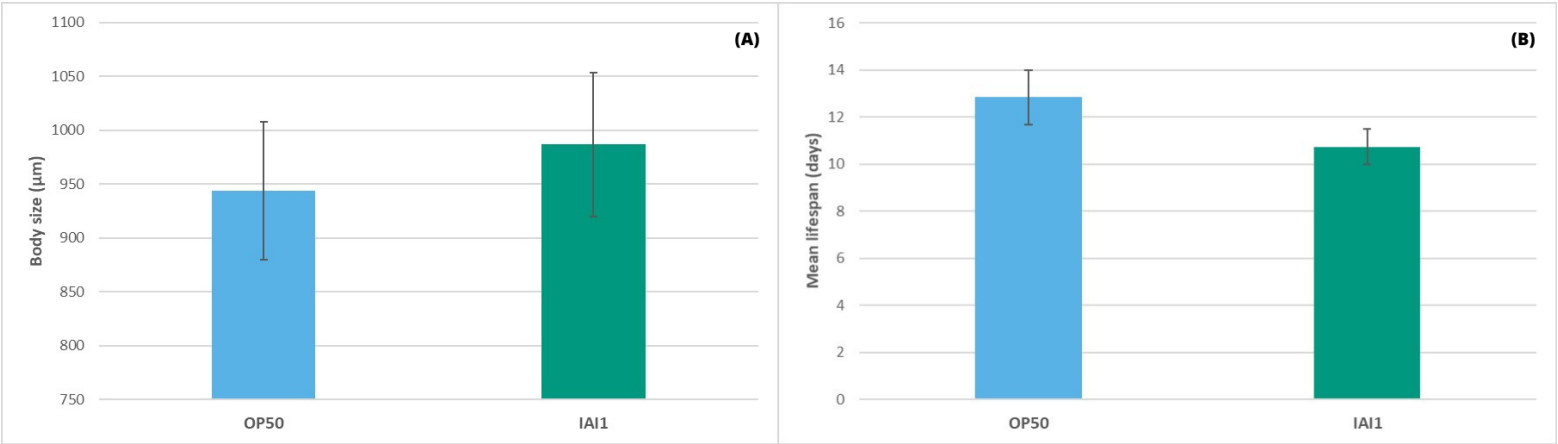


Figure 3.1. Average Mean Lifespan and Standard Deviation for *C. elegans*, on Two Bacteria. This figure illustrates the comparison of (**A**) Body size or (**B**) Average mean lifespan and standard deviation values between *C. elegans* feeding on two distinct bacteria strains, OP50 strain and IAI1 strain. The blue and green bars represent the two different bacteria, Bacterium OP50 strain and IAI1 strain, respectively. This visual comparison allows for insights into the variability in the average lifespan and body size between worms fed on different bacterial diets.

2. Correlation of phenotypes

Data on the body size and mean lifespan of worms, seeded on OP50 strain or IAI1 strain, was examined for correlations using a variety of statistical tests.

The influence of body size and bacterial strain on mean lifespan was studied using a two-way ANOVA, which began by considering the possibility of an interaction effect. The results indicated that there wasn't a statistically significant interaction between the effects of body size and bacteria ($F(1, 32) = 1.6161$, $p = 0.212796$). The analysis was repeated, now only considering an additive effect, and the results revealed that both body size ($F(1) = 11.281$, $p = 0.001988$) and bacteria ($F(1) = 31.248$, $p = 3.232e\text{-}06$) had a statistically significant influence on mean lifespan.

Figure 3.2 depicts the trend lines for mean lifespan versus body size for OP50 strain and IAI1 strain, where it is clear that life expectancy is negatively correlated to body size in both lines, meaning that there is a linear drop in mean lifespan with the increase of body size. The slopes are not significantly different between the different strains (-0.0037 days/µm), but the values are always higher in the OP50 strain compared to the IAI1 strain, as expected given the known pathogenic effects of IAI1 strain on *C. elegans*.

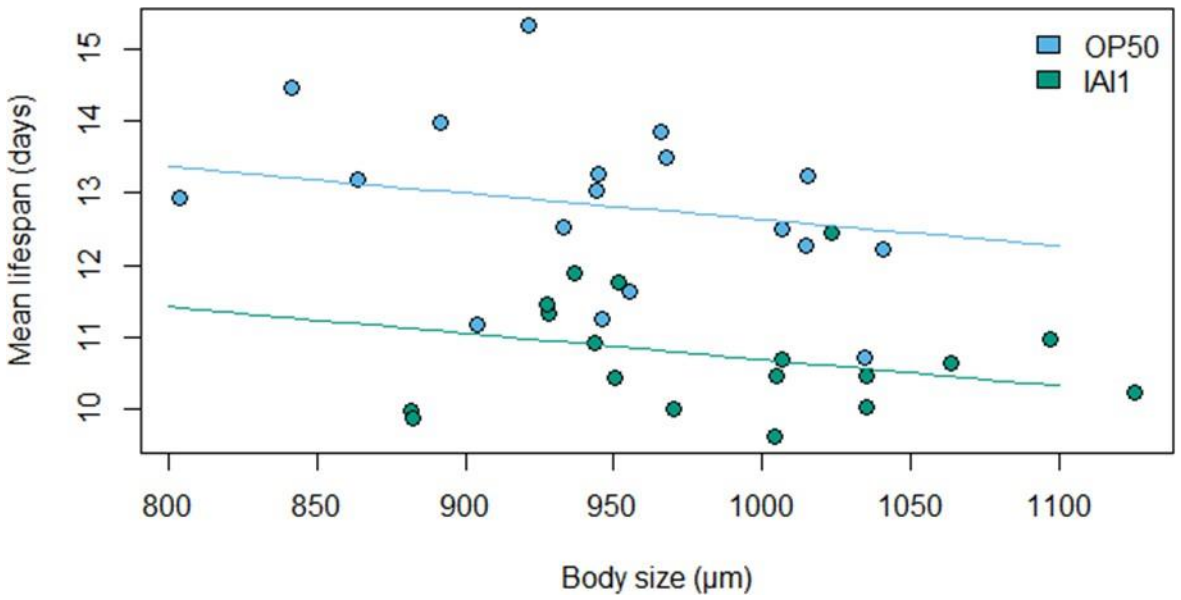


Figure 3.2. Correlation Between Mean Lifespan and Body Size, on Two Bacteria, in *C. elegans*. This figure presents a negative correlation between mean lifespan and body size for two different bacterial species. As body size increases (represented on the x-axis), the average mean lifespan (on the y-axis) decreases. Blue points and line represent the *E. coli* OP50 strain (slope −0.003666 days/µm, intercept 16.299710). Green points and line represent the *E. coli* IAI1 strain (slope −0.003666 days/µm, intercept 14.355186). It highlights this inverse relationship, shedding light on how variations in body size can influence the lifespan of *C. elegans* in both these bacterial strains.

3. Association Mapping

Manhattan plots for the six phenotypes were created to provide a visual representation of the genetic association results. Figures 3.3A and 3.3B represent Manhattan plots for the body size and lifespan composite phenotypes, respectively. Supplementary Figures 2A and 2B represent Manhattan plots for the body size phenotype on OP50 strain and IAI1 strain, whilst Supplementary Figures 3A and 3B represent Manhattan plots for the lifespan phenotype on OP50 strain and IAI1 strain. Each plot's x axis illustrates the genomic positions organized by chromosomes, while the p-values in their negative logarithmic form were plotted on the y axis. A horizontal threshold line was added, with significant markers (adjusted for multiple testing corrections) displayed above this line. Significant markers were only observed on the plots for the composite phenotypes.

Tables 1 and 2 show information concerning the regions where the SNPs that achieved the genome-wide significant level associated with body size and lifespan traits, respectively, are located. A total of 17 SNPs on five chromosomes were shown as significantly associated with lifespan and body size features across the genome. There were nine SNPs that were related to the body size trait and eight related to lifespan, but none of them were shared between the two phenotypes. As for chromosomes, chromosome 3 had the highest number of SNPs (6 SNPs) related to body size, followed by chromosome 2 (2 SNPs) and then chromosome 4 (1 SNP). In contrast, chromosome 5 had the most SNPs (3 SNPs)

that were significantly associated with lifespan, succeeded by chromosomes 1 and 4 with the same amount (2 SNPs), and finally, chromosome 3 (1 SNP).

