

NORTHWESTERN UNIVERSITY

Mechanisms of Larval Growth Control in *Caenorhabditis elegans*

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biological Sciences

By

Joy Nyaanga

EVANSTON, ILLINOIS

June 2022

© Copyright by Joy Nyaanga 2022

All Rights Reserved

## Abstract

Body size is one of the most discernible ways in which animal species vary. The blue whale (*Balaenoptera musculus*), the largest animal on earth, can reach up to 30 m in length and weigh up to 200 tonnes. At the other extreme, a species of frog called *Paedophryne amauensis* is among the smallest animals on earth with adults measuring just 8 mm in length, about the size of a pea. Understanding how organisms grow to their characteristic sizes is a fundamental biological question. Although a larger body size can increase an organism's competitive advantage, an increased body size also requires added time and nutrients to develop. As such, organisms must have mechanisms to both sense and adjust growth during development. Studies of single cells have revealed that growth regulation can be achieved using time or size sensing control methods. In multicellular organisms, however, regulatory mechanisms must not only control single cell growth but also integrate it across organs and tissues during development. In this dissertation, I investigate growth control in the roundworm nematode *Caenorhabditis elegans*. *C. elegans* is particularly suited for the study of growth control because of its rapid development and genetic tractability. I first optimize a high-throughput phenotyping platform that facilitates quantitative assessment of thousands of individuals at high precision. Using this platform, I quantify changes in animal size and shape throughout development, and explore how genetics and natural genetic variation contribute to differences in animal growth. The results of this work lay the foundation for a mechanistic dissection of organismal growth control.

## Acknowledgements

*Six years ago, inspired by the incredible scientists who'd mentored me and the research I'd been immersed in, I made the decision to pursue a graduate degree. Five years ago, I left undergrad feeling empowered and excited about a future studying science.*

*Four years ago, I left a graduate program at Princeton University feeling broken by the deceptively pristine atmosphere of the ivory tower. Uncertain about any future studying science.*

*Three years ago, I joined the Andersen Lab, rightfully cautious but hopeful about my future studying science.*

*Three years later, I write this dissertation as a testament the scientific journey that reinvigorated my love for science and shaped me into a stronger writer, researcher and person.*

---

It takes a village to raise a Ph.D candidate and this journey would not have been possible without the support I've received from my village.

Were it not for my undergraduate academic advisors, Drs. Jim Lissemore and Chrystal Bruce, and my undergraduate research advisor, Dr. Michael Nichols, I would never have considered a career in research. I am forever grateful to the patience they showed me as a young scientist, the incessant curiosity they imparted upon me, and the respect they taught me I deserved. They each, in their own way, showed me the very best of academia and for that I am thankful.

Above all, I would like to acknowledge my advisor, Dr. Erik Andersen. Throughout my four years as a graduate student at Northwestern University, Erik has not only been a constant source of support, but also has gone above and beyond to ensure I had what I needed to be

successful. I believe, without a doubt in my mind, that Erik has found his calling as a scientist, teacher, and mentor. His passion for science is infectious, his commitment to research inspiring, and his sheer determination unmatched. I look forward to seeing where his career takes him.

My time in graduate school would not have been the same without the many colleagues and friends I've made along the way. First, I would like to acknowledge Dr. Gaotian Zhang who helped me find my place in the laboratory during my rotation. Gaotian's secret power is his enduring positivity and confidence. I am so fortunate to have worked with him and witness his experimental skill firsthand – I am convinced there is no experiment too difficult for Gaotian. I will carry his calmness with me through all future stressful situations.

Two other people I would like to mention are Drs. Tim Crombie and Sam Widmayer. I had the pleasure of building the easyXpress package with these incredibly talented postdocs. They have both been a constant source of positive feedback and constructive criticism. My progress over the last four years would not have been possible without our daily collaboration and conversation. Unknowingly, they have pushed and mentored me to be a better programmer and scientist. Above all, they give me hope for the future of academia – I look forward to seeing the product of their hard work and dedication to science.

Next, I would like to thank Dr. Katie Evans. Katie's extensive knowledge and computational prowess intimidated me for the longest time. I soon found that we were incredibly similar and I have been fortunate enough to lean on her experience and advice in the last year as I explored careers in industry. As she makes her own transition to industry, I am certain she will continue to make great contributions to her field.

The data in this dissertation would not exist were it not for the help of two talented technicians, Nicole Roberto and Sophia Gibson. They both shared the burden of my 72-hour time course experiments, allowing me to get a few hours of shut eye while knowing my animals would be in good hands. I would also be remiss if I did not mention Dr. Robyn Tanny. Robyn's tireless management of all aspects of the laboratory, especially the CeNDR strain collection, is both impressive and inspiring. Her kindness and hospitality is infectious, and her birthday treats are delicious!

I am grateful to my committee members, Drs. John Marko, Marco Gallio, Shelby Blythe, Niall Mangan, and Erik Andersen, who have provided helpful comments and advice regarding my scientific and career journey. I have sincerely appreciated the support they've shown me as I've pursued a non-academic career path. In particular, I would like to thank Niall for the time I've spent as part of her group. The experience has shaped me into a better communicator and deepened my respect for the field of mathematics.

During my time in graduate school, I had the fortune of an incredibly supportive cohort. Loraina Stinson, Emily Pujadas, and Emily Czajkowski have shared in the joys and struggles of grad school from day one (\*cough\* Quant Bio \*cough\*). I am grateful for their continued support and friendship throughout this journey. In addition to my fellow scientists, I am thankful for my network of family and friends. My parents and three younger siblings remain a constant reminder of the privilege I have to pursue my passion for science. My close friend (and soon to be fellow Dr.!), Sara Martinko, was always a short call away when I needed a sympathetic ear to

vent about graduate school. And of course my cat, Spooky, who's unconditional love and soft purrs have brightened every morning and eased every evening.

Last, but certainly not least, my graduate school journey would not have been possible without the love and support of my partner, Austin McIlvaine. Austin has been my rock for the last seven years, keeping me sane amidst the trials of graduate school admissions, candidacy exams, and now this – the doctoral dissertation and defense. I am grateful for the countless hours spent listening to me talk on-and-on about science, time spent driving me to campus for late night experiments, and many many words of encouragement. All of this would not have been possible without him.

## Dedication

For my family, and the little black girl who always aspired to be great.

*Though she be but little, she is fierce.*  
- Shakespeare

# Table of Contents

<b>List of Tables</b>	<b>13</b>
<b>List of Figures</b>	<b>14</b>
<b>1. INTRODUCTION</b>	<b>17</b>
1.1. PHENOTYPIC VARIATION	18
1.1.1. From genotype to phenotype	19
1.2. QUANTITATIVE TRAIT LOCI MAPPING	20
1.3. <i>CAENORHABDITIS ELEGANS</i>	21
1.3.1. As a model organism	22
1.3.2. Post-embryonic development	24
1.3.3. Growth regulation	25
1.3.4. Genetic pathways regulate body size	25
1.3.5. Environmental factors modulate body size	26
1.3.6. Cuticle structure influences body size	27
1.3.7. Natural genetic variation	28
1.3.8. Linkage mapping	29
<b>2. A potential role for cuticle elasticity in developmental growth control</b>	<b>31</b>
2.1. ABSTRACT	32
2.2. INTRODUCTION	32
2.3. MATERIALS AND METHODS	36
2.3.1. Worm culture	36
2.3.2. Bacterial food	36
2.3.3. Growth of the animals	37
2.3.4. High-throughput measurements of body size and fluorescence	38
2.3.5. Image processing	38
2.3.6. Data processing	39

		10
2.3.7.	Molt analysis	40
2.3.8.	Comparison of model fits	41
2.3.9.	Stretcher model analysis	41
2.4.	RESULTS	42
2.4.1.	Quantitative measurements of <i>C. elegans</i> growth	42
2.4.2.	Fluorescence provides a quantitative measurement of animal feeding behavior and developmental progression	46
2.4.3.	Changes in <i>C. elegans</i> body shape occur at larval-stage transitions	49
2.4.4.	Modeling <i>C. elegans</i> cuticle stretch dynamics	52
2.4.5.	Sensing of cuticle stretch as a trigger for larval-stage transitions	52
2.4.6.	Discontinuities in animal growth rate driven by limits on cuticle stretch	57
2.5.	DISCUSSION	60
2.5.1.	A potential role for cuticle stretch in the timing of larval-stage transitions	60
2.5.2.	Development comprises complex interactions of growth regulation across diverse scales	63
2.6.	CONTRIBUTIONS	63
2.7.	SUPPLEMENT	64
<b>3.</b>	<b>Multiple genetic loci underlie differences in <i>C. elegans</i> growth</b>	<b>75</b>
3.1.	ABSTRACT	76
3.2.	INTRODUCTION	76
3.3.	MATERIALS AND METHODS	78
3.3.1.	Strains	78
3.3.2.	High-throughput growth assay	79
3.3.3.	High-throughput fitness assay for linkage mapping	79
3.3.4.	Linkage mapping	80
3.3.5.	Generation of chromosome substitution strains (CSSs) and near-isogenic lines (NILs)	81
3.3.6.	Statistical analysis of CSS and NIL results	81

		11
<b>3.4. RESULTS</b>		82
3.4.1. Growth dynamics of the N2 and CB4856 strains		82
3.4.2. Identification of QTL underlying variation in growth		85
3.4.3. Validation of loci associated with differences in animal size		87
3.4.4. Identification of candidate genes in the chromosome X QTL		89
<b>3.5. DISCUSSION</b>		91
3.5.1. A complex genetic architecture underlies differences in body size		91
3.5.2. Candidate genes for variation in body size		93
3.5.3. Comparison with previous QTL studies of <i>C. elegans</i> growth		93
<b>3.6. CONTRIBUTIONS</b>		94
<b>3.7. SUPPLEMENT</b>		95
<b>4. easyXpress: An R package to analyze and visualize <i>C. elegans</i> microscopy data generated using CellProfiler</b>	<b>106</b>	
4.1. ABSTRACT		107
4.2. INTRODUCTION		107
4.3. MATERIALS AND METHODS		109
4.3.1. Preparation of animals for imaging		109
4.3.2. Imaging		109
4.3.3. Paraquat dose response		110
4.3.4. Naming conventions		110
4.3.5. Data availability		111
4.4. RESULTS		111
4.4.1. Design and implementation		111
4.4.2. Data import and model assignment		112
4.4.3. Data pruning and summarization		114
4.4.4. Visualization		115
4.4.5. Application to <i>C. elegans</i> growth data		120
4.5. CONCLUSIONS		121
4.6. CONTRIBUTIONS		121

	12
4.7. SUPPLEMENT	122
<b>5. DISCUSSION</b>	<b>126</b>
5.1. SUMMARY	127
5.2. THE FUTURE OF DEVELOPMENTAL GROWTH STUDIES	128
5.2.1. The <i>C. elegans</i> cuticle and its role in developmental timing	128
5.2.2. Identifying genetic factors that contribute to variation in growth and body size	131
5.2.3. Alternatives to classical linkage mapping	132
5.2.4. Phenotypic plasticity and genotype-by-environment interaction	134
<b>References</b>	<b>136</b>
<b>Appendix</b>	<b>151</b>
A. Growth dynamics of <i>C. elegans</i> cuticle mutants	152
B. Food allocation during <i>C. elegans</i> growth	159
<b>Vita</b>	<b>172</b>

## List of Tables

### **Chapter 2:**

Table S2-1. Results of analysis of variance models fit to COPAS BIOSORT DATA	73
Table S2-2. Model fit criteria used to assess candidate growth models	74

### **Chapter 3:**

Table 3-1. Body size QTL	87
Table 3-2. Genes in QTL interval for chromosome X	91

### **Chapter 4:**

Table 4-1. Suggested naming conventions for CellProfiler metadata	111
---	-----

# List of Figures

## Chapter 1:

Figure 1-1. Photomicrographs of <i>C. elegans</i>	23
Figure 1-2. Illustration of the <i>C. elegans</i> life cycle	24
Figure 1-3. Illustration of the <i>C. elegans</i> cuticle	28
Figure 1-4. Overview of causal genes identified through QTL mapping approaches	30

## Chapter 2:

Figure 2-1. An overview of the quantitative growth assay	43
Figure 2-2. Quantitative measurements of animal size.	45
Figure 2-3. Fluorescence dynamics outline larval stages.	48
Figure 2-4. Changes in body shape occur during larval-stage transitions.	51
Figure 2-5. Cuticle stretch dynamics guide larval-stage transitions.	54
Figure 2-6. Experimental data are consistent with a length threshold in cuticle stretch.	59
Figure S2-1. Raw measurements of animal size.	64
Figure S2-2. Correlation analysis of body size measurements.	65
Figure S2-3. Mixture modeling of COPAS BIOSORT data was used to prune data.	66
Figure S2-4. Fluorescence measurements normalized by body size.	67
Figure S2-5. Cuticles identified during periods of decreased feeding.	68
Figure S2-6. Volume growth data fit with linear, exponential, and cubic models.	69
Figure S2-7. Density plots of population size dynamics across all larval transitions.	70

	15
Figure S2-8. Animals in all replicates, measured from images.	71
Figure S2-9. Stretcher model analysis of replicate 2 COPAS BIOSORT data for different stage thresholding.	72

### **Chapter 3:**

Figure 3-1. Quantitative measurements of growth for N2 and CB4856 animals	84
Figure 3-2. Linkage mapping identifies three QTL associated with body size	86
Figure 3-3. NILs validated the chromosome X QTL	88
Figure S3-1. Raw measurements of animal size	95
Figure S3-2. Mixture modeling of COPAS BIOSORT data was used to prune data	96
Figure S3-3. Pruned measurements of animal size	97
Figure S3-4. Comparison of means across developmental time points	98
Figure S3-5. Two dimensional genome scan for mean animal length	99
Figure S3-6. Two dimensional genome scan for mean animal width	100
Figure S3-7. Validating the chromosome IV length-associated QTL	101
Figure S3-8. Validating the chromosome V width-associated QTL	102

### **Chapter 4:**

Figure 4-1. easyXpress workflow.	112
Figure 4-2. Example plot generated by <i>viewPlate()</i>	116
Figure 4-3. Example plots generated by <i>viewWell()</i>	117
Figure 4-4. Example plots generated by <i>viewDose()</i>	119
Figure 4-5. easyXpress applied to <i>C. elegans</i> growth data	120

16  
Figure S4-1. Multiple model objects assigned to a single primary object 122

Figure S4-2. Uneven illumination along well edge hinders CellProfiler's ability to segment animals from background 124

Figure S4-3. Raw data from *C. elegans* growth experiment displayed by *viewWell()* 125

**Chapter 5:**

Figure 5-1. Schematic representation of reaction norms, illustrating forms of plasticity 135

**Appendix A:**

Figure A-1: Quantitative assessment of *C. elegans* larval growth 153

**Appendix B:**

Figure B-1: Visualization and analysis of food intake and utilization models 160

## 1

# INTRODUCTION

A fundamental objective in biology is to understand what makes individuals, populations, and species different from one another. The observable characteristics of an organism, such as morphology, development, and behavior, make up its *phenotype*. By contrast, *genotype* describes the inheritable genetic material that is responsible for determining a given trait [1]. The relationship between genotype and phenotype is often complex and multifaceted. Some traits, like color blindness, can be linked to single gene mutations [2]. Most traits, however, are influenced by a combination of genetic and environmental factors. Investigating the variation in complex traits, provides deeper insight into phenotypic diversity and its causes.

This introduction starts by providing an overview of phenotypic heterogeneity, focusing on the genetic variation that can influence its emergence (**Section 1.1**). I discuss how quantitative genetic mapping is used to study underlying variation in complex traits (**Section 1.2**). Next, *Caenorhabditis elegans* is presented as an appropriate model system for the study of phenotypic variation (**Section 1.3**). I discuss post-embryonic development and examine existing knowledge of how the processes of growth and body size are controlled in the nematode. Finally, I review principles of genetic mapping in *C. elegans*.

## 1.1. PHENOTYPIC VARIATION

The field of quantitative genetics classically postulates that variation among phenotypes arises from two major sources: genetics and environment [3,4]. Expressed as an equation, this can be written as,

$$V_p = V_G + V_E \quad (1-1)$$

where  $V_p$  represents variance in a phenotype,  $V_G$  is genetic variation, and  $V_E$  is variation in the environment. The genetic component of variation arises from genetic variants within a population [5]. Conversely, environmental variation arises when individuals of the same genotype experience changes in environmental conditions (*i.e.* food availability, temperature) and respond with multiple phenotypic differences [6]. Notably, genetic and environmental components of phenotypic variation are rarely independent; many traits exhibit a genetic component to plasticity. In this case, the observable change in a population's phenotype due to environmental differences is influenced by genetic variation. This form of variation indicates the presence of genotype-by-environment (GxE) interaction. Given this interaction term, total phenotypic variation is more appropriately expressed as,

$$V_p = V_G + V_E + V_{GxE} \quad (1-2)$$

where the contribution of genetic variance to plasticity is captured in the term,  $V_{GxE}$  [7].

Another important source of phenotype variation is stochastic variation. For a given environmental condition, random noise exists in the expression of the true phenotype. This uncontrolled variation manifests as differences among individuals of the same genotype raised in a common environment [8]. Two factors can drive this level of variation: randomness in the

developmental process and small fluctuations in the environment. Stochastic fluctuations in both the developmental process and an individual's experienced environment contributes to total variance in a population's phenotypic response [9]. Accordingly, the total phenotypic variation of a population should now be represented as,

$$V_p = V_G + V_E + V_{GxE} + V_{Noise} \quad (1-3)$$

where  $V_{Noise}$  represents intra-individual variation in response to an intrinsic factor. To address the contribution of this variance, both the environment and background genotype must be constrained. Measuring the variation of traits within an isogenic population increases the sampled phenotypic space and allows for the isolation of incoming sources of variation. Differences in genetically identical individuals reflects the intrinsic variance present [9]. For species in which genetic clones are available or selfing is possible, such as *C. elegans*, this is easily accomplished. Nonetheless, the interaction of the various sources of variation ( $V_G, V_E, V_{GxE}, V_{Noise}$ ), combined with the sheer magnitude of the number of phenotypic traits, makes the study of the phenotypic variation in a population strikingly complex.

### **1.1.1. From genotype to phenotype**

In 1865, an Austrian monk by the name of Gregor Mendel published work investigating the connection between genotype and phenotype. For eight years he studied the pea plants that grew in the garden of his monastery, observing patterns in their color, texture, and height from one generation to the next [10]. Mendel's experiments laid the foundation for our understanding of genetic inheritance and the correlation between underlying genetic makeup and physically

visible traits. More than 150 years later, scientists continue to study how genetic variation contributes to phenotypic differences.

Recent advances in sequencing technologies have accelerated the generation of vast amounts of genotype data [11–13] and enabled scientists to identify genetic perturbations causally linked to disease traits. However, discerning the relationship between genotype and phenotype remains a challenge. Although there is a strong correlation between the two, phenotype is not solely determined by the genotype of a few trait-associated genes but rather is the cumulative result of interactions among genetic and non-genetic factors over time [14]. This is particularly apparent in the context of disease. The clinical symptoms we observe in affected individuals are the result of interactions between their genetic background, mutations, and the environment over time. Disease phenotypes may follow Mendelian inheritance patterns and thus be predominantly driven by a single gene mutation, but oftentimes involve multiple genes along with additional environmental factors [15,16]. As the complexity of a trait increases, it becomes increasingly more difficult to resolve the genotype to phenotype relationship. However, the identification of genetic variants underlying phenotypes is key to understanding how genotype influences complex traits associated with growth, health, and disease.

## 1.2. QUANTITATIVE TRAIT LOCI MAPPING

Quantitative trait loci (QTL) mapping is a method for identifying the genetic basis of complex traits. With this technique, regions of the genome that contain loci that affect a measurable trait can be identified. Mapping QTL allows scientists to study the underlying genetic architecture of

complex traits. Many studies have mapped QTL that affect human disease and have uncovered new loci providing insights into the biology of disease [17–19].

There are two main approaches to identify QTL: linkage mapping and genome-wide association (GWA) mapping [20]. Both techniques rely on statistical methods to correlate genetic variation with phenotypic variation. Simply, these methods determine whether the genotype at a given position in the genome explains phenotypic differences among individuals. Linkage mapping uses populations of recombinants generated from a cross between phenotypically and genotypically diverse strains. In contrast, GWA mapping leverages the existing natural genetic diversity among genetically distinct individuals [20].

Oftentimes, QTL span large genomic regions containing many genes. To narrow the search space for potential genes contributing to phenotypic differences, the QTL region can be isolated in a clean genetic background with near-isogenic lines (NILs) [22,23]. The phenotypes of these NILs can then be used to refine the QTL interval by determining whether the introgressed genomic region recapitulates the QTL effect. Ultimately, in organisms with well-annotated genomes, causality can be tested with genome editing tools [24–26].

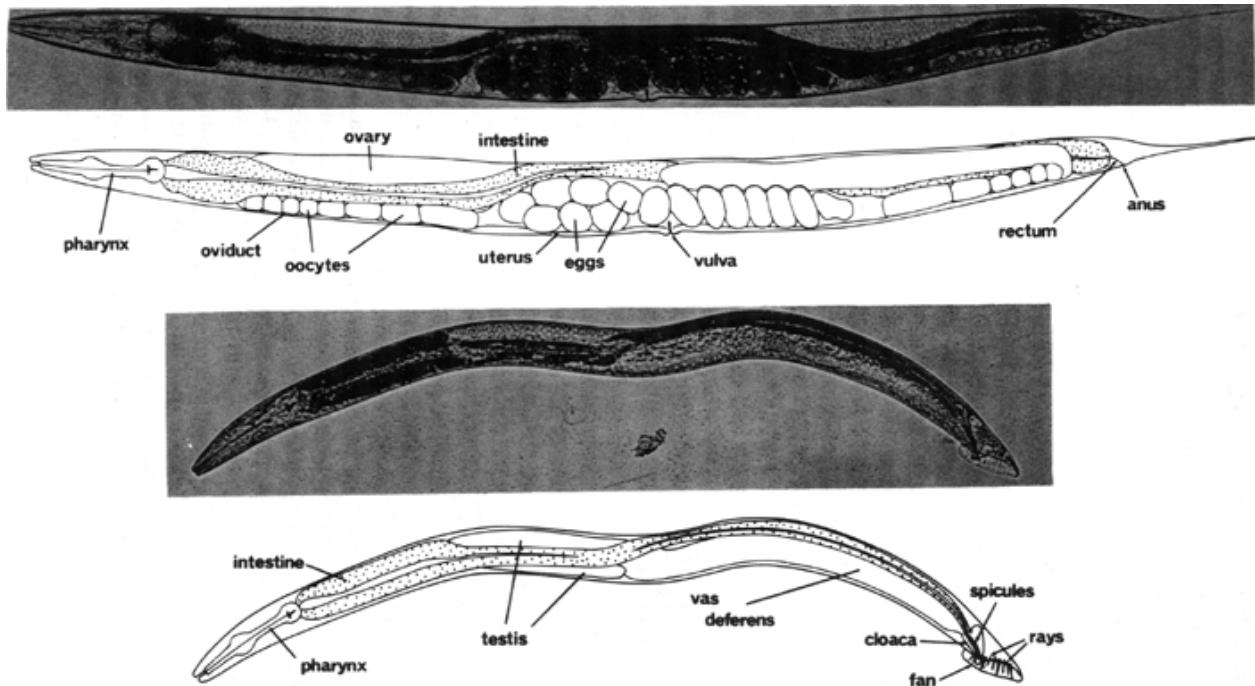
### **1.3. *CAENORHABDITIS ELEGANS***

French biologist, Emile Maupas, first described isolating *Rhabditis elegans* (now *Caenorhabditis elegans*) near the city of Algiers (Algeria) in 1900 [27]. He conducted some of the first experiments on the nematode, elaborately detailing its anatomy and development. Several decades later, *C. elegans* was again isolated for study by Victor Nigon and Ellsworth Dougherty. These scientists worked to refine culture conditions and experimental methods for propagating

animals in the laboratory [28]. However, it wasn't until Sydney Brenner published work on the nematode in the 1970s that *C. elegans* became an established model organism [29]. Since then, research on the roundworm has expanded to explore diverse areas of modern biology and is actively studied in over a thousand laboratories worldwide [30].

### 1.3.1. As a model organism

In 1963, well-known bacterial geneticist, Sydney Brenner, decided to turn his research efforts away from bacteria and towards an animal system. He sought an organism that could serve as a genetic model to probe how mutations disrupt eukaryotic processes. After considering a number of organisms, he ultimately settled on *C. elegans* for several reasons [31]. This free-living bacterivore grows rapidly on agar plates or in liquid culture, reaching an adult length of approximately 1.5 mm in three days [32]. Its small size and transparent body make the nematode an ideal organism for microscopic observation (**Figure 1-1**). As hermaphrodites, *C. elegans* can produce up to 300 genetically identical progeny in a lifetime. These individuals can be maintained nearly indefinitely through cryopreservation, providing a method for long-term storage and safeguarding against the accumulation of mutations [32,33].

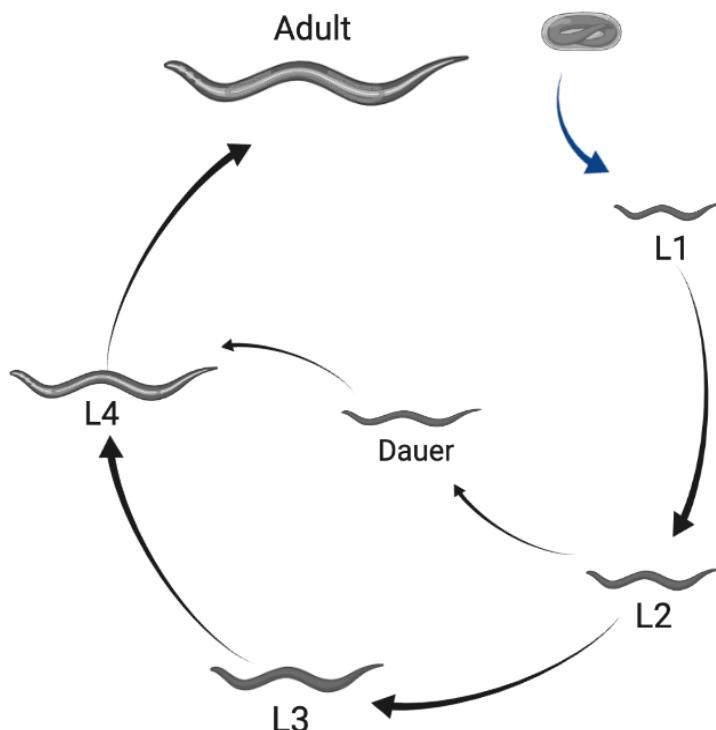


**Figure 1-1. Photomicrographs of *C. elegans*.** Major anatomical features of the *C. elegans* adult hermaphrodite (top) and male (bottom) under bright-field illumination. Bar, 20  $\mu\text{m}$ . (From [32]).

Since Brenner's first paper establishing *C. elegans* as a model organism, the animal has become a powerful system for genetic and molecular analysis. Notably, by the 1980s, scientists had documented the entire cell lineage of the nematode, from embryo to adult [34–37]. A decade later, *C. elegans* became the first multicellular organism to have its genome sequenced [38]. This 100 Mb genome, though relatively small, comprises six chromosomes and approximately 20,000 protein-coding genes that are homologous to protein-coding genes found in other organisms, including humans [39,40].

### 1.3.2. Post-embryonic development

*C. elegans* post-embryonic development consists of four larval stages (L1 - L4) that are punctuated by distinct developmental events called molts where animals shed an exoskeleton called the cuticle (**Figure 1-2**). Each larval stage begins with a period of active growth, and is followed by a period of quiescence (lethargus) where feeding and movement is temporarily halted [41]. During this time, a new cuticle is synthesized under the old. A molt is completed when animals shed the old cuticle (ecdysis) and transition to the next stage [42].



**Figure 1-2. Illustration of the *C. elegans* life cycle.** Embryos hatch to the first larval stage and proceed through four larval stages (L1 - L4) before becoming mature adults. Transitions between larval stages are marked by molt events where the animal's outer collagen-rich cuticle is shed. An

alternate developmental stage, dauer, occurs in response to unfavorable environmental conditions. Created with Biorender.com.

The rate of larval development in *C. elegans* is strongly dependent on environmental conditions. Dietary restriction has been shown to extend developmental progression up to ten-fold [43]. In extreme conditions, such as absence of food or high population density, *C. elegans* are able to halt development entirely by entering dauer, a long-lived morphologically distinct stage optimized for survival [44,45]. The decision to enter dauer is one of many checkpoints regulating developmental progression. Such checkpoints often occur early in the molting process and have been identified in every larval stage [46,47]. The well-coordinated timing of development is essential to the completion of stage-specific processes including cell-differentiation and tissue morphogenesis [48,49], and underscores the importance of growth regulation during *C. elegans* larval development.

### 1.3.3. Growth regulation

Organismal growth is regulated on a genetic level, as changes in gene expression patterns and signaling dictate much of development. However, environmental conditions (e.g. nutrient availability and temperature) also have strong impacts on growth. Therefore, it is important to account for how the processes of growth and development are controlled genetically, and how they can be influenced environmentally.

#### 1.3.3.1. Genetic pathways regulate body size

The control of *C. elegans* developmental progression has been studied extensively on the genetic level. The very first body size mutants were isolated in genetic screens by Sydney Brenner [33].

Subsequent screens for small (Sma) and long (Lon) mutants revealed a major regulator of growth and body size: DBL-1 [50–53]. DBL-1 is a ligand in the transforming growth factor- $\beta$  (TGF- $\beta$ ) superfamily, and homologous to members of the mammalian bone morphogenetic protein (BMP) family [54]. Secreted by neurons and body wall muscle, DBL-1 is necessary for body size regulation, and other developmental and homeostatic processes [50]. Mutants deficient in *dbl-1* have a marked reduction (40%) in both length and width compared to wild type throughout post-embryonic development [50,51]. In contrast, increased DBL-1 signaling causes animals to be longer (25%) than wild type [50,55].

### 1.3.3.2. Environmental factors modulate body size

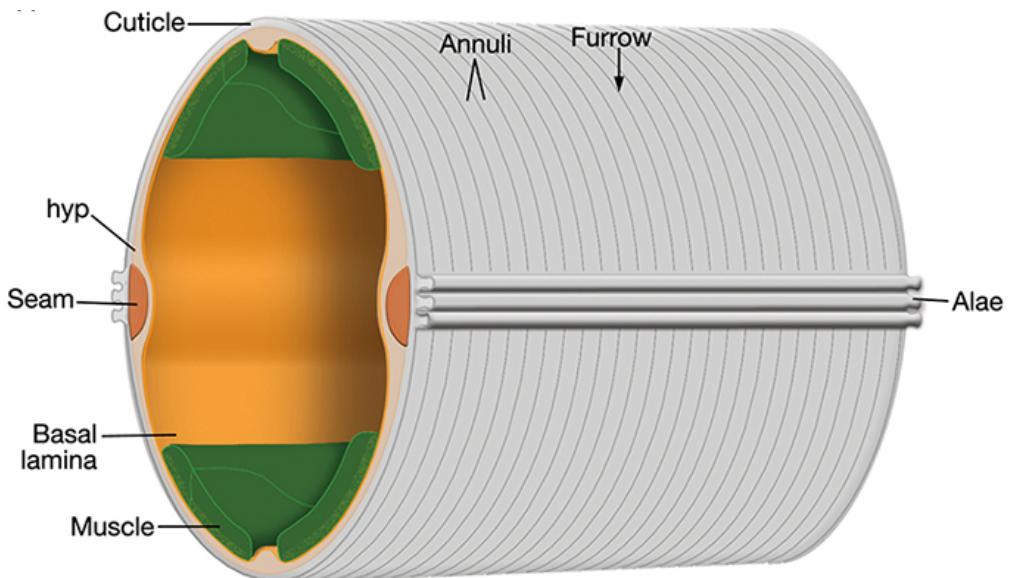
As previously mentioned, nutrient restriction is known to decrease *C. elegans* growth rate or, when extreme, induce complete developmental arrest [43,45,56]. Body size is also affected by the quality of food. As bacterivores, *C. elegans* obtain their nutrients primarily from bacteria. Historically, a strain of *Escherichia coli*, OP50, was chosen as the standard laboratory diet [57]. While most experimental studies use OP50 as the primary food source, the number of bacterial diets used to propagate *C. elegans* has expanded in recent years. Although these diets support development, research has identified several phenotypic effects that are altered. For example, when fed *Comamonas* DA1877 or *E. coli* HB101, animals develop faster and grow larger than those fed *E. coli* OP50 [58–60].

Aside from diet, other environmental stimuli can also regulate *C. elegans* body size. Standard wild-type animals used in most experiments grow larger when cultured at lower temperatures [61]. Additionally, research has shown that the *C. elegans* nervous system plays a key role in the regulation of body size by environmental stimuli. When raised in isolation, *C.*

*elegans* exhibit decreased growth rates and reach smaller adult sizes compared to animals raised in groups [62]. This result, however, is reversed with mechanosensory stimulation during development, demonstrating a role for mechanosensory neurons in body size regulation. Similarly, a number of mutations affecting ciliated sensory neurons through which *C. elegans* perceive environmental stimuli are shown to be required for animals to grow to the proper size [63,64]. This suggests that perception of environmental stimuli is crucial for the regulation of body size in *C. elegans*.

#### 1.3.3.3. Cuticle structure influences body size

The *C. elegans* cuticle is a complex, multi-layered structure primarily composed of collagens. The major surface structures of the cuticle include the circumferential ridges (annuli) and troughs (furrows) [65] (**Figure 1-3**). As animals progress through their life-cycle, the structure and thickness of the cuticle changes but its role in the maintenance of body morphology and integrity remains. To date, 21 cuticle collagen mutants have been identified that cause a range of body morphology defects [65]. Several of these mutants (*dpy-2, 3, 7, 8, 10*) lack annuli and exhibit a disproportionate reduction in body size [66]. These *dpy* animals are noticeably shorter in length and wider than wild type, clearly demonstrating the importance of the physical structure of the cuticle on growth.



**Figure 1-3. Illustration of the *C. elegans* cuticle.** The hypodermis (hyp) comprises hypodermal cells which secrete the cuticle (grey). Circumferential rings called annuli run the length of the cuticle. The valleys between annuli are furrows, which form immediately above circumferential bundles of actin (not shown) present in hypodermal cells during embryogenesis and molting. Seam cells (orange) that run along the left and right sides. (From [67]).

#### 1.3.4. Natural genetic variation

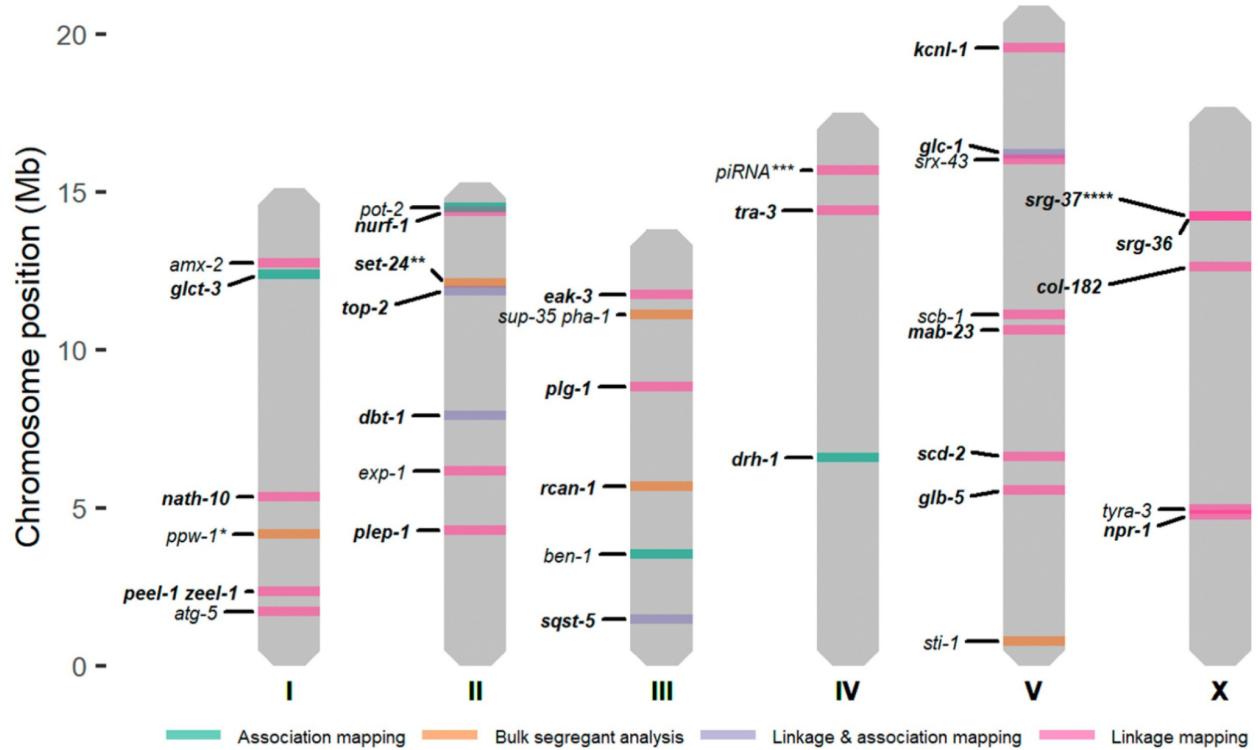
*C. elegans* present a scalable and tractable animal model to connect phenotypic differences to genetic variants. *C. elegans* are easily propagated as clonal cultures, eliminating genetic variance as a contributor to phenotypic variance in experimental studies. Additionally, hundreds of wild strains have been collected worldwide, providing access to extensive genetic diversity [68,69]. Two heavily studied strains are the laboratory-adapted strain, N2, and a wild isolate, CB4856. N2, derived from an isolate found in Bristol, UK in 1951, is recognized as the canonical laboratory wild-type strain with its genome serving as the reference [70]. Across the globe on the

Hawaiian island of Maui, CB4856 was isolated in 1972 [71]. Compared against the N2 reference, the CB4856 genome is highly divergent [72,73]. This extensive genetic diversity likely underlies the many phenotypic differences between the two strains, including social behavior, life-history traits (body size, fecundity, lifespan), and gene expression [74–79].

### 1.3.5. Linkage mapping

The N2 and CB4856 strains can be studied to better understand how genetic variation contributes to phenotypic differences. Crossing the N2 and CB4856 strains for many generations produces individuals with unique variants derived from each parental background. Performed at a large scale, these panels of recombinant inbred lines are a powerful tool for identifying genomic regions that are correlated with phenotypic variation.

Linkage mapping is a popular genetic mapping approach used to identify functional variants that contribute to phenotypic diversity. In *C. elegans*, linkage mapping leverages recombinant line panels, correlating genotype and phenotype to identify QTL. The Andersen Lab has developed an R package to facilitate such analyses [80]. By investigating the natural genetic variation underlying phenotypic differences, researchers are able to uncover genetic pathways involved in biological processes. Several studies have taken this approach to study the genetic underpinnings of complex traits like body size, social behavior, and drug response (**Figure 1-4**) [79,81–86].



**Figure 1-4. Overview of causal genes identified through QTL mapping approaches.** The colors represent the mapping technique(s) that were used for QTL mapping: bulk-segregant analysis (BSA) (orange); linkage mapping (pink); genome-wide association (GWA) mapping (green); linkage and GWA mapping (purple). (From [21])

**2**

# **A potential role for cuticle elasticity in developmental growth control**

In the spring quarter of 2019, I decided to complete my final first-year rotation in the Andersen Laboratory. At the time, I was coming off a rotation with Dr. Niall Mangan who was involved in a relatively new collaboration with Dr. Erik Andersen as part of the NSF-Simons Center for Quantitative Biology. The general goal of the project was to study the interplay between organismal growth and environmental variables such as diet and temperature. Needing a little extra experimental manpower, Erik recruited me to the team. Little did I know this project would become the heart of my thesis. What ensued was a series of daunting, exhausting, and incredibly rewarding time course experiments. The following chapter highlights the results of just one of these experiments, submitted as a first-author manuscript to Cells and Development in January 2022 [87].

## 2.1. ABSTRACT

Growth control establishes organism size, requiring mechanisms to sense and adjust growth during development. Studies of single cells revealed that size homeostasis uses distinct control methods. In multicellular organisms, mechanisms that regulate single cell growth must integrate control across organs and tissues during development to generate adult size and shape. We leveraged the roundworm *Caenorhabditis elegans* as a scalable and tractable model to collect precise growth measurements of thousands of individuals, measure feeding behavior, and quantify changes in animal size and shape during a densely sampled developmental time course. As animals transitioned from one developmental stage to the next, we observed changes in body aspect ratio while body volume remained constant. Then, we modeled a physical mechanism by which constraints on cuticle stretch could cause changes in *C. elegans* body shape. The model-predicted shape changes are consistent with those observed in the data. Theoretically, cuticle stretch could be sensed by the animal to initiate larval-stage transitions, providing a means for physical constraints to influence developmental timing and growth rate in *C. elegans*.

## 2.2. INTRODUCTION

Growth is a complex process fundamental to development. Individual cells and whole animals must reach an appropriate size to remain competitive in their environment. A larger body size conveys many selective advantages to an organism, including increased success in predation, defense against predation, success in mating, or successful intraspecific as well as interspecific competition. Offsetting these advantages, larger organisms require more food resources to grow, take longer to develop, and produce fewer offspring [88]. Therefore, it is critical for multicellular

organisms to effectively coordinate the growth of both individual cells and the whole body. Additionally, growth at both of these scales must be coupled with developmental progression to ensure the proper timing of irreversible developmental events.