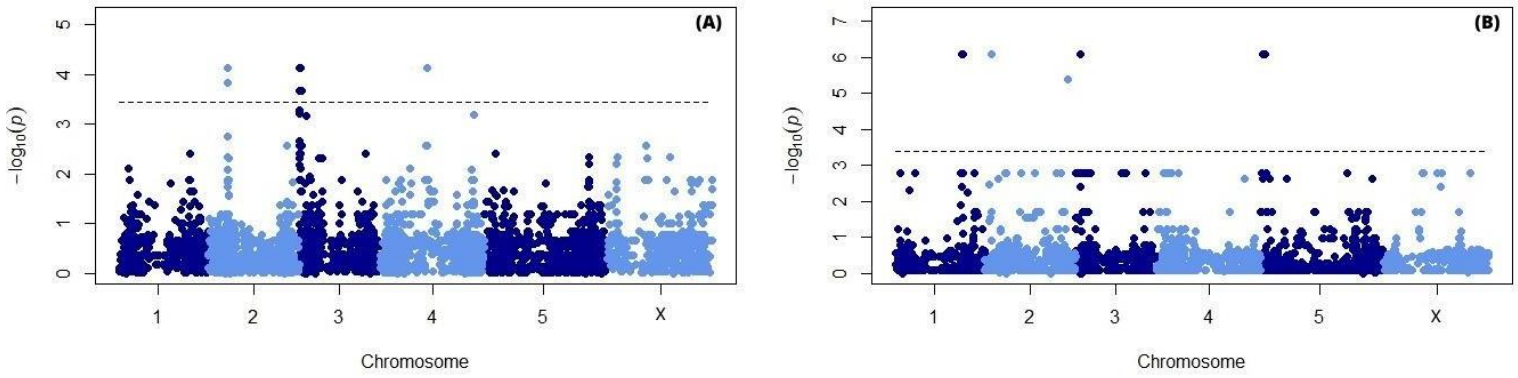


Figure 3.3. Manhattan plots of a genome-wide association analysis of Body size and Lifespan Composite Phenotypes. These Manhattan plots visualize the results of a genome-wide association study (GWAS) for the composite (A) body size and (B) lifespan phenotypes. Each point on the plot represents a genetic variant (SNP) across the genome, and the y-axis displays the statistical significance (represented as -log10 p-values) of their associations with the composite phenotype. The x-axis represents the chromosomal positions of these variants. The horizontal dotted line represents the threshold for genome-wide significance. Significant peaks indicate regions of the genome where genetic variants are strongly associated with the composite body size and lifespan phenotypes. This plot was a tool for visualizing the potential candidate regions contributing to phenotype variations.

Table 1. The list of the most significant SNP Association Results (FDR_BH<0.05) associated with the Body Size composite phenotype. This table provides a comprehensive summary of significant Single Nucleotide Polymorphism (SNP) association results related to body size, each row presenting detailed information about a specific SNP. The following information is given for each SNP: Chr is the chromosome; SNP is the identifier; Base Pair is the position along the chromosome; R-squared is the square of the correlation; p-value are the p-values for the genome-wide association study for the body size composite phenotype. The Closest Genes are the genes closest to our SNP in a range of 20 kb. Observations include several phenotypes associated with the nearest genes and other genes related to the closest ones.

Chromosome	SNP	Base pair	R2	p-value	Closest Genes (< 10 kb)	Observations
2	snp61804	3249292	0,6030	1.509e-04	C17F4.2; srz-68; C17F4.3; C17F4.12; C17F4.7; fbxc-50; C17F4.13; clec-124; gcy-19;	Main Phenotypes: small Other Phenotypes (from genes related to Body size): long; small; aging variant; organism morphology variant; body elongation defective; body length variant; developmental delay; slow growth; modulation of lifespan; larval arrest; dumpy; dauer development phenotype; dauer formation variant; organism pathogen response variant; organism stress response variant
2	snp62261	3268332	0,6364	7.306e-05	F02E11.7; gcy-19; F02E11.2; F02E11.t1; F02E11.3; F02E11.4;	Main Phenotypes: No phenotypes have been reported as observed Other Phenotypes (from genes related to Body size): small; organism morphology variant; body elongation defective; development phenotype; modulation of lifespan; larval arrest; slow growth; dumpy; organism pathogen response variant; embryonic arrest
3	snp106574	74192	0,6364	7.306e-05	attf-3; pck-1; spt-4; F54C4.9; F54C4.8; F54C4.6; F54C4.7; F54C4.5;	Main Phenotypes: larval arrest; slow growth; long; embryonic arrest; embryonic lethal; larval lethal; modulation of lifespan; mitochondria alignment variant; protein phosphorylation reduced; exploded through vulva Other Phenotypes (from genes related to Body size): small; long; modulation of lifespan; larval arrest; slow growth; organism development variant; dauer defective; dauer development phenotype; dauer formation variant
3	snp107488	121431	0,6364	7.306e-05	pals-24; C29F9.6; C29F9.15; C29F9.5; C29F9.12; pals-23; nhr-280; C29F9.2;	Main Phenotypes: growth variant; shortened lifespan; embryonic lethal; egg laying variant Other Phenotypes (from genes related to Body size): small; body elongation defective; developmental delay; growth rate variant; larval development retarded; body length variant; body elongation defective; modulation of lifespan; larval arrest; slow growth; organism morphology variant; aging variant
3	snp108205	146934	0,6364	7.306e-05	pals-25; fbxb-81; T17A3.14; fbxb-82; T17A3.12; fbxb-83; fbxb-84; sqst-5; ver-2; srz-7; adbp-2 ; zig-13; ver-1;	Main Phenotypes: organism pathogen response variant; transgene expression reduced Other Phenotypes (from genes related to Body size): small; developmental delay; growth rate variant; larval development retarded; body elongation defective; modulation of lifespan; larval arrest; slow growth; embryonic arrest; body size variant; aging variant; organism morphology variant; pathogen resistance increased
3	snp110248	225084	0,5871	2.088e-04	klp-20; Y50D7A.15; hpo-38;	Main Phenotypes: pore forming toxin hypersensitive; Bacillus thuringiensis toxin hypersensitive Other Phenotypes (from genes related to Body size): small; slow growth; developmental delay; body elongation defective; developmental delay; body size variant; organism development variant; aging variant; modulation of lifespan; larval arrest; pathogen susceptibility increased
3	snp111011	284144	0,5871	2.088e-04	taco-1; xpd-1; Y50D7A.13; Y50D7A.1;	Main Phenotypes: fat content reduced; larval arrest; slow growth; lethal; embryonic lethal; cell UV response variant Other Phenotypes (from genes related to Body size): modulation of lifespan; dauer lifespan shortened; drug-induced life span variant;dauer death increased developmental timing variant
3	snp112078	443235	0,5871	2.088e-04	affl-1; affl-2; Y55B1BR.10; Y55B1BR.8; Y55B1BR.7;	Main Phenotypes: embryonic lethal; engulfment variant Other Phenotypes (from genes related to Body size): organism development variant; developmental delay; aging variant; body length variant; modulation of lifespan; larval arrest
4	snp186924	7820839	0,6364	7.306e-05	ric-3; zgpa-1; cdgs-1; 21ur-15670; T14A8.2; 21ur-12328; C06A6.4; mvb-12;	Main Phenotypes: larval arrest; embryonic lethal; fat content reduced; levamisole resistant; shortened lifespan; larval lethal; early larval arrest; lysosome-related organelle localization variant; lysosome-related organelle morphology variant; endocytic transport variant; protein degradation variant; receptor mediated endocytosis defective Other Phenotypes (from genes related to Body size): small; slow growth; body elongation defective; developmental delay; body length variant; aging variant; modulation of lifespan; larval arrest

Table 2. The list of the most significant SNP Association Results (FDR<25%) associated with the Lifespan composite phenotype. This table provides a comprehensive summary of significant Single Nucleotide Polymorphism (SNP) association results related to lifespan, each row presenting detailed information about a specific SNP. The following information is given for each SNP: Chr is the chromosome; SNP is the identifier; Base Pair is the position along the chromosome; R-squared is the square of the correlation; p-value are the p-values for the genome-wide association study for the lifespan composite phenotype. The Closest Genes are the genes closest to our SNP in a range of 20 kb. Observations include several phenotypes associated with the nearest genes and other genes related to the closest ones.

Chromosome	SNP	Base pair	R2	p-value	Closest Genes (< 10 kb)	Observations
1	snp15910	11147275	0,7908	7.971e-07	clec-94; clec-93; ZK39.11; ZK39.13; ZK39.9; clec-95; clec-96; clec-97; clec-98; clec-99;	Main Phenotypes: lethal; fat content reduced Other Phenotypes (from genes related to Lifespan): modulation of lifespan; larval arrest; slow growth; larval lethal; drug-induced life span variant; pathogen susceptibility increased; aging variant
1	snp16321	11268082	0,7908	7.971e-07	F36D1.10; F36D1.8; F36D1.12; spex-2; F36D1.13; F36D1.14; F36D1.11; F36D1.5; F36D1.15; F36D1.6; F36D1.4; spex-1; F36D1.9; madf-2;	Main Phenotypes: No phenotypes have been reported as observed Other Phenotypes (from genes related to Lifespan): modulation of lifespan; larval arrest; aging variant; diet induced life span variant; developmental delay; organism development variant; growth rate variant; organism pathogen response variant; pathogen-induced death increased; pathogen susceptibility increased; organism stress response variant; dauer formation variant; dauer lifespan variant; adult lethal
2	snp33844	1067765	0,7908	7.971e-07	K02E7.6; btb-10; mir-4805; K02E7.5; K02E7.10; K02E7.12; linc-13; K02E7.4; K02E7.11;	Main Phenotypes: fat content reduced Other Phenotypes (from genes related to Lifespan): modulation of lifespan; larval arrest; aging variant; slow growth; diet-induced life span variant; organism development variant dauer lifespan variant; dauer defective; dauer formation variant; dauer death increased; stress-induced lethality increased; organism stress response variant; larval lethal
2	snp98718	13948089	0,7897	4.224e-06	W07G1.1; W07G1.24; W07G1.19; W07G1.25; W07G1.23; W07G1.17; W07G1.9; W07G1.8; sre-44; zip-3;	Main Phenotypes: dauer lifespan extended Other Phenotypes (from genes related to Lifespan): larval arrest; modulation of lifespan; slow growth; organism development variant; larval lethal; adult lethal
3	snp114277	871701	0,7908	7.971e-07	exos-4.2; arx-4; T24C4.2; T24C4.3; T24C4.8; T24C4.4; T24C4.5; ucr-2.3; zer-1;	Main Phenotypes: dauer lifespan extended; embryonic lethal; late larval lethal; lethal; slow embryonic development; larval arrest; chemical resistant; metabolic pathway variant; protein expression reduced; mitochondria alignment variant; protein aggregation variant; receptor mediated endocytosis defective Other Phenotypes (from genes related to Lifespan): modulation of lifespan; larval arrest; aging variant; diet-induced life span variant; pathogen-induced death increased; pathogen susceptibility increased; organism development variant; developmental delay; dauer lifespan variant; slow growth; adult lethal; larval lethal
5	snp216345	434980	0,7908	7.971e-07	B0554.2; B0554.5; B0554.4; 21ur-15502; B0554.1; B0554.7; srh-33; srh-34;	Main Phenotypes: dauer lifespan extended; fat content reduced Other Phenotypes (from genes related to Lifespan): modulation of lifespan; larval arrest; aging variant; diet-induced life span variant; developmental delay; growth rate variant; larval development retarded; organism development variant; pathogen-induced death increased; pathogen susceptibility increased; dauer lifespan variant; pathogen resistance increased; adult lethal; larval lethal
5	snp223612	688124	0,7908	7.971e-07	H25P19.1; srx-21; srx-22; srx-23; F31F4.1; srx-4; srj-53; srj-52; ugt-29;	Main Phenotypes: embryonic lethal; organism morphology variant Other Phenotypes (from genes related to Lifespan): modulation of lifespan; larval arrest; aging variant; slow growth; organism development variant; pathogen-induced death increased; pathogen susceptibility increased; organism stress response variant; adult lethal; larval lethal
5	snp225726	893041	0,7908	7.971e-07	srlf-23; srj-38; T02B11.6; srlf-21; srlf-22; srlf-20; nas-32; fmo-5;	Main Phenotypes: axon guidance variant; ectopic axon outgrowth; neuronal outgrowth variant; axon development variant Other Phenotypes (from genes related to Lifespan): extended lifespan; dumpy; embryonic lethal; larval lethal; organism morphology variant

4. Annotation of Candidate Genes

Genes containing or closest to the significant SNPs (within 20 kb, since the SNPs that appear as significant may be revealing the effect of nearby genes but not necessarily in their exact location) were proposed as prospective candidate genes for lifespan and body size. This method resulted in the identification of 141 genes in total, 64 for body size (Table 1) and 77 for lifespan (Table 2), as candidate genes in association with these traits. Candidate genes related to body size were distributed by Chromosomes 3 (41 genes), 2 (15 genes), and 4 (8 genes). For lifespan, the potential genes were found on Chromosomes 5 (25 genes), 1 (24 genes), 2 (19 genes), and 3 (9 genes). In both tables, all the candidate genes for each SNP were recorded, along with their associated traits information related to lifespan and body size. The most interesting genes related to lifespan and body size are discussed in the next section.

IV. Discussion

The main goal of this study was to identify possible genes linked to the response of *C. elegans* when growing in the presence of a pathogenic *Escherichia coli* strain IAI1, specifically those causing lifespan and body length variation. In addition to pinpointing these genes, the next step was to compare them to a set of genes acquired in the presence of a non-pathogenic *Escherichia coli* OP50, known as the control. Considering these goals, the body length and lifespan of about 18 inbred lines were assessed, followed by an association mapping analysis, employed to identify significant genes and gene interactions using a previously constructed mapping panel.

C. elegans are typically moved from one plate to another in survival studies using a single picking process, which is exceedingly time-consuming because only one or two individuals are transferred at a time, resulting in a restriction of the number of worms and experimental settings that may be used in a reasonable amount of time. Since a large number of experimental populations was required for this study, specifically regarding the survival assays, this prompted the creation of a filtering protocol as a faster way of transferring adult *C. elegans* rather than utilizing a picker. The primary benefit will be a significant increase in the total number of experimental individuals. Disadvantages included the necessity of plate inspection for the presence of a small number of larvae that were not filtered despite the high efficiency in separating adults from eggs and larvae.

1. Body Size and Lifespan Variation

The effects of different bacterial species on *C. elegans* lifespan and body size have proven challenging to interpret.

When *C. elegans* were grown on the bacteria strain *E. coli* OP50, they lived significantly longer than when grown on *E. coli* IAI1, hinting at the already known notion that *E. coli* IAI1, in line with its pathogenic nature, is more hazardous to *C. elegans* than *E. coli* OP50, a non-pathogenic bacterium. In other words, the non-pathogenic bacteria *E. coli* OP50 is typically associated with higher lifespans, which corresponds with prior observations (Portal-Celhay et al., 2012).

Additionally, the results regarding body size variation in worms exposed to the different bacterial strains yielded somewhat unexpected findings. While the overall differences in body size between the two bacteria were not statistically significant, contrary to our initial expectations of

smaller body sizes for individuals in the presence of a pathogenic bacteria (Balloux & van Dorp, 2017; Janikddfghx et al., 2020), worms exposed to *E. coli* IAI1 exhibited slightly higher average body sizes when compared to worms exposed to *E. coli* OP50. These intriguing findings align with emerging evidence from other research cases involving *E. coli* IAI1 and various pathogenic bacteria, extending beyond body size variation and including studies where worm hosts exposed to these pathogens exhibit a reproduction that was shifted to earlier performance (Baeriswyl et al., 2010). Also, despite different reports (Flemming et al., 2000; Woodruff et al., 2018) on the body size measurement of several *C. elegans* strains and other nematodes on OP50 strain, body size measurement on IAI1 strain and differences in body length between the two bacteria haven't been described.