In recent years, efforts have focused on understanding how organisms control growth to achieve size homeostasis [89–91]. Many of these studies are motivated by the decades-long debate about whether growth is linear or exponential; two separate models each having unique implications for size regulation. In a linear model with constant growth rate, smaller organisms must grow proportionally more than larger organisms to maintain size homeostasis. In this paradigm, organism size can be controlled simply by specifying growth duration. Subsequently, this method of growth control was named the “Timer” model [92,93]. *Caulobacter vibrioides*, a bacterium whose cells divide after a certain amount of time, has been reported to follow a “Timer” growth model [94]. Alternatively, organisms can instead monitor size and adjust duration of growth to reach an optimal size. This method of control is referred to as the “Sizer” model [95–97]. In an exponential model, growth rate is proportional to size. A time-based control mechanism alone would fail to maintain size homeostasis because larger organisms would grow proportionally more during a specified period of time. This difference in growth requires a size-based control mechanism to ensure that growth is halted once a maximum size is reached. Fission yeast, a rod-shaped eukaryote whose cells divide after growing to a certain size, have been reported to follow a “Sizer” growth regime [98]. Although “Timer” and “Sizer” are the most often proposed size-control models, other models have been suggested. The “Adder” model proposes that a fixed volume is added to a cell or organism during growth [99,100], whereas the “Folder” model specifies that an organism increases in volume by a fixed proportion

in order to control growth [101]. It is not trivial to determine which model most accurately describes growth of individual cells or whole organisms because quantitative measurements of growth must be collected at high precision and throughput under tightly controlled experimental conditions. In unicellular organisms, the development of high-throughput experimental techniques in combination with theoretical models have advanced the understanding of size control [102–106]. Progress has been slower for multicellular organisms because cell growth within tissues and tissue growth within organisms often progress at different rates, suggesting that they are likely not regulated in the same ways [107–109].

The nematode *Caenorhabditis elegans* presents both a scalable and tractable multicellular animal model to study growth control. With an adult body length of approximately 1 mm, hundreds of thousands of individuals are easily cultured in controlled laboratory conditions [31]. Moreover, *C. elegans* post-embryonic development is marked by several molts that provide clear developmental milestones [65]. Each molt is initiated by a period of quiescence (lethargus) and terminated once the animal successfully sheds its collagen-rich outer cuticle (ecdysis) [41]. Four molts separate the *C. elegans* life cycle into five distinct stages: four larval stages (L1-L4) and adult. The timing of these molts determines the completion of stage-specific development [48,49] and underscores the importance of growth regulation during *C. elegans* larval development.

A full description of an organism's development includes the assessment of how growth and body size are regulated. Initial studies of *C. elegans* development described whole-organism growth as a sigmoidal curve characterized by continuous larval growth in length that reaches saturation in adulthood [110]. These early studies hypothesized that molt events had little effect

on continuous growth as the *C. elegans* cuticle allowed for stretch during larval stages. Later work determined that larval progression was not continuous but rather piecewise in nature [111]. This study showed that *C. elegans* volumetric growth rate increased from stage to stage such that L1 animals had the slowest rate of growth and L4 animals had the fastest. This finding suggests that *C. elegans* have a mechanism for regulating growth rate, potentially at each molt. Next, researchers using single-animal imaging strategies observed that animals did not advance to the next developmental stage until a critical volume was reached [43]. This finding suggests that *C. elegans* growth follows a “Sizer” model with each molt decision controlled by a volume threshold and further implies that individual cells are able to communicate information about body size to precisely regulate growth. Most recently, live imaging and characterization of body volume heterogeneity revealed that with respect to the start of a larval stage, *C. elegans* relative change in volume within a stage is nearly invariant thereby preventing rapid divergence in volume between fast- and slow-growing animals [101]. A mechanism that maintains a constant relative change in volume within each larval stage relies on the coupling between growth rate and developmental timing. Notably, such coupling is consistent with recent observations of temporal scaling in *C. elegans* development where despite inter-individual variability in the absolute duration of a larval stage, relative timing of a stage is similar [112,113].

To understand *C. elegans* growth control at the whole-organism level, we used a combination of quantitative measurements and mathematical modeling. We performed a high-resolution longitudinal study of *C. elegans* larval progression and captured high-precision details about animal length, width, volume, and feeding dynamics. By investigating *C. elegans* feeding and growth in tandem for thousands of individual animals, we found decreases in

feeding behavior associated with each larval transition that were also correlated in time with changes in growth rate. We used our large-scale measurements of body shape to further analyze the periods of time surrounding each larval transition. At each molt, we observed simultaneous increases in length, decreases in width, and maintenance of volume. Based on the physical characteristics of the cuticle, we propose a “Stretcher” mechanism whereby physical constraints on cuticle stretch influence body shape. We find the Stretcher model-predicted shape changes are consistent with observed data. Animals may be able to detect when the cuticle reaches its maximum capacity for stretch providing a signal to initiate larval-stage transitions.

## 2.3. MATERIALS AND METHODS

### 2.3.1. Worm culture

The canonical laboratory strain N2 was obtained from the *C. elegans* Natural Diversity Resource [114]. Animals were cultured at 20°C on 6 cm plates of modified nematode growth media (NGMA), which contained 1% agar and 0.7% agarose seeded with *E. coli* OP50 bacteria [79].

### 2.3.2. Bacterial food

*E. coli* HB101 bacteria were prepared from cultures grown for 15 hours in Superbroth and then pelleted by centrifugation. HB101 bacteria were diluted to OD100 in K medium (51 mM NaCl, 32 mM KCl, 3 mM CaCl<sub>2</sub>, and 3 mM MgSO<sub>4</sub> in distilled water) and stored at -80°C. Bacteria were thawed and fed to animals at a concentration sufficient to sustain population growth from hatching to adulthood (OD20).

### 2.3.3. Growth of the animals

Populations of animals were propagated on NGMA plates for two generations without starvation. In the third generation, gravid adults were bleach-synchronized [115]. Embryos were resuspended in K medium, aliquoted into a 500 mL flask at a concentration of one embryo per  $\mu\text{L}$ , and allowed to hatch overnight. The following day, arrested L1s were fed HB101 bacteria at a final concentration of OD20 in a final flask volume of 100 mL K medium and HB101 food. Animals were grown for three days at 20°C with constant shaking. Following these three days, adult animals were bleach-synchronized once more and embryos were aliquoted to seven replicate 500 mL flasks at a concentration of one embryo per  $\mu\text{L}$  in 100 mL. The following morning, six flasks were fed HB101 bacterial food at a final concentration of OD20 in a final flask volume of 100 mL K medium and HB101 food. Two additional flasks were included to control for L1 animal size and possible clumping of bacterial food: one flask contained L1 larvae but did not have food added and one flask contained no larvae but the same concentration of HB101 bacteria as the six flasks containing L1 larvae. All replicate flasks were kept in an incubator at 20°C with shaking for the duration of the experiment. A small temperature gradient of 1.25°C was recorded in the shaking incubator with the highest temperature readings on the right side and lowest temperature readings on the left side. This slight variation in temperature contributed to variation in developmental rate among replicates based on position within the incubator (replicates were placed in numerical order with replicate 1 positioned on the far right side of the incubator).

#### **2.3.4. High-throughput measurements of body size and fluorescence**

Flasks were sampled each hour beginning one hour after feeding and continuing for 72 consecutive hours. At each hour, 500 µL was removed from each flask and transferred to a well of a deep 96-well plate. Each flask was sampled at each time point. Fluorescent polychromatic beads (Polysciences, 19507-5) with a 0.5 µm particle size were added to each well at a final concentration of  $3.64 \times 10^8$  beads/mL and incubated at 20°C for 10 minutes with shaking. Following the bead incubation, 30 µL from each well of the deep 96-well plate was aliquoted to a 96-well microtiter plate. The process was repeated 11 times to 11 separate wells of the same microtiter plate with pipetting to mix the well contents from the deep 96-well plate. Animals were then treated with sodium azide at a final concentration of 50 mM to paralyze and prevent defecation of the ingested beads. The 96-well plate was imaged with an ImageXpress Nano (Molecular Devices, SanJose, CA) using both 2x (Nikon MRD00025) and 10x (Nikon MRH00101) objectives. The ImageXpress Nano acquires brightfield images using a 4.7 megaPixel CMOS camera. Images are stored in 16-bit TIFF format. Finally, animals were scored using a large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston MA). The COPAS BIOSORT sheath flow rate was kept at a constant  $10.3 \pm 0.1$  mL per minute to reduce variability in length measurements.

#### **2.3.5. Image processing**

Manual measurements of animal size were performed using the free Java image-processing program ImageJ [116]. Well images for the six replicate flasks, excluding controls, were loaded into ImageJ software. Length was measured from head to tail, and width was measured at the widest point of the animal. Five animals were measured per well across thirty total wells for each

hour. Measurements were repeated for all 72 time points in the assay. Body length and width were used to estimate cross-sectional area (length\*width). This metric was used to describe animal area for the extent of the text. Volume was calculated from body length and width by approximating the animal as a cylinder. Pixels were converted to  $\mu\text{m}$  using a conversion factor of 3.2937  $\mu\text{m}/\text{pixel}$ .

### 2.3.6. Data processing

The COPAS BIOSORT was used to collect measurements of animal length (TOF), optical extinction (EXT), and fluorescence for every animal in each well. These traits measure properties of nematode development and, as such, increase as animals progress to adulthood [117]. Optical extinction measurements correspond to the amount of light absorbed over the full length of an animal as it passes through the instrument. An additional measurement (norm.EXT) can be calculated by normalizing optical extinction by length. The raw data collected were imported and processed using the *easysorter* R package [118].

The COPAS BIOSORT data were analyzed further using Gaussian finite mixture modeling as implemented in the *mclust* R package [119]. These probabilistic models assume that data are generated from a mixture of multivariate normal distributions and, therefore, can be used to classify unstructured data into meaningful groups. Specifically, the *mclust* package fits a mixture model to data and selects the optimal number of clusters using an expectation-maximization algorithm and Bayes Information Criteria. For model-based clustering, log transformed animal length (logTOF) and log transformed optical extinction (logEXT) were used as inputs for the *Mclust* function. Data from each hour of the experiment was analyzed by replicate and clusters that did not appear to include majority animal objects were identified and

removed as described previously [120]. This processing removed non-animal objects such as bacterial clumps, shed cuticles, and next generation larval animals from the time-course data.

We used a numpy polyfit regression of well-median data from the COPAS BIOSORT and image measurements to convert TOF and norm.EXT data to microns. Only the unit-corrected BIOSORT data were used for further analysis.

### 2.3.7. Molt analysis

Fluorescence data obtained from the COPAS BIOSORT was used as a proxy for feeding behavior to distinguish animals in a molt from growing animals. First, fluorescence was normalized by EXT to account for the ability of larger animals to consume more food and beads. Next, an analysis of variance statistical model was fit to the fluorescence data normalized by EXT to determine the amount of variance contributed by replicate and well (Table S1). A local kernel regression smoothing method was then applied to the residuals of the variance analysis using the *lokern* R package [121]. Residuals were used to address only the differences over time and ignore minor variation among replicates and wells. The local minima of the regression function were found by solving for where the first derivative of this function equaled zero. The time associated with each local minimum was used to represent the timing of each molt. Molts occurred at 14, 25, 36, and 48 hours.

To identify periods of time that contained a majority of growing animals, the inflection points of the regression function were calculated by solving for where the second derivative of the function equaled zero. Time points between inflection points that did not contain a local fluorescence minimum were considered as growth periods. These hours were 1-13, 17-22, 27-32, and 39-45 corresponding to L1, L2, L3, and L4 growth periods.

Each molt is initiated when animals enter quiescence: a behavioral state where animals cease active feeding. To classify individual animals as in a molt or growing, we set a quiescence threshold using fluorescence measurements at each local minimum. The fluorescence measurement at each local minimum was as follows: 0.07, 0.06, 0.06, 0.06. The average of these measurements (0.06) was used as the fluorescence threshold signifying quiescent behavior. Any individual animals that fell below this threshold fluorescence value were designated as in a molt and animals above this threshold value were classified as growing.

### **2.3.8. Comparison of model fits**

To determine the volume growth model, we fit linear, exponential, and cubic functions to data designated as growth periods for each larval stage. Both linear and nonlinear functions were fitted using least-squares regression. Akaike's information criterion (AIC) [122] and Bayesian information criterion (BIC) [123] were goodness of fit criteria used to evaluate candidate models. To assess the strength of evidence for each candidate model, we identified the model with the smallest AIC or BIC value and assessed the difference between this value and the AIC or BIC of the other two models. The magnitude of the difference was used to determine the level of support for each candidate model as previously described [124,125]. All model fits and analysis were performed using the *stats* R package.

### **2.3.9. Stretcher model analysis**

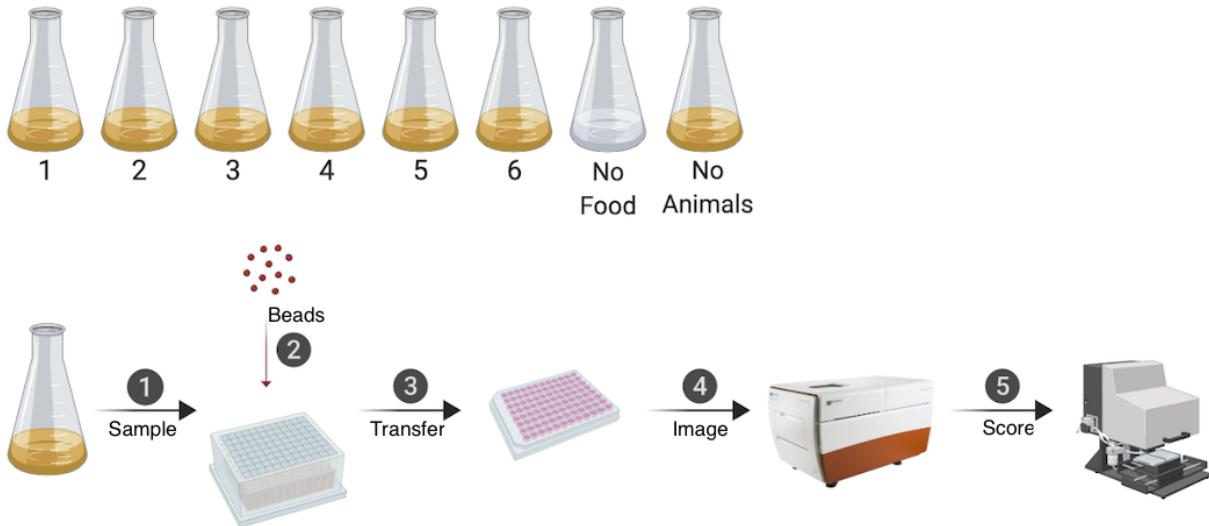
To analyze shape dynamics, length and width data from growth time periods were extracted from the full COPAS BIOSORT population data and analyzed from each replicate separately to avoid issues with replicate variability. For replicate 2, the hours defining growth periods were 1-13, 16.37-22.39, and 26.93-32.96; corresponding to L1, L2, and L3. Hours defining larval stages

were rounded as data was collected at exact hour increments. The L4 stage was excluded from the analysis because of the high variability within the population. This occurs because small changes in growth rate within each stage are amplified as the animals age. We applied a local kernel regression, *lokern* R package [126], to smooth the population dynamics of length and width. To calculate mean and standard deviation, the smoothed population measurements were bootstrapped using 2,000 samples with replacement. To determine cuticle properties throughout larval stages, we calculated the mean ratio of derivatives of regression width and length.

## 2.4. RESULTS

### 2.4.1. Quantitative measurements of *C. elegans* growth

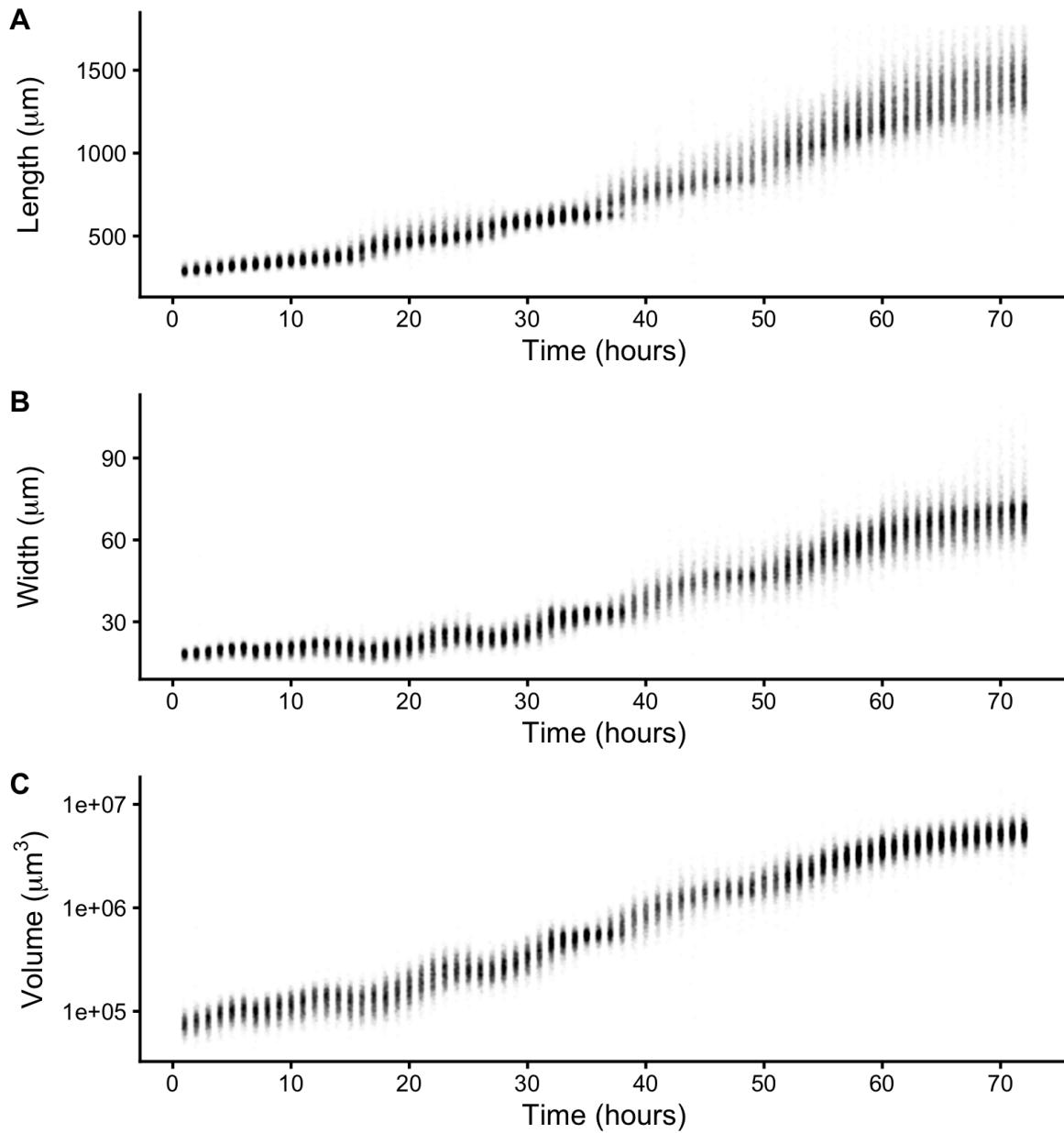
We have optimized a quantitative growth assay that reliably measures small changes in *C. elegans* body size throughout development (**Figure 2-1**). Our method provides both high-throughput and high-precision assessment of developmental growth. In brief, populations of 100,000 animals were cultured in flasks. We cultured six replicate populations of *C. elegans* for a total of 600,000 synchronized and growing animals. Every hour after feeding, a sample of the population from each flask (~300 animals/flask) was collected to measure animal length, width, and feeding rate. The ImageXpress system (Molecular Devices) was used to collect images of sampled animals. Feeding rate, examined using ingestion of fluorescent microspheres as a proxy, and body size were then measured using the COPAS BIOSORT (Union Biometrica). This platform enabled further analysis of life stage and body size, contributing added precision to our measurements.



**Figure 2-1. An overview of the quantitative growth assay.** Synchronized animals were cultured in flasks where six flasks contained replicate populations of nematodes, one flask had a population of unfed animals, and one flask only contained bacterial food. At each hour of the experiment, all eight flasks were sampled. In step 1, animals were transferred from each flask to a single well of a 96-well microtiter plate. In step 2, fluorescent beads were added to each well. Following a 10-minute incubation period, animals from each well of the deep-well plate were transferred to several wells of a 96-well microtiter plate for step 3. In step 4, animals in each well of the microtiter plate were imaged. In step 5, the same animals were measured using the COPAS BIOSORT. This process was repeated every hour after feeding for 72 consecutive hours (see Methods). Schematic of the experimental workflow was created with BioRender.com.

Two measurements of body size were collected from raw data taken from the COPAS BIOSORT: time of flight (TOF) and optical extinction (EXT) (**Figure S2-1**). Time of flight is a measurement of body length, and optical extinction corresponds to optical density, a measurement influenced by body length, thickness, and composition [117,127]. We investigated

whether optical extinction could be correlated to a different measure of body size using the collection of manual size measurements obtained from images (**see Methods**). We calculated the median length, width, area, and volume of animals in a subset of imaged wells from each hour of the experiment. We then compared these values to median measurements of animals in each well from the processed COPAS BIOSORT data. We found a strong correlation between manual measurements of animal length from the image analysis and TOF measurements from the COPAS BIOSORT (**Figure S2-2**). We also observed an equally strong correlation between manual measurements of animal area and EXT as well as animal width and EXT normalized by body length (norm.EXT). We then calculated animal volume using measurements from the COPAS BIOSORT by using a cylindrical approximation for *C. elegans* shape (**see Methods**). This result expanded the number of body size parameters that we were able to assess using the COPAS BIOSORT data, allowing us to investigate growth dynamics in length, width, and volume (**Figure 2-2A-C**). To disentangle nematode objects from non-animal objects (bacteria clumps, detritus, shed cuticles), we employed model-based clustering to remove unwanted objects and better examine growth of animals (**Figure S2-3**). Lastly, we converted unitless COPAS BIOSORT measurements into microns (**see Methods**).



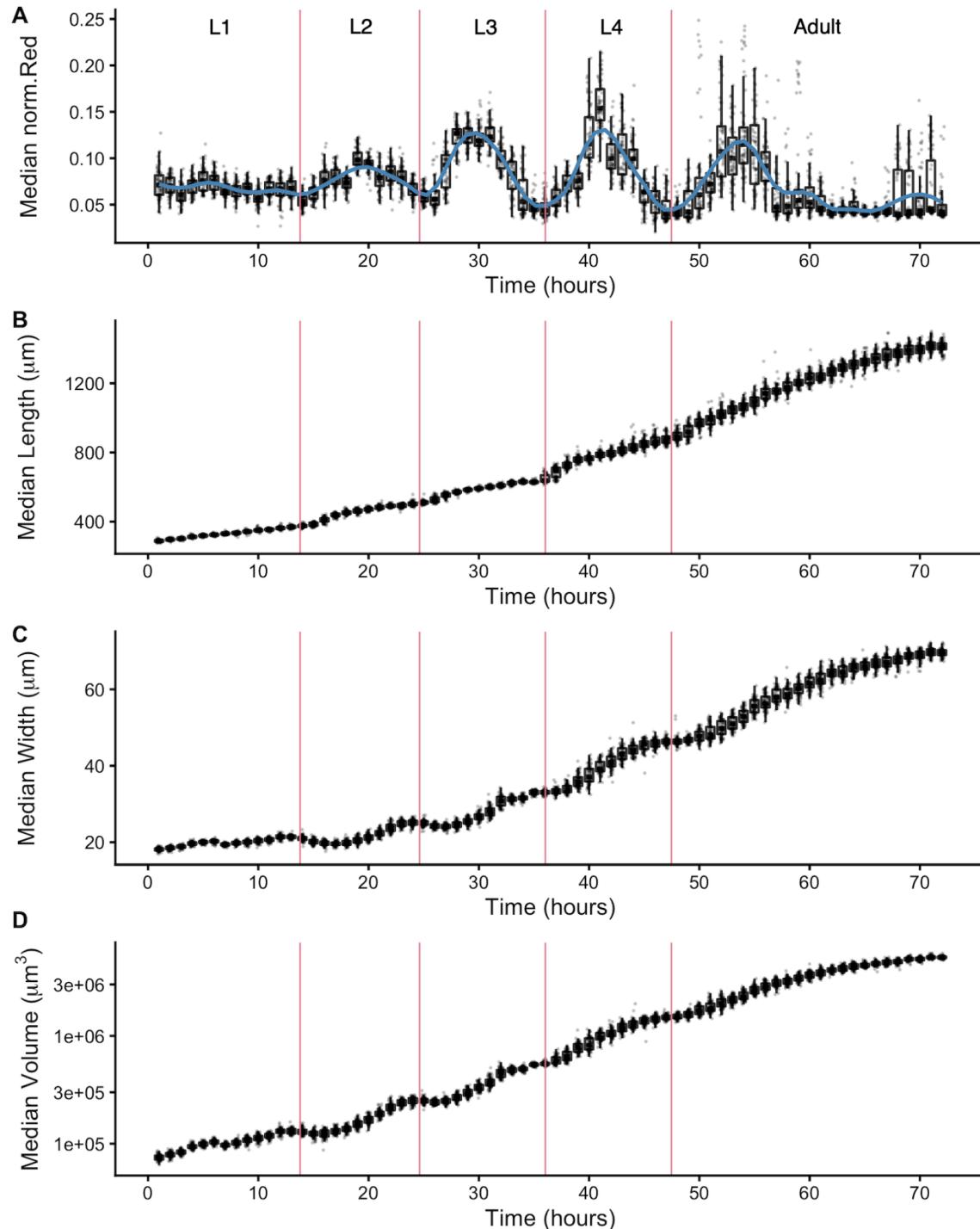
**Figure 2-2. Quantitative measurements of animal size.** COPAS BIOSORT data of animal length (A), width (B), and volume (C) after the removal of non-animal objects using model-based clustering methods (see Methods).

We report body length, width, and volume of animals at each hour of development from L1 to adult (**Figure S2-1** and **Figure 2-2**). Historically, growth of *C. elegans* has been shown as a sigmoidal curve where exponential growth during larval stages reaches a maximum rate in adulthood [110]. More recently, researchers have identified that growth curves are discontinuous during development punctuated by larval transitions [43,111]. Using our quantitative growth assay, we captured these small-scale discontinuities in larval growth rate as well as an apparent growth maximum during early adulthood. We noticed that all size variables (length, width, and volume) displayed these dynamics. Objects identified as animals appear to grow in size. However, in particular time windows during development, growth dynamics visibly shift, producing discontinuities in animal growth rate. With these data, we were able to further investigate *C. elegans* growth and size control.

#### **2.4.2. Fluorescence provides a quantitative measurement of animal feeding behavior and developmental progression**

In addition to body size and shape, the raw data from the quantitative growth assay included measurements of fluorescence for each animal. To readily assess the thousands of measurements acquired at each hour, we generated summary statistics of median well measurements (**Table S2-1**). With these summarized data, we investigated the relationship between feeding behavior and developmental stage. It is well established that temporary suspensions of *C. elegans* feeding occur during each molt [44,110]. As such, active feeding is frequently used to distinguish growing animals from individuals in a molt. We quantified feeding behavior by exposing animals to fluorescent beads the approximate size of bacteria and measuring fluorescence of animals [128]. Because larger animals are able to consume more food and therefore contain more

ingested food, we normalized fluorescence by animal area to account for increases in body size (**Figure S2-4**). The resulting fluorescence data showed a dynamic pattern (**Figure 2-3A**). At approximately 15 hours, fluorescence steadily increased to a peak before decreasing back to initial levels at approximately hour 27. This pattern, repeated three additional times, established clear time windows of local minimal fluorescence. These local minima represent periods of time where a large proportion of the population had reduced or ceased feeding and therefore suggests time windows where a majority of animals were likely not feeding because they were in a molt. We used a local kernel regression method to estimate a smooth curve and calculate the derivative to identify the time associated with each local minimum (**see Methods**). We then assessed images collected from the growth assay and found that periods of decreased feeding are concurrent with the presence of shed cuticles, supporting that animals are undergoing a molt during these periods of time (**Figure S2-5**). When we overlaid the timing of each local minimum on the population size data, we were able to outline the start and end of each larval stage (**Figure 2-3 B-D**). Notably, local minima occurred approximately every ten hours, consistent with well established observations of molt timing [110]. Furthermore, we observed a clear relationship between changes in feeding behavior and growth dynamics where decreases in feeding occurred simultaneously with discontinuous growth in length, width, and volume.



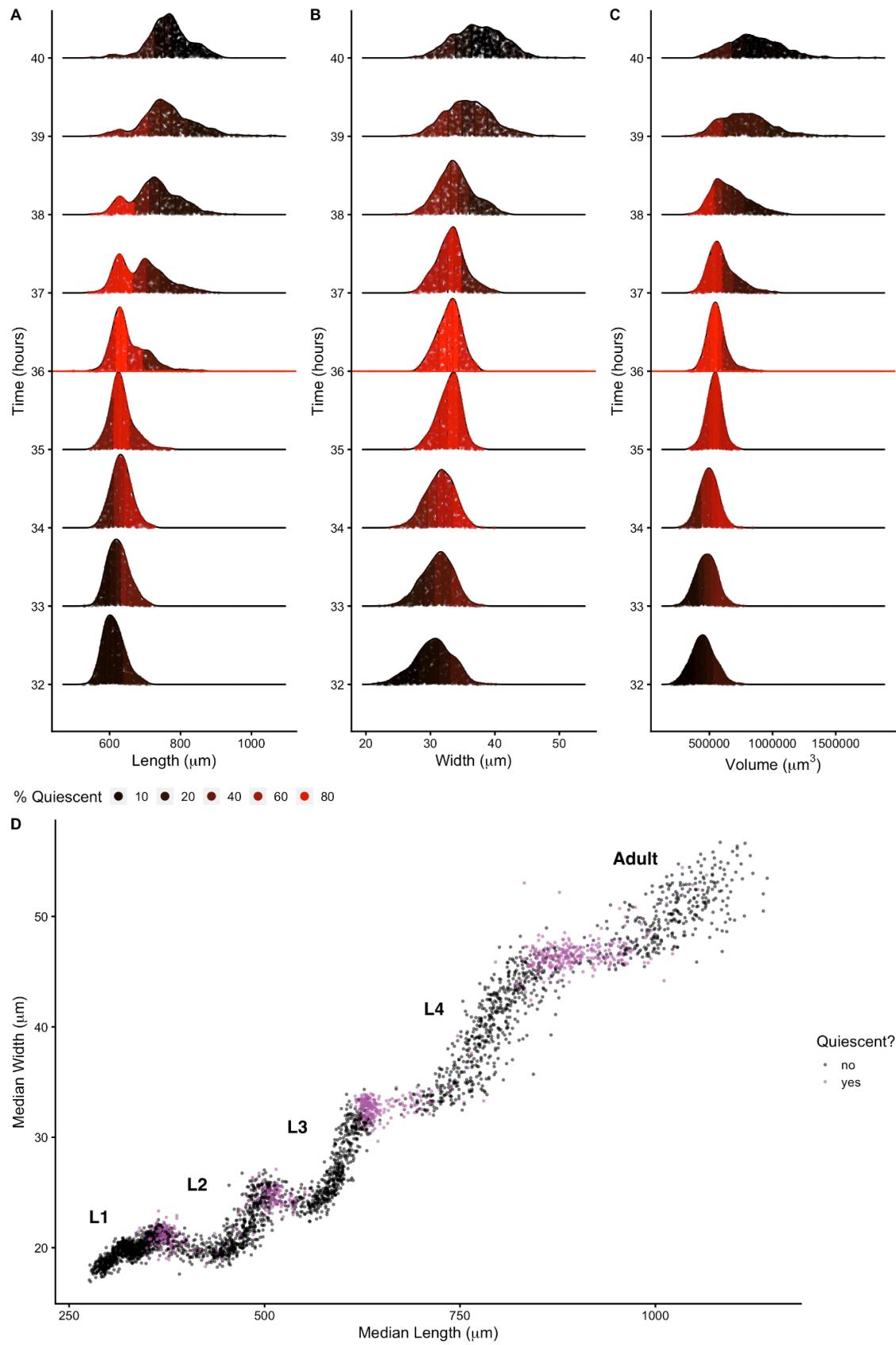
**Figure 2-3. Fluorescence dynamics outline larval stages.** (A) Median normalized red fluorescence (y-axis) over time (x-axis) is shown. The blue line represents the kernel regression

fit to the data. The red vertical lines correspond to the local minima of the regression and represent the transition between larval stages. Median length (B), median width (C), and median log volume (D) are shown with larval-stage transitions as well. Upper and lower bounds of the box plots correspond to the first and third quartiles. The upper and lower whiskers extend to 1.5 times the value of the interquartile range.

#### 2.4.3. Changes in *C. elegans* body shape occur at larval-stage transitions

Adult body size is ultimately determined by the coordination of developmental progression and rate of growth. To understand how *C. elegans* achieve final size, we must first examine how *C. elegans* grow. Quantitative studies of *C. elegans* growth frequently assess changes in length or volume over time; however, to fully characterize changes associated with growth, it is also important to consider the dynamics of width. Two general models were proposed for *C. elegans* growth in volume: linear and exponential [43,110,111]. Notably, these volume growth models require different dynamics in length and width. To achieve linear volume growth, length and width must increase at precise sublinear rates that together result in a linear increase in volume. If animal length and width increased at a constant linear rate, then volume would increase at a cubic rate. Alternatively, if both length and width grew exponentially, then volume would fit an exponential model. We sought to identify which model best described *C. elegans* growth behavior but were unable to consistently distinguish between linear, exponential, and cubic models using statistical information criterion because of the similarity in the shapes of the growth curves (**Figure S2-6** and **Table S2-2**). This result is not surprising because computational simulations have shown that increases in experimental noise, above 2% added noise, limit the correct identification of growth models [129].

Growth has important implications for how animals regulate size (length, width, or volume). Size homeostasis requires that growth rate and developmental rate are coordinated. Despite significant variation in individual growth rate, relative timing of *C. elegans* larval transitions is highly similar across individuals [112,113], implying a control mechanism to regulate developmental progression. Early work proposed a volume-based growth control model in *C. elegans* [43], although recent work suggests that volume homeostasis is achieved through a Folder mechanism [101]. To assess changes in body size and shape during a larval transition, we examined the dynamics of animal length, width, and volume in the hours before, during, and after each molt. We find that for each shape variable, larger animals enter molt first (**Figure 2-4**). We also observe differences in the distributions of lengths during a larval transition compared to widths and volumes. Measurements of animal width and volume remain unimodal throughout a molt, but length does not. As larger animals begin to exit the molt, a rapid increase in body length occurs that leads to the appearance of bimodality of lengths across the population. Importantly, volume remains constant while length increases and width decreases, indicating a change in body aspect ratio not size. Notably, the length increase occurs simultaneously with a decrease in width across the population (**Figure 2-4D**). These changes in the physical dimensions at each larval transition suggests that body shape may be involved in the control of *C. elegans* growth.



**Figure 2-4. Changes in body shape occur during larval-stage transitions.** Population density curves of length (A), width (B) and volume (C) for the hours surrounding the L3 - L4 larval transition (red horizontal line at 36 hours corresponds to the molt). Each distribution was divided into five quantiles. The percentage of quiescent animals present within each quantile was calculated (see Methods), and each quantile was colored to reflect this percentage. In all shape variables, quantiles that contain the largest animals displayed an increase in quiescence earlier than quantiles that contain the smallest animals. These dynamics were consistent across all larval-stage transitions (Figure S2-7). (D) Median width vs. median length for experimental hours 1 - 55. Red indicates measurements that fall above the quiescence threshold (see Methods). Simultaneous changes in length and width occur during periods of increased quiescence.

#### 2.4.4. Modeling *C. elegans* cuticle stretch dynamics

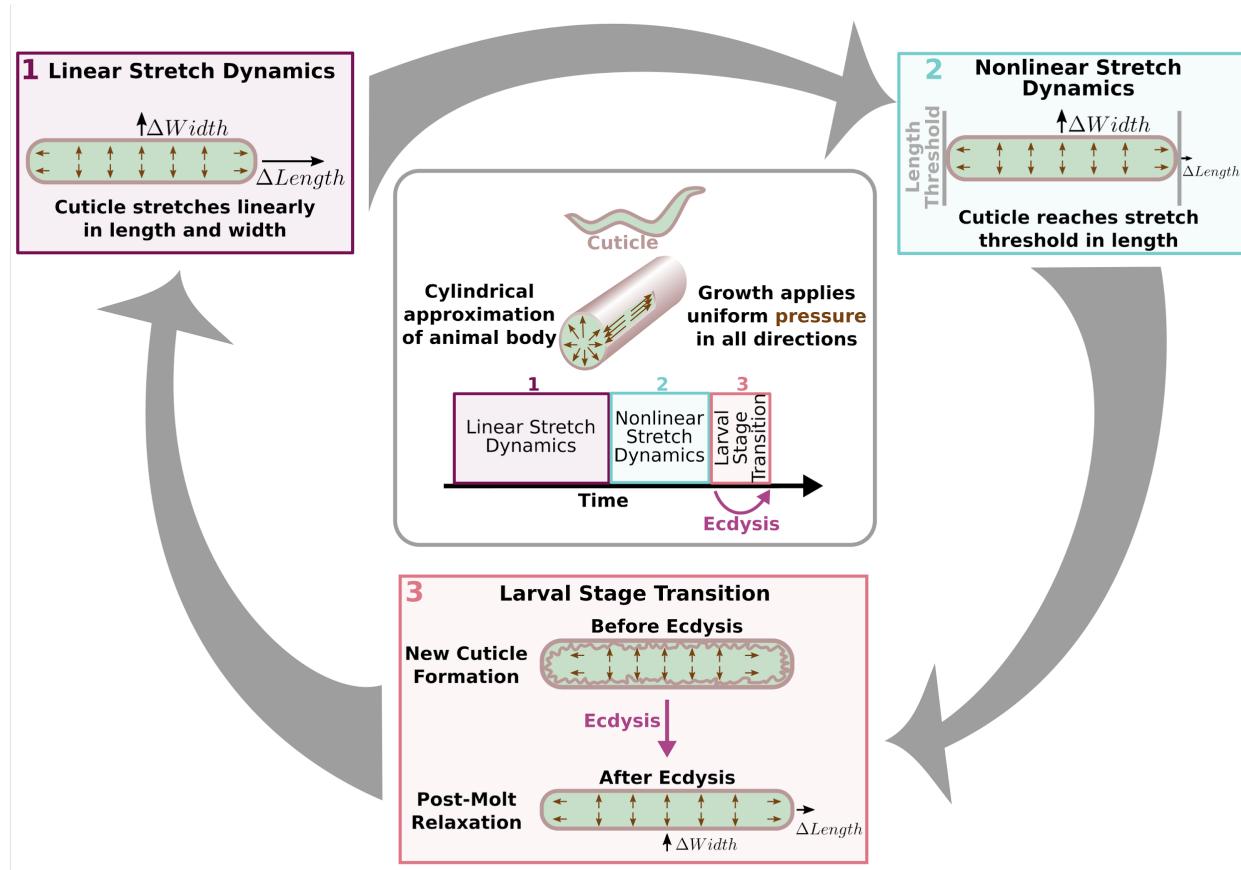
##### 2.4.4.1. Sensing of cuticle stretch as a trigger for larval-stage transitions

Previous studies theorized that the internal mechanism for sensing body size and triggering molts in *C. elegans* is driven, in part, by the properties of the collagen-rich cuticle [43,101]. Many cuticle collagen mutations cause morphological defects in nematode shape [130]. Some of these mutants are shorter than the wild type but do not have differences in animal width, implying that the cuticle affects length and width independently [33]. The *C. elegans* cuticle does not grow through the addition of new material but rather stretches to accommodate increases in animal body size. Cuticle stretch is likely limited by the material properties of the cuticle. The *C. elegans* cuticle is primarily made of cross-linked collagens organized into lateral ridges and circumferential bands [131]. Commonly found in many biological systems, collagen-based

materials are fairly flexible under low stress conditions. However, as stress increases, collagen fibrils become elongated and orient in the load bearing direction leading to a decrease in elasticity [132]. Previous work using atomic force microscopy revealed a high level of biomechanical stiffness at the circumferential bands [133], leading others to speculate that mechanical strain on these structures is likely adjusted as internal body pressure changes [134]. Additionally, in *nekl-3(sv3)* molting mutants, the cuticle is not properly removed from the middle part of their body, leaving the free head and tail to grow normally while the encased middle is constricted by the old cuticle to pre-molt dimensions [135]. Given this body restriction, we speculated that the old cuticle stretches beyond its tolerance, becomes stiff, and constricts the center of the nematode relative to the growing head and tail size. To our knowledge, it is unknown if the cuticle stretches enough during the course of normal development to become stiff in either the length or width directions. If the cuticle does become stiff before a molt, *C. elegans* may be able to sense the reduction of elasticity or "stretchiness," and use this signal, along with others, to determine when to initiate a molt. To detect possible changes in cuticle stretch during normal development, we asked how changes in cuticle elasticity will impact body shape.

To understand the impact of cuticle elasticity on body shape, we developed a "Stretcher" model independent of our measurements during development. We propose that the nematode passes through three distinct regimes related to cuticle stretch: linear stretch dynamics, nonlinear stretch dynamics, and the larval-stage transition (**Figure 2-5**). These regimes arise naturally from the following biologically supported assumptions. The cuticle structure is anisotropic, possibly leading to distinct stiffness properties in the length and width directions [136,137]. We approximated the cuticle as a hollow cylinder of negligible thickness filled by the body of the

nematode. Growth was modeled as internal pressure evenly applied to the cuticle in all directions. An anisotropic, elastic cuticle responds differently to pressure from growth during linear stretch, nonlinear stretch, and post-molt relaxation, leading to differences in shape during each stage of development.