Other possibilities may include variability in the nutritional content across bacteria, as well as the quality between different bacterial strains, which most likely influence the capacities of these species to support the growth and development of their host. These discoveries raise intriguing questions about the intricate interactions between these particular strains, as well as other non- and pathogenic bacteria, and the nematode worms that feed on them, suggesting that the relationship between host and pathogen may be more complex than initially thought. This new light on a more dynamic nature of host-microbe interactions warrants further investigation to elucidate the underlying mechanisms at play.

2. Correlation of phenotypes

The relationship between lifespan and body size is an interesting topic in biology, and it is often observed that there is a correlation between these two factors, observing that smaller organisms tend to live less when compared to their larger counterparts (Speakman, 2005).

There are situations when it's reasonable to believe that the two independent variables have an interaction effect rather than an additive effect. In our experiment, for example, it's plausible that different bacterial strains from which the worms feed influence the worms' growth and body length. Since it may impact the body length trait of the individual, the two-way ANOVA must account for that. Upon looking at the output table, the interaction variable had a low sum-of-squares value and a high p-value and, when the ANOVA analysis was repeated but only considering an additive effect, both body size and bacterial strain had p-values that are significant at the alpha level of 0.05. These results suggest that the individual's body size and the strain of bacteria from which it feeds have a consistent effect on the individual's lifespan, but also that there isn't much variation in lifespan that the interaction between body size and bacteria can explain.

Observation of Figure 2 showed a negative correlation between the increase in body size and the decline in longevity, verified in the presence of both *E. coli* OP50 and *E. coli* IAI1. These findings imply that these traits are perhaps causally connected; for example, smaller body size may contribute

to increased longevity for several reasons, including a shorter mouth opening. In these cases, a reduced mouth size limits the intake of potentially harmful bacteria and pathogens from the environment, minimizing the risk of infections and microbial-related diseases and ultimately promoting a longer and healthier lifespan. Alternatively, an independent but common mechanism may influence both of these traits, like stress during early development leading to shorter body size and extended lifespan. Individuals may undergo physiological changes after being exposed to stress early in life, like a resource allocation to survival and mechanisms to combat stress instead of growth and development, resulting in an increase in longevity and a decrease in body size. A hypothesis can be proposed to explain the negative correlation between lifespan and body size, which is based on the concept of developmental delays, which posits that smaller organisms may experience a slower pace in their development, which in turn contributes to their extended lifespan. One possible biological mechanism that connects delayed development to the negative correlation between lifespan and body size will be trade-offs. Evolutionarily, there may be trade-offs between growth and longevity (Hou, 2013), meaning smaller worms may benefit from longer lifespans as they are less likely to encounter predators or adverse environmental conditions. On the other hand, larger worms may prioritize rapid growth and reproduction to maximize their chances of passing on their genes.

This delay in development, which contributes to the negative correlation between lifespan and body size, may also be influenced by dietary factors, specifically the types of bacteria they interact with. For instance, some organisms might have a less nutrient-rich diet by feeding on *E. coli* OP50 (So et al., 2011), which can delay development as they apply resource allocation. This efficient mechanism allows smaller individuals to invest in maintenance processes and stress resistance, prioritizing survival and longevity instead of growth and reproduction, ultimately leading to extended lifespans. In contrast, worms that feed on a more harmful bacteria strain, like the pathogenic *E. coli* IAI1, will most likely prioritize growth and reproduction, resulting in larger bodies and shortened lifespans.

3. Association Mapping

The Manhattan plots served only as visual tools for portraying the results of the genetic association study, with these data typically obtained through specialized software like PLINK. Thus, it makes it easier to identify genomic regions or variants that may be significantly associated with, for example, a particular trait like body size or lifespan. The Manhattan plot results for body size and lifespan on both strains OP50 and IAI1 (Supplementary Figures 2 and 3) revealed a lack of statistical significance since no SNP was able to reach the conventional significance threshold. These results may be primarily attributed to the limited pool of genotypes used, considering that the restricted genetic

diversity inherent in the dataset might have limited the study's ability to detect meaningful associations between genetic markers and the trait of interest. The use of composite phenotypes was able to overcome the limitations imposed by the reduced genotypic dataset, uncovering significant genetic associations that may have been missed in the initial analysis.

These Manhattan plots offered a comprehensive visual summary of the genetic landscape for each phenotype, demonstrating that their genetic basis typically involves the intricate interplay of multiple genes. These complex traits are not governed by a single gene with a clear-cut cause-and-effect relationship, but instead, they arise from the cumulative effects of numerous genetic variants present throughout the genome.

4. Annotation of Candidate Genes

This research conducted a Genome-Wide Association Study (GWAS) on different *C. elegans* inbred lines to uncover the genetic determinants related to the influence of two distinct bacterial diets, OP50 strain and IAI1 strain, on lifespan and body size variation. Our study employed an approach to identify possible genetic loci by highlighting distinct rows in Tables 1 and 2, each corresponding to a single SNP. Due to the close proximity of the SNPs, a strong likelihood exists that they are associated with the same gene responsible for variations in the phenotype. By visually emphasizing these rows, we aimed to emphasize the potential significance of these genetic markers and provide a more precise representation of potential genetic loci of interest.

Our research identified several noteworthy genes close to the markers associated with lifespan differences in *C. elegans*. A first example can be the seven members of the CLEC gene family adjacent to an SNP (snp15910) located on chromosome 1, a family possessing many genes known to be involved in the response against pathogens (Pees et al., 2021). Although none of these genes have a direct connection to lifespan reported, aside from the clec-98's association with the lethal phenotype (premature death at any stage of the life cycle), several other genes were reported as associated with these seven genes and having that direct connection, including daf-2 and skn-1. Notably, other genes like spex-2 and F36D1.5, which once again have no direct link to lifespan, were reported to be significantly associated with daf-2, as well as daf-16 and pmk-1. The study also uncovered intriguing genes adjacent to the significant SNPs associated with body size variations in *C. elegans*. Similar to lifespan, there were four members of the F-box genes, all adjacent to an SNP (snp108205) located on chromosome 3, a gene family known to participate in the regulation of lifespan and development (Ma et al., 2021). Although no direct link between these genes and body size has been reported, multiple other genes were reported as both correlated with body size and with these four F-box genes, in which

daf-2 and nhr-25 are included. Furthermore, the attf-3 gene was described in association with traits such as larval arrest, slow growth and longer body, suggesting a possible association with body size variation despite the lack of information regarding this particular gene.

Among the associated genes is the daf-2 gene, one of the most well-known genes affecting *C. elegans* lifespan (Halaschek-Wiener et al., 2005; Kim, 2013). This gene encodes an insulin/IGF-1 receptor homolog, and mutations in it can significantly extend lifespan by activating pathways that promote stress resistance, pathogen resistance and longevity. Just as it plays a role in lifespan regulation, the daf-2 gene also influences body size in *C. elegans* by affecting growth and development. Mutations in daf-2 can result in smaller body sizes (Venz et al., 2021).

Additionally, the gene skn-1, involved in regulating oxidative stress and detoxification, is a homolog of the mammalian Nrf2 transcription factor known for its role in promoting longevity under stress conditions, extending lifespan by enhancing the organism's ability to combat oxidative stress. The activation of skn-1 is mediated by the pathway of another gene, the pmk-1 (Kim, 2013; Radeke & Herman, 2021). Another one is the daf-16 gene, a transcription factor that regulates the expression of genes involved in the host's defense, leading to extensions in C. elegans lifespan. The resistance to pathogenic bacteria exhibited by daf-2 mutants is also regulated by daf-16 (Kim, 2013).

The nhr-25 gene is a nuclear hormone receptor that controls the expression of genes involved in growth and development (Hada et al., 2010), thus regulating body size. It is a gene directly associated with the larva-to-adult transition.

These genes represent a subset of the many genes identified through the Genome-wide association study of *C. elegans* on lifespan and body size, with several of them being conserved across species, including humans. The insights gained from studying these genes in *C. elegans* may provide valuable clues about the genetic pathways underlying body size and lifespan regulations in more complex organisms. However, it is imperative to recognize that all genes identified through this research are merely potential candidates, and the genes that are actually responsible for the observed variations might not necessarily be among those discovered. The study provided statistical associations between specific genetic markers and traits, but it does not definitively pinpoint causative genes or variants. There is the possibility that none of these genes are responsible, and in fact, other genes close to the marker are the ones behind the variation in the phenotypes.

5. Final remarks

This research is the first to offer an overview of the genetic trade-offs that occur during the evolution of aging in the presence of pathogens. Genome-wide association studies (GWAS) have been instrumental in unravelling the genetic basis of various traits in the model organism *Caenorhabditis elegans*, including body size and lifespan. Several genes were proposed as candidate genes for longevity and body length within identified significant genomic regions linked to these traits. This study demonstrated the existence of a genetic base behind the influence of bacteria over the variation of body size and longevity. Specific genes in *C. elegans* play crucial roles in mediating interactions with bacteria, influencing processes like nutrient uptake, immune responses, and metabolism. These genetic interactions ultimately determine the worm's lifespan and body size, highlighting the intricate interplay between genetics and the microbiome in shaping the biology of this model organism. Additionally, these findings have implications for understanding developmental processes in more complex organisms, and they have opened up avenues for potential interventions to extend a healthy lifespan and combat age-related diseases in humans.

Despite diligent efforts to investigate potential genetic markers for phenotype variations, the study's findings regarding lifespan and body size on OP50 strain and IAI1 strain failed to reach the conventional significance threshold, only obtaining significant results when the analysis was applied to lifespan and body size as composite phenotypes, thus highlighting the critical importance of a larger and more diverse genotypic dataset. As such, future research endeavors should prioritize expanding the genotypic pool to encompass a more comprehensive range of genetic inbred lines, thereby enhancing the statistical power and robustness of the analysis to uncover potential genetic influences.

As previously mentioned, all genes identified using this Genome-Wide Association Study are only prospective candidates, and the genes responsible for the observed variation may not even be among those detected since this study can only offer statistical connections between genetic markers and phenotypes and is not capable of establishing a causal relation between them. The genetic landscape is extraordinarily complex, with countless genes and their interactions contributing to the phenotypic variation; thus, while GWAS was undoubtedly a valuable tool in uncovering genetic associations to these two traits, it served simply as a starting point. Further research and functional studies are necessary to discover the precise drivers of genetic variation. Among the various approaches, a very interesting one is based on the use of introgression lines, which have emerged as invaluable tools in genetic research, particularly in validating Genome-Wide Association Study (GWAS) findings. These specialized lines are generated through repeated backcrossing and selfing between two distinct parental populations, leading to the introgression of specific genomic segments from one parent into the genetic background

of the other, replacing chromosome fragments of the recipient with chromosome fragments of the donor (B. Zhang et al., 2022). Researchers can precisely isolate and assess the impact of individual genomic regions or candidate genes implicated in GWAS by utilizing introgression lines. As exemplified in some studies, this approach offers a powerful means of confirming the association between a genetic variant and a specific trait or phenotype and elucidating the underlying mechanisms governing these associations (Frézal et al., 2023). The use of introgression lines thus enhances the robustness and reliability of GWAS results, ultimately advancing our understanding of complex trait genetics and aiding in developing more targeted and effective interventions in fields ranging from agriculture to human health.

V. Bibliography

Aguirre-Sánchez, J. R., Valdez-Torres, J. B., del Campo, N. C., Martínez-Urtaza, J., del Campo, N. C., Lee, B. G., Quiñones, B., & Chaidez-Quiroz, C. (2022). Phylogenetic group and virulence profile classification in Escherichia coli from distinct isolation sources in Mexico. *Infection, Genetics and Evolution*, *106*, 105380. https://doi.org/10.1016/J.MEEGID.2022.105380

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Introduction to Pathogens*. https://www.ncbi.nlm.nih.gov/books/NBK26917/

Antell, D. E., & Taczanowski, E. M. (1999). How environment and lifestyle choices influence the aging process. *Annals of Plastic Surgery*, *43*(6), 585–588. https://doi.org/10.1097/00000637-199912000-00001

Arata, Y., Oshima, T., Ikeda, Y., Kimura, H., & Sako, Y. (2020). OP50, a bacterial strain conventionally used as food for laboratory maintenance of C. elegans, is a biofilm formation defective mutant. *MicroPublication Biology*, *2020*. https://doi.org/10.17912/MICROPUB.BIOLOGY.000216

Badal, V. D., Vaccariello, E. D., Murray, E. R., Yu, K. E., Knight, R., Jeste, D. V., & Nguyen, T. T. (2020). The Gut Microbiome, Aging, and Longevity: A Systematic Review. *Nutrients*, *12*(12), 1–25. https://doi.org/10.3390/NU12123759

Baeriswyl, S., Diard, M., Mosser, T., Leroy, M., Manière, X., Taddei, F., & Matic, I. (2010). Modulation of aging profiles in isogenic populations of Caenorhabditis elegans by bacteria causing different extrinsic mortality rates. *Biogerontology*, *11*(1), 53–65. https://doi.org/10.1007/S10522-009-9228-0/FIGURES/6

Balloux, F., & van Dorp, L. (2017). Q&A: What are pathogens, and what have they done to and for us? *BMC Biology*, *15*(1). https://doi.org/10.1186/S12915-017-0433-Z

Braz, V. S., Melchior, K., & Moreira, C. G. (2020). Escherichia coli as a Multifaceted Pathogenic and Versatile Bacterium. *Frontiers in Cellular and Infection Microbiology*, *10*, 548492. https://doi.org/10.3389/FCIMB.2020.548492/BIBTEX

Brenner, S. (1974). THE GENETICS OF CAENORHABDITIS ELEGANS. *Genetics*, *77*(1), 71–94. https://doi.org/10.1093/GENETICS/77.1.71

Charles A Janeway, J., Travers, P., Walport, M., & Shlomchik, M. J. (2001). *Pathogens have evolved various means of evading or subverting normal host defenses*. https://www.ncbi.nlm.nih.gov/books/NBK27176/

Chelo, I. M. (2014). Experimental determination of invasive fitness in Caenorhabditis elegans. *Nature Protocols*, *9*(6), 1392–1400. https://doi.org/10.1038/NPROT.2014.098

Collins, J. J., Huang, C., Hughes, S., & Kornfeld, K. (2008). The measurement and analysis of age-related changes in Caenorhabditis elegans. *WormBook : The Online Review of C. Elegans Biology*, 1–21. https://doi.org/10.1895/WORMBOOK.1.137.1

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.1111/J.2517-6161.1972.TB00899.X

Cressler, C. E., Nelson, W. A., Day, T., & Mccauley, E. (2014). Disentangling the interaction among host resources, the immune system and pathogens. *Ecology Letters*, *17*(3), 284–293. https://doi.org/10.1111/ELE.12229

de Sousa Figueiredo, M. B., Pradel, E., George, F., Mahieux, S., Houcke, I., Pottier, M., Fradin, C., Neut, C., Daniel, C., Bongiovanni, A., Foligné, B., & Titécat, M. (2021). Adherent-Invasive and Non-Invasive Escherichia coli Isolates Differ in Their Effects on Caenorhabditis elegans' Lifespan. *Microorganisms 2021, Vol. 9, Page 1823*, *9*(9), 1823. https://doi.org/10.3390/MICROORGANISMS9091823

Diard, M., Baeriswyl, S., Clermont, O., Gouriou, S., Picard, B., Taddei, F., Denamur, E., & Matic, I.