**Figure 2-5. Cuticle stretch dynamics guide larval-stage transitions.** The Stretcher model describes each larval stage as a cycle. Nematodes are modeled as a cylindrical object with a thin cuticle epidermis. (Box 1) Linear Stretch Dynamics: uniform growth pressure stretches the cuticle linearly in both length and width. (Box 2) Nonlinear Stretch Dynamics: the cuticle has reached a stretch threshold in length, and under uniform growth pressure the length stretches less

(sub-linear) and width stretches linearly. (Box 3) Larval Stage Transition: a new cuticle is formed and the old cuticle is shed (ecdysis), removing constraints in length. The nematode body “relaxes” in length, causing an increase in length, a decrease in width, and constant volume.

In the linear stretch regime (**Figure 2-5**), the cuticle would be linearly elastic in both the length and width directions, stretching proportionally to the pressure exerted on the cuticle. Previous work found evidence for a linearly elastic cuticle [138,139] in animals expanded in a negative external pressure environment or after positive force was applied to the cuticle. Gilpin *et al.* found evidence of linear elasticity in the nematode body. We conjecture that this linear elasticity is caused by the constraints applied by the cuticle [138,139]. A linearly elastic cuticle will have  $\Delta L$  stretch in the length direction and  $\Delta W$  stretch in the width direction, each related to growth-applied pressure  $\Delta p$  by

$$\Delta L = a_L \Delta p \quad (1)$$

$$\Delta W = a_W \Delta p. \quad (2)$$

The “stretch coefficients” in length,  $a_L$ , and width,  $a_W$ , measure the stiffness of the cuticle (File S5, Eq. S4-S13). Smaller values correspond to a stiffer material, which is less able to stretch in response to pressure. The stretch coefficients are constant in the linearly elastic regime and are determined by geometric constants and material properties. The ratio of the change in length (Eq. 1) and width (Eq. 2) produces a pressure-independent relationship that depends only on the ratio of the geometric and material properties, which can be verified using measurements of length and width (**Figure 2-3**). During the linearly elastic regime, the ratio of growth in width to growth in length is constant throughout a larval stage where the cuticle properties are fixed as in

$$\frac{\Delta W}{\Delta L} = \frac{a_w}{a_L} = \text{constant.} \quad (3)$$

In the nonlinear stretch regime (**Figure 2-5**), growth continues to apply pressure to the cuticle uniformly in all directions. As observed in *nek1-3(sv3)* mutants, the cuticle can restrict body growth [135]. Once outside of the linearly elastic regime, the cuticle would hardly stretch, even under large forces. We hypothesize that this shift from linear to nonlinear regimes can provide a mechanism for size-sensing and cues the larval-stage transition (**Figure 2-5**). In principle, this transition could occur in either the width or length directions. For simplicity, we illustrate a transition from linear to nonlinear stretch in the length direction while linear stretch in the width direction is maintained. In the nonlinear regime, the stretch in the length direction in response to pressure becomes

$$\Delta L \approx \tilde{a}_L(p) \Delta p. \quad (4)$$

The nonlinear “stretch coefficient,”  $\tilde{a}_L(p)$ , is no longer constant and decreases with increasing pressure. It is smaller than  $a_L$  because the cuticle has become less elastic than in the linear regime. If the length-direction enters the nonlinear regime and has reduced stretch response and width has the same constant stretch response, then we expect the  $\Delta W/\Delta L$  ratio to increase

$$\frac{\Delta W}{\Delta L}|_{\text{non-linear}} = \frac{a_w}{\tilde{a}_L(p)} > \frac{a_w}{a_L} = \frac{\Delta W}{\Delta L}|_{\text{linear}}. \quad (5)$$

If the width-direction enters the nonlinear regime first, then the  $\Delta W/\Delta L$  would decrease. It is possible that both length and width could enter the nonlinear regime at around the same time, in which case the change in the  $\Delta W/\Delta L$  ratio would depend on the differences in the pressure-dependent stretch coefficients. This prediction motivates an analysis of the  $\Delta W/\Delta L$  ratio

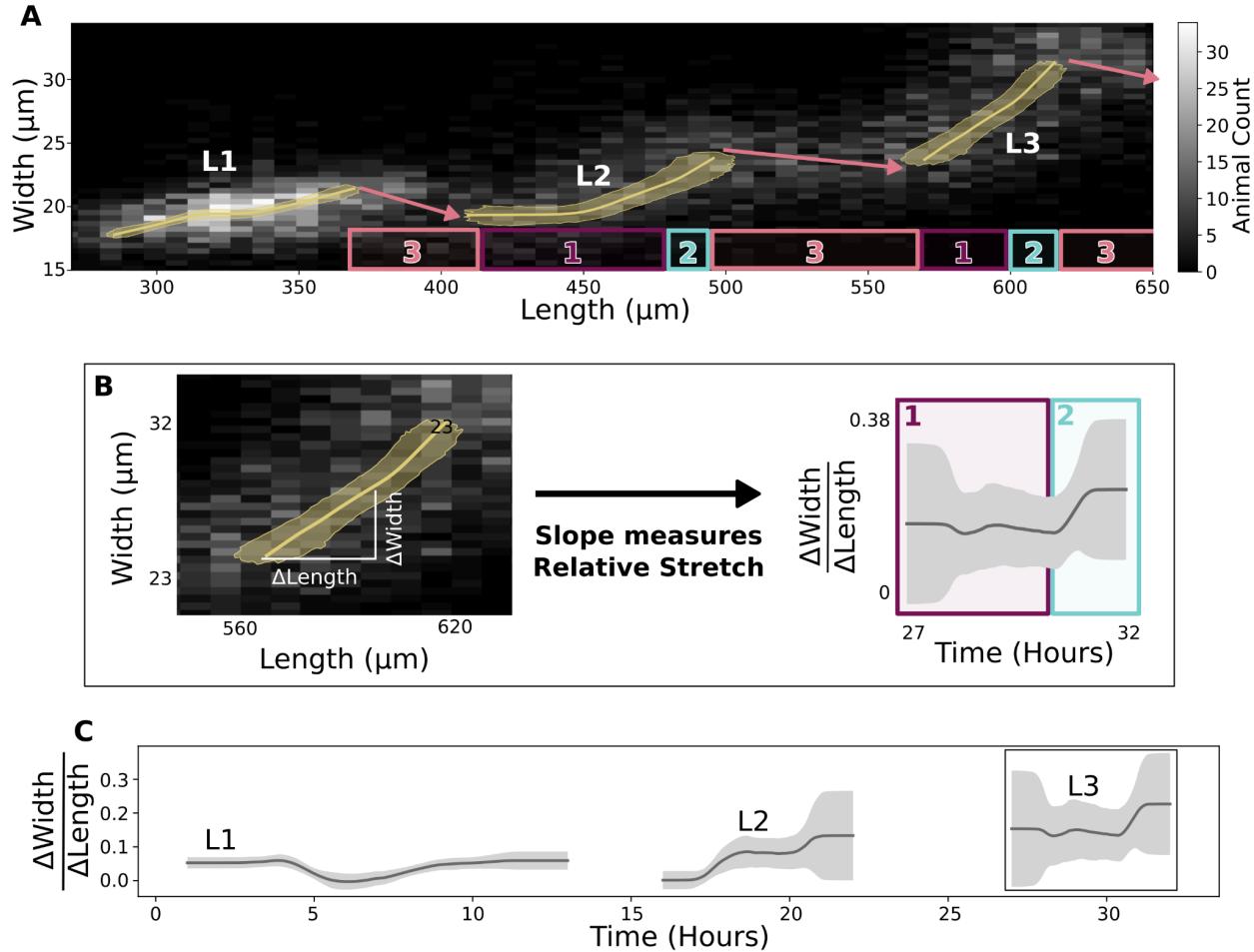
of the shape data (**Figure 2-4**), to determine if any sign of a nonlinear stretch regime in either direction is observed.

During the larval-stage transition (**Figure 2-5**), a new, larger cuticle is formed beneath the old cuticle that is shed during ecdysis. Because the old cuticle constrained growth in length, we predict a rapid increase in the length direction when the constraint is removed. Animal volume is conserved as growth does not occur during this process. Therefore, the relaxation in length is accompanied by a corresponding decrease in width.

#### **2.4.4.2. Discontinuities in animal growth rate driven by limits on cuticle stretch**

To test whether the shape dynamics predicted by the Stretcher model are consistent with the data, we analyzed the relationship between measured nematode length and width over developmental time. This analysis requires no model fitting or assumptions about which direction may enter a nonlinear regime first, and it involves only calculating  $\Delta W/\Delta L$ . Although the shape relaxation for an individual animal is expected to happen at a much shorter time scale than data collection (seconds to minutes), the measured shape change within the population occurs on a larger time scale (several hours). As such, we can observe the sudden shape change at transitions by assessing changes in length and width of the population (**Figure 2-6**). Doing so, we detected all three regimes predicted by the Stretcher model in the COPAS BIOSORT data: linear stretch, nonlinear stretch, and relaxation (**Figure 2-6A**). In all larval stages, we observed an approximately constant instantaneous  $\Delta W/\Delta L$  ratio, consistent with a linear stretch regime (**Figure 2-6B** and **Figure 2-6C**). We also detected a large slope decrease during the L1 stage, which could correspond in time to the metabolic decision for entry into the stress-resistant, developmentally arrested dauer stage [140] divisions of the seam cells, a population of epidermal

stem cells in *C. elegans* [34]. Although transitions in the slope ( $\Delta W/\Delta L$  ratio) are difficult to detect at all larval stages because of noise amplification and population effects (**Figure S2-9**), we observed a sharp slope increase, consistent with a nonlinear stretch regime in length prior to larval stage transitions (see end of L2 and L3 stages). Following the nonlinear stretch regime, we noted a simultaneous increase in length and decrease in width at the transition between larval stages, consistent with a length threshold in the Stretcher model (**Figure 2-5D**, **Figure 2-6A**, and **Figure S2-8**). As a whole, the changes in animal shape are consistent with the hypothesis that the cuticle reaches the nonlinear elastic regime before larval-stage transitions during normal development. The ability to sense when a critical cuticle stiffness is reached would allow animals to monitor relative changes in body growth within each larval stage, serving as a connection between growth rate and developmental timing.



**Figure 2-6. Experimental data are consistent with a length threshold in cuticle stretch.** (A) A grayscale histogram of the width (y-axis) vs length (x-axis) of all sampled animals in replicate 2. The range of all bootstrap regressions is in gold. (B) Demonstration of calculating the ratio of width-to-length stretch as the local slope using L3. Left panel is a repetition of L3 data from **Figure 2-6A**. Right panel is a repetition of results from **Figure 2-6C**. (C) Within a larval stage, the ratio of width to length stretch varies over time. The standard deviation captures population variation (grey).

## 2.5. DISCUSSION

Using an integrated image-based and flow-based phenotyping strategy to precisely evaluate feeding, growth, and molt dynamics at high replication, we detected oscillations in feeding behavior consistent with larval progressions and used these dynamics to define larval stages. We observed changes in body shape at each larval-stage transition that can be driven by differences in physical cuticle properties along length or width (anisotropy). These results suggest a mechanism by which animals sense their size and control molt timing by detecting the physical stretch of the cuticle, demonstrating how physical constraints can influence developmental timing and growth rate.

### 2.5.1. A potential role for cuticle stretch in the timing of larval-stage transitions

Measurement of both animal length and width allowed us to observe changes in body shape as well as body size. We propose a mechanism in which a stretch threshold along the body length axis acts as a trigger to larval-stage transitions. Previously, a Folder mechanism for *C. elegans* growth has been suggested [101]. Mechanical stretch sensing could provide organisms a way to couple the rate of growth and development to maintain a constant relative change, or fold change, in volume within a larval stage. In this way, smaller animals would reach a stretch limit at a smaller size as the cuticle would only stretch by a percentage of its original size before reaching a threshold. For cuticle stretch to trigger larval-stage transitions, animals must either have the ability to measure the amount the cuticle has stretched or the stiffness of the cuticle. Across biological systems, cells can respond to the stiffness of their environment using mechanosensitive components [141,142], but few examples in tissues or whole-organisms are

known. In *C. elegans*, it has been demonstrated that hemidesmosomes, which connect the cuticle and the epidermis, are mechanosensitive during embryogenesis [143,144]. Additionally, dense bodies, which connect the epidermis and muscles, are hypothesized to be mechanosensitive as well [49,145–147]. Changes in cuticle composition, and presumably stiffness, have been shown to also affect well known growth controlling pathways such as the BMP signaling pathway [148]. These possible mechanosensitive components could monitor the stiffness of the cuticle and be part of the signaling pathways that regulate the initiation of molts. Further experiments are required to explicitly test whether these components control larval-stage transitions but disentangling epistatic growth effects in mutants from specific mechanosensation effects might be difficult. It is also important to note that evidence for a stretch-based mechanism for growth control does not preclude the possibility of a developmental timer. It is likely that physical constraint-based events are part of a larger regulatory system coordinating developmental decisions.

Our analysis of  $\Delta W/\Delta L$  ratio changes over larval stages provides a first approximation of the timing of cuticle stretch properties. The increase we observed in the  $\Delta W/\Delta L$  ratio, followed by sudden relaxation to a longer thinner animal is consistent with a decrease in the elasticity of the cuticle in the length direction before a molt. These observations of shape change also indicate that *C. elegans* do not preempt the shape change by molting before they hit the “stretch limit”, meaning that the decrease in elasticity could be detected by the animal and this stretch-based physical constraint mechanism could play a role in developmental decisions. Interestingly, when observing the L4 to adult transition, others have detected anisotropic constriction on the transverse (width) axis followed by gradual relaxation driven by rearrangements in cortical actin

networks [149]. Single-worm, high frequency measurements targeting hours surrounding the sudden width-to-length ratio increase, are needed to better resolve cuticle shape dynamics. Although analysis of the larval growth dynamics for *C. elegans* body shape mutants (*dpy*, *lon*, or *sma*) may provide insights into  $\Delta W/\Delta L$  ratio variation, measurements of whole-animal length and width only provide a total stiffness estimate, leaving us unable to distinguish the contributions of cuticle stiffness from other tissues. For example, previous work has shown that bodies of *dpy-10(e128)* mutants are twice as stiff as the wild type [150], and *dpy-5(e61)* mutants are softer than wild-type animals [138]. Authors have speculated that this difference in stiffness is caused by an increased internal glycerol concentration in *dpy-10(e128)* animals that is absent in *dpy-5(e61)* animals [150]. Therefore, although these mutations impact body shape similarly, they do not have the same effect on body stiffness, indicating that body shape alone does not predict body stiffness. To assess cuticle properties independent of other nematode tissues and organs, future experiments probing the stiffness of free cuticles are necessary.

Additionally, within the L1 stage, the relative stretch measured in width and length did not follow the pattern observed in other larval stages. We observed a mid-stage dip in the width-to-length ratio that is otherwise approximately constant throughout the L1 stage. Experiments exploring the structural properties of cuticles at all larval stages might help to determine where the L1 shape changes originate.

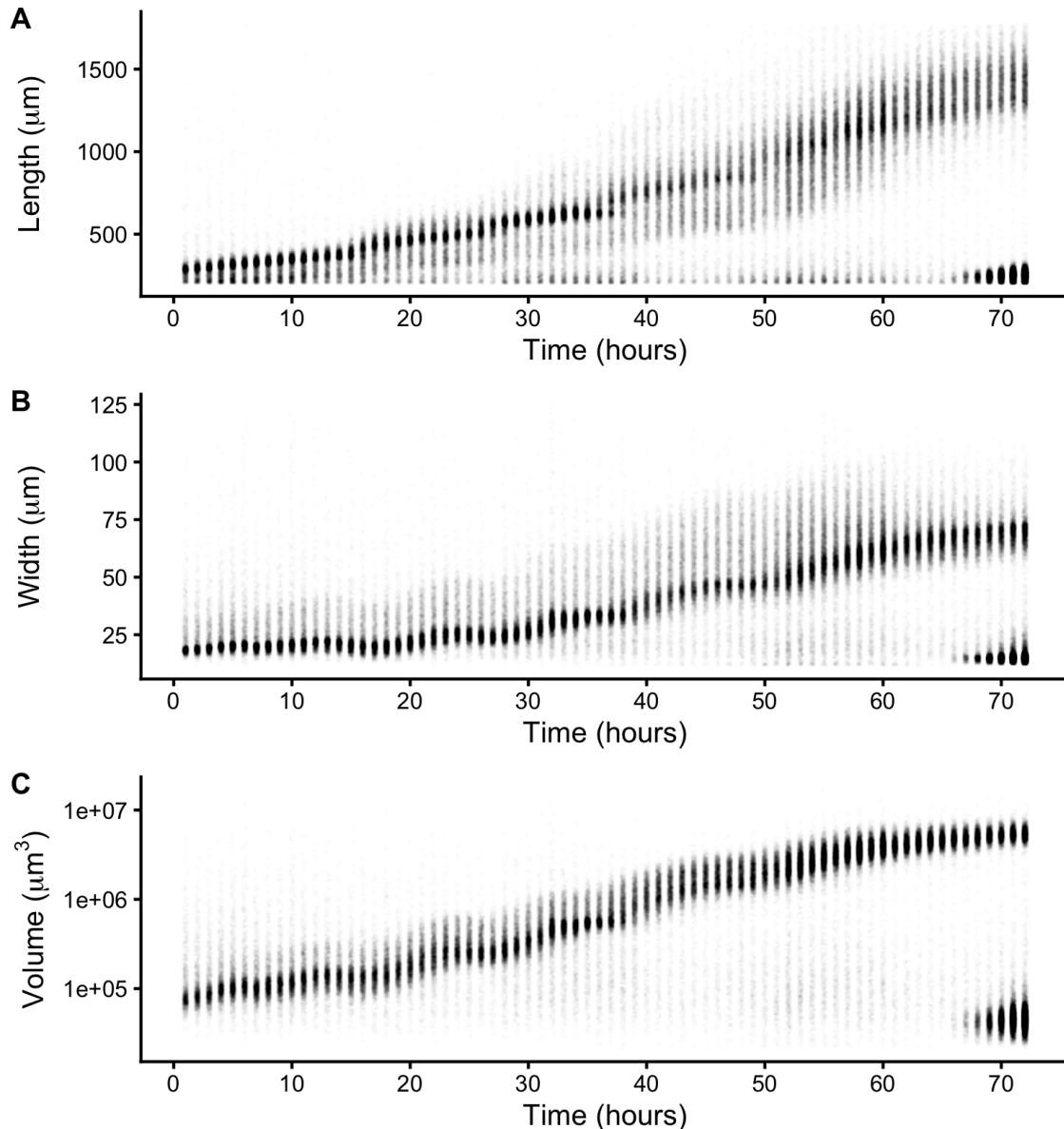
### 2.5.2. Development comprises complex interactions of growth regulation across diverse scales

Our results are consistent with a model in which *C. elegans* use physical constraints on animal size, in tandem with other regulatory mechanisms, to control growth rate and determine developmental transitions. This type of regulation could be applicable to organisms with stiff cuticles or other physical barriers to growth, like many species of Ecdysozoa. The control of whole-organism growth requires cells, tissues, and organs to orchestrate final size and cell number. In *C. elegans*, cell number is precisely defined and invariant from animal to animal [151], so the final adult size of an individual must come from cell size as opposed to number. Future studies should focus on how whole-organism size is determined by the integration of cell, tissue, and organ size. By incorporating these different developmental scales, the Stretcher model can be refined to completely describe how physical constraints on parts of the organism impact the whole. *C. elegans* gives investigators a powerful system to investigate animal-to-animal variation in developmental trajectories across each of these scales.

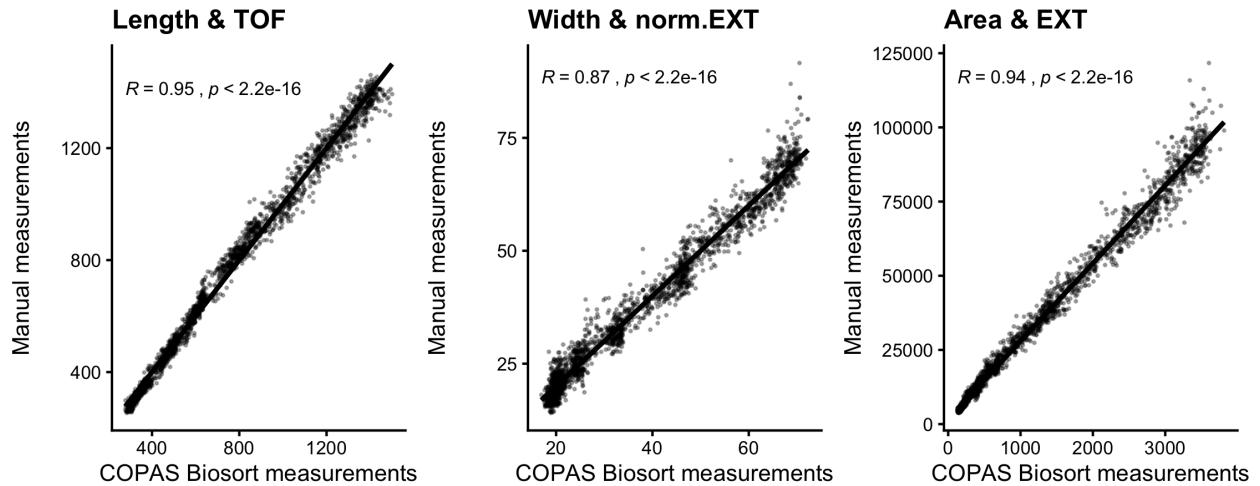
## 2.6. CONTRIBUTIONS

This work was supported by the NSF-Simons Center for Quantitative Biology at Northwestern University. Dr. Erik Andersen, Dr. Gaotian Zhang, and Nicole Roberto assisted with the high-throughput sampling. Dr. Sasha Shirman and Christina Goss built and tested the mathematical models. Hannah N. Ahmed, Elliot J. Andersen, Isabella R. Miller, Justine K. Rozenich, Iris L. Swarthout, and Jordan A. Vaughn helped collect manual size measurements of animals from images. Drs. Erik Andersen and Niall Mangan supervised the project.

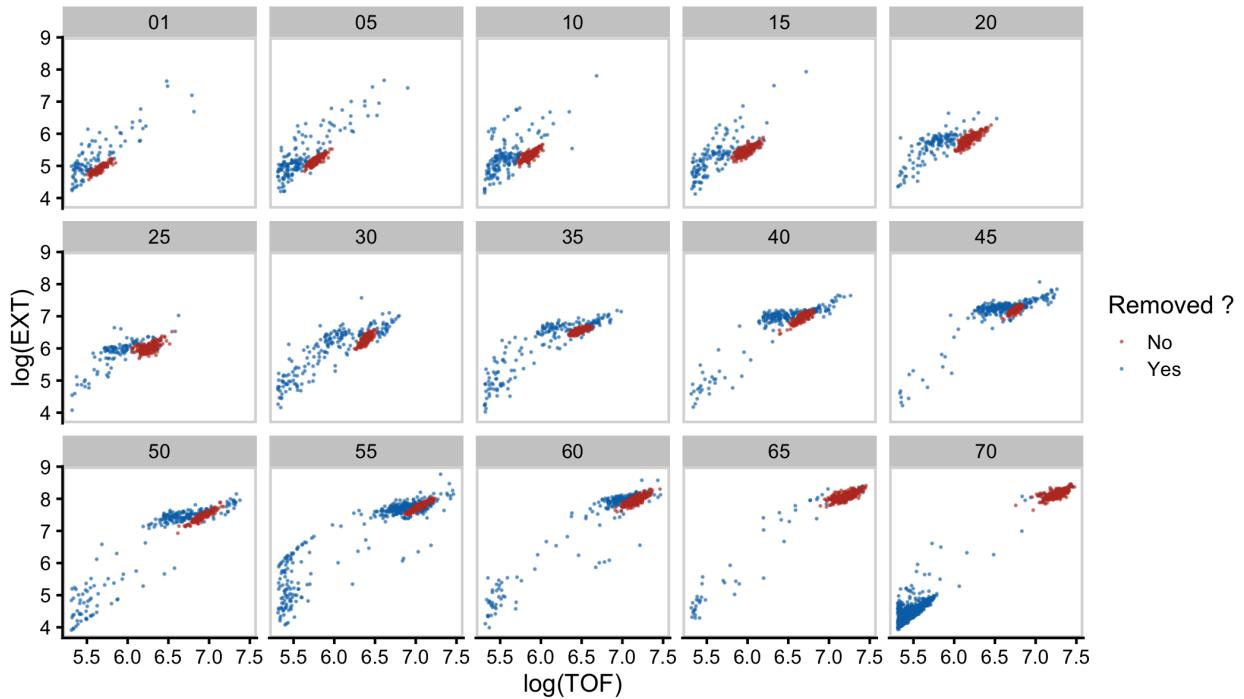
## 2.7. SUPPLEMENT



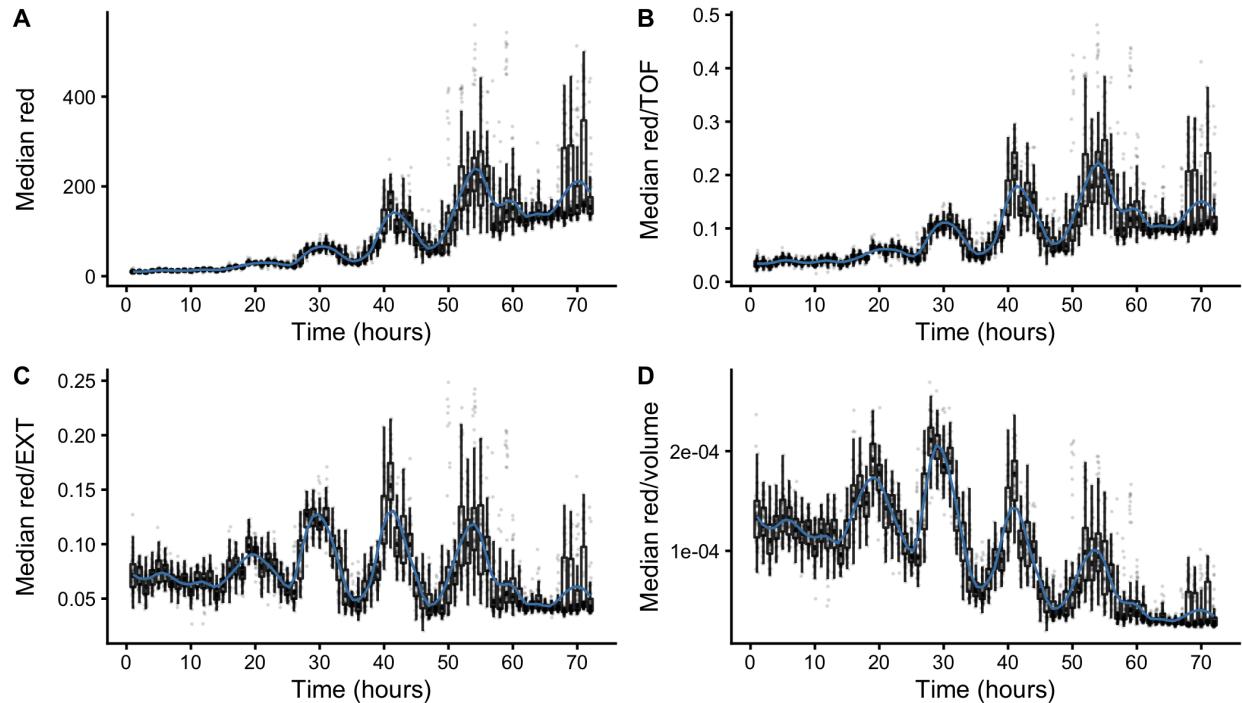
**Figure S2-1. Raw measurements of animal size.** Raw COPAS BIOSORT data of animal length (A), width (B), and volume (C) are shown here. After 60 hours, animals have developed to the adult stage. Smaller objects observed after 65 hours were the next generation of newly hatched L1 larvae laid by the animals that developed during the time course.



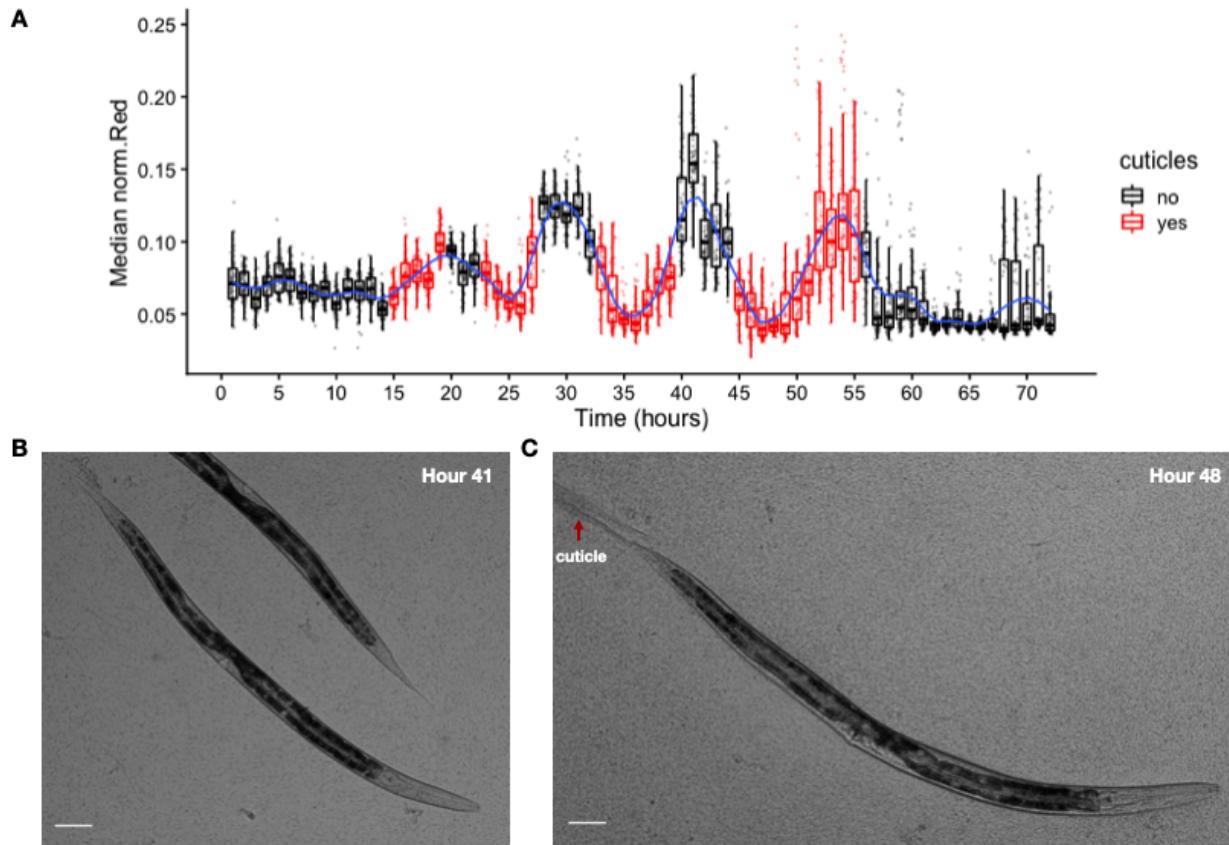
**Figure S2-2. Correlation analysis of body size measurements.** Manual measurements of animal length, width, and estimated area were compared to COPAS BIOSORT measurements of TOF (time of flight), norm.EXT (extinction normalized by TOF), and EXT (extinction), respectively. A Kendall correlation value is shown in each plot.



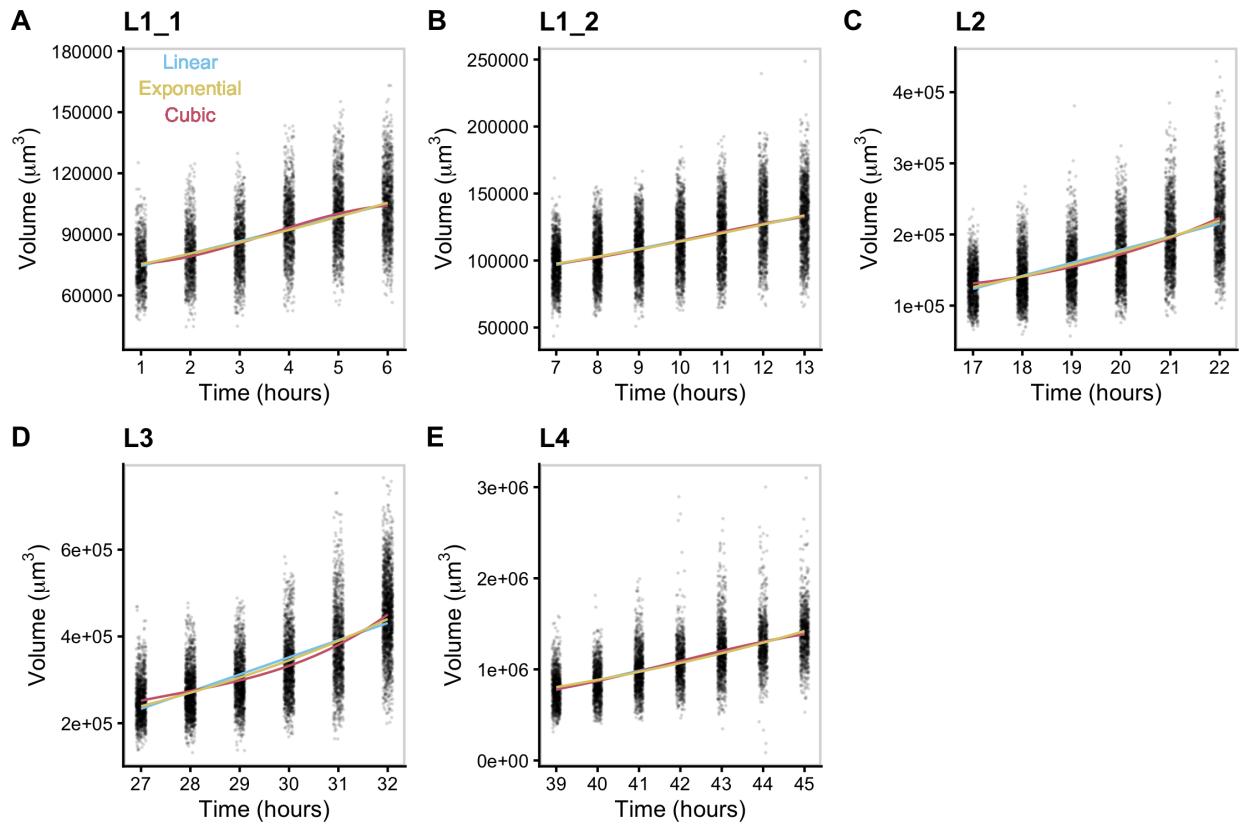
**Figure S2-3. Mixture modeling of COPAS BIOSORT data was used to prune data.** Mixture models of Gaussian distributions were fit to log transformed animal length (x-axis) and log transformed optical extinction (y-axis). Data from each hour of the experiment was analyzed and processed to remove clusters that did not include animal objects. All replicates were pruned independently; a subset of data from replicate 2 is shown here. Panels indicate experimental hours from which data were taken. Across the entire dataset, approximately 50% of objects were identified as non-animal objects and removed for all further analysis.



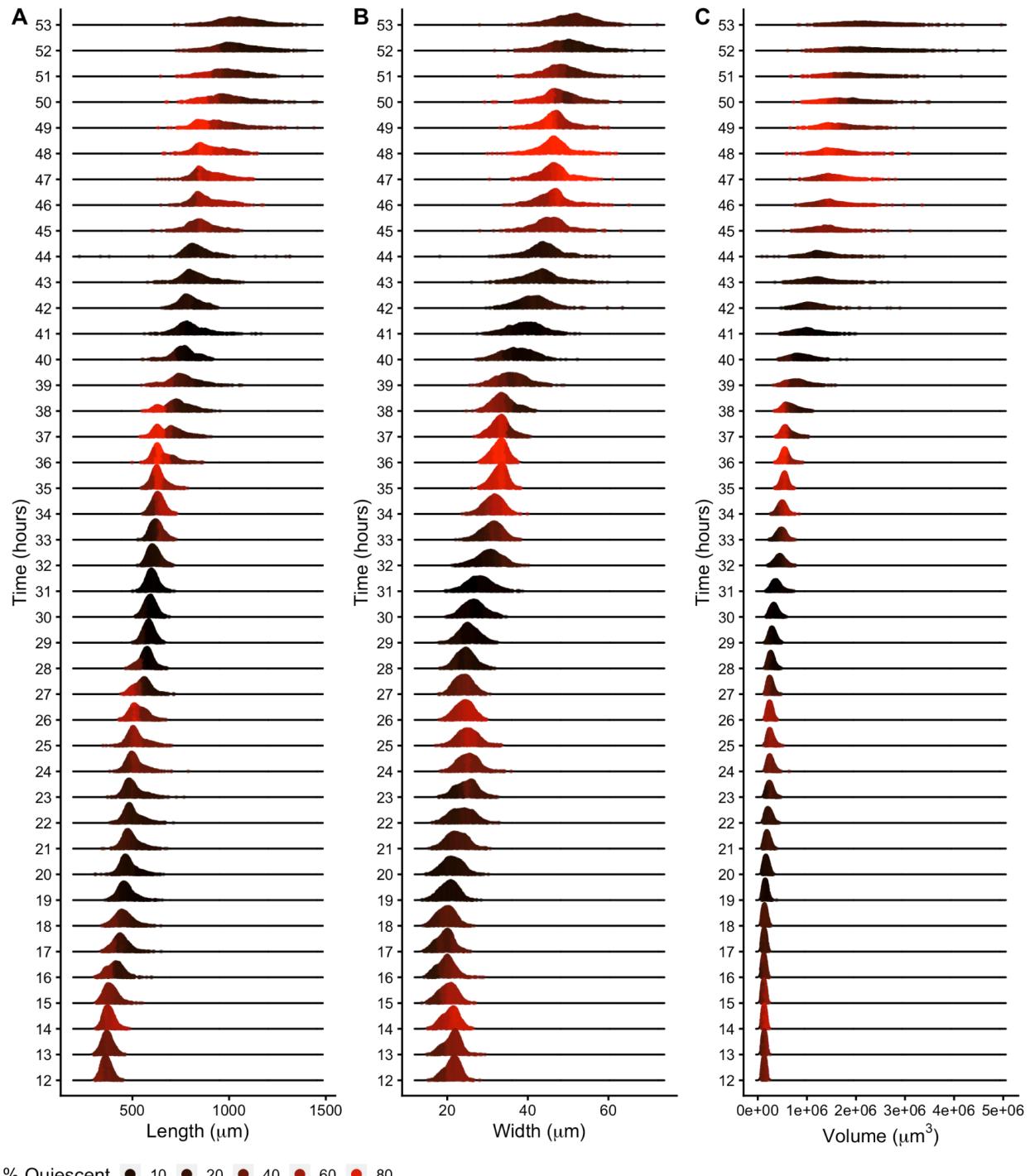
**Figure S2-4. Fluorescence measurements normalized by body size.** Red fluorescent beads were fed to animals during experimentation and fluorescence data was collected by the COPAS BIOSORT. Raw data were summarized by well to obtain median red fluorescence measurements (A). Fluctuations in fluorescence indicate fluctuations in feeding behavior. Median red fluorescence data was normalized by body size measurements (TOF - length in (B), EXT - area in (C), and volume in (D)) to account for increases in body size. Dividing fluorescence by area resulted in dynamics where local minima in fluorescence reached a baseline and therefore was most successful in normalizing fluorescence dynamics to account for changes in animal size over time.