(2007). Caenorhabditis elegans as a simple model to study phenotypic and genetic virulence determinants of extraintestinal pathogenic Escherichia coli. *Microbes and Infection*, *9*(2), 214–223. https://doi.org/10.1016/J.MICINF.2006.11.009

Flemming, A. J., Shen, Z. Z., Cunha, A., Emmons, S. W., & Leroi, A. M. (2000). Somatic polyploidization and cellular proliferation drive body size evolution in nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(10), 5285. https://doi.org/10.1073/PNAS.97.10.5285

Frézal, L., Saglio, M., Zhang, G., Noble, L., Richaud, A., & Félix, M.-A. (2023). Genome-wide association and environmental suppression of the mortal germline phenotype of wild C. elegans. *BioRxiv*, 2023.05.17.540956. https://doi.org/10.1101/2023.05.17.540956

Ghosh, T. S., Shanahan, F., & O'Toole, P. W. (2022). The gut microbiome as a modulator of healthy ageing. *Nature Reviews Gastroenterology & Hepatology 2022 19:9*, *19*(9), 565–584. https://doi.org/10.1038/s41575-022-00605-x

Goldstein, B. (2016). Sydney Brenner on the Genetics of Caenorhabditis elegans. *Genetics*, *204*(1), 1. https://doi.org/10.1534/GENETICS.116.194084

Hada, K., Asahina, M., Hasegawa, H., Kanaho, Y., Slack, F. J., & Niwa, R. (2010). The nuclear receptor gene nhr-25 plays multiple roles in the C. elegans heterochronic gene network to control the larva-to-adult transition. *Developmental Biology*, *344*(2), 1100. https://doi.org/10.1016/J.YDBIO.2010.05.508

Halaschek-Wiener, J., Khattra, J. S., McKay, S., Pouzyrev, A., Stott, J. M., Yang, G. S., Holt, R. A., Jones, S. J. M., Marra, M. A., Brooks-Wilson, A. R., & Riddle, D. L. (2005). Analysis of long-lived C. elegans daf-2 mutants using serial analysis of gene expression. *Genome Research*, *15*(5), 603. https://doi.org/10.1101/GR.3274805

Hamazaki, K., & Iwata, H. (2020). RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method. *PLOS Computational Biology*, *16*(2), e1007663. https://doi.org/10.1371/JOURNAL.PCBI.1007663

Hou, C. (2013). The energy trade-off between growth and longevity. *Mechanisms of Ageing and Development*, *134*(9), 373–380. https://doi.org/10.1016/J.MAD.2013.07.001

Huang, Y., & Kammenga, J. E. (2020). Genetic Variation in Caenorhabditis elegans Responses to Pathogenic Microbiota. *Microorganisms*, *8*(4). https://doi.org/10.3390/MICROORGANISMS8040618

Iser, W. B., & Wolkow, C. A. (2007). DAF-2/Insulin-Like Signaling in C. elegans Modifies Effects of Dietary Restriction and Nutrient Stress on Aging, Stress and Growth. *PLoS ONE*, *2*(11), 1240. https://doi.org/10.1371/JOURNAL.PONE.0001240

James, C. G., Morah, O., Panwala, V., & Yarmand, A. (2018). *Survival of Caenorhabditis elegans Infected with Escherichia coli DFB1655 is not Affected by a Missense Mutation in dop-1 or Treatment with Chlorpromazine Hydrochloride*. *4*, 1–9. https://jemi.microbiology.ubc.ca/1

Janikddfghx, E., Ceremuga, M., Niemcewicz, M., & Bijak, M. (2020). Dangerous Pathogens as a Potential Problem for Public Health. *Medicina*, *56*(11), 1–23. https://doi.org/10.3390/MEDICINA56110591

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G., Carbonnelle, E., Lortholary, O., Clermont, O., Denamur, E., Picard, B., Nassif, X., & Brisse, S. (2008). Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. *BMC Genomics*, *9*(1), 1–14. https://doi.org/10.1186/1471-2164-9-560/TABLES/4

Jnani, D., & Ray, S. D. (2022). Escherichia coli. *Reference Module in Biomedical Sciences*. https://doi.org/10.1016/B978-0-12-824315-2.00190-1

Jura, M., & Kozak, L. P. (2016). Obesity and related consequences to ageing. *Age*, *38*(1). https://doi.org/10.1007/S11357-016-9884-3

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*(282), 457–481. https://doi.org/10.1080/01621459.1958.10501452

Kim, D. H. (2013). Bacteria and the aging and longevity of Caenorhabditis elegans. *Annual Review of*

Genetics, *47*, 233–246. https://doi.org/10.1146/ANNUREV-GENET-111212-133352
Kong, F., & Yang, L. (2023). Pathogen-triggered changes in plant development: Virulence strategies or host defense mechanism? *Frontiers in Microbiology*, *14*, 1122947. https://doi.org/10.3389/FMICB.2023.1122947/BIBTEX
Laranjeira, A. C. (2016). *Can evolution of gut microbiota alter C. elegans longevity*.
Lin, K., Hsin, H., Libina, N., & Kenyon, C. (2001). Regulation of the Caenorhabditis elegans longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nature Genetics*, *28*(2), 139–145. https://doi.org/10.1038/88850
Ma, F., Lau, C. Y., & Zheng, C. (2021). Large Genetic Diversity and Strong Positive Selection in F-Box and GPCR Genes among the Wild Isolates of Caenorhabditis elegans. *Genome Biology and Evolution*, *13*(5). https://doi.org/10.1093/GBE/EVAB048
Marsh, E. K., & May, R. C. (2012). Caenorhabditis elegans, a Model Organism for Investigating Immunity. *Applied and Environmental Microbiology*, *78*(7), 2075. https://doi.org/10.1128/AEM.07486-11
Meneely, P. M., Dahlberg, C. L., & Rose, J. K. (2019). Working with Worms: Caenorhabditis elegans as a Model Organism. *Current Protocols Essential Laboratory Techniques*, *19*(1), e35. https://doi.org/10.1002/CPET.35
Merkx-Jacques, A., Coors, A., Brousseau, R., Masson, L., Mazza, A., Tien, Y. C., & Topp, E. (2013). Evaluating the Pathogenic Potential of Environmental Escherichia coli by Using the Caenorhabditis elegans Infection Model. *Applied and Environmental Microbiology*, *79*(7), 2435. https://doi.org/10.1128/AEM.03501-12
Mordecai, E. A. (2011). Pathogen impacts on plant communities: unifying theory, concepts, and empirical work. *Ecological Monographs*, *81*(3), 429–441. https://doi.org/10.1890/10-2241.1
Noble, L. M., Chelo, I., Guzella, T., Afonso, B., Riccardi, D. D., Ammerman, P., Dayarian, A., Carvalho, S., Crist, A., Pino-Querido, A., Shraiman, B., Rockman, M. V., & Teotónio, H. (2017). Polygenicity and Epistasis Underlie Fitness-Proximal Traits in the Caenorhabditis elegans Multiparental Experimental Evolution (CeMEE) Panel. *Genetics*, *207*(4), 1663–1685. https://doi.org/10.1534/GENETICS.117.300406
Noble, L. M., Rockman, M. V., & Teotónio, H. (2021). Gene-level quantitative trait mapping in Caenorhabditis elegans. *G3 Genes|Genomes|Genetics*, *11*(2). https://doi.org/10.1093/G3JOURNAL/JKAA061
Pakbin, B., Brück, W. M., & Rossen, J. W. A. (2021). Virulence Factors of Enteric Pathogenic Escherichia coli: A Review. *International Journal of Molecular Sciences*, *22*(18). https://doi.org/10.3390/IJMS22189922
Pees, B., Yang, W., Kloock, A., Petersen, C., Peters, L., Fan, L., Friedrichsen, M., Butze, S., Zárate-Potes, A., Schulenburg, H., & Dierking, K. (2021). Effector and regulator: Diverse functions of C. elegans C-type lectin-like domain proteins. *PLoS Pathogens*, *17*(4). https://doi.org/10.1371/JOURNAL.PPAT.1009454
Picard, B., Garcia, J. S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., & Denamur, E. (1999). The link between phylogeny and virulence in Escherichia coli extraintestinal infection? *Infection and Immunity*, *67*(2), 546–553. https://doi.org/10.1128/IAI.67.2.546-553.1999/ASSET/75146A0C-8510-4AE7-8265-F2A289EAD41F/ASSETS/GRAPHIC/II0290560002.JPEG
PLINK: Whole genome data analysis toolset. (n.d.). Retrieved June 21, 2023, from https://zzz.bwh.harvard.edu/plink/
Porta-de-la-Riva, M., Fontrodona, L., Villanueva, A., & Cerón, J. (2012). Basic Caenorhabditis elegans Methods: Synchronization and Observation. *Journal of Visualized Experiments : JoVE*, *64*. https://doi.org/10.3791/4019
Portal-Celhay, C., Bradley, E. R., & Blaser, M. J. (2012). Control of intestinal bacterial proliferation in regulation of lifespan in Caenorhabditis elegans. *BMC Microbiology*, *12*(1), 1–17. https://doi.org/10.1186/1471-2180-12-49/TABLES/2
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P.,