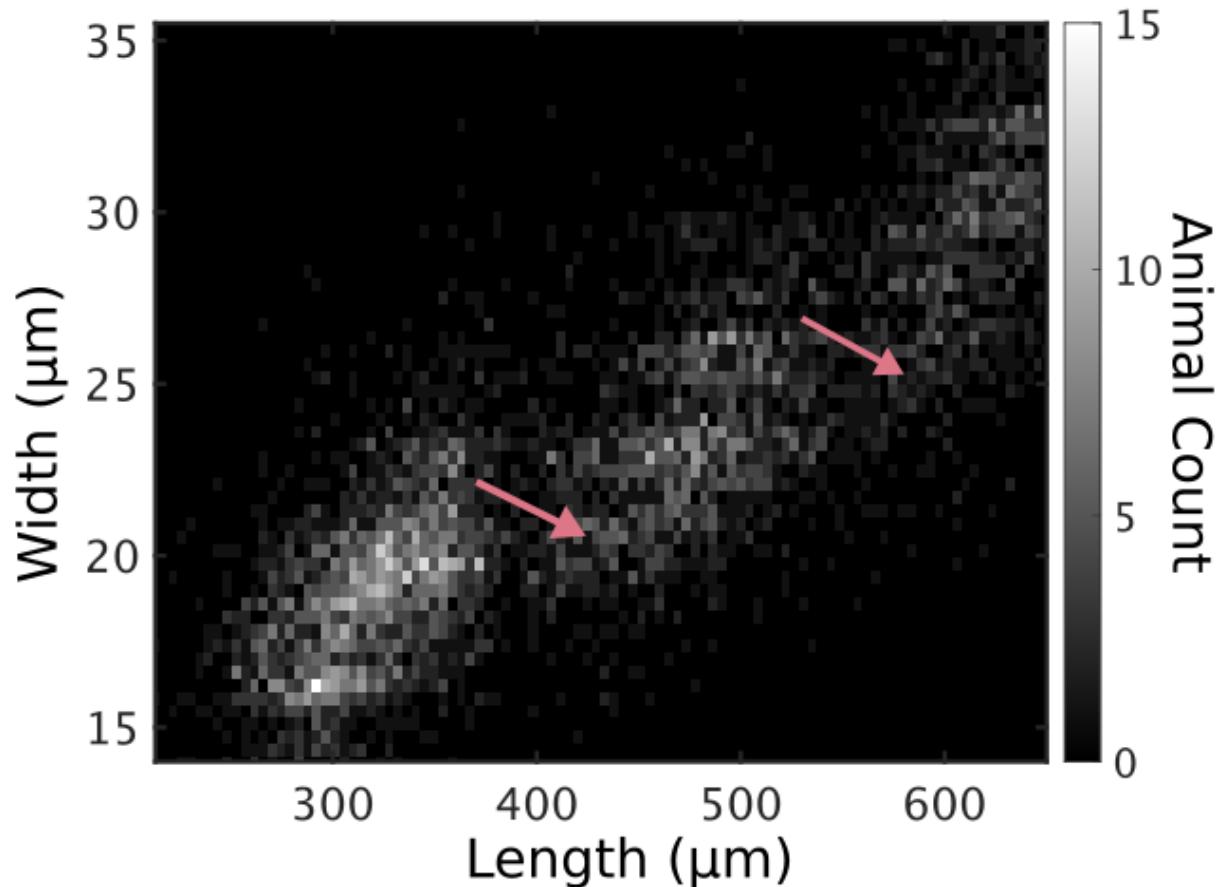
**Figure S2-5. Cuticles identified during periods of decreased feeding.** Images of wells collected during the experiment were examined for evidence of shed cuticles. (A) Experimental hours where cuticles were identified from images overlap with hours where population feeding behavior is low. Cuticles shed from the L4-Adult molt persisted longer than previous larval stage cuticle debris. (B) Example image of animal without visible cuticle during a period of elevated feeding. (C) Example image of an animal with visible cuticle indicating completion of molt during a period of decreased feeding. Scale bar, 263.5  $\mu$ m.



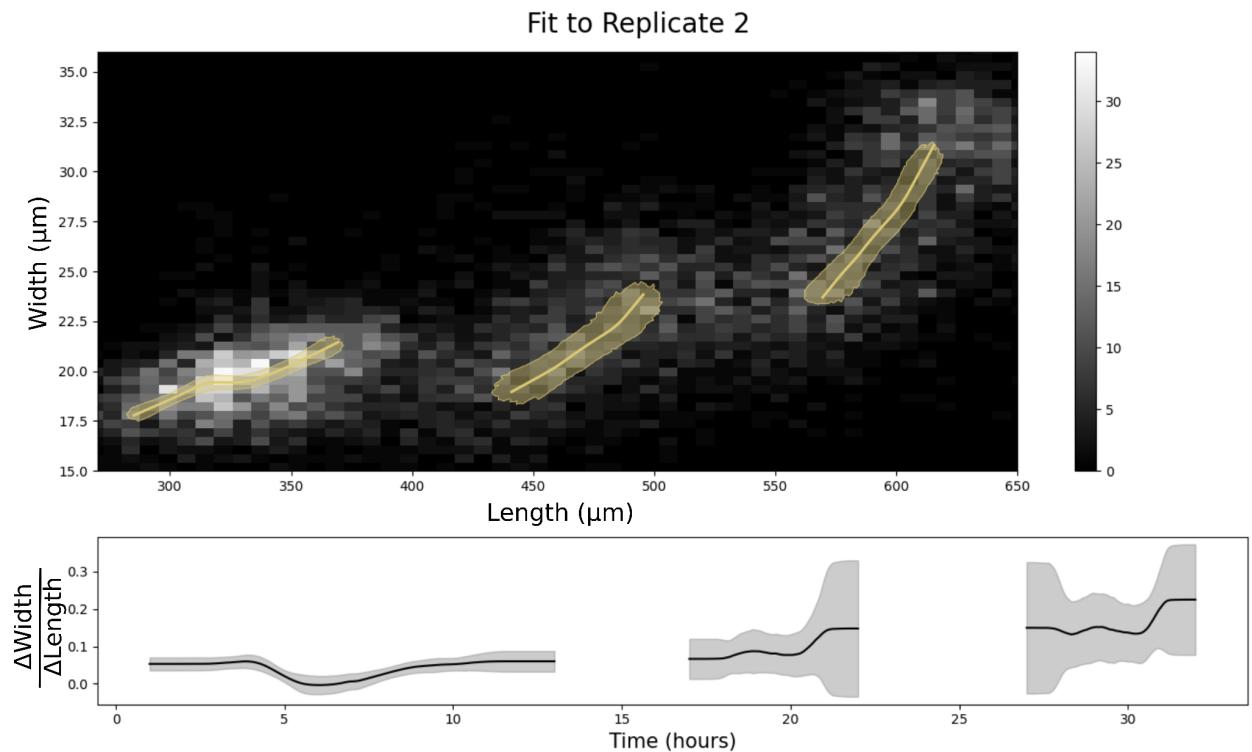
**Figure S2-6. Volume growth data fit with linear, exponential, and cubic models.** Volume data of individuals in time points defined as growth periods are analyzed for each stage. L1 stage was further separated into two periods to account for the volume dip that occurs mid-stage. Shapes of the three growth curves all appear relatively similar due to the high relative noise in the data. The coefficient of variation is between 15-16% over the course of the experiment. Computational simulations have shown that increases in experimental noise limit the correct identification of growth models [129].



**Figure S2-7. Density plots of population size dynamics across all larval transitions.** Density curves of length (A), width (B) and volume (C). Curves are divided into five quantiles and colored by the percentage of quiescent animals present within that quantile. Molts are estimated to occur at experimental hours 14, 25, 36, and 48 (see Methods).



**Figure S2-8. Animals in all replicates, measured from images.** Animal length and width over *C. elegans* development captured from image data. Higher noise levels in these measurements preclude accurate regressions to individual larval stages. Length jumps and width dips are still apparent. Compare with Fig 6.



**Figure S2-9. Stretcher model analysis of replicate 2 COPAS BIOSORT data for different stage thresholding.** Compare to Figure 2-5. Larval hours were defined by taking the ceiling of the lower boundary and the floor of the upper boundary. This rounding method for larval stage definition demonstrates the sensitivity of the analysis to edge effects. The unexpected step in the L2 larval stage (Fig 6) was significantly reduced with this rounding method.

**Table S2-1. Results of analysis of variance models fit to COPAS BIOSORT data.** Analysis of variance tests were used to quantify the amount of variance in our data contributed by the sampling technique. The sampling technique involved unbiased sampling of animals from six replicate populations and subsequent distribution into multiple wells of a microtiter plate for analysis. We quantified the amount of variance contributed by replicate and well. We find that the variance explained by well is nearly negligible whereas replicate contributes minor variance in some measurements. Given this information, we deem the generated summary statistics an appropriate representation of the population. Df = degrees of freedom, Sum Sq = sum of squares, Mean Sq = mean squares (or  $\frac{\text{sum of squares error}}{\text{degrees of freedom}}$ ), F value = F-statistic, Pr(>F) = p-value, % Var Explained = percent variance explained.

**Response = Norm.Red**

Terms	Df	Sum Sq	Mean Sq	F value	Pr(>F)	% Var Explained
hour	1	439.46	439.46	217762.01	0	54.34
replicate	6	165.82	27.64	13694.48	0	20.51
well	10	0.32	0.03	15.62	0	0.04
Residuals	100619	203.06	0	NA	NA	25.11

**Response = Length**

Terms	Df	Sum Sq	Mean Sq	F value	Pr(>F)	% Var Explained
hour	1	86190107879	86190107879	8223506	0	98.4
replicate	6	349438944	58239824	5557	0	0.4
well	10	834970	83497	8	0	0
Residuals	100619	1054582098	10481	NA	NA	1.2

**Response = Width**

Terms	Df	Sum Sq	Mean Sq	F value	Pr(>F)	% Var Explained
hour	1	209090438.47	209090438.47	6950760.3	0	97.99
replicate	6	1266985.87	211164.31	7019.7	0	0.59
well	10	1495.32	149.53	4.97	0	0
Residuals	100619	3026786.99	30.08	NA	NA	1.42

**Response = Volume**

Terms	Df	Sum Sq	Mean Sq	F value	Pr(>F)	% Var Explained
hour	1	621712474191714688	621712474191714688	843342	0	84.08
replicate	6	43356968850687072	7226161475114512	9802	0	5.86
well	10	152458715405862	15245871540586	21	0	0.02
Residuals	100619	74176459811167296	737201321929	NA	NA	10.03

**Table S2-2. Model fit criteria used to assess candidate growth models.** To determine the level of support for each model, the candidate model with the smallest raw AIC was identified and compared to other AIC values. The same assessment was performed for BIC. The magnitude of the  $\Delta AIC$  or  $\Delta BIC$  value was then used to determine the level of support for each candidate model given previously established significance thresholds [124,125]. If the delta value was greater than 6, the model with the smallest raw AIC or BIC value was denoted as the best model. If the delta value was less than 6 but greater than 2, the model with the smallest AIC or BIC value was determined to likely be the best model. If the delta value was less than 2, we are unable to distinguish the model of best fit.

Stage	$\Delta AIC$			$\Delta BIC$			Best model by AIC	Best model by BIC
	Linear	Exponential	Cubic	Linear	Exponential	Cubic		
L1_1	17	21	0	4	9	0	Cubic	Likely Cubic
L1_2	2	4	0	0	2	12	Can't distinguish	Can't distinguish
L2	142	43	0	128	28	0	Cubic	Cubic
L3	374	145	0	360	131	0	Cubic	Cubic
L4	4	44	0	0	40	10	Likely Cubic	Linear

## 3

## Multiple genetic loci underlie differences in *C. elegans* growth

The Andersen Lab is particularly adept at connecting phenotypic variation in a population to unique genetic variants. As a follow-up to the work described in the previous chapter, I was interested in leveraging this expertise to explore how natural genetic variation influenced growth in *C. elegans*. This meant running another long time course experiment (on my golden birthday no less!), digging into a treasure trove of “throw-away” data collected in 2014, and working with a few new strains other than our old friend, N2. The following chapter is based off of my work on this project. It is currently in preparation for submission as a first-author manuscript to G3.

### 3.1. ABSTRACT

Growth rate and body size are complex traits that contribute to the fitness of organisms. The identification of loci that underlie differences in these traits provides insights into the genetic contributions to development. Leveraging *Caenorhabditis elegans* as a tractable metazoan model for quantitative genetics, we can identify genomic regions that underlie differences in growth. We measured post-embryonic growth of the laboratory-adapted wild-type strain (N2) and a wild strain from Hawaii (CB4856), and found differences in body size. Using linkage mapping, we identified three distinct quantitative trait loci (QTL) on chromosomes IV, V, and X that are associated with variation in body size. We further examined these size-associated QTL using chromosome substitution strains and near-isogenic lines, and validated the chromosome X QTL. Additionally, we generated a list of candidate genes for the chromosome X QTL. These genes could potentially contribute to differences in animal growth and should be evaluated in subsequent studies. Our work reveals the genetic architecture underlying animal growth variation and highlights the genetic complexity of body size in *C. elegans* natural populations.

### 3.2. INTRODUCTION

Precise regulation of final body size is essential to the development and fitness of organisms. Although a larger body size can increase competitive advantages, it also requires added time and nutrients to develop [88]. For this reason, mechanisms that control developmental growth rate and ultimate body size are likely under strong natural selection.

The robustness and precision with which animal development is choreographed is still not well understood. Developing systems coordinate the organization and interaction among

cells, tissues, and organs at high reproducibility even in the presence of genetic and environmental perturbations. The early developmental biologist C.H. Waddington coined the term “canalization” to describe this biological robustness [152]. Developmental canalization has been widely studied in human growth [153–155], and shown that to achieve an adult height within normal range, shorter individuals tend to undergo accelerated growth whereas taller individuals experience a decreased rate of growth [156]. In this way, the growth curves of individuals, though variable, converge on a narrow range.

To study the phenomenon of organismal size uniformity, considerable precision and throughput is needed, which can be a challenge when working with multicellular organisms. The nematode *Caenorhabditis elegans* presents a powerful model organism to study developmental growth. *C. elegans* has a quick generation time, produces large numbers of genetically identical offspring, and is easily cultured in controlled laboratory conditions [31]. Furthermore, *C. elegans* post-embryonic development is well characterized and marked by four larval-stage transitions (molts) that separate the *C. elegans* life cycle into five distinct stages: four larval stages (L1-L4) and adult [41]. The timing of these molts determines the completion of stage-specific development [48,49], underscoring the importance of developmental growth regulation in *C. elegans*.

We can leverage *C. elegans* natural genetic diversity [21,68,69,157,158] to connect phenotypic differences to genetic variants. Two particular strains of interest are the laboratory-adapted wild-type strain, N2, and a wild strain from Hawaii, CB4856. The genetic diversity between these two strains was shown to underlie multiple phenotypic differences, including aggregation behavior, life history traits, and gene expression [21]. Recombinant inbred

lines constructed from crosses between the N2 and CB4856 strains each have unique variants derived from each parental background. Performed at a large scale, these populations of recombinant individuals are a powerful tool to identify genomic regions that are correlated with phenotypic variation. Mapping the natural variation underlying phenotypic differences allows for the dissection of genetic networks involved in important biological processes. Many others have taken this approach to study the genetic underpinnings of complex traits [21,158].

To characterize the genetic basis for variation in body size and growth in *C. elegans*, we first performed a longitudinal study of post-embryonic growth in N2 and CB4856 animals. Although we observed similar patterns in overall growth dynamics, we also noticed small differences in body size at individual time points across development. To study these differences in size, we used linkage mapping to identify three distinct QTL that influence animal size variation. We further assessed each QTL independently using chromosome substitution strains and near-isogenic lines. Doing so, we validated the chromosome X QTL and identified promising candidate genes that could contribute to the differences in size between the N2 and CB4856 strains. Our work provides a framework for future studies to investigate the genetic mechanisms controlling developmental growth and body size in natural populations of *C. elegans*.

### **3.3. MATERIALS AND METHODS**

#### **3.3.1. Strains**

Animals were grown at 20°C on 6 cm plates of modified nematode growth media (NGMA), containing 1% agar and 0.7% agarose seeded with *E. coli* OP50 bacteria. Recombinant inbred advanced intercross lines (RIAILs) used for linkage mapping were constructed previously [117].

The construction of chromosome substitution strains (CSSs) and near isogenic lines (NILs) used for validation is detailed below. Strains are available upon request.

### 3.3.2. High-throughput growth assay

Measurements of body size and fluorescence were measured as previously described [87]. Briefly, the N2 and CB4856 strains were propagated for three generations, bleach-synchronized, and titered at a concentration of 1 embryo per  $\mu\text{L}$  into six replicate 500 mL flasks for a final volume of 25 mL. The following day, arrested L1s were fed HB101 food at a final concentration of OD20 in a final flask volume of 100 mL K medium and HB101 food. Animals were then grown at 20°C with constant shaking. Flasks were sampled each hour beginning one hour after feeding and continuing for 51 consecutive hours. At each hour, animals were sampled from each flask, treated with sodium azide, imaged with an ImageXpress Nano (Molecular Devices, San Jose, CA) and scored using a large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston MA). The COPAS BIOSORT platform was used to collect measurements of animal length (TOF) and optical extinction (EXT). Normalized optical extinction (norm.EXT) was previously established as a proxy for animal width. The raw data collected were imported and processed using the *easysorter* R package [118]. Processing removed non-animal objects such as bacterial clumps, shed cuticles, and next generation larval animals from the time-course data using the *mclust* R package [119].

### 3.3.3. High-throughput fitness assay for linkage mapping

For RIAIL phenotyping, we used a high-throughput fitness assay previously described [117]. In brief, populations of each strain were propagated on NGMA plates for four generations after which gravid adults were bleach-synchronized and embryos from each strain were aliquoted at a

concentration of 25-50 embryos/ $\mu$ L into 96-well microtiter plates for a final volume of 50  $\mu$ L K medium. The next day, arrested L1s were fed HB101 bacterial lysate (Pennsylvania State University Shared Fermentation Facility, State College, PA; [159]) at a final concentration of 5 mg/mL in K medium and grown to the L4 larval stage for 48 hours at 20°C with constant shaking. Animals were then sorted using a COPAS BIOSORT platform during which time animal length and width were collected. Measurements collected by the COPAS BIOSORT were processed and analyzed using the *easysorter* R package [118]. Well populations of recombinant strains that contained more than 100 or fewer than three individuals were removed from further processing, resulting in an average of 25 independent replicate wells per strain. Differences among strains tested on different days were controlled using a linear model (*animal\_size ~ experiment\_date*). In this way, we address only the differences among strains caused by growth and the day-to-day experimental variance is controlled. These residual values are used for plotting.

### 3.3.4. Linkage mapping

310 RIAILs (set 2 RIAILs) were phenotyped using the high-throughput assay described above. Linkage mapping was performed for body size traits using the R package *linkagemapping* ([www.github.com/AndersenLab/linkagemapping](https://www.github.com/AndersenLab/linkagemapping)) as previously described [82]. The genotypic data and residual phenotypic data were merged using the *merge\_pheno* function with the argument *set = 2*. Quantitative trait loci (QTL) were detected using the *fsearch* function. This function calculates the logarithm of the odds (LOD) scores for each genetic marker and each trait as  $-n(\ln(1-R^2)/2\ln(10))$  where R is the Pearson correlation coefficient between the RIAIL genotypes at the marker and trait values [160]. A significance threshold based on a 5%

genome-wide error rate was calculated by permuting the phenotypic values of each RIAIL 1000 times. QTL were identified as the marker with the highest LOD score above the significance threshold. This marker was then integrated into the model as a cofactor and mapping was repeated iteratively until no further significant QTL were identified. Finally, the *annotate\_lods* function was used to calculate the effect size of each QTL. 95% confidence intervals were defined by a 1.5-LOD drop from the peak marker.

### **3.3.5. Generation of chromosome substitution strains (CSSs) and near-isogenic lines (NILs)**

CSSs were generated from a cross of the N2 and CB4856 strains. These strains were crossed and heterozygous hermaphrodite progeny were mated to each parental genotype for four generations followed by three generations of selfing to ensure homozygosity of the genome. For each cross, PCR amplicons for insertion-deletions (indels) on the left and right sides of chromosomes IV and V were used to confirm progeny genotypes and select non-recombinants within the introgressed region (**S1 Text**). CSSs were whole-genome sequenced to confirm their genotypes.

NILs were generated as previously described [82,86,161,162] by either backcrossing a selected RIAIL or NIL for six generations or *de novo* by crossing the parental strains N2 and CB4856 to create a heterozygous individual that was then backcrossed for six generations. PCR amplicons for indel variants were used to track the genomic interval (**S1 Text**). NILs were whole-genome sequenced to verify introgressions.

### **3.3.6. Statistical analysis of CSS and NIL results**

Growth dynamics for CSSs were tested using a modified version of the high-throughput fitness assay for linkage mapping. Animals were propagated on NGMA plates for two generations

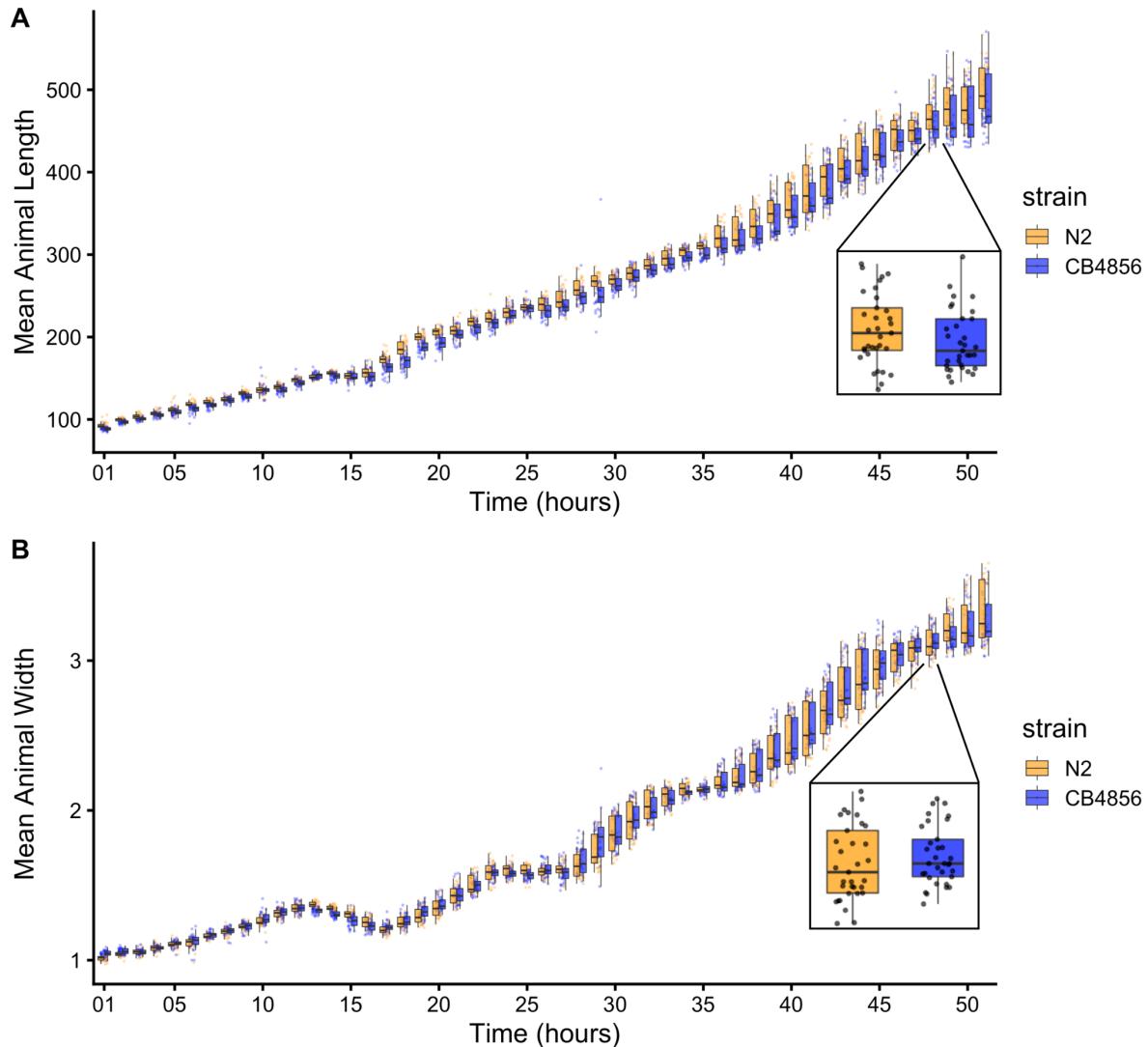
before gravid adults were bleach-synchronized and embryos from each strain were aliquoted at a concentration of 1 embryo/ $\mu$ L into 12-well, flat bottom culture plates. After three days, gravid adults were bleach-synchronized and embryos were titered into 96-well microtiter plates at a concentration of 50 embryos/ $\mu$ L for a final volume of 50  $\mu$ L K medium. The next day, arrested L1s were fed HB101 live bacterial food at a final concentration of OD20. Animals were grown for 48 hours at 20°C with constant shaking and then scored using the COPAS BIOSORT platform as before. The raw data collected were again imported and processed using the *easysorter* R package [118]. Processing removed non-animal objects such as bacterial clumps, shed cuticles, and next generation larval animals from the time-course data using the *mclust* R package [119]. Complete pairwise strain comparisons were performed using the *TukeyHSD* function [163] on an ANOVA model with the formula *phenotype ~ strain*. A *p*-value of *p* < 0.05 was used as a threshold for statistical significance. Recapitulation was defined by the significance and direction of effect the CSS or NIL had compared to the parental strains.

## 3.4. RESULTS

### 3.4.1. Growth dynamics of the N2 and CB4856 strains

To precisely evaluate *C. elegans* growth dynamics, we previously developed a high-throughput growth assay that integrates image-based and flow-based devices to quantify the growth of thousands of animals over developmental time [87]. We used this assay to collect body size measurements of N2 and CB4856 animals over the course of larval development from the L1 stage through the L4 stage. Briefly, populations of 100,000 animals were cultured in flasks in triplicate for each strain. Every hour after feeding, we sampled the population from each flask

(approximately 300 animals per flask), collected images, and measured length (TOF) and width (norm.EXT) of sampled animals using the COPAS BIOSORT platform (**Figure S3-1**). From these raw body size measurements, we removed non-animal objects using model-based clustering and generated summary statistics to study population changes (**Figures S3-2 and S3-3**, see **Methods**). Here, we report mean length and mean width of animals over 51 consecutive developmental time points (**Figure 3-1**). Overall, we observed little divergence in growth behavior between the two strains. As previously reported, we detected continuous growth punctuated by periods of discontinuous growth rate, resulting in visible shifts in length and width over time. Although growth behavior is consistent in both N2 and CB4856 animals, we capture significant differences in animal length and width at individual time points, particularly early in development (**Figure S3-4**). As animals age we identify fewer time points with significant differences likely due to increased population variance. We find that across all instances where there is a significant difference in animal length, N2 is consistently longer than CB4856. However, this is not the case in animal width as we observe time points where CB4856 is wider than N2 and others where it is thinner.

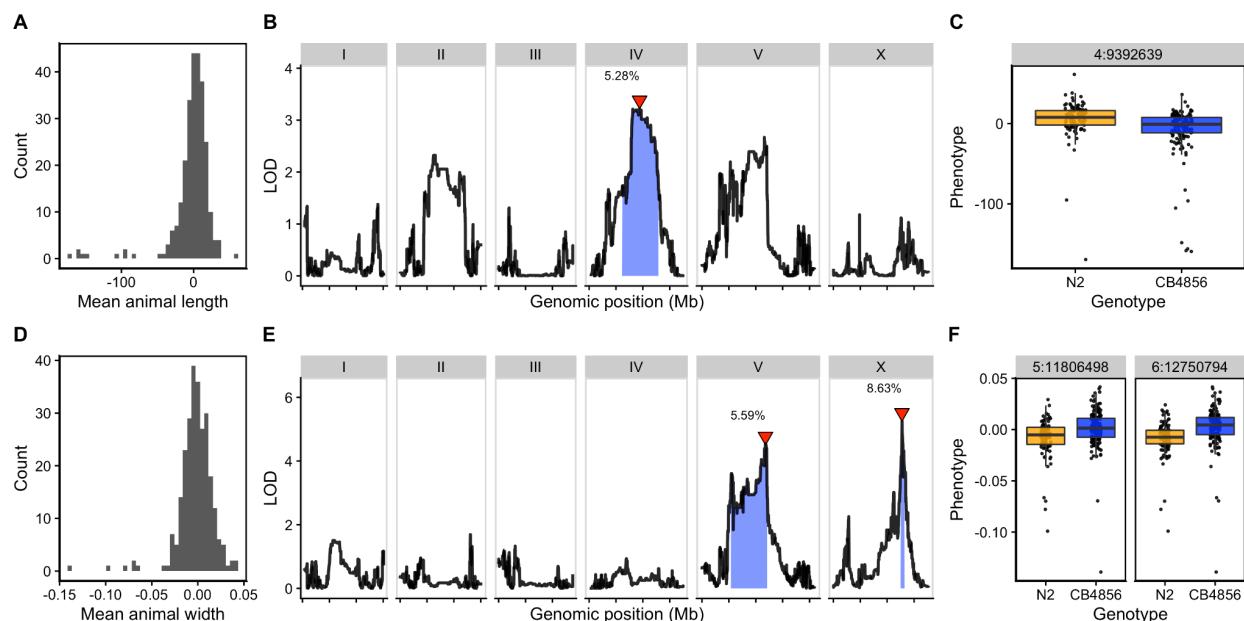


**Figure 3-1. Quantitative measurements of growth for N2 and CB4856 animals.** Tukey boxplots of mean length (A) and mean width (B) for the N2 (orange) and CB4856 (blue) strains over developmental time. The horizontal line in the middle of the box is the median, and the box denotes the 25th to 75th quantiles of the data. The vertical line represents the 1.5 interquartile range. Inset plots magnify mean animal size measurements from hour 48. Each point corresponds to the means length or mean width of a population of animals in each well.

### 3.4.2. Identification of QTL underlying variation in growth

As a complex trait, developmental growth is likely influenced by many genes as well as the interactions among them. To investigate the genetic basis of differences in growth, we assessed the development of a panel of 310 RIAILs derived from a cross between the N2 and CB4856 strains (set 2 RIAILs, see **Methods**). In lieu of collecting measurements throughout development, we used body size as a convenient proxy for developmental progression, where fast growth corresponds to large size and slow growth corresponds to small size. After 48 hours of growth, we collected measurements of length and width using a high-throughput fitness assay, and removed wells containing more than 100 or fewer than three animals from downstream processing (see **Methods**). Doing so, we observed a distribution of both length (**Figure 3-2A**) and width (**Figure 3-2D**) among the RIAILs, indicating that growth rate varies in the strain population. Next, we mapped body length and width separately and obtained three significant QTL (**Figure 3-2B,E**). The length-associated QTL, spanning the center of chromosome IV, and the width-associated QTL on the center of chromosome V independently explain approximately 5% of the phenotypic variation among the RIAILs. The third QTL on the right arm of chromosome X explains slightly more variation at 8.6% (**Table 3-1**). Notably, not only did we find distinct QTL for length and width, we also observed QTL with opposite effects on body shape. Strains with the N2 allele on chromosome IV were longer than strains with the CB4856 allele at this location (**Figure 3-2C**). By contrast, strains with the CB4856 alleles on chromosomes V and X were wider than strains with the N2 allele at these loci (**Figure 3-2F**). The identification of distinct QTL for length *vs.* width indicates that body shape is influenced by multiple genetic mechanisms. Additionally, we scanned the genome for interactions between

pairs of genomic markers that could affect the phenotypic distribution of length or width in the RIAILs and identified no significant interactions (**Figures S3-5 and S3-6**). These data suggest that the three identified loci contain variants that uniquely influence growth rate along multiple axes, where each locus independently affects the longitudinal or circumferential growth of animals.



**Figure 3-2. Linkage mapping identifies three QTL associated with body size.** Histogram of normalized mean body length (**A**) and mean body width (**D**) of the RIAIL population. (**B, E**) Linkage mapping results for mean body length or mean body width are shown with genomic position (x-axis) plotted against the logarithm of the odds (LOD) score (y-axis). X-axis tick marks denote every 5 Mb. Significant QTL are denoted by a red triangle at the peak marker, and blue shading shows the 95% confidence interval around the peak marker. The percentage of the total phenotypic variance in the RIAIL population that is explained by each QTL is shown above the peak marker. (**C, F**) Tukey box plots show the normalized mean length or width (y-axis) of

RIAILs split by genotype at the marker with the maximum LOD score (x-axis). Populations of recombinant strains were grown in independent wells. Each point corresponds to the mean value of all well means. Boxes for data from strains with the N2 allele are colored orange, and boxes for data from strains with the CB4856 allele are shown in blue.

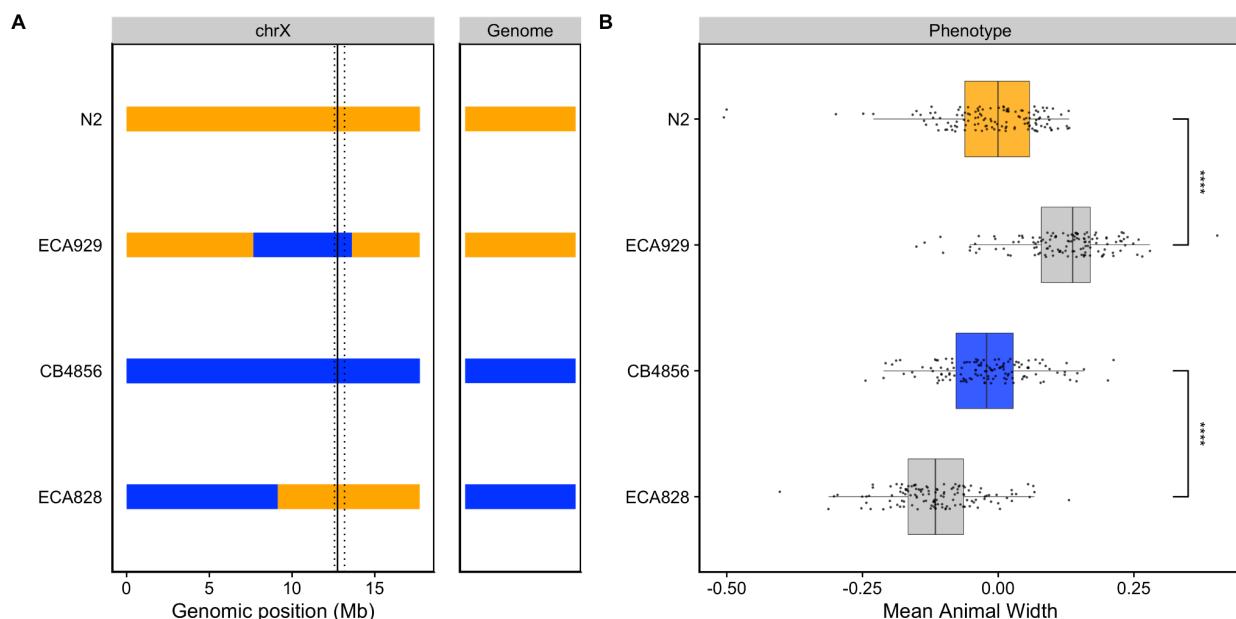
**Table 3-1. Body size QTL**

Trait	Chromosome	Interval (bp)	Peak	LOD	Variance explained (%)	Effect size
Length	IV	6,211,685 - 12,868,784	9,392,639	3.22	5.28	-0.229
Width	V	5,371,124 - 12,112,105	11,806,498	4.54	5.59	0.236
Width	X	12,565,734 - 13,173,080	12,750,794	5.25	8.63	0.293

### 3.4.3. Validation of loci associated with differences in animal size

To validate whether genetic variation between the N2 and CB4856 strains contributes to differences in animal size, we generated chromosome substitution strains (CSSs) for chromosomes IV and V in which the entire chromosome from the N2 strain was introgressed into the CB4856 genetic background and vice versa. We also constructed reciprocal near-isogenic lines (NILs) for chromosomes IV, V, and X. These NILs contain a small genomic segment derived from one parent strain introgressed into the genetic background of the other parent strain. We then measured the length and width of animals after 48 hours of growth and calculated statistical significance in a pairwise manner for each strain (see **Methods**). For the chromosomes IV and V QTL, we were unable to recapitulate the results observed in the linkage mapping (**Figures S3-7 and S3-8**). These two QTL each explain only 5% of the total phenotypic variation among the RIAILs and have the smallest effect sizes among the three detected QTL

(**Table 3-1**). The inability to validate these QTL suggests a complex genetic architecture that cannot be explained by isolating these loci using CSSs and NILs, or a lack of power to detect differences driven by these QTL in the CSSs and NILs. By contrast, we successfully validated the chromosome X QTL by observing that genotype significantly contributed to differences in body width of NILs (**Figure 3-3**). The strain with the CB4856 allele on chromosome X crossed into the N2 genetic background was significantly wider than the N2 strain (Tukey's HSD,  $p$ -value =  $1.29e^{-10}$ ). Similarly, the strain with the N2 chromosome X region introgressed into the CB4856 genetic background was significantly thinner than the CB4856 strain (Tukey's HSD,  $p$ -value =  $1.29e^{-10}$ ). These results confirmed that genetic variation between the N2 and CB4856 strains on chromosome X contributes to the difference in body width between these strains.



**Figure 3-3. NILs validated the chromosome X QTL. (A)** Strain genotypes are displayed as colored rectangles (N2: orange, CB4856: blue) for chromosome X (left) and in general for the rest of the chromosomes (right). The solid vertical line represents the peak marker of the QTL.

The dashed vertical lines represent the confidence interval. **(B)** Residual mean animal width (x-axis) is plotted as Tukey box plots against strain (y-axis). Each point corresponds to the mean width of a population of animals from each well. The boxes for the parental strains are colored: N2, orange; CB4856, blue. Statistical significance was calculated by Tukey's HSD ( $**** = p\text{-value} < 0.0001$ ).

#### 3.4.4. Identification of candidate genes in the chromosome X QTL

To identify candidate genes that could underlie variation in body width, we investigated the genes in the chromosome X interval in the N2 strain. We found 151 genes present in this interval and eliminated 96 genes that had no genetic variation in the CB4856 strain (**Table 3-2**). Of the remaining 55 genes, 18 have genetic variation in the amino-acid sequence of a protein (protein-coding variation), and 34 have genetic variation that is not protein-coding (non-coding variation). However, protein-coding variation is just one way in which genetic variation can cause phenotypic variation. We also considered instances where genetic variation causes a change in gene expression. Using an expression QTL (eQTL) dataset that mapped expression differences in another panel of RIAILs (set 1) derived from N2 and CB4856 [83,164], we identified five genes with eQTL that map to our region of interest. Additionally, we found 17 other genes outside this genomic interval with eQTL that map to this interval, resulting in a total of 72 candidate genes, none of which are located within a hyper-divergent region [165].

To further narrow our list of genes, we inspected the functional descriptions and gene ontology (GO) annotations for the remaining 72 candidate genes. When considering the 21 genes with protein-coding variation and/or eQTL, one candidate (*ppk-3*) stood out. *ppk-3*, or phosphatidylinositol phosphate kinase 3, is an ortholog of the mammalian PIKfyve. These

kinases play important roles in cell communication and membrane trafficking [166]. Notably, mutations in *ppk-3* are responsible for a range of developmental defects, including embryonic lethality, developmental arrest, and larval growth delay [167]. Investigating the sequence read alignments of the N2 and CB4856 strains at the *ppk-3* locus using the Variant Browser on CeNDR [69], we observed a missense variant in the second exon predicted to encode a serine-to-threonine substitution (S43T). Although this variant is not in a predicted functional domain, it could alter protein function thereby contributing to the observed phenotypic difference. Aside from *ppk-3*, we identified two additional candidate genes when assessing the functional description for the 34 genes with non-coding genetic variation. The first, *nhr-25*, encodes a nuclear hormone receptor orthologous to Ftz-F1 in *Drosophila* and is required for proper molting and developmental control [168]. We observed a splice-site variant in the *nhr-25* locus that could disrupt proper RNA splicing. Interestingly, disruption of *nhr-25* often causes embryonic arrest; however, mutants that survive hatching display a squat body stature (Dpy phenotype), suggesting that *nhr-25* could play a role in body size and shape [169]. The second, *bcat-1*, encodes a branched-chain amino acid aminotransferase that, by RNAi screens, is shown to be required for embryonic and larval development [170]. In the *bcat-1* locus we observe a variant in an intron and in the three prime untranslated region. Together, these results suggest that one or more genes on chromosome X are candidates that need additional study to explain the variation that we observe in animal growth.

**Table 3-2. Genes in QTL interval for chromosome X**

No variation <sup>a</sup>	Protein-coding variation and/or eQTL <sup>b</sup>	Non-coding variation <sup>c</sup>	Other eQTL that map to interval <sup>d</sup>	Total
96	21	34	17	168

<sup>a</sup>Genes within genomic interval with no genetic variation

<sup>b</sup>Genes within genomic interval with protein-coding variation and/or an eQTL that maps to this interval

<sup>c</sup>Genes within genomic interval with non-coding variation and no eQTL that maps to this interval

<sup>d</sup>Genes outside genomic interval with eQTL that maps to this interval

### 3.5. DISCUSSION

Here, we investigated larval growth of N2 and CB4856 animals from the L1 stage to the L4 stage. Although we observed similarities in the dynamics of growth, we also saw differences in the size of animals across developmental time. We used linkage mapping to investigate these differences and identified three small-effect QTL associated with variation in body size. Two QTL underlie variation in animal width, and a single non-overlapping QTL contributes to differences in animal length. Using NILs, we validated the width-associated QTL on chromosome X and identified candidate genes that could underlie variation in width. Taken together, our results demonstrate the power of leveraging natural genetic variation to examine the genetic architecture of complex traits such as body size and shape.

#### 3.5.1. A complex genetic architecture underlies differences in body size

As a complex life history trait, developmental growth could be influenced by several loci [171]. In this study, we report three size-associated QTL. Strikingly, we find loci that decouple components of body size, revealing a complex genetic system that influences growth along different axes of the body. Evidence for genetically separate modules underlying distinct aspects of a single trait has been observed in studies of *C. elegans* behavioral patterns where linkage

mapping studies using a panel of RIAILs (set 1, [164]) identified distinct loci underlying separate aspects of avoidance response to thermal stimuli [172]. Here, we identify distinct QTL for length and width, suggesting that different genetic mechanisms control animal growth along the length *vs.* width directions. This finding is particularly interesting given the differences in general growth dynamics of length compared to width that we found here (**Figure 3-1**) and previously [87] a simultaneous increase in length and decrease in width at the transition between larval stages.

The results of the linkage mapping experiment identified two broad peaks on chromosomes IV and V associated with length and width respectively, as well as a narrow peak on chromosome X for width. Although we successfully validated the width-associated QTL on chromosome X (**Figure 3-3**), we were unable to validate the other two QTL (**Figures S3-6** and **S3-7**). Our inability to recapitulate the results observed in the linkage mapping might be driven by several factors. First, many loci spread across the genome could underlie variation in body size. Under this polygenic model, any region can harbor variants driving our observed phenotypic difference through additive and/or non-additive effects. The contribution of polygenicity to phenotypic variance has previously been explored in *C. elegans*. Studies of fertility and body size in the *C. elegans* multiparental experimental evolution (CeMEE) panel found that a significant fraction of phenotypic variance, nearly 40% for fertility, can be explained by polygenicity [173]. Second, the intervals could contain QTL of opposing effects, making it difficult to recapitulate the results observed in the mapping using NILs. Notably, researchers have observed patterns of polygeny and antagonistic-effect loci when investigating *C. elegans* growth and reproduction in nickel stress [174]. Third, it is possible that the QTL effects are

smaller than 5% and we are underpowered to detect differences driven by these QTL in the CSSs and NILs.