De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, *81*(3), 559. https://doi.org/10.1086/519795

R: The R Project for Statistical Computing. (n.d.). Retrieved September 29, 2023, from https://www.r-project.org/

Radeke, L. J., & Herman, M. A. (2021). Take a Walk to the Wild Side of Caenorhabditis elegans - Pathogen Interactions . *Microbiology and Molecular Biology Reviews*, *85*(2). https://doi.org/10.1128/MMBR.00146-20/ASSET/1C8AD47D-E276-4B71-8F15-F7CFC4DB4E35/ASSETS/IMAGES/LARGE/MMBR.00146-20-F0003.JPG

Robbens, J., Devriese, L., Verstraete, K., & Heyndrickx, M. (2014). Escherichia coli. *Encyclopedia of Toxicology: Third Edition*, 459–461. https://doi.org/10.1016/B978-0-12-386454-3.01006-X

Rodríguez-Rodero, S., Fernández-Morera, J. L., Menéndez-Torre, E., Calvanese, V., Fernández, A. F., & Fraga, M. F. (2011). Aging Genetics and Aging. *Aging and Disease*, *2*(3), 186. /pmc/articles/PMC3295054/

Schaible, U. E., & Kaufmann, S. H. E. (2005). A nutritive view on the host-pathogen interplay. *Trends in Microbiology*, *13*(8), 373–380. https://doi.org/10.1016/j.tim.2005.06.009

Sen, R., Nayak, L., & De, R. K. (2016). A review on host-pathogen interactions: classification and prediction. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology*, *35*(10), 1581–1599. https://doi.org/10.1007/S10096-016-2716-7

So, S., Miyahara, K., & Ohshima, Y. (2011). Control of body size in C. elegans dependent on food and insulin/IGF-1 signal. *Genes to Cells*, *16*(6), 639–651. https://doi.org/10.1111/J.1365-2443.2011.01514.X

Speakman, J. R. (2005). Body size, energy metabolism and lifespan. *The Journal of Experimental Biology*, *208*(Pt 9), 1717–1730. https://doi.org/10.1242/JEB.01556

Stiernagle, T. (2006). Maintenance of C. elegans. *WormBook : The Online Review of C. Elegans Biology*, 1–11. https://doi.org/10.1895/WORMBOOK.1.101.1

Therneau, T. M. (2023). *Survival Analysis [R package survival version 3.5-7]*. https://CRAN.R-project.org/package=survival

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., El Karoui, M., Frapy, E., … Denamur, E. (2009). Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths. *PLOS Genetics*, *5*(1), e1000344. https://doi.org/10.1371/JOURNAL.PGEN.1000344

Uno, M., & Nishida, E. (2016). Lifespan-regulating genes in C. elegans. *Npj Aging and Mechanisms of Disease 2016 2:1*, *2*(1), 1–8. https://doi.org/10.1038/npjamd.2016.10

Uno, M., Tani, Y., Nono, M., Okabe, E., Kishimoto, S., Takahashi, C., Abe, R., Kurihara, T., & Nishida, E. (2021). Neuronal DAF-16-to-intestinal DAF-16 communication underlies organismal lifespan extension in C. elegans. *IScience*, *24*(7). https://doi.org/10.1016/J.ISCI.2021.102706

Vellai, T., Takacs-Vellai, K., Zhang, Y., Kovacs, A. L., Orosz, L., & Müller, F. (2003). Influence of TOR kinase on lifespan in C. elegans. *Nature 2003 426:6967*, *426*(6967), 620–620. https://doi.org/10.1038/426620a

Venz, R., Pekec, T., Katic, I., Ciosk, R., & Ewald, C. Y. (2021). End-of-life targeted degradation of DAF-2 insulin/IGF-1 receptor promotes longevity free from growth-related pathologies. *ELife*, *10*. https://doi.org/10.7554/ELIFE.71335

Woodruff, G. C., Willis, J. H., & Phillips, P. C. (2018). Dramatic evolution of body length due to postembryonic changes in cell size in a newly discovered close relative of Caenorhabditis elegans. *Evolution Letters*, *2*(4), 427–441. https://doi.org/10.1002/EVL3.67

WormBase : Nematode Information Resource. (n.d.). Retrieved August 12, 2023, from https://wormbase.org//#012-34-5

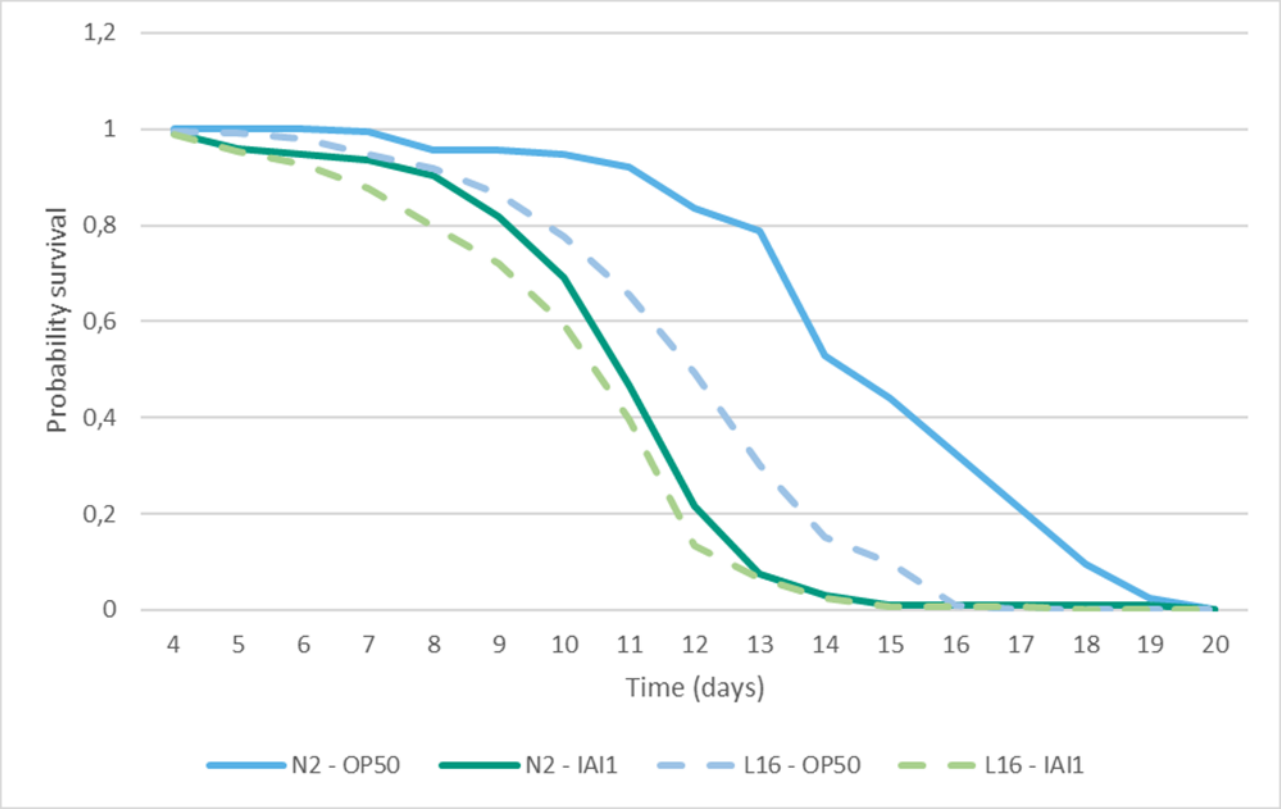
Zhang, B., Ma, L., Wu, B., Xing, Y., & Qiu, X. (2022). Introgression Lines: Valuable Resources for