### 3.5.2. Candidate genes for variation in body size

Genetic variants underlying complex traits are often elusive [21,175]. Ultimately, when searching for QTL, we aim to identify genes contributing to the variation in phenotypes among individuals. Here, we identified candidate genes located in the interval of the chromosome X QTL (**Table 3-2**). However, complex traits, such as body size, are likely affected by many genes. For example, recent studies of human genetic variation using data from 5.4 million individuals report finding over 12,000 independent loci associated with height [176]. In the laboratory strain of *C. elegans*, we know many loci that quantitatively affect body size and shape. Mutations in these genes span various classes, including abnormal pharyngeal pumping (Eat), egg-laying defective (Egl), uncoordinated (Unc), abnormal dauer formation (Daf), and several cuticle and body shape classes (Dpy, Lon, Sma, Rol, Sqt) [59,130,177]. The polygenic nature of complex traits is a recognized barrier in identifying the genes contributing to phenotypic variation in a population [174,178]. However, we believe that molecular analysis of loci that underlie variation in development-associated traits in natural populations of *C. elegans* is essential to deciphering the evolutionary significance of developmental canalization.

### 3.5.3. Comparison with previous QTL studies of *C. elegans* growth

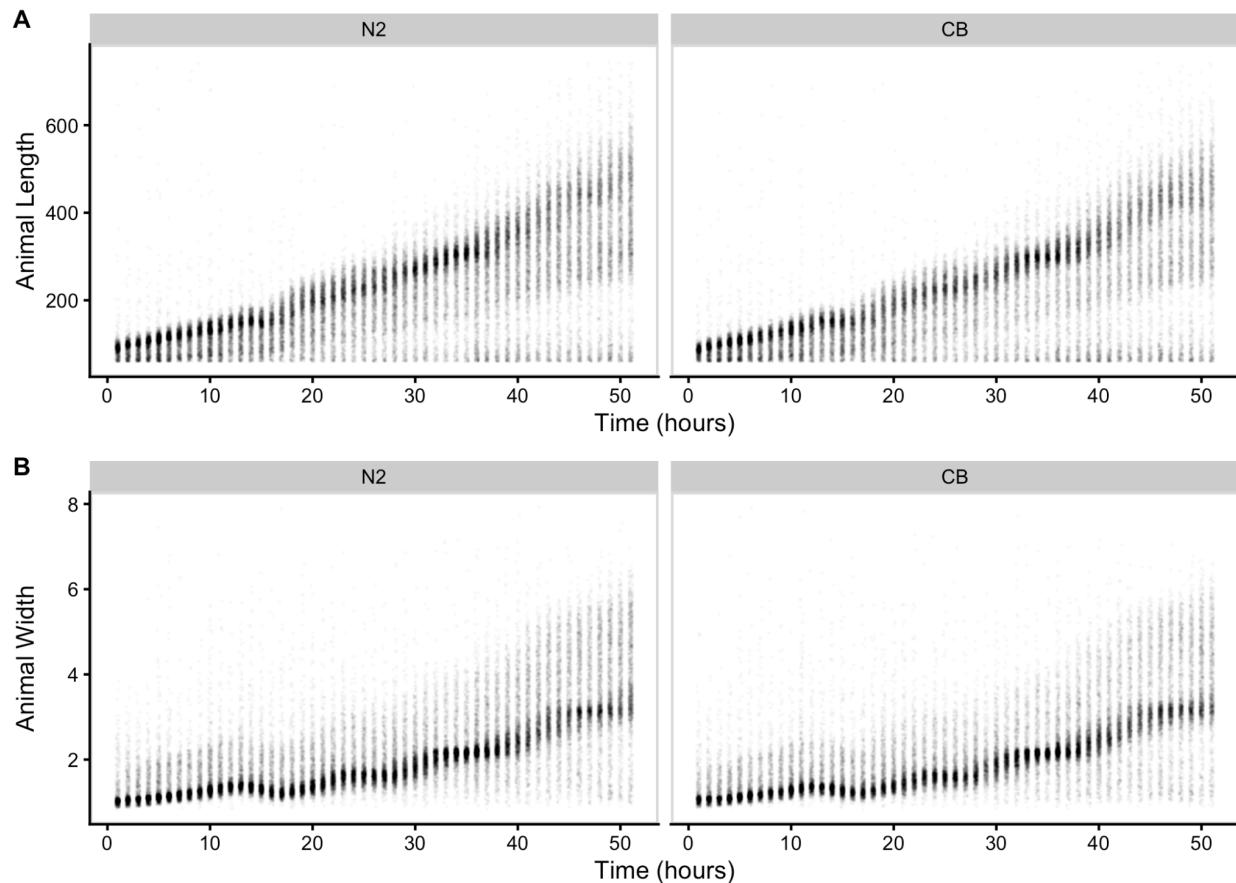
Our mapping results both recapitulate and expand upon previous QTL studies of growth in *C. elegans*. Previously, median body length of mixed-stage animals was mapped using the same panel of RIAILs (set 2) [117]. A single small-effect (5.7%) QTL in the center of chromosome IV

was found, consistent with our findings. Also in this study, the authors mapped median body width (norm.EXT) to three QTL on chromosomes III, IV, and X. We detected an overlapping genomic region on chromosome X in our current study. Discrepancy in the other QTL is likely caused by differences in the body size assay as the previous study measured mixed-stage animals and we focused on synchronized L4 animals. Additionally, others have mapped variation in animal length for a collection of introgression lines produced from the N2 and CB4856 strains at 48 hours after L1 arrest [179]. Here, investigators found five separate QTL on chromosome IV affecting body size. This result suggests the presence of several independent loci on chromosome IV each contributing to variation in length. Further investigation is necessary to determine whether the overlapping genomic region detected in our current study is in fact separate loci that independently contribute to variation in animal length. Most recently, a group using a *C. elegans* RIL population identified 18 QTL influencing various body-size traits at a range of temperatures, with the majority clustering on chromosome X [180]. This work not only demonstrates the genetic complexity underlying body-size phenotypes, but also suggests the presence of co-regulatory loci underlying plasticity. *C. elegans* gives investigators a powerful system to better our understanding of the genetic mechanisms that shape growth and environmental sensitivity in natural populations.

### **3.6. CONTRIBUTIONS**

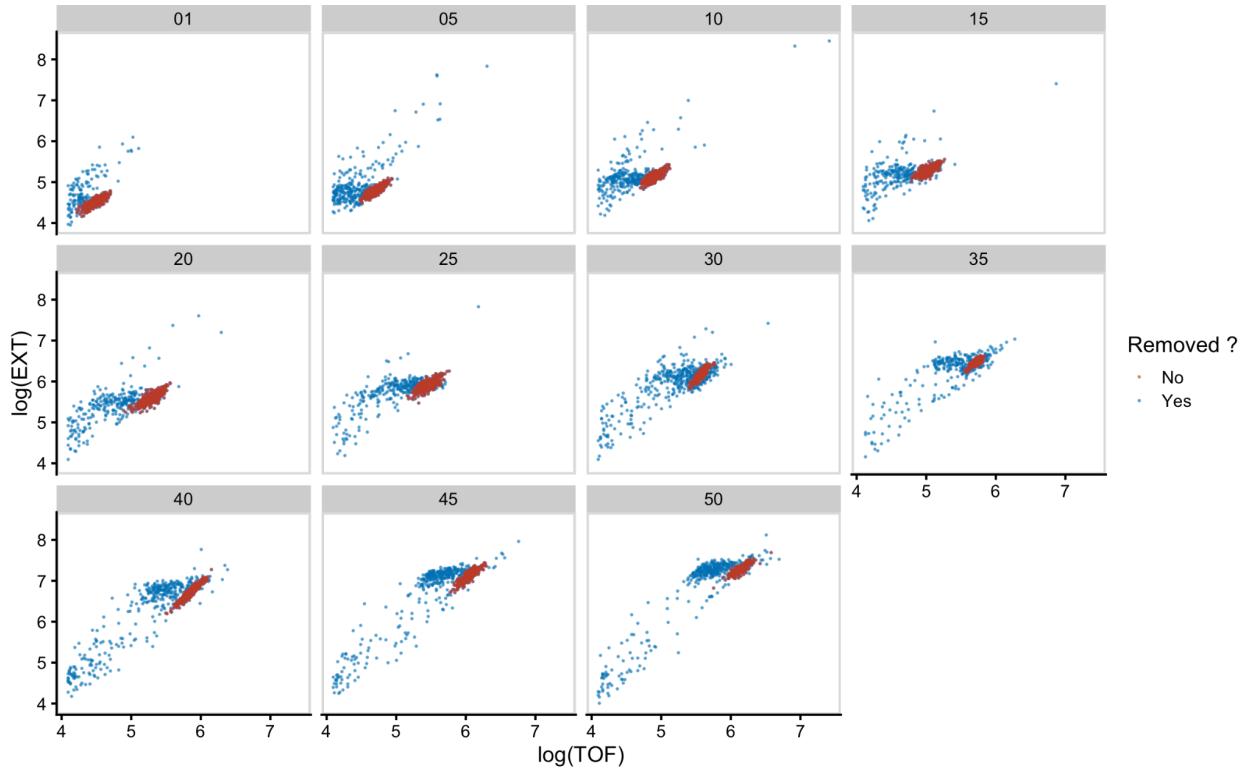
This work was supported by the NSF-Simons Center for Quantitative Biology at Northwestern University. Dr. Erik Andersen, Dr. Gaotian Zhang, and Sophia Gibson assisted with the high-throughput sampling.

### 3.7. SUPPLEMENT



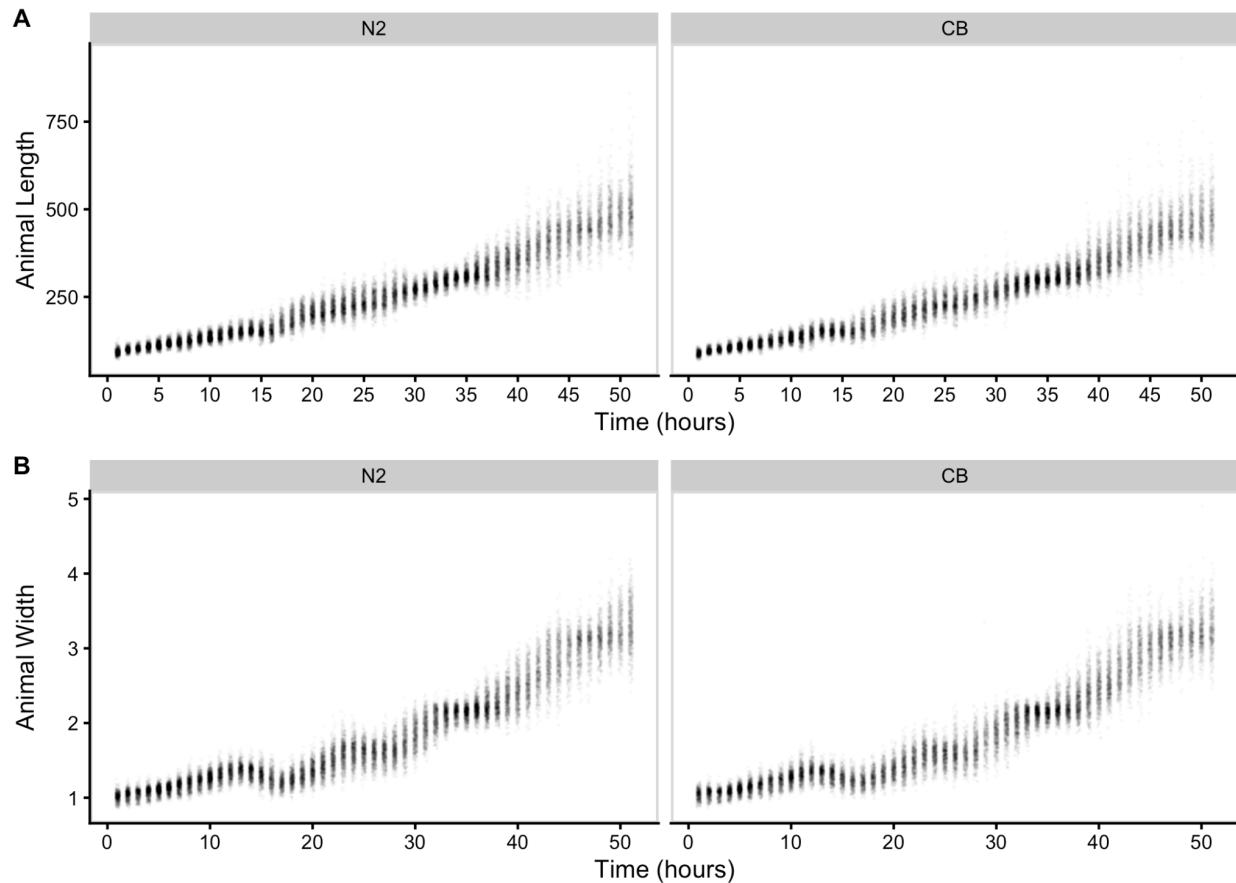
**Figure S3-1. Raw measurements of animal size.**

Raw COPAS BIOSORT of animal length (A) and width (B) for N2 and CB4856 animals are shown here.



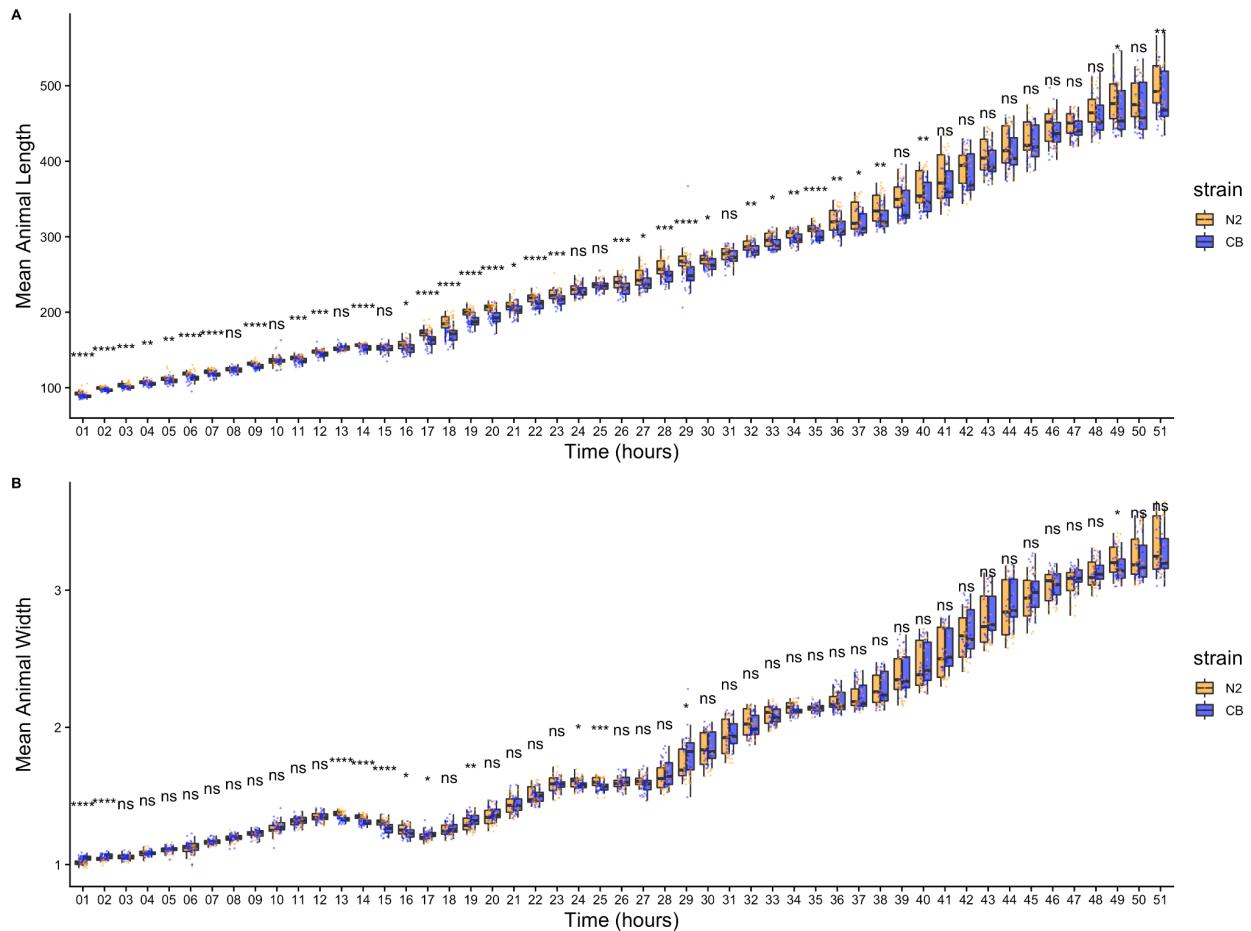
**Figure S3-2. Mixture modeling of COPAS BIOSORT data was used to prune data.**

Mixture models of Gaussian distributions were fit to log transformed animal length (x-axis) and log transformed optical extinction (y-axis). Data from each hour of the experiment was analyzed and processed to remove clusters that did not include animal objects. Panels indicate experimental hours from which data were taken.



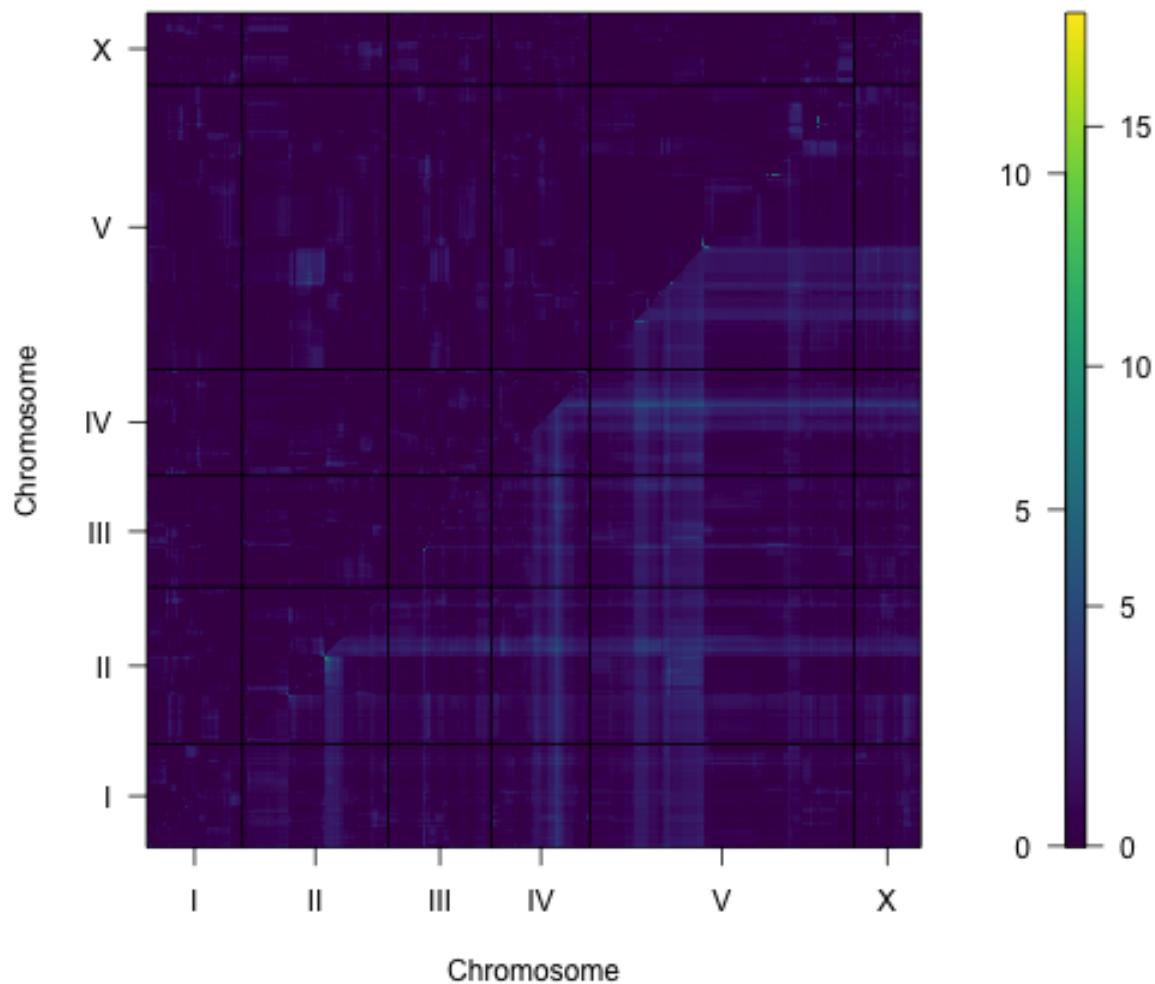
**Figure S3-3. Pruned measurements of animal size.**

COPAS BIOSORT data of animal length (A) and width (B) after the removal of non-animal objects using model-based clustering methods.



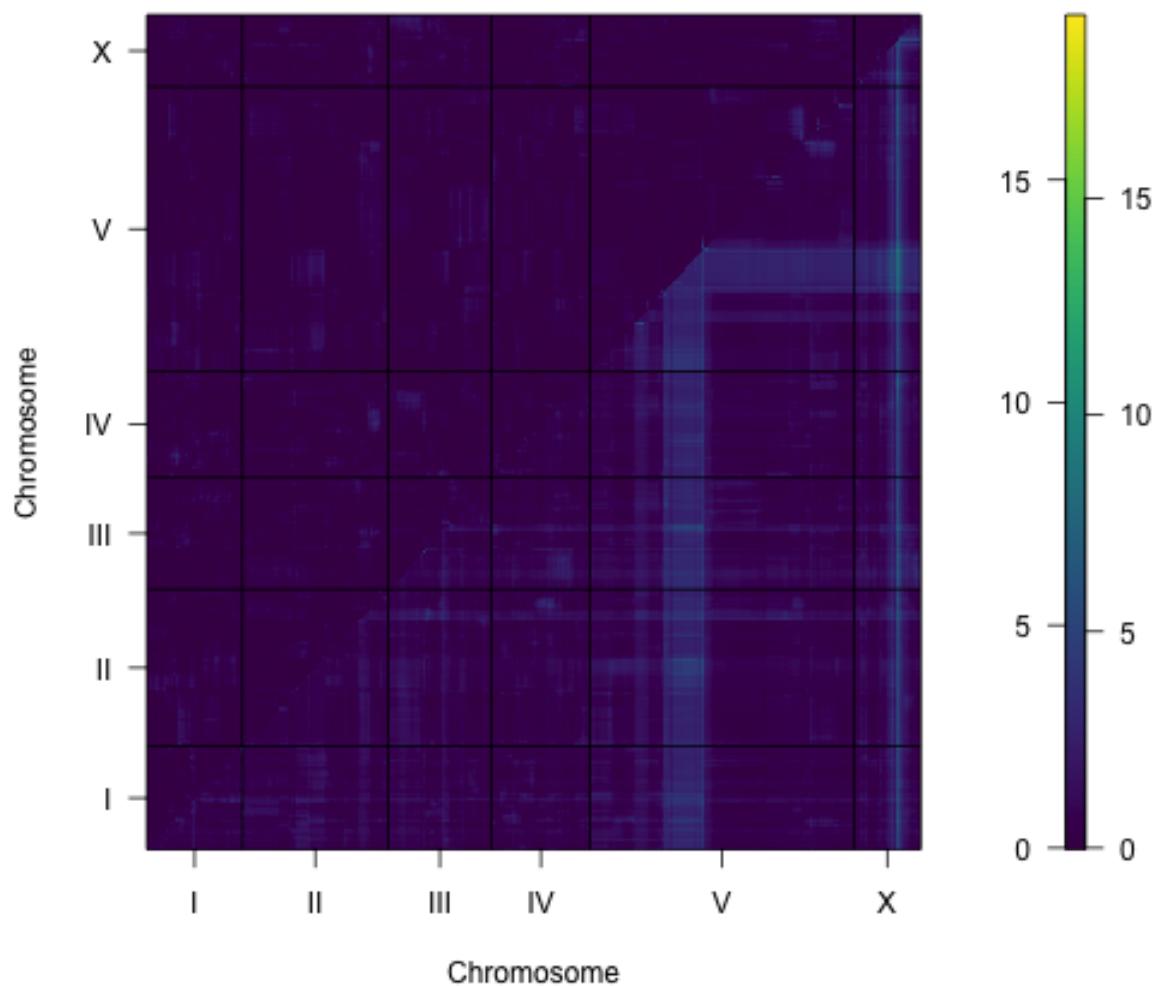
**Figure S3-4. Comparison of means across developmental time points.**

Tukey boxplots of mean length (**A**) and mean width (**B**) for N2 (orange) and CB4856 (blue) over developmental time. The horizontal line in the middle of the box is the median, and the box denotes the 25th to 75th quantiles of the data. The vertical line represents the 1.5 interquartile range. Inset plots magnify mean animal size measurements from hour 48. Each point corresponds to the mean of a population from each well. Statistical significance was calculated using a Wilcoxon test (ns = non-significant ( $p$ -value  $> 0.05$ ); \*, \*\*, \*\*\*, and \*\*\*\* = significant ( $p$ -value  $< 0.05, 0.01, 0.001$ , or  $0.0001$ , respectively).



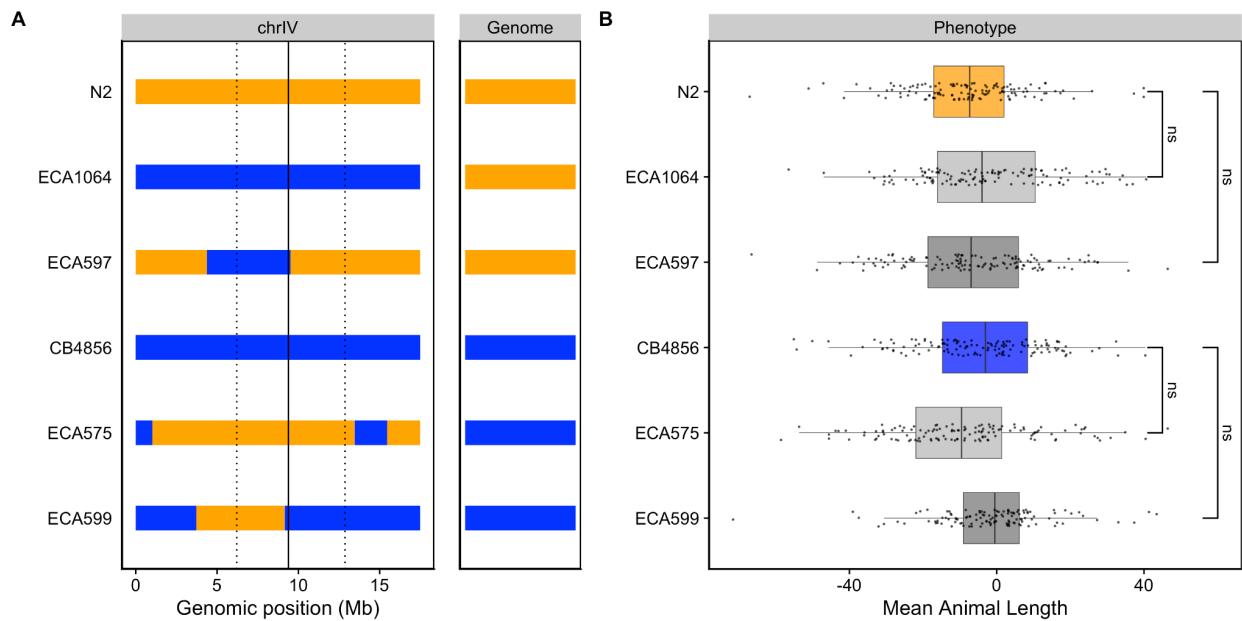
**Figure S3-5. Two dimensional genome scan for mean animal length.**

Log of the odds (LOD) scores are shown for each pairwise combination of loci, split by chromosome. The upper-left triangle contains the epistasis LOD scores and the lower-right triangle contains the LOD scores for the full model. LOD scores are colored, increasing from purple to green to yellow. The LOD scores for the epistasis model are shown on the left of the color scale and the LOD scores for the full model are shown on the right.



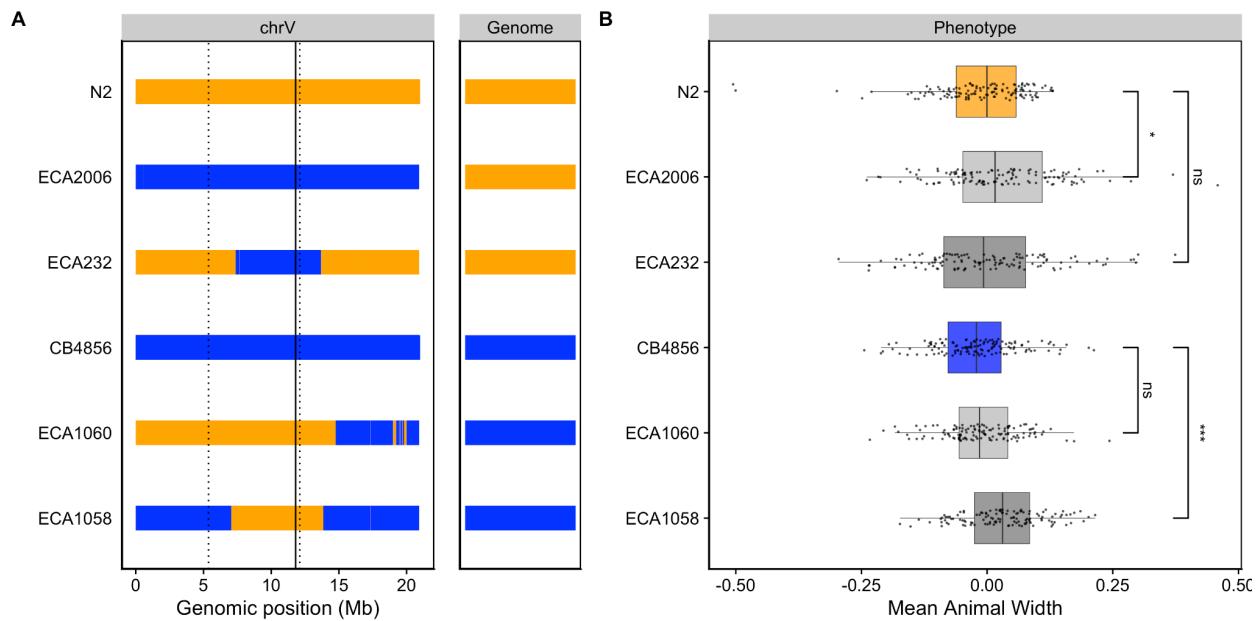
**Figure S3-6. Two dimensional genome scan for mean animal width.**

Log of the odds (LOD) scores are shown for each pairwise combination of loci, split by chromosome. The upper-left triangle contains the epistasis LOD scores and the lower-right triangle contains the LOD scores for the full model. LOD scores are colored, increasing from purple to green to yellow. The LOD scores for the epistasis model are shown on the left of the color scale and the LOD scores for the full model are shown on the right.



**Figure S3-7. Validating the chromosome IV length-associated QTL.**

Strain genotypes are displayed as colored rectangles (N2: orange, CB4856: blue) for chromosome IV (left) and in general for the rest of the chromosomes (right). The solid vertical line represents the peak marker of the QTL. The dashed vertical lines represent the confidence interval. (B) Residual mean animal length (x-axis) is plotted as Tukey box plots against strain (y-axis). Statistical significance was calculated by Tukey's HSD (ns = non-significant,  $p$ -value > 0.05).



**Figure S3-8. Validating the chromosome V width-associated QTL.**

Strain genotypes are displayed as colored rectangles (N2: orange, CB4856: blue) for chromosome V (left) and in general for the rest of the chromosomes (right). The solid vertical line represents the peak marker of the QTL. The dashed vertical lines represent the confidence interval. (B) Residual mean animal width (x-axis) is plotted as Tukey box plots against strain (y-axis). Statistical significance was calculated by Tukey's HSD (ns = non-significant,  $p$ -value  $>$  0.05; \* and \*\*\* = significant,  $p$ -value  $<$  0.05 or 0.001 respectively).

**S3-1 Text. Reagents used to generate CSSs and NILs.****List of RIAILs used:**Set 2:

QX115, QX123, QX134, QX144, QX160, QX161, QX175, QX226, QX227, QX231, QX240, QX241, QX242, QX243, QX244, QX245, QX248, QX250, QX252, QX253, QX254, QX258, QX261, QX263, QX264, QX265, QX266, QX267, QX268, QX269, QX270, QX271, QX272, QX273, QX274, QX275, QX276, QX277, QX278, QX279, QX280, QX281, QX282, QX283, QX284, QX285, QX286, QX287, QX288, QX289, QX290, QX291, QX293, QX294, QX295, QX296, QX297, QX298, QX299, QX300, QX301, QX302, QX303, QX304, QX305, QX306, QX307, QX309, QX310, QX311, QX314, QX315, QX316, QX318, QX319, QX320, QX321, QX322, QX323, QX324, QX325, QX326, QX327, QX328, QX329, QX330, QX331, QX332, QX333, QX334, QX335, QX336, QX337, QX338, QX339, QX340, QX341, QX343, QX345, QX346, QX347, QX348, QX349, QX350, QX352, QX353, QX354, QX355, QX356, QX357, QX358, QX359, QX360, QX361, QX362, QX363, QX364, QX365, QX366, QX367, QX368, QX369, QX370, QX371, QX372, QX373, QX374, QX375, QX376, QX377, QX378, QX379, QX380, QX381, QX382, QX383, QX384, QX385, QX386, QX387, QX390, QX391, QX392, QX393, QX394, QX395, QX396, QX397, QX398, QX399, QX400, QX401, QX402, QX403, QX404, QX405, QX406, QX407, QX408, QX409, QX410, QX411, QX412, QX413, QX414, QX416, QX417, QX418, QX419, QX420, QX421, QX423, QX424, QX426, QX427, QX428, QX429, QX430, QX431, QX432, QX433, QX434, QX435, QX436, QX437, QX438, QX439, QX440, QX441, QX443, QX444, QX445, QX446, QX447, QX448, QX449, QX450, QX451, QX452, QX453, QX454, QX455, QX456, QX457, QX458, QX459, QX460, QX461, QX463, QX464, QX465, QX466, QX467, QX468, QX469, QX470, QX471, QX472, QX473, QX474, QX475, QX476, QX477, QX478, QX479, QX480, QX481, QX482, QX483, QX484, QX485, QX486, QX487, QX488, QX489, QX49, QX490, QX491, QX492, QX493, QX494, QX495, QX496, QX497, QX498, QX500, QX501, QX503, QX506, QX508, QX511, QX512, QX513, QX514, QX515, QX517, QX520, QX521, QX523, QX524, QX525, QX526, QX527, QX528, QX529, QX530, QX531, QX533, QX534, QX538, QX539, QX540, QX542, QX545, QX549, QX550, QX551, QX553, QX554, QX555, QX556, QX557, QX559, QX560, QX561, QX563, QX564, QX565, QX570, QX572, QX573, QX574, QX579, QX580, QX583, QX585, QX587, QX588, QX59, QX594, QX596, QX597, QX598, QX61, QX62, QX72, QX78, QX85,

**Reagents to generate CSSs and NILs:**

Strain	Genotype	Constructed from	Left primer	Right primer
ECA232	<i>eanIR152</i> [V, CB4856>N2]	QX450xN2	oECA799 & oECA800	oECA745 & oECA746
ECA575	<i>eanIR324</i> [IV, N2>CB4856]	N2xCB4856	oECA1132 & oECA1133	oECA1135 & oECA1136
ECA597	<i>eanIR330</i> [IV, CB4856>N2]	ECA231xN2	oECA781 & oECA782	oECA857 & oECA858
ECA599	<i>eanIR332</i> [IV, N2>CB4856]	ECA598xCB4856	oECA781 & oECA782	oECA857 & oECA858
ECA828	<i>eanIR359</i> [X, N2>CB4856]	N2xCB4856	oECA1313 & oECA1314	oECA1246 & oECA1247
ECA929	<i>eanIR411</i> [X, CB4856>N2]	N2xCB4856	oECA1313 & oECA1314	oECA1246 & oECA1247
ECA1058	<i>eanIR433</i> [V, N2 > CB4856]	ECA1029xCB4856	oECA1408 & oECA1409	oECA1341 & oECA1342
ECA1060	<i>eanIR435</i> [V, N2 > CB4856]	ECA554xCB4856	oECA745 & oECA746	oECA763 & oECA764
ECA1064	<i>eanIR439</i> [IV, CB4856>N2]	N2xCB4856	oECA1131 & oECA1132	oECA1135 & oECA1136
ECA2006	<i>eanIR446</i> [V, CB4856>N2]	N2xCB4856	oECA1141 & oECA1142	oECA1147 & oECA1148

**Primers:**

Primer	Genomic position	Sequence
oECA745	V:13,110,045	tgcagagggtggagtaaccct
oECA746	V:13,110,045	ctcggtctctccccactaa
oECA763	V:15,121,356	cgcacattcttattctggcg
oECA764	V:15,121,356	atcggccgtttcacctga
oECA781	IV:5,110,734	gagcacttggcgacttcg
oECA782	IV:5,110,734	tccgggcaaatttagtgtggc
oECA799	V:7,862,556	ttctcgctactggAACACGC
oECA800	V:7,862,556	tcaagaAGCGTGGAAAGTCT
oECA1131	IV:1,039,851	taccCACCGcatcaaaacca
oECA1132	IV:1,039,851	acaggcgttcaaAGACACCA
oECA1135	IV:17,317,014	ttcagacaggAAAGCGCCT
oECA1136	IV:17,317,014	gttgagAGATCCGGACCgAC
oECA1141	V:144,547	ctcatggagtaacctggc
oECA1142	V:144,547	cggtgacaACGGAGAATCCA
oECA1147	V:20,622,851	gtttAGTACCAAGCGGGCAT
oECA1148	V:20,622,851	tgcattCCGACCCAAGAGAC
oECA1246	X:11,696,902	tgcggTGGACTTTCTTGT
oECA1247	X:11,696,902	gtcccAGCATGTAACCGTCT
oECA1313	X:8,038,337	gctgtcaggACTGGATGTA
oECA1314	X:8,038,337	tgctttCTGATCTGTGCCGT
oECA1341	V:7,104,674	cccatccccacaATGTTCG
oECA1342	V:7,104,674	aatcgacgAGTGGCACTTGT
oECA1408	V:3,778,859	caCgtGCCCTTGCAATGA
oECA1409	V:3,778,859	gagctcccgaaaACTCGAA

# 4

## **easyXpress: An R package to analyze and visualize *C. elegans* microscopy data generated using CellProfiler**

What I feel to be the most rewarding project of my doctoral work forms the basis of this chapter. I joined the Andersen Lab with an introductory level of coding experience in Python and little knowledge of R. Making quick study of the new programming language, I have since fallen head over heels in love with it. Now, I use R to not only analyze data but also generate reports, build websites, and even write this thesis (thank you, bookdown! [181]). This chapter describes the R package I helped develop for handling high-throughput *C. elegans* microscopy data which was published in PLoS ONE in 2021 [182]. I am beyond ecstatic to contribute to the open-source and *#rstats* worlds and hope to continue to do so in my future.

## 4.1. ABSTRACT

High-throughput imaging techniques have become widespread in many fields of biology. These powerful platforms generate large quantities of data that can be difficult to process and visualize efficiently using existing tools. We developed easyXpress to process and review *C. elegans* high-throughput microscopy data in the R environment. The package provides a logical workflow for the reading, analysis, and visualization of data generated using CellProfiler's WormToolbox. We equipped easyXpress with powerful functions to customize the filtering of noise in data, specifically by identifying and removing objects that deviate from expected animal measurements. This flexibility in data filtering allows users to optimize their analysis pipeline to match their needs. In addition, easyXpress includes tools for generating detailed visualizations, allowing the user to interactively compare summary statistics across wells and plates with ease. Researchers studying *C. elegans* benefit from this streamlined and extensible package as it is complementary to CellProfiler and leverages the R environment to rapidly process and analyze large high-throughput imaging datasets.

## 4.2. INTRODUCTION

Developments in high-throughput imaging techniques have led to a rapid increase in these data. Researchers are able to move away from the laborious manual collection of images that typically limits large-scale analyses [183]. Furthermore, these advances have enabled scientists to collect data of intact cells, tissues, and whole-organisms with increased temporal and spatial resolution [184]. However, typical users require software methods for efficient handling, analysis, and visualization to make the most of these extensive image datasets.

*C. elegans* is a globally distributed, free-living roundworm nematode that is amenable to many types of experimental biology. The *C. elegans* cell lineage is completely characterized [34], and the *C. elegans* connectome is completely mapped [185], making these animals an exemplary model for developmental biology and neurobiology. The species can also be rapidly reared in large, genetically diverse populations in laboratory settings, providing unparalleled statistical power for experimental biology compared to any other metazoan [186]. Furthermore, metabolic and developmental pathways in *C. elegans* are conserved in humans [40]. High-throughput imaging technologies can improve *C. elegans* studies by increasing experimental efficiency, scalability, and quality. Existing systems for automated image acquisition, such as the Molecular Devices ImageXpress platforms generate images of nematodes that can be analyzed with software like CellProfiler's WormToolbox [187] to extract nematode phenotype information. This software uses probabilistic nematode models trained on user-selected animals to automate the segmentation of nematodes from the background of images in high-throughput. As a result, CellProfiler's WormToolbox is able to measure hundreds of phenotypes related to animal shape, intensity, and texture. Implementing this software for large-scale imaging experiments can generate large quantities of data that requires additional analysis software for reliable and reproducible handling, processing, and visualization. CellProfiler Analyst was developed to offer tools for the analysis of image-based datasets, but this software is not integrated with modern statistical environments. We sought to design a resource that facilitates the exploration of CellProfiler data in the R environment [188], where this limitation can be eliminated. The R language provides extensive open-source statistical and data visualization tools that are well supported by the user community. In leveraging R, we are

able to create a flexible tool that can be rapidly integrated with other statistical R packages to suit project-specific analysis needs.

We developed easyXpress, a software package for the R statistical programming language, to assist in the processing, analysis, and visualization of *C. elegans* data generated using CellProfiler. easyXpress provides tools for quality control, summarization, and visualization of image-based *C. elegans* phenotype data. Built to be complementary to CellProfiler, this package provides a streamlined workflow for the rapid quantitative analysis of high-throughput imaging datasets.

## 4.3. MATERIALS AND METHODS

### 4.3.1. Preparation of animals for imaging

Bleach-synchronized animals were fed *E. coli* HB101 bacteria suspended and allowed to develop at 20°C with continuous shaking. Animals in 96-well microtiter plates were titered to approximately 30 animals per well. Prior to imaging, animals were treated with sodium azide (50 mM in 1X M9) for 10 minutes to paralyze and straighten their bodies.

### 4.3.2. Imaging

Animals in microtiter plates were imaged at 2X magnification with an ImageXpress Nano (Molecular Devices, San Jose, CA). The ImageXpress Nano acquires brightfield images with a 4.7 megaPixel CMOS camera and are stored in 16-bit TIFF format. The images were processed using CellProfiler software (for details see <https://github.com/AndersenLab/CellProfiler>).

#### 4.3.3. Paraquat dose response

A 1.5 M solution of paraquat (Methyl viologen dichloride, Sigma, 856177-1G) was prepared in sterile water, aliquoted, and frozen at -20°C until used. Experimental animals were grown at 20°C and fed OP50 bacteria spotted on modified nematode growth medium, containing 1% agar and 0.7% agarose to prevent animals from burrowing. After three generations of passaging, animals were bleach-synchronized and embryos were transferred to the wells of 96-well microplates. Each well contained approximately 30 embryos in 50 µL of K medium [189]. Microplates were incubated overnight at 20°C with continuous shaking. The following day, arrested L1 animals were fed HB101 bacteria suspended in K medium. At the time of feeding, the animals were also exposed to paraquat at one of six concentrations (0, 7.81, 31.25, 125, 500, 2000 µM) by serial dilution of a freshly thawed aliquot of 1.5 M paraquat solution. The final volume in each well after dosing and feeding was 75 µL. The animals were then grown for 48 hours at 20°C with continuous shaking, afterwards the microplates were imaged to assess the effects of paraquat exposure on nematode development.

#### 4.3.4. Naming conventions

Several functions in the *easyXpress* package require specific naming conventions to work properly. For full details regarding essential file naming and directory structure see the package repository (<https://github.com/AndersenLab/easyXpress>). Importantly, when using the Metadata module in CellProfiler to extract information describing your images, specific column names are suggested (Table 1).

**Table 4-1. Suggested naming conventions for CellProfiler metadata.**

Image_FileName_RawBF	Image_PathName_RawBF	Metadata_Date	Metadata_Experiment	Metadata_Plate	Metadata_Magnification	Metadata_Well
20191119-growth-p05-m2X_C03.TIF	/CellProfiler/example/raw_images	20191119	growth	p05	m2X	C03
20191119-growth-p06-m2X_C09.TIF	/CellProfiler/example/raw_images	20191119	growth	p06	m2X	C09
20191119-growth-p09-m2X_C06.TIF	/CellProfiler/example/raw_images	20191119	growth	p09	m2X	C06

The naming of “Metadata\_Plate” and “Metadata\_Well” are essential to the *setflags()*, *viewPlate()*, *viewWell()*, and *viewDose()* functions. Additionally, “Image\_fileName\_RawBF” and “Image\_PathName\_RawBF” are necessary for the proper function of *viewDose()*.

#### 4.3.5. Data availability

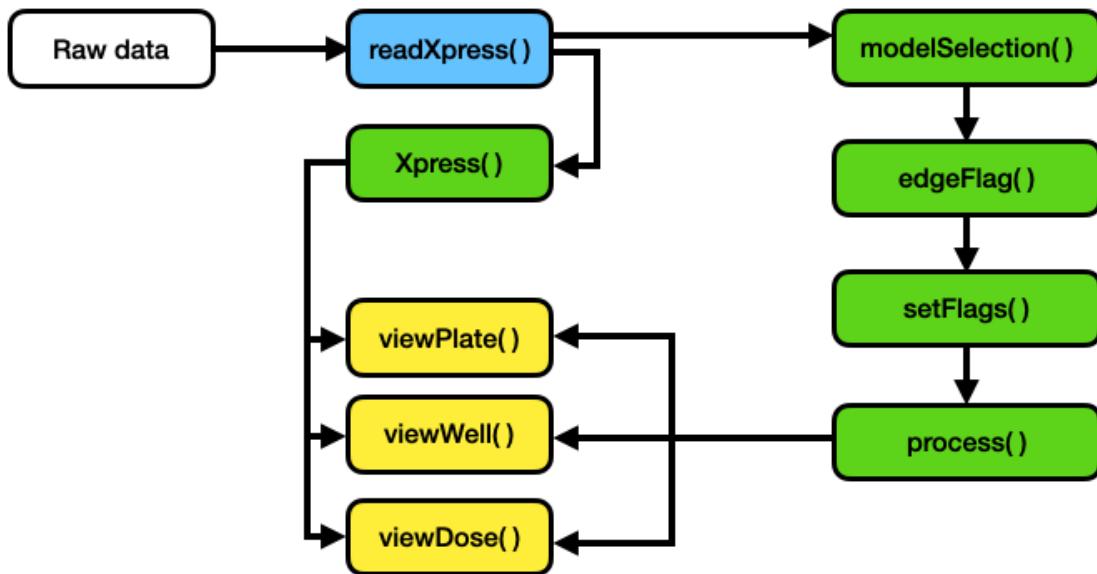
The entirety of the easyXpress package is written in the R language and is free to install across any system supporting R, including Linux, MacOSX, and Windows. The complete source code, example data, extensive documentation, and installation details are available on GitHub. A tutorial on the usage of easyXpress and the available functions, can be found at <https://rpubs.com/jnyaanga/765641>. This package is open-source; for updates and to submit comments, visit <https://github.com/AndersenLab/easyXpress>.

### 4.4. RESULTS

#### 4.4.1. Design and implementation

The easyXpress package is designed to be simple and accessible to users familiar with the R environment. The easyXpress package comprises nine functions for reading, processing, and visualizing large high-throughput image-based datasets acquired from microplate-based assays processed with CellProfiler (**Figure 4-1**). Because our software is built to handle CellProfiler

data as input, we suggest users review the overview and applications of CellProfiler as a prerequisite description of data generation [187]. Below, we describe the workflow for users to analyze their image data with easyXpress.



**Figure 4-1. easyXpress workflow.** The suggested workflow for using the easyXpress package starts with raw data generated from CellProfiler. For more information on implementing CellProfiler to generate data, see <https://github.com/AndersenLab/CellProfiler>. Reading steps are shown in blue, processing steps are shown in green, and visualization steps are shown in yellow.

#### 4.4.2. Data import and model assignment

To read in CellProfiler data files, we provide *readXpress()*. Measurements calculated by CellProfiler can be exported in a comma-separated value (csv) file and accessed using *readXpress()*. For large-scale, high-throughput experiments, users can employ a computing cluster for increased analysis speed (<https://github.com/AndersenLab/CellProfiler>). In this case,

CellProfiler data stored in .RData format is accessed using *readXpress()*. Additionally, the function can optionally import a design file created by the user containing experimental treatments and conditions. This design file is joined to the CellProfiler data and output as a single dataframe.

CellProfiler's WormToolbox detects and measures the phenotypes of individual animals based on user-calibrated models of variability in body size and shape [187]. To effectively detect animals in a mixed-stage population, multiple worm models must be used. However, using multiple worm models creates a one-to-many relationship between real animals and their measured phenotype (**Figure S4-1**). We have included the function *modelSelection()* to annotate this information for downstream analysis. In instances where multiple worm model objects are assigned to a single primary object, *modelSelection()* will identify the best fitting model. Models are first ranked by frequency in the dataset such that the smallest model is classified as the most frequently occurring and the largest model is the least frequently occurring. In our experience, the most frequently occurring model in the dataset has the smallest size because it is often repeatedly assigned to a single primary object. Conversely, the least frequently occurring model in the dataset has the largest size as it is able to define the entire size of an animal, and is thus assigned to a primary object only once. The largest ranked model occurring within a single primary object is then selected as the best fitting model for that primary object. If necessary, *modelSelection()* will also specify whether the selected model object was repeatedly assigned to the same primary object and flag this event as a cluster. This problem occurs in instances where a model object is repeatedly assigned to a single primary object. If the largest model object is found to occur repeatedly in a single primary object, this model will be selected and a cluster flag

will be added (**Figure S4-1C**). The *modelSelection()* step is essential to resolve cases where multiple instances of a selected model object are assigned to a single primary object, thus contributing to inaccurate phenotype measurements.

#### 4.4.3. Data pruning and summarization

Once the data are read into the R statistical environment, it is crucial to optimize data quality before in-depth analysis. Uneven well illumination can hinder the performance of CellProfiler's object identification and phenotype extraction. Despite correcting for uneven illumination within a well, discerning foreground objects from background can be especially challenging near the periphery of the well and can add noise to nematode phenotype data (**Figure S4-2**). The function *edgeFlag()* was written to identify and flag animals located near the edge of circular wells using the centroid coordinates of the selected model object. By default, the function sets the radius of even illumination from the image center to 825 pixels, but this parameter can be adjusted by the user to serve project-specific analyses.

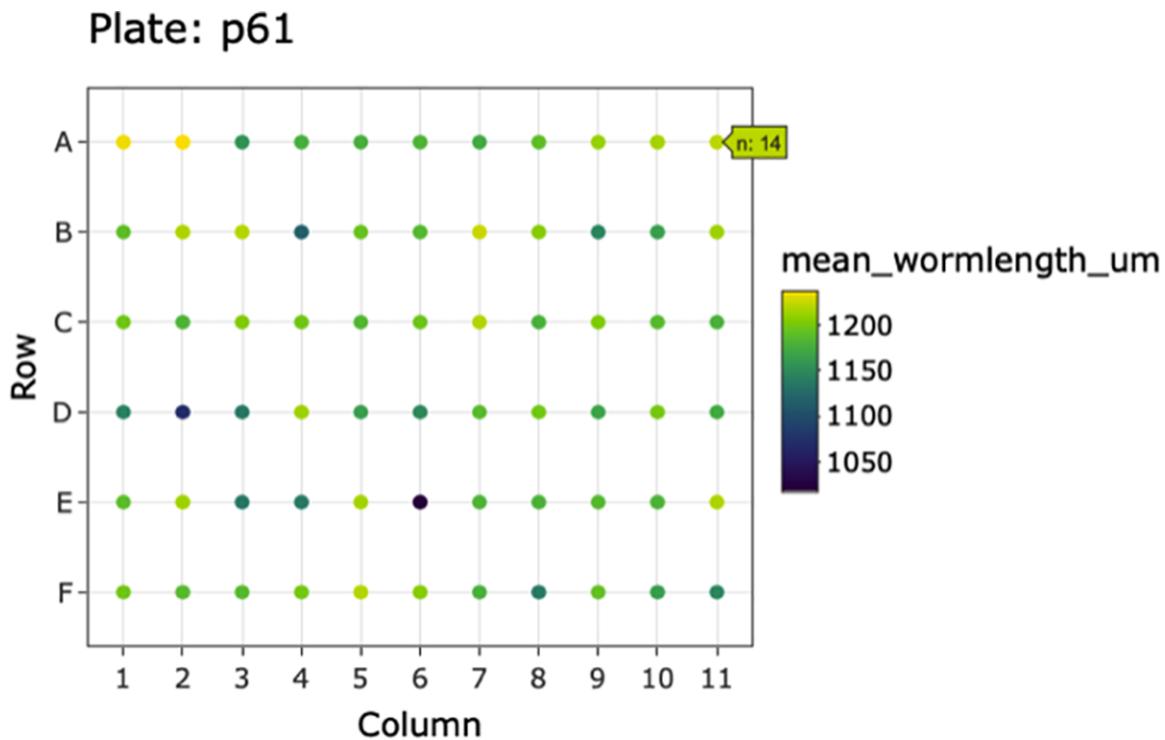
We also developed *setFlags()* in conjunction with *edgeFlag()* to further address data points that deviate from the expected animal measurements. The function *setFlags()* takes the output of *edgeFlags()* and detects outlier measurements among all measurements within a well using Tukey's fences [190]. By default, outlier calculations are performed by excluding data identified by *modelSelection()* as part of a cluster as well as data in close proximity to the well edge. However, *setFlags()* is customizable, allowing the user to specify which filters to include. *edgeFlag()* and *setFlags()* were designed to allow for analysis-specific optimization when handling various experimental datasets. This flexibility in data filtering makes easyXpress extensible to many unique projects.

Once data are adequately flagged, the function *process()* organizes the data into a list containing four elements: raw data, processed data, and summaries for both datasets. The raw data element is the CellProfiler data following *modelSelection()* and flag annotation. The processed data are generated by default after subsequent removal of all cluster, edge, and outlier flags. If a user includes data annotated as clusters or edge cases in *setFlags()*, cluster and edge cases will be retained in the processed data output. Finally, it is often useful to summarize data by well to interpret patterns specific to experimental variables. Alternatively, measurements may be summarized by other experimental factors according to the individual experimenter's plate design. *process()* aids in the summarization of both the raw and processed data elements. This function comprehensively calculates the means, variances, quantiles, minimum, and maximum values of animal length for any experimental unit (*e.g.* well). We have also included the wrapper function *Xpress()* to accelerate the import and processing of CellProfiler data. *Xpress()* will perform the above functions with all default settings, but a user can alter input arguments to better suit project specific needs.

#### 4.4.4. Visualization

The easyXpress package provides several plotting functions to allow users to explore the data through detailed and elegant visualizations. After data summarization, it is often useful to inspect the values of the summary statistics in order to recognize patterns or identify potential outlier data. We provide *viewPlate()* to assist with the visualization of mean animal length within each well across a microtiter plate (**Figure 4-2**). This function accepts either raw or processed data to

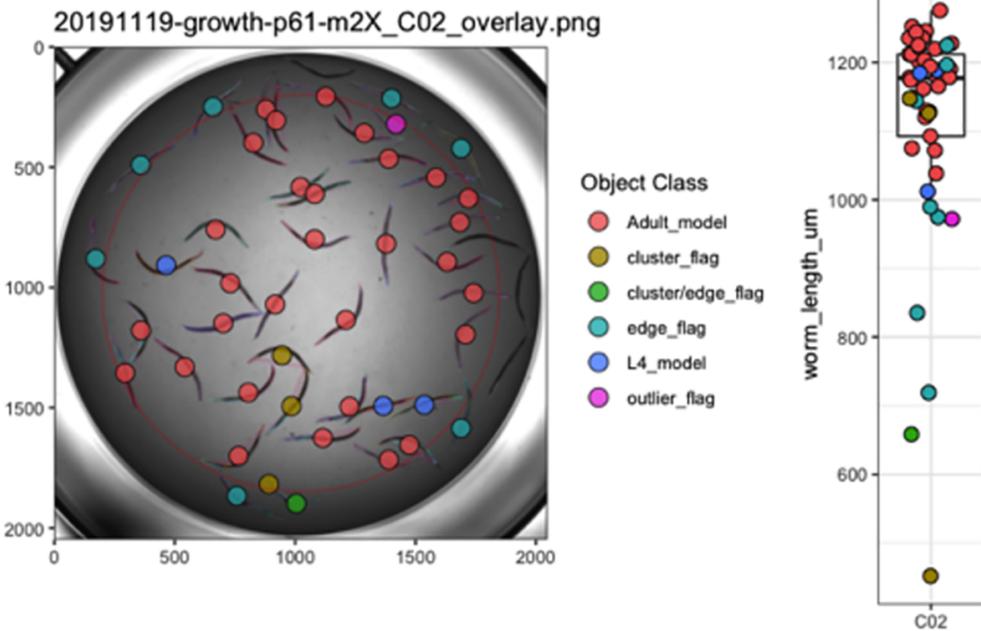
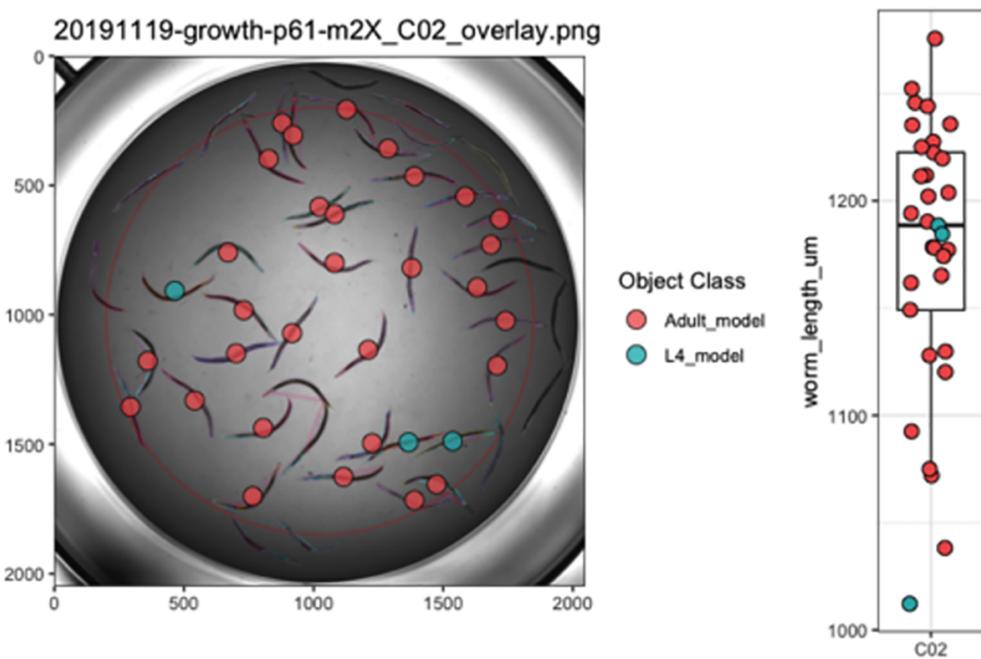
generate an interactive plot that allows users to scan across a plate to determine the number of objects present within individual wells.



**Figure 4-2. Example plot generated by *viewPlate()*.** Well-wise plot of mean animal length ( $\mu\text{m}$ ) from the summarized processed data. Interactive feature enables the assessment of the number of animals per well.

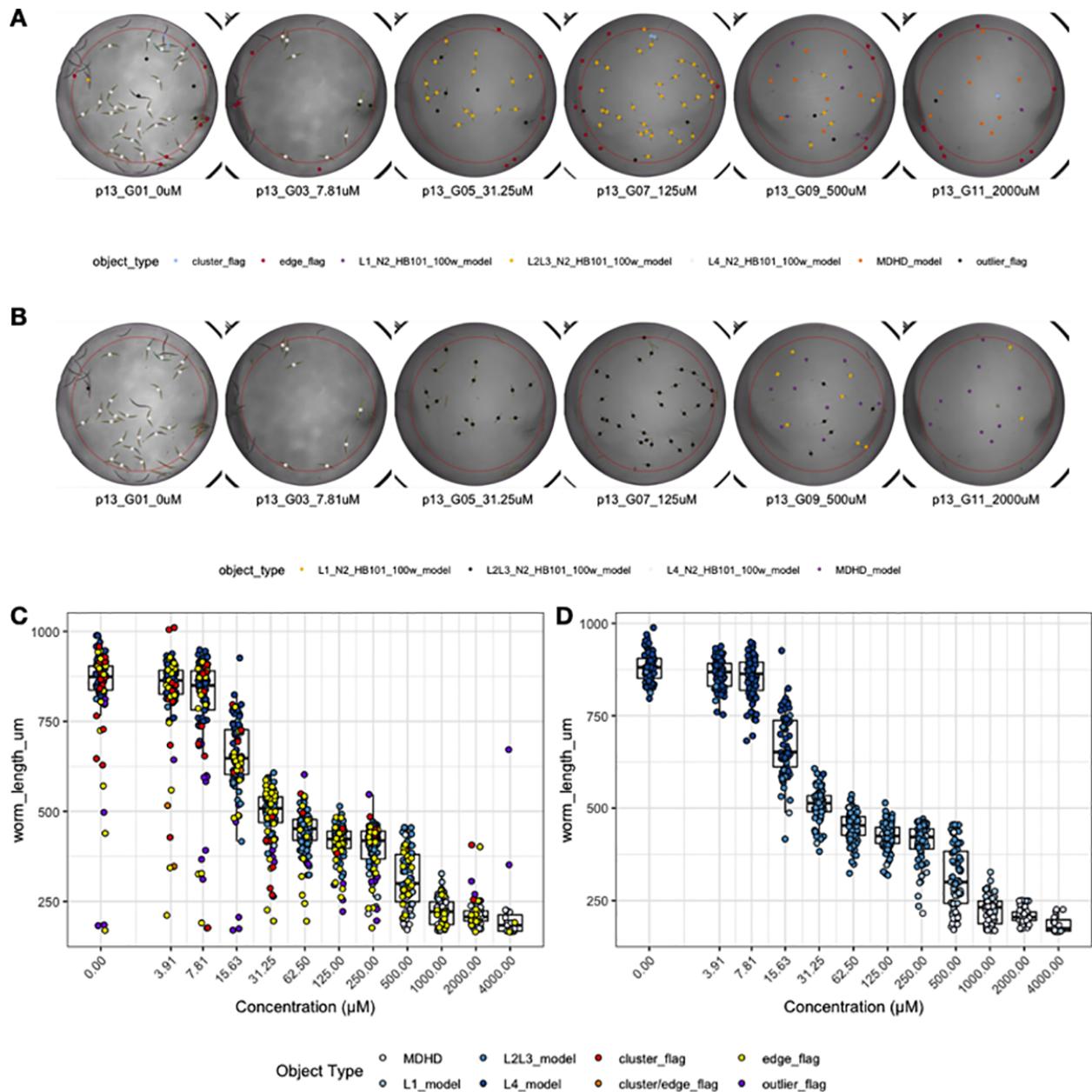
To complement the top-level data visualization provided by *viewPlate()*, we have included *viewWell()* to allow users to deeply explore data within individual wells. This function generates a plot of the well image following CellProfiler analysis with all objects annotated with their assigned class (**Figure 4-3**). Additionally, *viewWell()* can optionally generate a boxplot of the length values for each object. This plotting function is especially useful because it enables rapid qualitative assessment of object classification performance. By overlaying the model object

classifications on the well image, users can quickly determine whether CellProfiler classified objects as expected or whether errors in model selection or data flags occurred.

**A****B**

**Figure 4-3. Example plots generated by *viewWell()*.** The function *viewWell()* facilitates the exploration of data within an individual well. Well images displaying easyXpress raw (A) and processed (B) data are annotated with the location of each model object centroid (circles) and are colored by object class in the legend (left). Animals are outlined in different colors to indicate the model object(s) identified for each primary object (see **Figure S4-1**). The length of each object is displayed as a boxplot (right). Well edge circumference defined by the function *edgeFlag()* is shown in red.

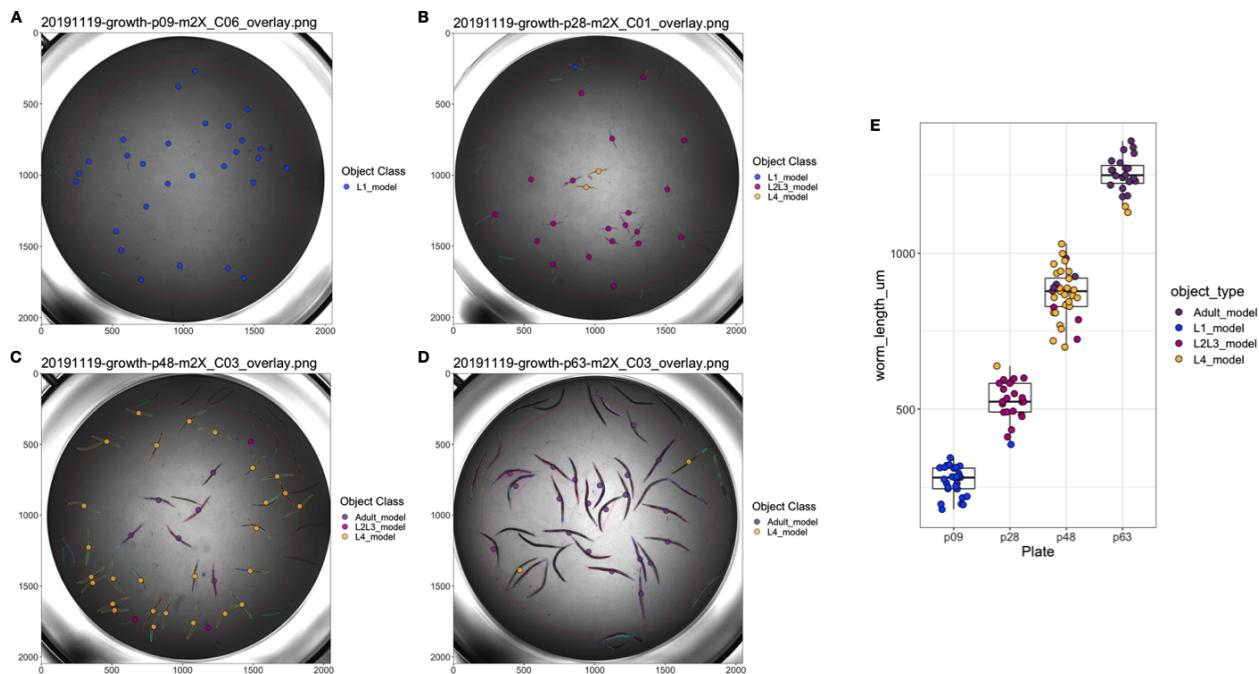
Lastly, we have developed the function *viewDose()* to allow for the visualization of dose response data. *C. elegans* are often used to study conserved responses to various compounds [83,85,117,191,192]. *viewDose()* allows a user to visually examine the effect of a compound on animal size and shape over a range of concentrations (**Figure 4-4**). By specifying the strain and compound of interest, a plot of representative wells will be generated that includes labels for each identified object.



**Figure 4-4. Example plots generated by *viewDose()*.** The function *viewDose()* plots representative raw (A) or processed (B) well images with objects annotated by model class for each dose of a selected drug and strain. The length measurements of raw (C) and processed (D) are also shown.

#### 4.4.5. Application to *C. elegans* growth data

We evaluated easyXpress using data collected from a *C. elegans* growth experiment [87]. Animals were imaged throughout the entire life cycle, beginning at the first larval (L1) stage and continuing until adulthood. Images were then processed with CellProfiler's WormToolbox and analyzed using easyXpress. During the implementation of easyXpress, four unique worm models representing *C. elegans* life stages were calibrated and applied: L1, L2/L3, L4, and Adult. These worm models do not designate stage assignments but rather represent the approximate sizes of animals that fall within the respective age groups (**Figure S4-1**). The function *modelSelection()* assigned the appropriate model object to animals at each life stage, *edgeFlag()* and *setFlags()* identified outlier data points, and *viewWell()* provided clear visualizations of both the processed (**Figure 4-5**) and raw (**Figure S4-3**) data.



**Figure 4-5. easyXpress applied to *C. elegans* growth data.** A subset of well images acquired during *C. elegans* development displaying easyXpress processed data are shown here. Images taken at (A) 9 hours indicating the L1 stage, (B) 28 hours indicating the L2/L3 stage, (C) 46 hours indicating the L4 stage, and (D) 63 hours indicating the adult stage were analyzed with CellProfiler using four worm models. The easyXpress workflow was then used to process and visualize the data. The length of each object identified after processing is shown in (E).

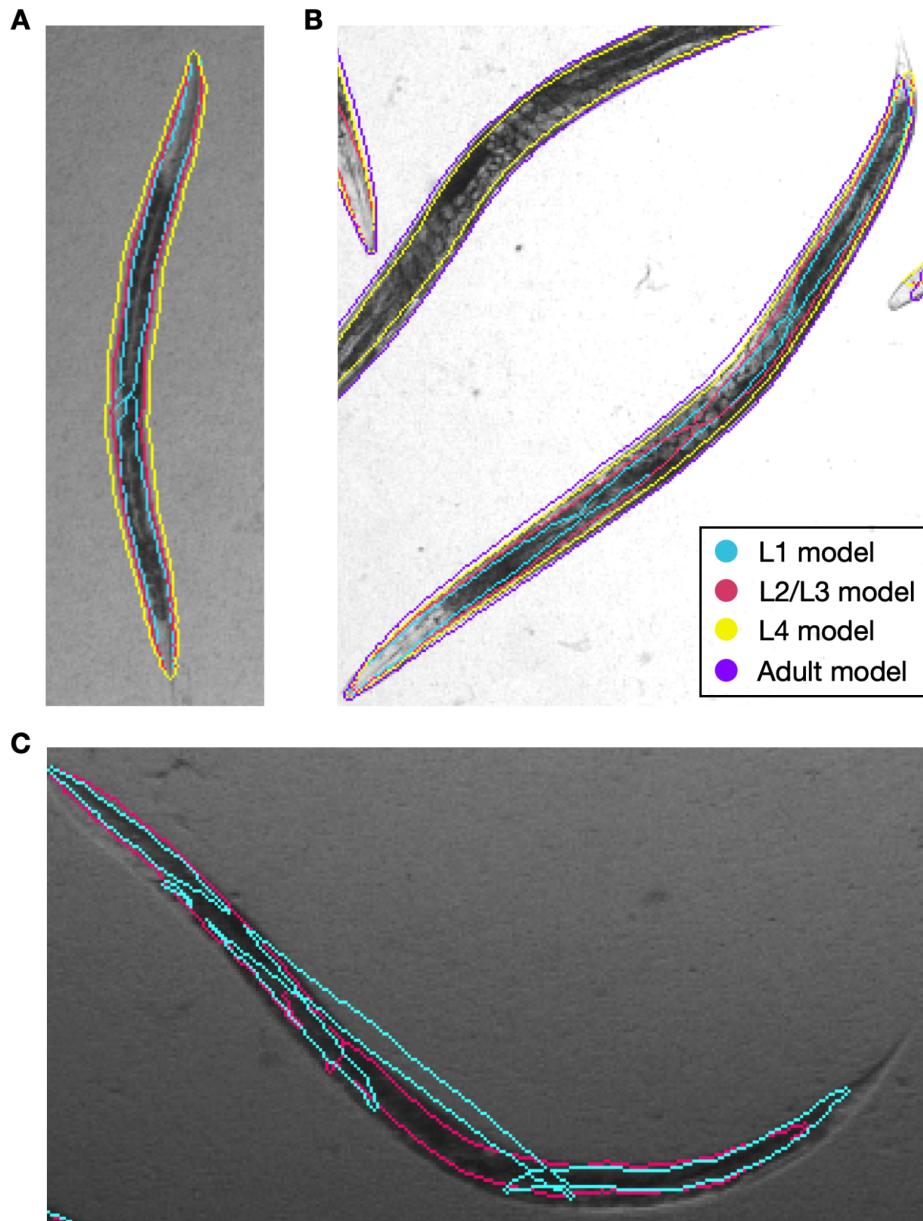
## 4.5. CONCLUSIONS

The easyXpress package presents an organized workflow for managing *C. elegans* phenotype data generated using CellProfiler. This package provides tools for the reading, processing, and visualization of these data in a simple and efficient way. By leveraging existing R infrastructure, easyXpress enables reproducible analysis, integration with other statistical R packages, and extensibility to many research projects using an open-source analysis pipeline.

## 4.6. CONTRIBUTIONS

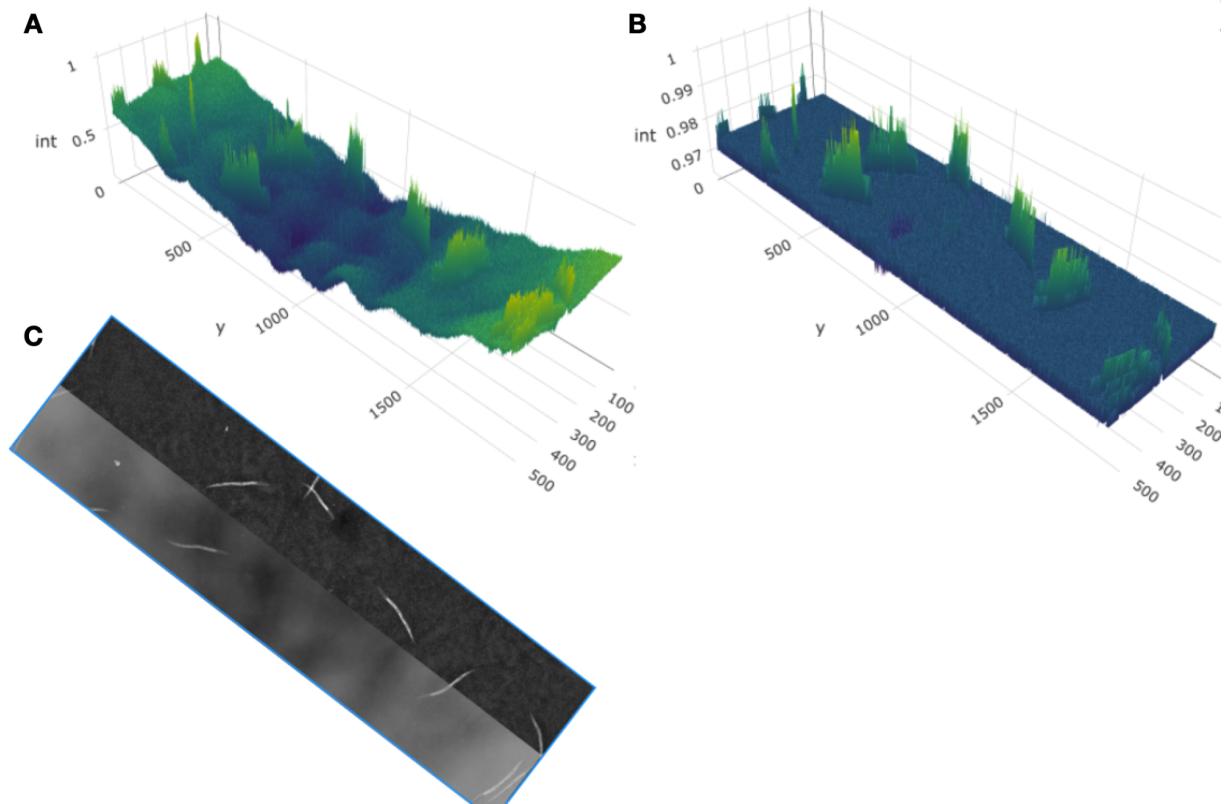
This package was built by Joy Nyaanga and Drs. Timothy Crombie and Sam Widmayer. Erik Andersen supervised the construction and development of the package.

#### 4.7. SUPPLEMENT

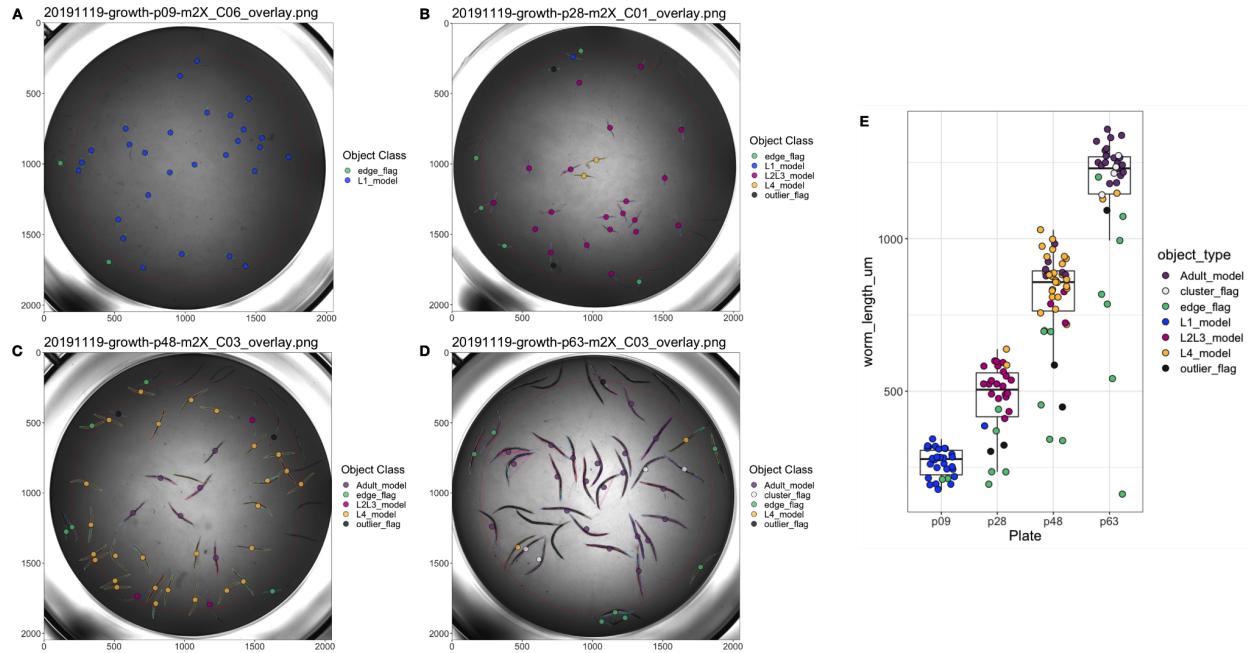


**Figure S4-1. Multiple model objects assigned to a single primary object.** When running CellProfiler's WormToolbox with multiple worm models, multiple model objects can be assigned to a single primary object (real animal). Different colors are used to outline each worm model

object. Here four unique models were used: L1, L2/L3, L4, and Adult. These worm models represent the approximate sizes of animals at each life stage. For example, some mutant or diverse wild genetic backgrounds might have differently sized adult animals as compared to the laboratory-adapted N2 strain. We have included this “soft matching” to account for small differences in the sizes of developmental stages across different genetic backgrounds, laboratories, and environmental conditions. (A) An animal detected by CellProfiler as a primary object has been assigned three unique worm models: two L1 model objects, one L2/L3 model object, and one L4 model object. *modelSelection()* classifies this animal as an L4 model object. (B) An animal detected as a primary object has been assigned four unique worm models: three L1 model objects, two L2/L3 model objects, one L4 model object, and one Adult model object. Here, *modelSelection()* identifies the Adult model as the best fitting model object. (C) An animal detected as a primary object has been assigned two unique worm models: three L1 model objects, and two L2/L3 model objects. In this case, *modelSelection()* classifies this animal as an L2/L3 model object and adds a cluster flag annotation to indicate the repeated assignment of the selected model object to the primary object.



**Figure S4-2. Uneven illumination along well edge hinders CellProfiler's ability to segment animals from background.** (A) Left is raw intensity values across well. (B) Right is with background correction. Intensities of object illumination are displayed on each z-axis. Objects near the edge of the well ( $y < 500$  and  $y > 1500$ ) have similar raw detected intensities (int) to more medial objects ( $y \sim 1000$ ) in (A) but lower corrected intensities in (B) because of uneven background correction. Raw and background-corrected image segments are displayed in (C). Notice animals on the edges of the well do not stand out from the background as much as animals in the center of the well and therefore are more challenging to discern.



**Figure S4-3. Raw data from *C. elegans* growth experiment displayed by the function `viewWell()`.** Similar to Fig 5, well images taken at (A) 9 hours indicating the L1 stage, (B) 28 hours indicating the L2/L3 stage, (C) 46 hours indicating the L4 stage, and (D) 63 hours were analyzed. Here, the raw data results are displayed. The length of each identified object identified is shown in (E).

**5**

## DISCUSSION

Extensive phenotypic diversity exists in the animal kingdom. A goal of biological research is to decipher the source of this diversity and study its impact on individuals, populations, and species. Body size, in particular, is a trait that varies strikingly across animal phylogeny, and has critical ramifications on an animal's ability to survive and reproduce. For this reason, scientists aim to understand how organisms grow to their characteristic sizes. Organismal development comprises a sequence of genetically programmed and intricately linked events that follow precise temporal and spatial ordering. Remarkably, despite variation in environmental conditions (*e.g.* temperature, nutrient availability, *etc*), developing systems coordinate the organization and interaction among cells, tissues, and organs with high reproducibility. My graduate work focused on investigating mechanisms of growth regulation in developing *Caenorhabditis elegans*. Here, I will elaborate on the implications of my data and provide insights that will hopefully propel future research.