Functional Genomics Research and Breeding in Rice (Oryza sativa L.). *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/FPLS.2022.863789

Zhang, R., & Hou, A. (2013). Host-Microbe Interactions in Caenorhabditis elegans. *ISRN Microbiology*, *2013*, 1–7. https://doi.org/10.1155/2013/356451

Zhang, S., Li, F., Zhou, T., Wang, G., & Li, Z. (2020). Caenorhabditis elegans as a Useful Model for Studying Aging Mutations. *Frontiers in Endocrinology*, *11*, 731. https://doi.org/10.3389/FENDO.2020.554994/BIBTEX

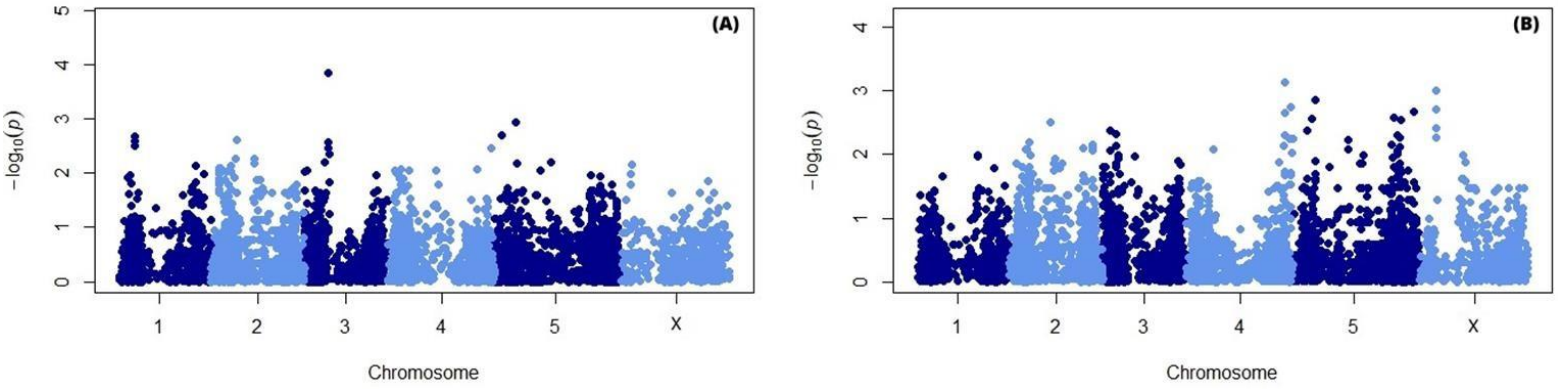
Supplementary Information



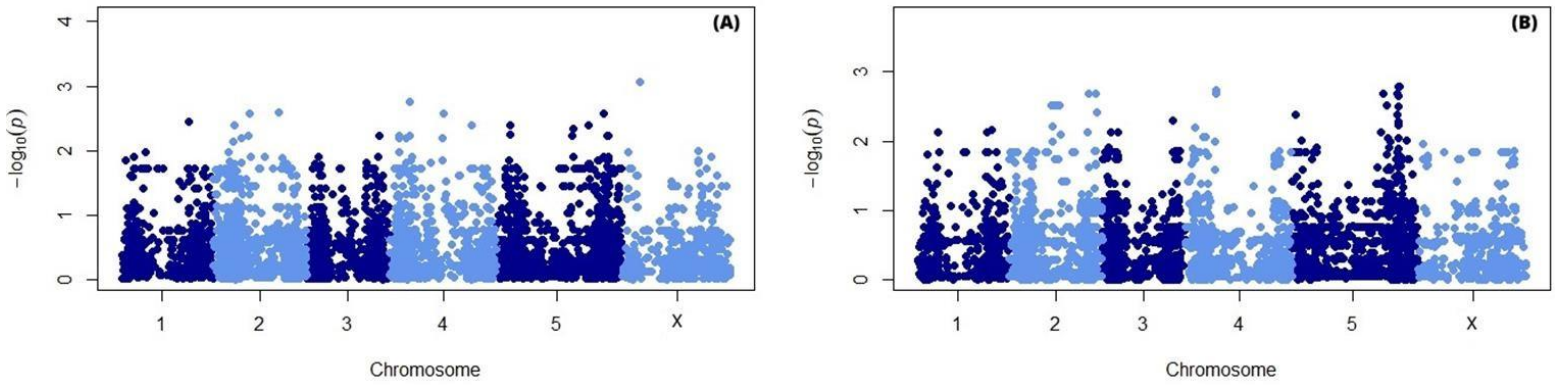
Supplementary Figure 1. Example of a Survival Curve. An example of a graphical representation used in survival analysis to visualize the probability of individuals or subjects surviving over time. The x-axis typically represents time, while the y-axis represents the estimated probability of survival. In a survival curve, each step or drop represents an event (e.g., death, failure, or event of interest) occurring in the study population. This curve provides insights into the survival of two distinct *C. elegans* strains, N2 (represented by a full line) and L16 (represented by a dashed line), on two different bacteria, OP50 strain and IAI1 strain.

Supplementary Table 1. Table of Mean Lifespan and Body Size, on Two Bacteria, for Each Inbred Line. This table compiles data on mean lifespan and body size corrected measurements for each inbred line on two different bacterial strains. Each row corresponds to a specific inbred line, providing information on both mean lifespan and body size values on OP50 strain and IAI1 strain, which have suffered a correction based on the N2 line. They are essential variables for understanding the characteristics and variations within each line. This table serves as a valuable reference for the impact of genetic diversity and bacterium on lifespan and body size.

Genotype	Body size_OP50 (µm)	Body size_IAI1 (µm)	Mean Lifespan_OP50 (days)	Mean Lifespan_IAI1 (days)
L1	955,3	950,6	11,6	10,4
L13	933,0	1004,1	12,5	9,6
L16	1034,1	1125,4	10,7	10,2
L17	946,1	970,0	11,3	10,0
L2	1015,2	951,5	13,2	11,8
L22	921,2	1063,5	15,3	10,6
L25	841,9	1035,0	14,5	10,0
L27	944,0	1096,6	13,0	11,0
L30	1006,3	937,0	12,5	11,9
L35	1014,9	928,2	12,3	11,3
L4	891,8	1035,2	14,0	10,5
L5	863,7	881,7	13,2	10,0
L61	1040,5	1004,9	12,2	10,5
L62	804,0	882,4	12,9	9,9
L63	944,8	1006,8	13,3	10,7
L7	904,2	1022,9	11,2	12,5
L9	966,0	927,5	13,9	11,4
N2	967,8	943,3	13,5	10,9



Supplementary Figure 2. Manhattan plots of a genome-wide association analysis of the Body size Phenotype in OP50 strain and IAI1 strain. These Manhattan plots visualize the results of a genome-wide association study (GWAS) for the body size phenotype, either on (**A**) OP50 strain or (**B**) IAI1 strain. Each point on the plot represents a genetic variant (SNP) across the genome, and the y-axis displays the statistical significance (represented as -log10 p-values) of their associations with the phenotype. The x-axis represents the chromosomal positions of these variants. There is no horizontal dotted line representing the threshold, because all values remained below the genome-wide significance. This plot was a tool for visualizing the potential candidate regions contributing to phenotype variations.



Supplementary Figure 3. Manhattan plots of a genome-wide association analysis of the Lifespan Phenotype in OP50 strain and IAI1 strain. These Manhattan plots visualize the results of a genome-wide association study (GWAS) for the lifespan phenotype, either on (**A**) OP50 strain or (**B**) IAI1 strain. Each point on the plot represents a genetic variant (SNP) across the genome, and the y-axis displays the statistical significance (represented as -log10 p-values) of their associations with the phenotype. The x-axis represents the chromosomal positions of these variants. There is no horizontal dotted line representing the threshold, because all values remained below the genome-wide significance. This plot was a tool for visualizing the potential candidate regions contributing to phenotype variations.