## 5.1. SUMMARY

This dissertation aimed to present mechanisms by which developmental growth may be controlled using the roundworm nematode *Caenorhabditis elegans*. Studying *C. elegans* larval development is relatively straightforward, however, doing so at high temporal resolution while maintaining high precision and throughput can be a challenge. In **Chapter 2**, I introduce a high-throughput phenotyping platform I optimized to facilitate the quantitative assessment of size and feeding behavior of thousands of individuals under tightly controlled environment conditions. Using this platform to collect growth data over time, I find that as animals transitioned from one developmental stage to the next, changes in body shape occurred. Given this result, I worked with mathematical scientists to model a physical mechanism by which constraints on cuticle elasticity could cause changes in *C. elegans* body shape, and found the model-predicted shape changes to be consistent with those observed in the experimental data. This work puts forth a novel “Stretcher” mechanism for how *C. elegans* could use physical constraints (*i.e.* restriction of cuticle stretch) as a sensor for growth and as a means to regulate developmental timing. Noting the importance of the cuticle in *C. elegans* growth, I further examine the larval development of three strains with mutations in distinct cuticle collagen proteins in **Appendix A**. Two of the three mutant strains exhibit delayed growth and maintain a consistently shorter and wider shape when compared to wild type, emphasizing the importance of cuticle structure in development. Importantly, all three strains exhibited dynamics consistent with a “Stretcher” growth model.

The process of development is highly reproducible even in the presence of genetic and environmental perturbations. In **Chapter 3**, I leverage *C. elegans* natural genetic diversity to

explore how genetic variation contributes to differences in animal growth. Using a quantitative genetic mapping approach, I identify distinct genomic regions that separately explain variation in animal length and width. This work highlights the complex genetic architecture of body size, identifies promising candidate genes that may underlie variation in animal width, and provides a framework for future studies to investigate genetic mechanisms controlling developmental growth and body size in natural populations of *C. elegans*.

Like many high-throughput imaging techniques, the experimental platform introduced in Chapter 2 generates large quantities of data that can be difficult to process and visualize efficiently. In **Chapter 4**, I discuss the development of easyXpress: an R package that provides tools for the reading, processing, and visualization of *C. elegans* microscopy data in a simple and efficient way. This software package enables reproducible analysis, integration with other statistical R packages, and extensibility to many research projects using an open-source analysis pipeline.

## 5.2. THE FUTURE OF DEVELOPMENTAL GROWTH STUDIES

### 5.2.1. The *C. elegans* cuticle and its role in developmental timing

Molting is a unique characteristic that defines the life cycles of many animal species including, arthropods (insects, arachnids, crustaceans), nematodes (roundworms), and other members of the Ecdysozoa group [193,194]. Despite its prevalence, the primary function of molting varies across these different groups. Arthropods, covered by rigid chitin-based cuticles, require molting for body size expansion [195,196]. By contrast, the nematode body is encased in a flexible

collagen-based cuticle that allows for growth and expansion between molts [197,198], leaving the question: *Does the C. elegans cuticle play a role in developmental growth and molting?*

Few longitudinal studies of *C. elegans* growth have been performed. Initial research described *C. elegans* development as a sigmoidal curve characterized by continuous larval growth [199,200]. These early studies reasoned that molt events had little effect on continuous growth as the *C. elegans* flexible cuticle allowed for stretch during development. Later work clarified that growth rate was not continuous but rather piecewise in nature and increased from stage to stage, suggesting that *C. elegans* may contain a mechanism for regulating growth rate, potentially through each molt event [111]. Next, researchers found that animals only advance to the next developmental stage once a critical size is reached, implying that each molt decision is controlled by a size threshold [43]. Recent work shows that *C. elegans* development is likely controlled not by size-dependent regulation of growth but rather by relative change, or fold change, in body volume, suggesting that animals trigger molting events in ways that maintain a constant fold change in body volume [101]. However, two questions remain: *How do animals sense their body size?* And most importantly, *how is this information disseminated to precisely regulate growth?*

In Chapter 2, I have proposed a novel mechanism by which physical constraints can influence developmental timing and growth rate in *C. elegans*. Although elasticity of the cuticle permits growth during each larval stage, cuticle “stretchiness” is limited. By sensing the reduction of cuticle elasticity, *C. elegans* could use this as a mechanical signal to trigger molting events. To validate this proposed mechanism, one must first determine whether cuticle elasticity changes during a larval stage. Previous atomic force microscopy (AFM) studies of *C. elegans*

suggest that animals experience a loss of stiffness as they age [201], however no studies have been performed on developing animals. Capturing AFM images of the cuticle over the course of a single larval stage would be a potential method to investigate cuticle stiffness during growth. AFM experiments, however, are limited in that they apply small highly localized indentation forces that may not adequately represent the large stretching loads experienced during growth. To confirm elasticity changes in the cuticle, direct measurements of the stiffness of isolated cuticles is essential. Recently published work using microfluidic technology to stretch isolated cuticles found significant increases in longitudinal stiffness at large stretching loads [202]. Performing similar tests to isolated cuticles of animals preceding and following a molt could provide support for our hypothesis that molting occurs as a result of decreased cuticle elasticity. Additionally, if we find that cuticle elasticity changes during development, it could be interesting to collect growth measurements of animals cultured under varying osmotic conditions. Given our stretch-based growth model, I would expect to observe accelerated developmental timing in animals exposed to hypoosmotic conditions because of the relatively high stretching loads on the cuticle throughout growth.

Identifying changes in cuticle elasticity during growth is just one part of understanding the role the cuticle may play in molt timing. We must also consider how information originating from the cuticle is sensed and propagated to the rest of the animal. In *C. elegans*, attachment complexes called hemidesmosomes connect the cuticle to the epidermis and are shown to respond to mechanical tension exerted by muscle contraction [144]. It is possible that *C. elegans* hemidesmosomes act not only as an attachment structure but also as a mechanosensor that responds to forces at the epidermis and at the cuticle. Additionally, motor neurons may serve as

another potential mechanosensor. Mechanosensory neurons are typically used to regulate various behaviors including, locomotion, egg laying, pharyngeal pumping, and defecation [203]. Notably, many motor neurons extend along the longitudinal body axis and some also extend circumferential axons [204], opening up the possibility that these neurons might provide signals proportional to body length and width. Future studies of how perturbations of hemidesmosomes and motor neurons influence *C. elegans* body size and developmental timing would be worthwhile.

### **5.2.2. Identifying genetic factors that contribute to variation in growth and body size**

Discerning how complex traits are genetically controlled is essential to understanding the evolution of phenotypic diversity. Genetic regulators of trait variation can be mapped using QTL analysis of RIAILs derived from genetically and phenotypically divergent strains. In my work, I used linkage mapping to identify genetic loci underlying differences in *C. elegans* body size, and found three small-effect QTL explaining 5-8% of variation in the RIAIL phenotypes. Our ability to detect and validate one of these QTL emphasizes the power of using *C. elegans* to correlate small-effect genetic loci to quantitative phenotypes. In Chapter 3, I postulate that the inability to recapitulate the other two QTL effects could be attributed to insufficient power or the presence of additional undetected QTL with additive or non-additive effects. The CSSs and NILs I used for validation experiments each contained a single region of introgression in a genetic background of the opposite genotype. Future work should aim to construct double CSSs or multi-region NILs where pairwise combinations of two genomic regions are introgressed in the opposite genotype. Phenotyping these additional strains could reveal additional loci contributing to the chromosome IV and V effects.

### 5.2.2.1. Alternatives to classical linkage mapping

#### ***Genome-wide association (GWA) mapping***

High levels of genotypic and phenotypic diversity exist in natural populations. In contrast to linkage mapping where genetic diversity is limited to the two parental strains, GWA mapping leverages the existing natural diversity found in a population. By correlating phenotypic variation among wild strains with whole-genome data, GWA studies identify functional variants that contribute to phenotypic diversity. In *C. elegans*, a growing collection of wild strains and genome-wide variation data are readily available (CeNDR, [69]). Using this resource, association mapping has led to the discovery of genes and variants that underlie variation in complex traits (**Figure 1-4**) [21].

Recently, the Andersen Lab completed an extensive GWA experiment, collecting body length measurements for 195 wild strains across 54 different drug conditions. Exploring this data to identify size-associated QTL of animals exposed to the control condition (water), and comparing the results to the linkage mapping presented in Chapter 3, would be the next step for subsequent investigations of genetic factors that underlie differences in *C. elegans* growth and body size. With data from the linkage and association mappings, we could look for QTL that overlap between the two methods. Overlapping QTL would suggest that a common variant present in the CB4856 strain contributes to differences in body size of animals in both mapping populations. Alternatively, non-overlapping QTL would suggest that the QTL identified from linkage mapping is driven by a rare variant in the CB4856 strain, whereas the QTL identified from association mapping is driven by a common variant not found in the CB4856 strain. The

combination of both association and linkage mapping provides an approach to both narrow genomic intervals and generate refined lists of potential causal genes.

### ***Bulk-segregant analysis***

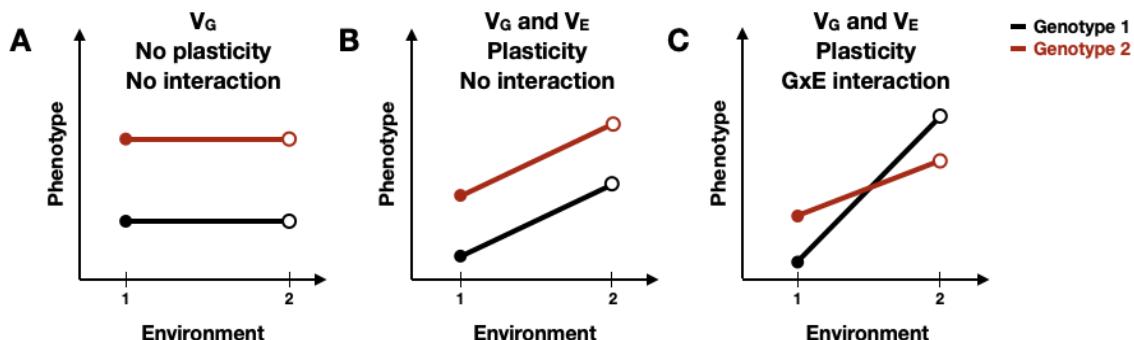
QTL mapping has been central to the identification of loci underlying complex traits in *C. elegans*. However, QTL mapping typically requires the generation of large panels of cross progeny that must be individually genotyped and phenotyped, a lengthy and often costly task. An alternative to traditional QTL mapping is bulk-segregant analysis (BSA). In contrast to constructing, sequencing, and phenotyping individuals for a RIL/RIAIL panel, the BSA approach uses pools of recombinant individuals generated from genetically and phenotypically divergent parents. Recombinant individuals are then selected based on the phenotype of interest and allele frequencies are compared between the unselected and selected pools. Over recent years, researchers have demonstrated the power of using BSA to connect phenotypic differences to specific genes (**Figure 1-4**) [21]. Leveraging BSA to identify QTL associated with developmental growth would be a worthwhile endeavor. Recombinant individuals from an N2xCB4856 cross could be cultured as described in Chapter 2. Then by selecting the tails of the population distribution, the allele frequency changes of the “large-sized” and “small-sized” individuals during the experiment can be compared. Repeating this over various stages during development could potentially reveal unique QTL for independent larval stages. Alternatively, one could make use of population-wide diversity by performing growth and selection on a pool of wild isolates. The application of a powerful mapping method like BSA to identify QTL will expand our understanding of the molecular mechanisms underlying *C. elegans* growth and development.

### 5.2.3. Phenotypic plasticity and genotype-by-environment interaction

In my graduate work I study the growth and development of *C. elegans* under tightly controlled environmental conditions. However, environmental factors in natural habitats are rarely constant, and this variation can impact developmental progression and growth rate. Research has shown that *C. elegans* rearing temperature, as well as both food quality and quantity alter the speed of larval development [58,60,205]. The term phenotypic plasticity is used to describe this phenomenon where an individual can generate multiple phenotypes based on external conditions [206]. Notably, distinct genotypes of a single species can show unique patterns of plasticity. Genotype-by-environment (GxE) interactions represent this contribution of genetic variation to environmental sensitivity.

Reaction norms are often used to visually compare plasticity among different genotypes by displaying phenotype differences as a function of environmental change [207]. A phenotype may be constant across genotypes and environments, or it may be affected in several ways. First, phenotype could differ between genotypes, indicating genetic variation, but remain unaffected across multiple environments. Displayed as horizontal lines between environmental parameters, this case represents robustness to environmental perturbation (**Figure 5-1A**). Second, regardless of genotype, there could be a general plastic response to environmental change. Here, diagonal lines between environmental conditions reflect plasticity and the slope of the line represents the magnitude of plasticity (**Figure 5-1B**). Third, the magnitude of plasticity for a given trait could vary among genotypes, represented as non-parallel reaction norms (**Figure 5-1C**), indicating the presence of GxE interaction. Although the occurrence of GxE interactions is well studied, the

genomic regions involved are not because of the complex interaction between genetics and environmental factors that drives plasticity [208–210].



**Figure 5-1.** Schematic representation of reaction norms, illustrating forms of plasticity.

In the last month, research was published that explored the genetic architecture of body-size plasticity in *C. elegans* [180]. Researchers used 40 RILs derived from the N2 and CB4856 strains to investigate the effects of temperature and developmental stage on the reaction norms of various body-size traits (animal length, width, and size of some internal organs). Not only did they find clear evidence of GxE interactions, but they also noticed that the shape of the reaction norm was affected by developmental stage, potentially stemming from differences in interaction between stage-specific gene expression patterns and environmental temperature. Authors also performed QTL mapping and found 18 loci underlying body-size traits and only five plasticity loci. Interestingly, three plasticity QTL colocalized with body-size QTL, suggesting the possibility of co-regulatory loci influencing both plasticity and the traits themselves. Future work should build on these findings by collecting size measurements for wild *C. elegans* strains under different environmental conditions. This research is vital to our understanding of the complex interplay between genetic mechanisms and environmental forces that shape phenotypic variation in natural populations.

## References

1. Johannsen W. The Genotype Conception of Heredity. *Am Nat.* 1911;45: 129–159.
2. Nathans J, Piantanida TP, Eddy RL, Shows TB, Hogness DS. Molecular genetics of inherited variation in human color vision. *Science.* 1986;232: 203–210.
3. Falconer DS, Others. Introduction to quantitative genetics. *Introduction to quantitative genetics.* 1960. Available: <https://www.cabdirect.org/cabdirect/abstract/19601603365>
4. Lynch M, Walsh B, Others. *Genetics and analysis of quantitative traits.* Sinauer Sunderland, MA; 1998.
5. Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. *Genetic variation.* W. H. Freeman; 2000.
6. Davis MB, Shaw RG, Etterson JR. Evolutionary responses to changing climate. *Ecology.* 2005;86: 1704–1714.
7. Gutteling EW, Doroszuk A, Riksen JAG, Prokop Z, Reszka J, Kammenga JE. Environmental influence on the genetic correlations between life-history traits in *Caenorhabditis elegans*. *Heredity.* 2007;98: 206–213.
8. Félix M-A, Barkoulas M. Pervasive robustness in biological systems. *Nat Rev Genet.* 2015;16: 483–496.
9. Viney M, Diaz A. Phenotypic plasticity in nematodes: Evolutionary and ecological significance. *Worm.* 2012;1: 98–106.
10. Druery CT, Bateson W. Experiments in plant hybridization. *Journal of the Royal Horticultural Society.* 1901;26: 1–32.
11. Telenti A, Pierce LCT, Biggs WH, di Julio J, Wong EHM, Fabani MM, et al. Deep sequencing of 10,000 human genomes. *Proc Natl Acad Sci U S A.* 2016;113: 11901–11906.
12. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467: 1061–1073.
13. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536: 285–291.
14. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet.* 2013;132: 1077–1130.

15. Antonarakis SE, Beckmann JS. Mendelian disorders deserve more attention. *Nat Rev Genet.* 2006;7: 277–282.
16. Barton NH, Keightley PD. Understanding quantitative genetic variation. *Nat Rev Genet.* 2002;3: 11–21.
17. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science.* 2008;322: 881–888.
18. Donnelly P. Progress and challenges in genome-wide association studies in humans. *Nature.* 2008;456: 728–731.
19. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet.* 2009;10: 241–251.
20. Mackay TFC, Stone EA, Ayroles JF. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet.* 2009;10: 565–577.
21. Evans KS, van Wijk MH, McGrath PT, Andersen EC, Sterken MG. From QTL to gene: *C. elegans* facilitates discoveries of the genetic mechanisms underlying natural variation. *Trends Genet.* 2021;37: 933–947.
22. Doroszuk A, Snoek LB, Fradin E, Riksen J, Kammenga J. A genome-wide library of CB4856/N2 introgression lines of *Caenorhabditis elegans*. *Nucleic Acids Res.* 2009;37: e110.
23. Wong GT. Speed congenics: applications for transgenic and knock-out mouse strains. *Neuropeptides.* 2002;36: 230–236.
24. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007;315: 1709–1712.
25. Paix A, Wang Y, Smith HE, Lee C-YS, Calidas D, Lu T, et al. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics.* 2014;198: 1347–1356.
26. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science.* 2014;346: 1258096.
27. Maupas E. Modes et formes de reproduction des nematodes. 1900.
28. Nigon V. Le déterminisme du sexe chez un Nématode libre hermaphrodite, *Rhabditis elegans* Maupas. *C R Séances Soc Biol Fil.* 1943;137: 40–41.
29. Corsi AK, Wightman B, Chalfie M. A Transparent Window into Biology: A Primer on *Caenorhabditis elegans*. *Genetics.* 2015;200: 387–407.

30. All Registered Worm Labs. In: The WBG [Internet]. 13 Apr 2016 [cited 28 Jan 2022]. Available: <http://wbg.wormbook.org/the-worm-lab-project/all-labs/>
31. Wood WB. The Nematode *Caenorhabditis Elegans*. Cold Spring Harbor Laboratory; 1988.
32. Riddle DL, Blumenthal T, Meyer BJ, Priess JR, editors. *C. elegans II*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2011.
33. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics*. 1974;77: 71–94.
34. Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol*. 1977;56: 110–156.
35. Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol*. 1983;100: 64–119.
36. Deppe U, Schierenberg E, Cole T, Krieg C, Schmitt D, Yoder B, et al. Cell lineages of the embryo of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*. 1978;75: 376–380.
37. Kimble J, Hirsh D. The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev Biol*. 1979;70: 396–417.
38. C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. 1998;282: 2012–2018.
39. Kaletta T, Hengartner MO. Finding function in novel targets: *C. elegans* as a model organism. *Nat Rev Drug Discov*. 2006;5: 387–398.
40. Kim W, Underwood RS, Greenwald I, Shaye DD. OrthoList 2: A New Comparative Genomic Analysis of Human and *Caenorhabditis elegans* Genes. *Genetics*. 2018;210: 445–461.
41. Singh RN, Sulston JE. Some Observations On Moulting in *Caenorhabditis Elegans*. *Nematologica*. 1978;24: 63–71.
42. Monsalve GC, Frand AR. Toward a unified model of developmental timing: A “molting” approach. *Worm*. 2012. Available: <https://www.tandfonline.com/doi/abs/10.4161/worm.20874>
43. Uppaluri S, Brangwynne CP. A size threshold governs *Caenorhabditis elegans* developmental progression. *Proc Biol Sci*. 2015;282: 20151283.
44. Cassada RC, Russell RL. The dauerlarva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. *Dev Biol*. 1975;46: 326–342.

45. Hu PJ. *Dauer*. WormBook; 2018.
46. Schindler AJ, Baugh LR, Sherwood DR. Identification of late larval stage developmental checkpoints in *Caenorhabditis elegans* regulated by insulin/IGF and steroid hormone signaling pathways. *PLoS Genet*. 2014;10: e1004426.
47. Baugh LR, Sternberg PW. DAF-16/FOXO regulates transcription of cki-1/Cip/Kip and repression of lin-4 during *C. elegans* L1 arrest. *Curr Biol*. 2006;16: 780–785.
48. Monsalve GC, Van Buskirk C, Frand AR. LIN-42/PERIOD controls cyclical and developmental progression of *C. elegans* molts. *Curr Biol*. 2011;21: 2033–2045.
49. Zaidel-Bar R, Miller S, Kaminsky R, Broday L. Molting-specific downregulation of *C. elegans* body-wall muscle attachment sites: the role of RNF-5 E3 ligase. *Biochem Biophys Res Commun*. 2010;395: 509–514.
50. Suzuki Y, Yandell MD, Roy PJ, Krishna S, Savage-Dunn C, Ross RM, et al. A BMP homolog acts as a dose-dependent regulator of body size and male tail patterning in *Caenorhabditis elegans*. *Development*. 1999;126: 241–250.
51. Morita K, Chow KL, Ueno N. Regulation of body length and male tail ray pattern formation of *Caenorhabditis elegans* by a member of TGF-beta family. *Development*. 1999;126: 1337–1347.
52. Savage-Dunn C, Maduzia LL, Zimmerman CM, Roberts AF, Cohen S, Tokarz R, et al. Genetic screen for small body size mutants in *C. elegans* reveals many TGFbeta pathway components. *Genesis*. 2003;35: 239–247.
53. Gumienny TL, Savage-Dunn C. TGF- $\beta$  signaling in *C. elegans*. WormBook; 2018.
54. Wu MY, Hill CS. Tgf-beta superfamily signaling in embryonic development and homeostasis. *Dev Cell*. 2009;16: 329–343.
55. Gumienny TL, MacNeil LT, Wang H, de Bono M, Wrana JL, Padgett RW. Glypican LON-2 is a conserved negative regulator of BMP-like signaling in *Caenorhabditis elegans*. *Curr Biol*. 2007;17: 159–164.
56. Baugh LR. To grow or not to grow: nutritional control of development during *Caenorhabditis elegans* L1 arrest. *Genetics*. 2013;194: 539–555.
57. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics*. 1974;77: 71–94.
58. MacNeil LT, Watson E, Arda HE, Zhu LJ, Walhout AJM. Diet-induced developmental acceleration independent of TOR and insulin in *C. elegans*. *Cell*. 2013;153: 240–252.
59. So S, Miyahara K, Ohshima Y. Control of body size in *C. elegans* dependent on food and

- insulin/IGF-1 signal. *Genes Cells.* 2011;16: 639–651.
60. Stuhr NL, Curran SP. Bacterial diets differentially alter lifespan and healthspan trajectories in *C. elegans*. *Commun Biol.* 2020;3: 653.
  61. Kammenga JE, Doroszuk A, Riksen JAG, Hazendonk E, Spiridon L, Petrescu A-J, et al. A *Caenorhabditis elegans* Wild Type Defies the Temperature–Size Rule Owing to a Single Nucleotide Polymorphism in *tra-3*. *PLoS Genet.* 2007;3: e34.
  62. Rose JK, Sangha S, Rai S, Norman KR, Rankin CH. Decreased sensory stimulation reduces behavioral responding, retards development, and alters neuronal connectivity in *Caenorhabditis elegans*. *J Neurosci.* 2005;25: 7159–7168.
  63. Fujiwara M, Sengupta P, McIntire SL. Regulation of body size and behavioral state of *C. elegans* by sensory perception and the EGL-4 cGMP-dependent protein kinase. *Neuron.* 2002;36: 1091–1102.
  64. Fujiwara M, Ishihara T, Katsura I. A novel WD40 protein, CHE-2, acts cell-autonomously in the formation of *C. elegans* sensory cilia. *Development.* 1999;126: 4839–4848.
  65. Page AP, Johnstone IL. The cuticle. *WormBook.* 2007; 1–15.
  66. McMahon L, Muriel JM, Roberts B, Quinn M, Johnstone IL. Two Sets of Interacting Collagens Form Functionally Distinct Substructures within a *Caenorhabditis elegans* Extracellular Matrix. *MBoC.* 2003;14: 1366–1378.
  67. Lints R, Hall DH. WormAtlas hermaphrodite handbook - the cuticle. *WormAtlas.* 2003. doi:10.3908/wormatlas.1.12
  68. Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, et al. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet.* 2012;44: 285–290.
  69. Cook DE, Zdraljevic S, Roberts JP, Andersen EC. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* 2017;45: D650–D657.
  70. Nicholas WL, Dougherty EC, Hansen EL. AXENIC CULTIVATION OF CAENORHABDITIS BRIGGSAE (NEMATODA: RHABDITIDAE) WITH CHEMICALLY UNDEFINED SUPPLEMENTS; COMPARATIVE STUDIES WITH RELATED NEMATODES\*. *Ann N Y Acad Sci.* 2006;77: 218–236.
  71. Hodgkin J, Doniach T. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics.* 1997;146: 149–164.
  72. Thompson OA, Snoek LB, Nijveen H, Sterken MG, Volkers RJM, Brenchley R, et al. Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis*

- elegans Hawaiian Strain CB4856. *Genetics*. 2015;200: 975–989.
73. Kim C, Kim J, Kim S, Cook DE, Evans KS, Andersen EC, et al. Long-read sequencing reveals intra-species tolerance of substantial structural variations and new subtelomere formation in *C. elegans*. *Genome Res*. 2019;29: 1023–1035.
  74. Capra EJ, Skrovanek SM, Kruglyak L. Comparative developmental expression profiling of two *C. elegans* isolates. *PLoS One*. 2008;3: e4055.
  75. Rockman MV, Skrovanek SS, Kruglyak L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science*. 2010;330: 372–376.
  76. Viñuela A, Snoek LB, Riksen JAG, Kammenga JE. Aging Uncouples Heritability and Expression-QTL in *Caenorhabditis elegans*. *G3* . 2012;2: 597–605.
  77. de Bono M, Bargmann CI. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell*. 1998;94: 679–689.
  78. Gutteling EW, Riksen JAG, Bakker J, Kammenga JE. Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* . 2007;98: 28–37.
  79. Andersen EC, Bloom JS, Gerke JP, Kruglyak L. A variant in the neuropeptide receptor npr-1 is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet*. 2014;10: e1004156.
  80. Shimko T, Brady SC, Cook DE, Zdraljevic S, Andersen EC. linkagemapping. Github; Available: <https://github.com/AndersenLab/linkagemapping>
  81. Bendesky A, Bargmann CI. Genetic contributions to behavioural diversity at the gene–environment interface. *Nat Rev Genet*. 2011;12: 809–820.
  82. Brady SC, Zdraljevic S, Bisaga KW, Tanny RE, Cook DE, Lee D, et al. A Novel Gene Underlies Bleomycin-Response Variation in *Caenorhabditis elegans*. *Genetics*. 2019;212: 1453–1468.
  83. Evans KS, Andersen EC. The Gene scb-1 Underlies Variation in *Caenorhabditis elegans* Chemotherapeutic Responses. *G3 Genes|Genomes|Genetics*. 2020;10: 2353–2364.
  84. Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC. Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog*. 2018;14: e1007226.
  85. Zdraljevic S, Fox BW, Strand C, Panda O, Tenjo FJ, Brady SC, et al. Natural variation in *C. elegans* arsenic toxicity is explained by differences in branched chain amino acid metabolism. 2019 [cited 5 May 2021]. doi:10.7554/eLife.40260

86. Zdraljevic S, Strand C, Seidel HS, Cook DE, Doench JG, Andersen EC. Natural variation in a single amino acid substitution underlies physiological responses to topoisomerase II poisons. *PLoS Genet.* 2017;13: e1006891.
87. Nyaanga J, Goss C, Zhang G, Ahmed HN, Andersen EJ, Miller IR, et al. Changes in body shape implicate cuticle stretch in *C. elegans* growth control. *bioRxiv*. 2022. p. 2021.04.01.438121. doi:10.1101/2021.04.01.438121
88. Hone DWE, Benton MJ. The evolution of large size: how does Cope's Rule work? *Trends Ecol Evol.* 2005;20: 4–6.
89. Björklund M. Cell size homeostasis: Metabolic control of growth and cell division. *Biochim Biophys Acta Mol Cell Res.* 2019;1866: 409–417.
90. Willis L, Huang KC. Sizing up the bacterial cell cycle. *Nat Rev Microbiol.* 2017;15: 606–620.
91. Turner JJ, Ewald JC, Skotheim JM. Cell size control in yeast. *Curr Biol.* 2012;22: R350–9.
92. Donnan L, John PC. Cell cycle control by timer and sizer in Chlamydomonas. *Nature*. 1983;304: 630–633.
93. Wang P, Hayden S, Masui Y. Transition of the blastomere cell cycle from cell size-independent to size-dependent control at the midblastula stage in *Xenopus laevis*. *J Exp Zool.* 2000;287: 128–144.
94. Iyer-Biswas S, Wright CS, Henry JT, Lo K, Burov S, Lin Y, et al. Scaling laws governing stochastic growth and division of single bacterial cells. *Proc Natl Acad Sci U S A.* 2014;111: 15912–15917.
95. Tzur A, Kafri R, LeBleu VS, Lahav G, Kirschner MW. Cell growth and size homeostasis in proliferating animal cells. *Science.* 2009;325: 167–171.
96. Pavelescu I, Vilarrasa-Blasi J, Planas-Riverola A, González-García M-P, Caño-Delgado AI, Ibañes M. A Sizer model for cell differentiation in *Arabidopsis thaliana* root growth. *Mol Syst Biol.* 2018;14: e7687.
97. Sveiczer A, Novak B, Mitchison JM. The size control of fission yeast revisited. *J Cell Sci.* 1996;109 ( Pt 12): 2947–2957.
98. Fantes PA. Control of cell size and cycle time in *Schizosaccharomyces pombe*. *J Cell Sci.* 1977;24: 51–67.
99. Campos M, Surovtsev IV, Kato S, Paintdakhi A, Beltran B, Ebmeier SE, et al. A constant size extension drives bacterial cell size homeostasis. *Cell.* 2014;159: 1433–1446.

100. Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, Paulsson J, et al. Cell-size control and homeostasis in bacteria. *Curr Biol.* 2015;25: 385–391.
101. Towbin BD, Grosshans H. A folder mechanism ensures size uniformity among *C. elegans* individuals by coupling growth and development. *bioRxiv*. 2021. Available: <https://www.biorxiv.org/content/10.1101/2021.03.24.436858v1.abstract>
102. Soifer I, Robert L, Amir A. Single-Cell Analysis of Growth in Budding Yeast and Bacteria Reveals a Common Size Regulation Strategy. *Curr Biol.* 2016;26: 356–361.
103. Osella M, Nugent E, Cosentino Lagomarsino M. Concerted control of *Escherichia coli* cell division. *Proc Natl Acad Sci U S A.* 2014;111: 3431–3435.
104. Jorgensen P, Tyers M. How cells coordinate growth and division. *Curr Biol.* 2004;14: R1014–27.
105. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, et al. Robust growth of *Escherichia coli*. *Curr Biol.* 2010;20: 1099–1103.
106. Cadart C, Monnier S, Grilli J, Sáez PJ, Srivastava N, Attia R, et al. Size control in mammalian cells involves modulation of both growth rate and cell cycle duration. *Nat Commun.* 2018;9: 3275.
107. Moss-Taylor L, Upadhyay A, Pan X, Kim M-J, O'Connor MB. Body Size and Tissue-Scaling Is Regulated by Motoneuron-Derived Activin $\beta$  in *Drosophila melanogaster*. *Genetics.* 2019;213: 1447–1464.
108. Spence AJ. Scaling in biology. *Curr Biol.* 2009;19: R57–61.
109. Uppaluri S, Weber SC, Brangwynne CP. Hierarchical Size Scaling during Multicellular Growth and Development. *Cell Rep.* 2016;17: 345–352.
110. Byerly L, Cassada RC, Russell RL. The life cycle of the nematode *Caenorhabditis elegans*. I. Wild-type growth and reproduction. *Dev Biol.* 1976;51: 23–33.
111. Knight CG, Patel MN, Azevedo RBR, Leroi AM. A novel mode of ecdysozoan growth in *Caenorhabditis elegans*. *Evol Dev.* 2002;4: 16–27.
112. Faerberg DF, Gurarie V, Ruvinsky I. Inferring temporal organization of postembryonic development from high-content behavioral tracking. *Dev Biol.* 2021;475: 54–64.
113. Filina O, Haagmans R, van Zon JS. Temporal scaling in *C. elegans* larval development. *bioRxiv*. 2020. p. 2020.09.21.306423. doi:10.1101/2020.09.21.306423
114. Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, et al. The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics.* 2016;204:

- 371–383.
115. Stiernagle T. Maintenance of *C. elegans*. WormBook; 2006.
  116. Abràmoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. *Biophotonics international*. 2004;11: 36–42.
  117. Andersen EC, Shimko TC, Crissman JR, Ghosh R, Bloom JS, Seidel HS, et al. A Powerful New Quantitative Genetics Platform, Combining *Caenorhabditis elegans* High-Throughput Fitness Assays with a Large Collection of Recombinant Strains. *G3* . 2015;5: 911–920.
  118. Shimko TC, Andersen EC. COPASutils: an R package for reading, processing, and visualizing data from COPAS large-particle flow cytometers. *PLoS One*. 2014;9: e111090.
  119. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J*. 2016;8: 289–317.
  120. Smith MV, Boyd WA, Kissling GE, Rice JR, Snyder DW, Portier CJ, et al. A discrete time model for the analysis of medium-throughput *C. elegans* growth data. *PLoS One*. 2009;4: e7018.
  121. Hermann E. lokern: Kernel Regression Smoothing with Local or Global Plug-in Bandwidth. R package version 1.1-8. 2016. Available: <https://CRAN.R-project.org/package=lokern>
  122. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Dordrecht, The Netherlands: D Reidel. 1986;81: 26853.
  123. Schwarz G. Estimating the Dimension of a Model. *aos*. 1978;6: 461–464.
  124. Kass RE, Raftery AE. Bayes Factors. *J Am Stat Assoc*. 1995;90: 773.
  125. Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media; 2007.
  126. Lokern: Kernel regression smoothing with local or global plug-in bandwidth. [cited 16 Mar 2021]. Available: <https://CRAN.R-project.org/package=lokern>
  127. Pulak R. Techniques for Analysis, Sorting, and Dispensing of *C. elegans* on the COPAS™ Flow-Sorting System. In: Strange K, editor. *C elegans: Methods and Applications*. Totowa, NJ: Humana Press; 2006. pp. 275–286.
  128. Nika L, Gibson T, Konkus R, Karp X. Fluorescent Beads Are a Versatile Tool for Staging *Caenorhabditis elegans* in Different Life Histories. *G3* . 2016;6: 1923–1933.

129. Vuaridel-Thurre G, Vuaridel AR, Dhar N, McKinney JD. Computational Analysis of the Mutual Constraints between Single-Cell Growth and Division Control Models. *Adv Biosyst.* 2020;4: e1900103.
130. Cho JY, Choi T-W, Kim SH, Ahnn J, Lee S-K. Morphological Characterization of small, dumpy, and long Phenotypes in *Caenorhabditis elegans*. *Mol Cells.* 2021;44: 160–167.
131. Page AP, Johnstone IL. The cuticle. *WormBook*; 2007.
132. Holzapfel GA. Similarities between soft biological tissues and rubberlike materials. *Constitutive Models for Rubber IV.* 2017; 607–617.
133. Essmann CL, Elmi M, Shaw M, Anand GM, Pawar VM, Srinivasan MA. In-vivo high resolution AFM topographic imaging of *Caenorhabditis elegans* reveals previously unreported surface structures of cuticle mutants. *Nanomedicine.* 2017;13: 183–189.
134. Dodd W, Tang L, Lone J-C, Wimberly K, Wu C-W, Consalvo C, et al. A Damage Sensor Associated with the Cuticle Coordinates Three Core Environmental Stress Responses in *Caenorhabditis elegans*. *Genetics.* 2018;208: 1467–1482.
135. Yochem J, Lažetić V, Bell L, Chen L, Fay D. *C. elegans* NIMA-related kinases NEKL-2 and NEKL-3 are required for the completion of molting. *Dev Biol.* 2015;398: 255–266.
136. Petzold BC, Park S-J, Ponce P, Roozeboom C, Powell C, Goodman MB, et al. *Caenorhabditis elegans* body mechanics are regulated by body wall muscle tone. *Biophys J.* 2011;100: 1977–1985.
137. Cox GN, Staprans S, Edgar RS. The cuticle of *Caenorhabditis elegans*. II. Stage-specific changes in ultrastructure and protein composition during postembryonic development. *Dev Biol.* 1981;86: 456–470.
138. Park S-J, Goodman MB, Pruitt BL. Analysis of nematode mechanics by piezoresistive displacement clamp. *Proc Natl Acad Sci U S A.* 2007;104: 17376–17381.
139. Gilpin W, Uppaluri S, Brangwynne CP. Worms under Pressure: Bulk Mechanical Properties of *C. elegans* Are Independent of the Cuticle. *Biophysical Journal.* 2015. pp. 1887–1898. doi:10.1016/j.bpj.2015.03.020
140. Golden JW, Riddle DL. The *Caenorhabditis elegans* dauer larva: developmental effects of pheromone, food, and temperature. *Dev Biol.* 1984;102: 368–378.
141. Schiller HB, Fässler R. Mechanosensitivity and compositional dynamics of cell-matrix adhesions. *EMBO Rep.* 2013;14: 509–519.
142. Wolfenson H, Bershadsky A, Henis YI, Geiger B. Actomyosin-generated tension controls the molecular kinetics of focal adhesions. *J Cell Sci.* 2011;124: 1425–1432.

143. Suman SK, Daday C, Ferraro T, Vuong-Breder T, Tak S, Quintin S, et al. The plakin domain of VAB-10/plectin acts as a hub in a mechanotransduction pathway to promote morphogenesis. *Development*. 2019;146. doi:10.1242/dev.183780
144. Zhang H, Landmann F, Zahreddine H, Rodriguez D, Koch M, Labouesse M. A tension-induced mechanotransduction pathway promotes epithelial morphogenesis. *Nature*. 2011;471: 99–103.
145. Moerman DG, Williams BD. Sarcomere assembly in *C. elegans* muscle. *WormBook*. 2006; 1–16.
146. Costa M, Draper BW, Priess JR. The role of actin filaments in patterning the *Caenorhabditis elegans* cuticle. *Dev Biol*. 1997;184: 373–384.
147. Broday L, Hauser CA, Kolotuev I, Ronai Z ’ev. Muscle-epidermis interactions affect exoskeleton patterning in *Caenorhabditis elegans*. *Dev Dyn*. 2007;236: 3129–3136.
148. Madaan U, Faure L, Chowdhury A, Ahmed S, Ciccarelli EJ, Gumienny TL, et al. Feedback regulation of BMP signaling by cuticle collagens. *Mol Biol Cell*. 2020;31: 825–832.
149. Katz SS, Maybrun C, Maul-Newby HM, Frand AR. Non-canonical apical constriction shapes emergent matrices in *C. elegans*. *bioRxiv*. 2018. p. 189951. doi:10.1101/189951
150. Fechner S, Loizeau F, Nekimken AL, Pruitt BL, Goodman MB. The bodies of *dpy-10(e128)* are twice as stiff as wild type. *MicroPubl Biol*. 2018;2018. doi:10.17912/ecsm-mp67
151. Horvitz HR, Sulston JE. Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. *Genetics*. 1980;96: 435–454.
152. Waddington CH. CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS. *Nature*. 1942;150: 563–565.
153. Tanner JM. THE REGULATION OF HUMAN GROWTH. *Child Dev*. 1963;34: 817–847.
154. Cameron N, Bogin B. Human Growth and Development. Academic Press; 2012.
155. Desmond C, Casale D. Catch-up growth in stunted children: Definitions and predictors. *PLoS One*. 2017;12: e0189135.
156. Hector KL, Nakagawa S. Quantitative analysis of compensatory and catch-up growth in diverse taxa. *J Anim Ecol*. 2012;81: 583–593.
157. Crombie TA, Zdraljevic S, Cook DE, Tanny RE, Brady SC, Wang Y, et al. Deep sampling

- of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. *eLife*. 2019;8. doi:10.7554/eLife.50465
158. Andersen EC, Rockman MV. Natural genetic variation as a tool for discovery in *Caenorhabditis* nematodes. *Genetics*. 2022;220. doi:10.1093/genetics/iyab156
159. García-González AP, Ritter AD, Shrestha S, Andersen EC, Yilmaz LS, Walhout AJM. Bacterial Metabolism Affects the *C. elegans* Response to Cancer Chemotherapeutics. *Cell*. 2017;169: 431–441.e8.
160. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. *Nature*. 2013. pp. 234–237. doi:10.1038/nature11867
161. Zdraljevic S, Fox BW, Strand C, Panda O, Tenjo FJ, Brady SC, et al. Natural variation in *C. elegans* arsenic toxicity is explained by differences in branched chain amino acid metabolism. *eLife*. 2019. doi:10.7554/elife.40260
162. Evans KS, Brady SC, Bloom JS, Tanny RE, Cook DE, Giuliani SE, et al. Shared Genomic Regions Underlie Natural Variation in Diverse Toxin Responses. *Genetics*. 2018;210: 1509–1525.
163. R Core Team R, Others. R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria; 2013. Available: [https://hero.epa.gov/hero/index.cfm/reference/details/reference\\_id/7469942](https://hero.epa.gov/hero/index.cfm/reference/details/reference_id/7469942)
164. Rockman MV, Kruglyak L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet*. 2009;5: e1000419.
165. Lee D, Zdraljevic S, Stevens L, Wang Y, Tanny RE, Crombie TA, et al. Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nat Ecol Evol*. 2021;5: 794–807.
166. Gillooly DJ, Simonsen A, Stenmark H. Cellular functions of phosphatidylinositol 3-phosphate and FYVE domain proteins. *Biochem J*. 2001;355: 249–258.
167. Nicot A-S, Fares H, Payrastre B, Chisholm AD, Labouesse M, Laporte J. The phosphoinositide kinase PIKfyve/Fab1p regulates terminal lysosome maturation in *Caenorhabditis elegans*. *Mol Biol Cell*. 2006;17: 3062–3074.
168. Asahina M, Ishihara T, Jindra M, Kohara Y, Katsura I, Hirose S. The conserved nuclear receptor Ftz-F1 is required for embryogenesis, moulting and reproduction in *Caenorhabditis elegans*. *Genes Cells*. 2000;5: 711–723.
169. Gissendanner CR, Sluder AE. *nhr-25*, the *Caenorhabditis elegans* ortholog of *ftz-f1*, is required for epidermal and somatic gonad development. *Dev Biol*. 2000;221: 259–272.

170. Maeda I, Kohara Y, Yamamoto M, Sugimoto A. Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr Biol.* 2001;11: 171–176.
171. Houle D. Comparing evolvability and variability of quantitative traits. *Genetics.* 1992;130: 195–204.
172. Ghosh R, Bloom JS, Mohammadi A, Schumer ME, Andolfatto P, Ryu W, et al. Genetics of Intraspecies Variation in Avoidance Behavior Induced by a Thermal Stimulus in *Caenorhabditis elegans*. *Genetics.* 2015;200: 1327–1339.
173. Noble LM, Chelo I, Guzella T, Afonso B, Riccardi DD, Ammerman P, et al. Polygenicity and Epistasis Underlie Fitness-Proximal Traits in the *Caenorhabditis elegans* Multiparental Experimental Evolution (CeMEE) Panel. *Genetics.* 2017;207: 1663–1685.
174. Bernstein MR, Zdraljevic S, Andersen EC, Rockman MV. Tightly linked antagonistic-effect loci underlie polygenic phenotypic variation in *C. elegans*. *Evol Lett.* 2019;3: 462–473.
175. Rockman MV. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution.* 2012;66: 1–17.
176. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A Saturated Map of Common Genetic Variants Associated with Human Height from 5.4 Million Individuals of Diverse Ancestries. *bioRxiv.* 2022. p. 2022.01.07.475305.  
doi:10.1101/2022.01.07.475305
177. Mörck C, Pilon M. *C. elegans* feeding defective mutants have shorter body lengths and increased autophagy. *BMC Dev Biol.* 2006;6: 39.
178. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnipotent. *Cell.* 2017;169: 1177–1186.
179. Snoek LB, Orbidans HE, Stastna JJ, Aartse A, Rodriguez M, Riksen JAG, et al. Widespread genomic incompatibilities in *Caenorhabditis elegans*. *G3.* 2014;4: 1813–1823.
180. Maulana MI, Riksen JAG, Snoek BL, Kammenga JE, Sterken MG. The genetic architecture underlying body-size traits plasticity over different temperatures and developmental stages in *Caenorhabditis elegans*. *Heredity.* 2022.  
doi:10.1038/s41437-022-00528-y
181. Xie Y. *bookdown: Authoring Books and Technical Documents with R Markdown.* CRC Press; 2016.
182. Nyaanga J, Crombie TA, Widmayer SJ, Andersen EC. *easyXpress: An R package to analyze and visualize high-throughput C. elegans microscopy data generated using CellProfiler.* *PLoS One.* 2021;16: e0252000.

183. Swedlow JR. Innovation in biological microscopy: current status and future directions. *Bioessays*. 2012;34: 333–340.
184. Cassidy PJ, Radda GK. Molecular imaging perspectives. *J R Soc Interface*. 2005;2: 133–144.
185. White JG, Southgate E, Thomson JN, Brenner S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci*. 1986;314: 1–340.
186. Teotónio H, Estes S, Phillips PC, Baer CF. Experimental Evolution with Nematodes. *Genetics*. 2017;206: 691–716.
187. Wählby C, Kamentsky L, Liu ZH, Riklin-Raviv T, Conery AL, O'Rourke EJ, et al. An image analysis toolbox for high-throughput *C. elegans* assays. *Nat Methods*. 2012;9: 714–716.
188. Team RC. R: A Language and Environment for Statistical Computing <http://www.R-project.org>. 2014.
189. Boyd WA, Smith MV, Freedman JH. *Caenorhabditis elegans* as a Model in Developmental Toxicology. In: Harris C, Hansen JM, editors. *Developmental Toxicology: Methods and Protocols*. Totowa, NJ: Humana Press; 2012. pp. 15–24.
190. Tukey JW, Others. Exploratory data analysis. Reading, Mass.; 1977.
191. Dilks CM, Hahnel SR, Sheng Q, Long L, McGrath PT, Andersen EC. Quantitative benzimidazole resistance and fitness effects of parasitic nematode beta-tubulin alleles. *Int J Parasitol Drugs Drug Resist*. 2020;14: 28–36.
192. Wit J, Rodriguez BC, Andersen EC. Natural variation in *Caenorhabditis elegans* responses to the anthelmintic emodepside. *Int J Parasitol Drugs Drug Resist*. 2021;16: 1–8.
193. Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, et al. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*. 1997;387: 489–493.
194. Telford MJ, Bourlat SJ, Economou A, Papillon D, Rota-Stabelli O. The evolution of the Ecdysozoa. *Philos Trans R Soc Lond B Biol Sci*. 2008;363: 1529–1537.
195. Alpatov WW. Growth and variation of the larvae of *Drosophila melanogaster*. *J Exp Zool*. 1929;52: 407–437.
196. Rice AL. Growth “rules” and the larvae of decapod crustaceans. *J Nat Hist*. 1968;2: 525–530.

197. Howells RE, Blainey LJ. The moulting process and the phenomenon of intermoult growth in the filarial nematode *Brugia pahangi*. *Parasitology*. 1983;87 (Pt 3): 493–505.
198. Wilson PAG. Nematode growth patterns and the moulting cycle: the population growth profile. *J Zool*. 1976;179: 135–151.
199. Byerly L, Scherer S, Russell RL. The life cycle of the nematode *Caenorhabditis elegans*. *Developmental Biology*. 1976. pp. 34–48. doi:10.1016/0012-1606(76)90120-2
200. Malakhov UV, Others. Nematodes: Structure, development, classification and phylogeny. *Nematodes: structure, development, classification and phylogeny*. 1986. Available: <https://www.cabdirect.org/cabdirect/abstract/19891130545>
201. Essmann CL, Martinez-Martinez D, Pryor R, Leung K-Y, Krishnan KB, Lui PP, et al. Mechanical properties measured by atomic force microscopy define health biomarkers in ageing *C. elegans*. *Nature Communications*. 2020. doi:10.1038/s41467-020-14785-0
202. Rahimi M, Sohrabi S, Murphy CT. Novel elasticity measurements reveal *C. elegans* cuticle stiffens with age and in a long-lived mutant. *Biophys J*. 2022;121: 515–524.
203. Riddle DL, Blumenthal T, Meyer BJ, Priess JR. *Introduction: the Neural Circuit For Locomotion*. Cold Spring Harbor Laboratory Press; 1997.
204. Altun ZF, Hall DH. WormAtlas hermaphrodite handbook - nervous system - general description. WormAtlas. 2005. doi:10.3908/wormatlas.1.18
205. Altun ZF, Hall DH. WormAtlas Hermaphrodite Handbook - Introduction. WormAtlas. 2006. doi:10.3908/wormatlas.1.1
206. Pigliucci M, Murren CJ, Schlichting CD. Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol*. 2006;209: 2362–2367.
207. Lafuente E, Beldade P. Genomics of Developmental Plasticity in Animals. *Front Genet*. 2019;10: 720.
208. Pigliucci M. Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology & Evolution*. 2005. pp. 481–486. doi:10.1016/j.tree.2005.06.001
209. Hodgins-Davis A, Adomas AB, Warringer J, Townsend JP. Abundant gene-by-environment interactions in gene expression reaction norms to copper within *Saccharomyces cerevisiae*. *Genome Biol Evol*. 2012;4: 1061–1079.
210. Rocha F, Medeiros HF, Klaczko LB. The reaction norm for abdominal pigmentation and its curve in *Drosophila mediopunctata* depend on the mean phenotypic value. *Evolution*. 2009;63: 280–287.

# Appendix

I had the pleasure of collaborating with members of the Mangan group in the Engineering Science and Applied Math department at Northwestern to investigate *C. elegans* growth using mathematical models. The following are two manuscripts that came out of this collaboration and directly relate to data and thoughts presented in Chapter 2 of this thesis. They are both currently in preparation for submission as microPublications.

## A. Growth dynamics of *C. elegans* cuticle mutants

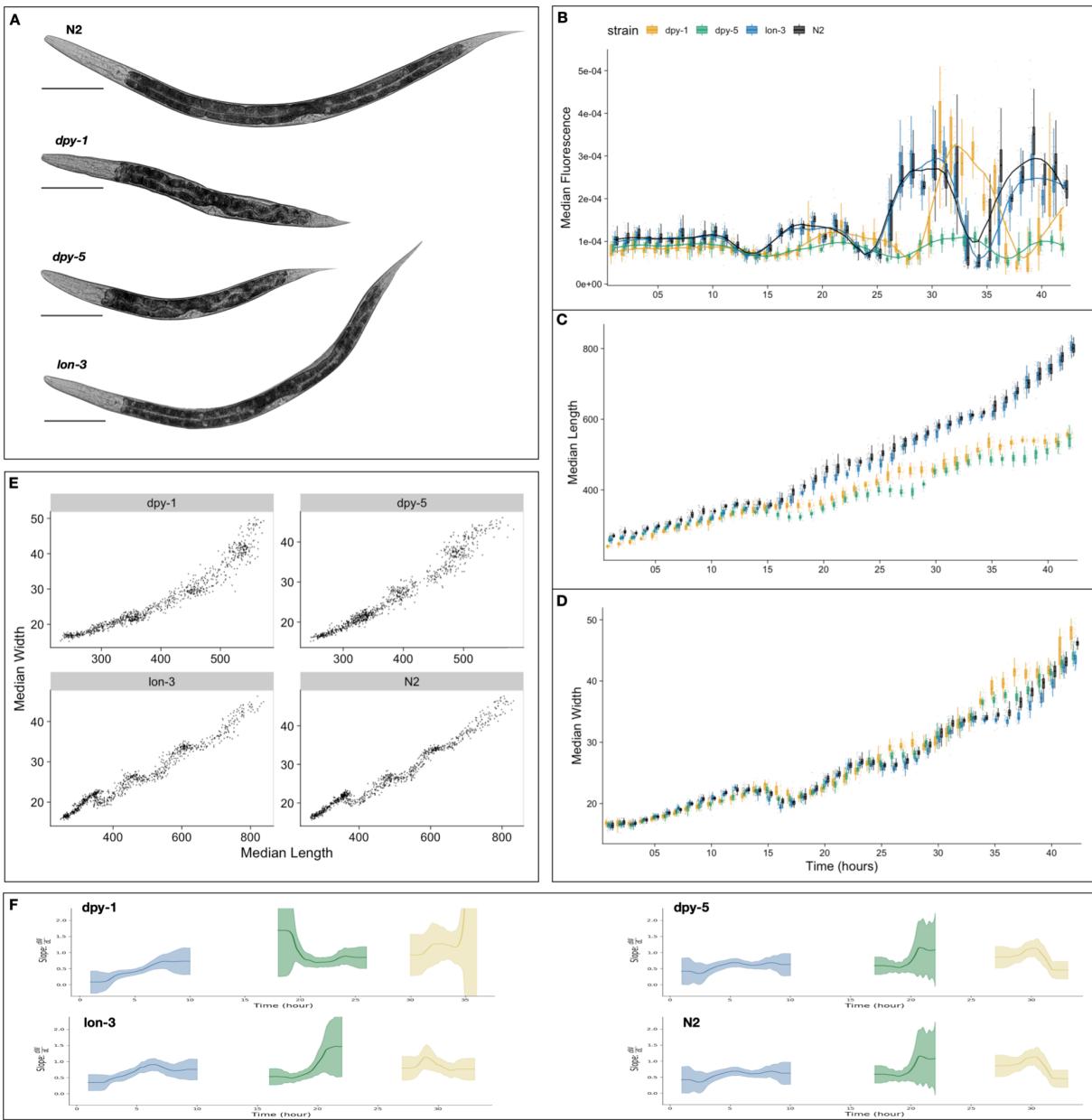
### Characterization of larval growth in *C. elegans* cuticle mutants

Joy Nyaanga<sup>1,2</sup>, Sasha Shirman<sup>3</sup>, Niall M. Mangan<sup>3</sup>, and Erik C. Andersen<sup>1</sup>

1. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA
2. Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA
3. Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208, USA

### Abstract

In *Caenorhabditis elegans*, many genes involved in the formation of the cuticle are also known to influence body size and shape. We assessed post-embryonic growth of both long and short *C. elegans* body size mutants from the L1 to L4 stage. We found similar developmental trajectories of N2 and *lon-3* animals. By contrast, we observed overall decreases in body length and increases in body width of tested *dpy* mutants compared to N2, consistent with the Dpy phenotype. We further show that the dynamics of animal shape in the mutant strains are consistent with a previously proposed “Stretcher” growth model.



**Figure A-1: Quantitative assessment of *C. elegans* larval growth.**

(A) Representative images of strains used in this study taken at the L4 stage. Tukey boxplots for median animal fluorescence normalized by area (B), median animal length ( $\mu\text{M}$ ) (C), and median animal width ( $\mu\text{M}$ ) (D). The horizontal line in the middle of the box is the median, and

the box denotes the 25th to 75th quantiles of the data. The vertical line represents the 1.5 interquartile range. Each point corresponds to the median value of a population of animals in each well. **(E)** Median length (x-axis) plotted against median length (y-axis). **(F)** The ratio of the change in width to length over time. Calculated from the local slope of data in panel (E). The standard deviation captures population variation (grey).

## Description

Body size largely influences an organism's functional characteristics: growth, reproduction, metabolism, lifespan. As a result, the determinant factors of organism size, particularly during development, have been explored in many systems. *Caenorhabditis elegans*, a free-living nematode, presents a versatile genetic model system to study how the processes of growth and development are regulated. *C. elegans* matures to an adult after multiple molting events during which time animals synthesize a new exoskeleton (cuticle) and expel their old one. The *C. elegans* cuticle is a complex, multi-layered structure primarily composed of collagens. As animals progress through their life-cycle, the structure and thickness of the cuticle changes but its role in the maintenance of body morphology and integrity remains. To date, 21 cuticle collagen mutants have been identified that cause a range of body morphology defects (Page and Johnstone 2007). Some of these mutants exhibit a disproportionate reduction in body size, while others are noticeably larger than wild type (Cho et al. 2021), clearly demonstrating the importance of the physical structure of the cuticle on growth. Analyzing the characteristics of size during development in various *C. elegans* body shape mutants is central to understanding the role these genetic pathways have on body growth.

We performed a high-resolution longitudinal study of growth in a selection of *C. elegans* cuticle mutants. We selected mutants that were both reportedly shorter (*dpy-1(e1)*, *dpy-5(e61)*) and longer (*lon-3(e2175)*) than wild type (Cho et al. 2021) (**Figure A-1A**). We then collected high-precision measurements of animal fluorescence (**Figure A-1B**), length (**Figure A-1C**), and width (**Figure A-1D**) from the L1 stage through the L4 stage. As we previously demonstrated (Nyaanga et al. 2022), we can use oscillations in fluorescence as a proxy for feeding behavior to characterize larval progression by associating periods of decreased feeding with molt events. Doing so, we notice that *lon-3(e2175)* animals follow similar developmental trajectories to the N2 wild type. By contrast, we detect a delay in the molt timing of both *dpy* mutant strains, with *dpy-1(e1)* undergoing each larval transition later than all other tested strains. We also observe a marked decrease in animal length and increase in animal width noticeable after the L1 stage in both *dpy* strains, consistent with their characteristic *umpy* phenotype. Interestingly, we note little size divergence between *lon-3(e2175)* and N2 animals during our time course.

Measurements of both animal length and width allow us to assess changes in body shape as well as size. Previously, motivated by changes in the body aspect ratio of animals we observed at larval stage transitions, we modeled a physical mechanism by which constraints on cuticle stretch could cause changes in *C. elegans* body shape (Nyaanga et al. 2022). We found that model-predicted shape changes were consistent with those seen in our data of N2 animals. Given this result, we proposed a “Stretcher” model for growth wherein *C. elegans* sense changes in cuticle elasticity, in tandem with other regulatory mechanisms, to control growth rate and determine developmental transitions. Given the structural impacts of *dpy-1(e1)*, *dpy-5(e61)* and *lon-3(e2175)* mutants, we sought to determine whether the shape dynamics predicted by the

Stretcher model would be consistent with the mutant data. By analyzing the relationship between measured animal length and width ( $\Delta W/\Delta L$ ) over time, we are able to detect the linear and nonlinear stretch regimes predicted by the Stretcher model (**Figure 1F**). In all strains, we observe an approximately constant  $\Delta W/\Delta L$  ratio in all larval stages, consistent with a linear stretch regime. Next, we observe a shape slope increase, consistent with a nonlinear stretch regime in length preceding larval stage transitions.

## Methods

### Worm culture

The laboratory strain N2 was obtained from the *C. elegans* Natural Diversity Resource (Cook et al. 2017). All other strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). Animals were cultured at 20°C on 6 cm plates of modified nematode growth media (NGMA), which contained 1% agar and 0.7% agarose seeded with *E. coli* OP50 bacteria.

### High-throughput growth assay

Measurements of body size and fluorescence were measured as previously described (Nyaanga et al. 2022). Briefly, strains were propagated for three generations, bleach-synchronized, and titered at a concentration of 1 embryo per  $\mu\text{L}$  into 250 mL flasks. The following day, arrested L1s were fed HB101 food at a final concentration of OD20 in a final flask volume of 100 mL K medium and HB101 food. Animals were grown with constant shaking at 20°C. Flasks were sampled each hour beginning one hour after feeding and continuing for 42 consecutive hours. At each hour, 800  $\mu\text{L}$  was removed from each flask and incubated with fluorescent polychromatic beads

(Polysciences, 19507-5) for 10 minutes with shaking. Following the bead incubation, animals were transferred to a 96-well microtiter plate, treated with sodium azide, imaged with an ImageXpress Nano (Molecular Devices, SanJose, CA), and scored using a large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston MA). COPAS BIOSORT was used to collect measurements of animal length (TOF), optical extinction (EXT), and red fluorescence for every animal in each well.

### **Data processing**

COPAS BIOSORT data were processed as previously described (Nyaanga et al. 2022). To remove non-animal objects such as bacterial clumps, shed cuticles, and next generation larval animals from the time-course data. Data for each well was summarized to obtain median well measurements. TOF and norm.EXT data were then converted to microns. Only the unit-corrected data were used for further analysis. “Stretcher” model analysis of shape dynamics was performed as previously described (Nyaanga et al. 2022).

### **Reagents**

STRAIN	GENOTYPE	AVAILABLE FROM
N2	<i>Caenorhabditis elegans</i>	CeNDR
CB1	<i>dpy-1(e1)</i>	CGC
CB61	<i>dpy-5(e61)</i>	CGC
CB4123	<i>lon-3(e2175)</i>	CGC

### **Funding**

For this work, S.S., J.N., E.C.A., and N.M.M received support from the NSF-Simons Center for Quantitative Biology at Northwestern University (awards Simons Foundation/SFARI

597491-RWC and the National Science Foundation 1764421). S.S. and N.M.M. received support from the National Science Foundation RTG: Interdisciplinary Training in Quantitative Biological Modeling, award 1547394.

## Author Contributions

Joy Nyaanga: Conceptualization, Methodology, Investigation, Formal Analysis, Visualization, Writing - original draft, Writing - review and editing

Sasha Shirman: Methodology, Formal Analysis, Visualization, Writing - review and editing

Niall M. Mangan: Methodology, Funding acquisition, Supervision, Writing - review and editing

Erik C. Andersen: Conceptualization, Methodology, Investigation, Funding acquisition, Supervision, Writing - review and editing.

## References

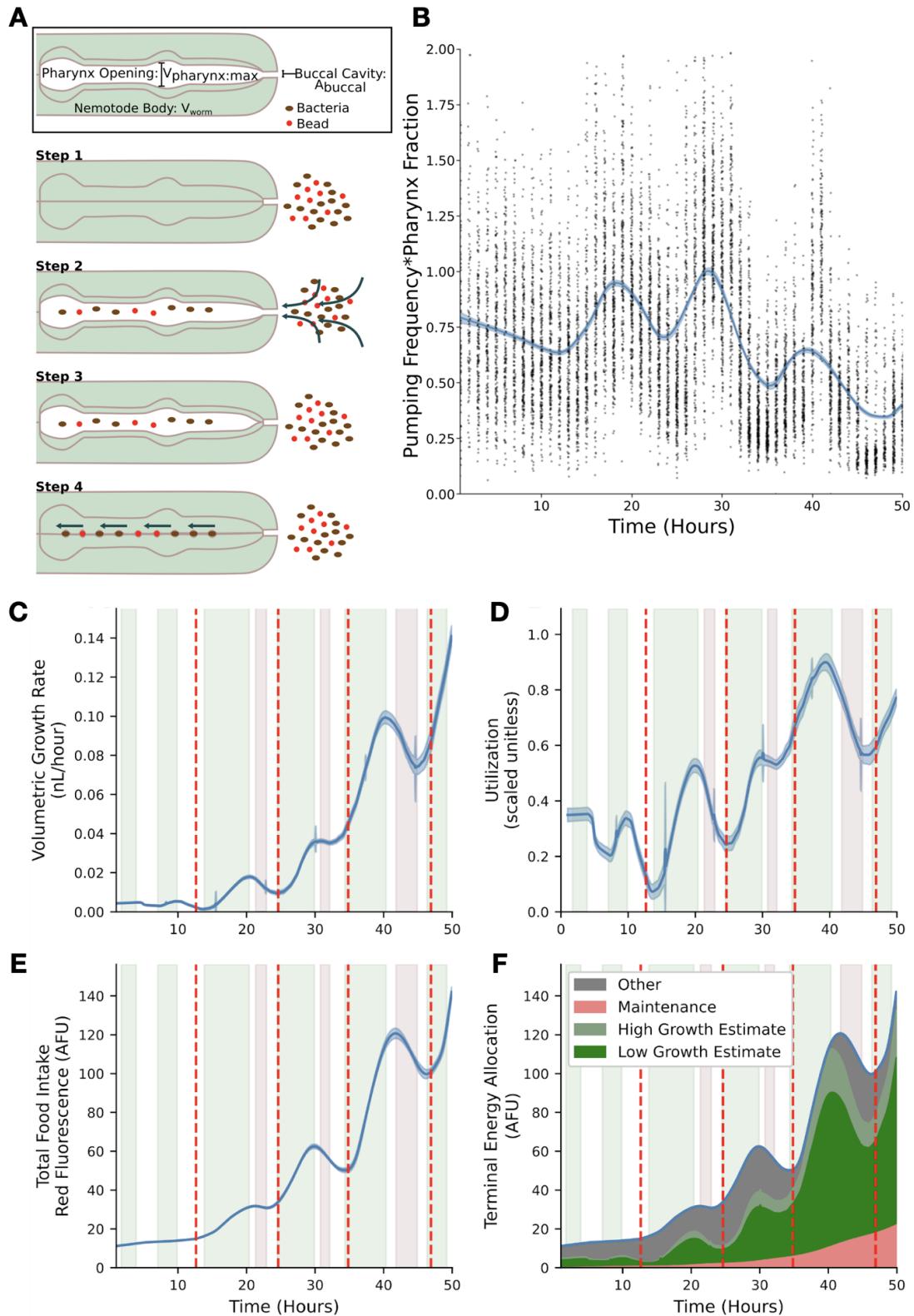
1. Page AP, Johnstone IL. The cuticle. WormBook. 2007 Mar 19;1–15. PMCID: PMC4781593
2. Cho JY, Choi T-W, Kim SH, Ahnn J, Lee S-K. Morphological Characterization of small, dumpy, and long Phenotypes in *Caenorhabditis elegans*. Mol Cells. Korean Society for Molecular and Cellular Biology; 2021 Mar 31;44(3):160–167. PMCID: PMC8019597
3. Nyaanga J, Goss C, Zhang G, Ahmed HN, Andersen EJ, Miller IR, Rozenich JK, Swarthout IL, Vaughn JA, Mangan NM, Shirman S, Andersen EC. Changes in body shape implicate cuticle stretch in *C. elegans* growth control [Internet]. bioRxiv. 2022 [cited 2022 Mar 28]. p. 2021.04.01.438121. Available from: <https://www.biorxiv.org/content/10.1101/2021.04.01.438121v3>
4. Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga JE, Wang J, Kruglyak L, Félix M-A, Lee J, Andersen EC. The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. Genetics. 2016 Sep;204(1):371–383. PMCID: PMC5012401

**B. Food allocation during *C. elegans* growth****Mechanical control of feeding and allocation of food energy likely constrain growth dynamics in *C. elegans***

Sasha Shirman<sup>1\*</sup>, Joy Nyaanga<sup>2,3\*</sup>, Erik C. Andersen<sup>2</sup>, and Niall M. Mangan<sup>1</sup>

1. Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208, USA
2. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA
3. Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA

\*Co-first authors



**Figure B-1: Visualization and analysis of food intake and utilization models.** **(A)** A single cycle in which animals take up and transport food to the gut consists of four general steps. The cycle begins with relaxed pharyngeal muscles and a closed pharyngeal lumen (step 1). The animal opens the pharyngeal lumen so that fluid and food flow through the buccal cavity and into the pharyngeal lumen (step 2). The animal relaxes the muscles that control the opening, stopping the flow of fluid through the buccal cavity and trapping a volume of media approximately equal to the maximum volume of the pharyngeal lumen (step 3). Finally, the pharynx closes, extruding excess media and trapping bacteria and beads that are then ‘swallowed’ (step 4). **(B)** Product of pumping frequency and pharyngeal lumen fraction. **(C)** Dynamics of volume growth rate. **(D)** Dynamics of the estimated food utilization towards growth. **(E)** Dynamics of red fluorescence. **(F)** Dynamics of food allocation breakdown. Food allocated toward maintenance is proportional to animal size (pink). Food allocated toward growth is calculated from food intake (C) and utilization (D) to produce low (dark green) and high (combined green) estimates. Food allocated to other metabolic processes consists of remaining food resources (grey). In panels C-E, the solid blue line represents the mean bootstrap regression. Standard deviation on the regression is marked by the shaded blue region. In panels C-F, Vertical red lines mark molt times. Errors on shaded regions are  $\pm 0.7$  hours. These data are from replicate 2 and are representative of all replicates.

## Description

We have previously shown that changes in physical properties of the *C. elegans* cuticle might serve as a cue for developmental timing (Nyaanga et al. 2021). *C. elegans* must have

mechanisms to control growth throughout development, particularly in response to these cues. Like most species, *C. elegans* do not increase their growth rate indefinitely in response to increased food availability (Uppaluri and Brangwynne 2015). The animals could control growth entirely using feeding rate, as they actively control the feeding rate (Fang-Yen et al. 2009) and stop feeding at the initiation of a molt (Singh and Sulston 1978). In addition to this mechanical control, they could use metabolic control to preemptively divert ingested resources toward or away from growth. Metabolic processing of stored resources could be especially useful if animals complete their molt and enter a food-limited environment. Using a quantitative feeding model, we examine the *C. elegans* growth data (Nyaanga et al. 2021) to investigate the possible mechanical control of feeding and the metabolic control of the allocation of ingested food toward organismal growth and development.

*C. elegans* is a filter feeder that pumps its food through a cycle of pharyngeal muscle contractions and relaxations that alternatively open and close the pharyngeal cavity. *C. elegans* are capable of actively modulating the length of time used to contract or open a subset of the pharyngeal muscles (Fang-Yen et al. 2009). Our feeding model describes how control of this time length and thus the control of the overall pumping period, defined as the length of time of one full pumping cycle, translates to changes in the rate of food uptake (**Figure B-1A**). The results of our analysis show both an oscillation in pumping rates and a slow change in pharyngeal lumen size throughout development (**Figure B-1B**). The minima of the product of pumping frequency and pharynx fraction occurred during molt times, consistent with prior knowledge (Singh and Sulston 1978; Byerly et al. 1976). The relative length of the pharynx to animal length has been shown to decrease over development (Avery 2003) and a similar trend in

the volume of the pharyngeal lumen could explain the slow decrease in the product of pumping frequency and pharynx fraction.

To understand whether animals control growth rate primarily using the mechanical feeding process or metabolic regulation, we describe how food is utilized once it is ingested by *C. elegans*. We assumed three general categories of food utilization: maintenance, observed volumetric growth, and all other processes. We assume that resources required for maintenance are proportional to the volume of the animal. Other processes may include the development of tissues and structures, such as reproductive components, within the worm which are resource intensive, but are not detected during the measurements of the width, length, or volume. Food intake rate (using red fluorescence as a proxy) and volume growth rate (numerically differentiated from the volume regression) follow similar dynamics throughout much of *C. elegans* development (**Figure 1C** and **Figure 1E**). Each larval stage consists of three types of dynamics: steady growth during which food intake and growth rate are both increasing (green), preparation for molt during which food intake and growth rate are both decreasing (red), and transition regions where food intake and growth rate do not vary together (white). During steady growth and molt initiation, mechanical food intake rate and allocation of food resources towards observable growth have similar dynamics. At transition points the allocation of food resources towards growth, reaches a local minimum or maximum. These local extrema in food utilization at transition points likely indicate large changes in metabolic regulation (**Figure 1D**).

Our analysis of the relationship between growth rate and food intake rate quantified the interplay of metabolic regulation and mechanical food intake (**Figure 1F**). Mechanical food intake provides the upper bound of available food resources for growth, but metabolic regulation

substantially changes how much of this available food is utilized for volumetric growth across development. Local extrema in metabolic regulation dynamics coincide with the start of transition times. Within the L2 and L3 stages we observed a decoupling between growth rate and food intake dynamics (**Figure 1F** white region) twice. The first of these time periods corresponds to the times at which the width-to-length ratio drastically changes and the second corresponds to ecdysis. We estimated that throughout larval stages metabolic resources allocated to non-growth processes varies much less than food resources required for growth and maintenance. The oscillatory behavior of food utilization motivates the need for further metabolomic experiments to probe metabolic regulatory dynamics particularly in around larval-stage transitions.

## Methods

### Worm culture

The canonical laboratory strain N2 was obtained from the *C. elegans* Natural Diversity Resource (Cook et al. 2017) and animals were cultured at 20°C on 6 cm plates of modified nematode growth media (NGMA), which contained 1% agar and 0.7% agarose seeded with *E. coli* OP50 bacteria.

### High-throughput growth assay

Measurements of body size and fluorescence were measured as previously described (Nyaanga et al. 2021). Briefly, strains were propagated for three generations, bleach-synchronized, and titered at a concentration of 1 embryo per  $\mu\text{L}$  into six replicate 500 mL flasks. The following day, arrested L1s were fed HB101 food at a final concentration of OD20 in a final flask volume of 100 mL K medium and HB101 food. Animals were allowed to grow at 20°C with constant

shaking. Flasks were sampled each hour beginning one hour after feeding and continuing for 72 consecutive hours. At each hour, 500  $\mu$ L was removed from each flask and incubated with fluorescent polychromatic beads (Polysciences, 19507-5) for 10 minutes with shaking. Following the bead incubation, animals were aliquoted to a 96-well microtiter plate, treated with sodium azide, imaged with an ImageXpress Nano (Molecular Devices, SanJose, CA), and scored using a large-particle flow cytometer (COPAS BIOSORT, Union Biometrica, Holliston MA). COPAS BIOSORT was used to collect measurements of animal length (TOF), optical extinction (EXT), and red fluorescence for every animal in each well.

### **Feeding model analysis**

To analyze volumetric growth dynamics and feeding dynamics, volume regression was calculated using a cylindrical approximation for animal shape and the same local kernel regression previously described (Nyaanga et al. 2021) was applied to red fluorescence data. Volume growth rate was calculated using the python numpy gradient function applied to the volume regressions. An additional moving time window average (1.4 hours) was applied to smooth numerical errors in the derivative when determining feeding and growth regime transition points. Model derivations can be found in the Extended Data file.

### **Funding**

For this work, S.S., J.N., E.C.A., and N.M.M received support from the NSF-Simons Center for Quantitative Biology at Northwestern University (awards Simons Foundation/SFARI 597491-RWC and the National Science Foundation 1764421). S.S. and N.M.M. received support from the National Science Foundation RTG: Interdisciplinary Training in Quantitative Biological Modeling, award 1547394).

## Author Contributions

Sasha Shirman: Methodology, Formal Analysis, Visualization, Writing - review and editing

Joy Nyaanga: Investigation, Data curation, Writing - original draft, Writing - review and editing

Erik C. Andersen and Niall M. Mangan: Conceptualization, Funding acquisition, Supervision, Writing - review and editing.

## References

1. Nyaanga J, Goss C, Zhang G, Ahmed HN, Andersen EJ, Miller IR, Rozenich JK, Swarthout IL, Vaughn JA, Andersen EC, Mangan NM, Shirman S. Highly scaled measurements of *C. elegans* development suggest that physical constraints guide growth trajectories and animal shape [Internet]. bioRxiv. 2021 [cited 2021 Sep 8]. p. 2021.04.01.438121. Available from: <https://www.biorxiv.org/content/10.1101/2021.04.01.438121v2>
2. Uppaluri S, Brangwynne CP. A size threshold governs *Caenorhabditis elegans* developmental progression. Proc Biol Sci [Internet]. 2015 Aug 22;282(1813):20151283. Available from: <http://dx.doi.org/10.1098/rspb.2015.1283> PMID: PMC4632629
3. Fang-Yen C, Avery L, Samuel ADT. Two size-selective mechanisms specifically trap bacteria-sized food particles in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A [Internet]. 2009 Nov 24;106(47):20093–20096. Available from: <http://dx.doi.org/10.1073/pnas.0904036106> PMCID: PMC2785297
4. Singh RN, Sulston JE. Some Observations On Moulting in *Caenorhabditis Elegans*. Nematologica [Internet]. Brill; 1978 Jan 1 [cited 2020 Apr 18];24(1):63–71. Available from: [https://brill.com/view/journals/nema/24/1/article-p63\\_7.xml](https://brill.com/view/journals/nema/24/1/article-p63_7.xml)
5. Byerly L, Cassada RC, Russell RL. The life cycle of the nematode *Caenorhabditis elegans*. I. Wild-type growth and reproduction. Dev Biol [Internet]. 1976 Jul 1;51(1):23–33. Available from: [http://dx.doi.org/10.1016/0012-1606\(76\)90119-6](http://dx.doi.org/10.1016/0012-1606(76)90119-6) PMID: 988845
6. Avery L. Food transport in the *C. elegans* pharynx [Internet]. Journal of Experimental Biology. 2003. p. 2441–2457. Available from: <http://dx.doi.org/10.1242/jeb.00433>
7. Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga JE, Wang J, Kruglyak L, Félix M-A, Lee J, Andersen EC. The

Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. Genetics. 2016 Sep;204(1):371--383. PMCID: PMC5012401

## Extended Data

### Derivation of eating model

We express the volume of the pharyngeal lumen as a fraction,  $g(t)$ , of nematode volume,  $V_{lumen:max}(t) = g(t) \cdot V_{worm}(t)$ . As the pumping period is much faster (ms in adults) than the rate of measured growth (hours), we averaged the food intake rate over a single pumping period. We begin by defining the instantaneous rate of food intake as a function of the flow rate of media through the buccal cavity and the cross sectional area of the buccal cavity

$$\frac{dV_{food}}{dt} = A_{buccal} \quad (1)$$

We then make the assumption that the uptake of media fills the pharynx cavity we have

$$\frac{dV_{food}}{dt} = C \frac{dV_{pharynx}}{dt} \quad (2)$$

We average both sides of the equation under the assumption that the pumping period is significantly shorter than the time scale of growth

$$\frac{1}{T} \int_0^T \frac{dV_{food}}{dt} dt = \frac{1}{T} \int_0^T CA_{buccal} dt \quad (3)$$

The integral on the right hand side is the total food intake during a single pumping period.

$$\frac{1}{T} \int_0^T \frac{dV_{food}}{dt} dt = \frac{1}{T} \Delta V_{food} \quad (4)$$

We then take into account that food is not transported to the gut in the same step as its uptake. Thus the total food intake during a single pump can be calculated by the amount of food that fills the fully opened pharynx

$$\frac{1}{T} \int_0^T \frac{dV_{food}}{dt} dt = \frac{C}{T} \Delta V_{pharynx:max} \quad (5)$$

We then replace the average on the right hand side with the average food intake rate over the pumping period. For simplicity and because we will deal entirely with the average food intake rate, we do not use a different notation for this average rate.

$$\frac{dV_{food}}{dt} dt = \frac{C}{T} \Delta V_{pharynx:max} = Cf(t)V_{pharynx:max} \quad (6)$$

Here,  $C$  is the concentration of bacteria in the culture where animals are grown and  $f(t)$  is the pumping frequency and the inverse of the period of a single pump. Both pumping frequency,  $f(t)$ , and the fractional pharyngeal lumen size,  $g(t)$ , vary over the animal's life cycle. Though any individual bacterium may take multiple pumps to travel down to the gut, we assume that, once food has entered the pharyngeal lumen, it will eventually be metabolized. With this assumption, we are only interested in the rate at which food enters so we do not model the dynamics of food traveling through the gut.

## Defecation analysis

We verify that red fluorescence (denoted as “Red”) can be treated as a proxy measurement for food intake rate given defecation rates. We use the defecation results found in (Liu and Thomas 1994) to determine if red fluorescent measurements can be used as a proxy for food intake rate as opposed to the instantaneous food volume in the gut. Defecation in adults happens very

regularly, with a period of  $T_d = 45 \pm 3s$ , with  $h_d = 43 \pm 10\%$  of their intestinal volume being expelled each time. The volume expelled is well mixed. Defining  $V_f$  as the current amount of food in the nematode gut. We can use conservation of mass to state that the rate of change in the amount of food in the gut is equal to the rate of food intake through eating less the defecation rate and the rate at which food volume is metabolized into cell products:

$$\frac{dV_f}{dt} = \left. \frac{dV_f}{dt} \right|_{eating} - \left. \frac{dV_f}{dt} \right|_{defecating} - \left. \frac{dV_f}{dt} \right|_{metabolized} \quad (7)$$

Using red fluorescence as a proxy for food intake we can ignore the metabolism term as the fluorescent beads were not metabolized.

$$\frac{dV_f}{dt} = \left. \frac{dV_f}{dt} \right|_{eating} - \left. \frac{dV_f}{dt} \right|_{defecating} \quad (8)$$

The above equation states that the rate of change of the volume of fluorescent beads in the gut is equal to the difference between the intake of red fluorescent beads minus the defecation rate of red fluorescent beads. Using the results of (Liu and Thomas 1994) for the second term, and defining  $V_{f:\max}$  as the volume of food in the gut just prior to defecation,  $h_d$  as the fraction of food expelled during a single defecation cycle, and  $T_d$  as the period of defecation. We average Equation (S8) over short time periods to remove the pumping and defecation period oscillations.

$$\frac{dV_{red:\max}}{dt} = \left. \frac{dV_{red}}{dt} \right|_{eating} - \frac{V_{red:\max} h_d}{T_d} \quad (9)$$

On the left hand side, the instantaneous rate of the gut red fluorescence in Equation (S8) is replaced by the rate of change of the maximum or “full” gut fluorescence. The second term on the right hand side of Equation (S8) has been replaced by the average defecation rate over a

cycle calculated by multiplying the full gut fluorescence by the fraction expelled and dividing by the defecation period. We solve Equation (S9) for the average eating rate.

$$\frac{dV_{red}}{dt}|_{eating} = \frac{dV_{red:max}}{dt} + \frac{V_{red:max} h_d}{T_d} \quad (10)$$

We take the local regression of the red fluorescence to determine  $V_{red:max}(t)$ . This value is plugged into the second term on the right hand side of Equation (S10) and its derivative is used to approximate the first term on the right hand side of Equation (S10). We take the adult values of  $h_d$  and  $T_d$  (Liu and Thomas 1994) as a first approximation. Figure (S6) demonstrates a comparison of the red fluorescence and the red intake rate with defecation taken into account at the constant adult rates and quantities. We have scaled both the pink curve denoting red fluorescence and the blue curve denoting red intake rate by their maximum. This scaling allows us to see that the two curves are only a multiplicative factor apart up to error bars. This allows us to use the red measurement as a proxy for both red and food intake rates.

We use red fluorescence and calculated volume as the product of pumping rate and pharyngeal lumen fraction,  $f(t)g(t) \propto \frac{Red}{V_{worm}}$ , over time.

### Derivation of food utilization model

The rate of food conversion to growth can be described by:

$$\frac{dV_{worm}}{dt}(t) = \eta(t)\alpha(t) \frac{dV_{food}}{dt}(t). \quad (11)$$

Here,  $\frac{dV_{food}}{dt}$  is the instantaneous rate of food intake (averaged over a single pump),  $\eta(t)$  is the metabolic efficiency of converting food to growth, and  $\alpha(t)$  is food utilization, meaning the fraction of food used for volumetric growth as opposed to maintenance or other metabolic

processes. If both  $\eta(t)$  and  $\alpha(t)$  are constant, then the equation predicts the rate of volumetric growth is directly proportional to the rate of food consumption, which would be equivalent to controlling growth through feeding with no changes in metabolic regulation (through enzyme expression or other direction of flux between pathways). Efficiency captures the maximum efficiency of metabolism, and we assume it does not vary significantly when the food source is constant,  $\eta(t) = \eta$ . Therefore, utilization captures the dynamics of metabolic regulation. If the fraction of food utilized for growth is not constant, then volumetric growth rate is no longer proportional to food consumption rate and metabolic control must play a role in driving the growth rate.

We calculate the value of the product of metabolic efficiency and food utilization,  $\alpha(t) \eta$ , from the data by manipulating Eq. (11):

$$\alpha(t) \eta = \left( \frac{dV_{\text{worm}}}{dt}(t) \right) / \left( \frac{dV_{\text{food}}}{dt}(t) \right). \quad (12)$$

By using the red fluorescence as a proxy for food intake rate in Eq. (12), we can estimate the dynamical behavior up to scaling factors for the metabolic efficiency times food utilization,  $\alpha(t) \eta \propto \left( \frac{dV_{\text{worm}}}{dt}(t) \right) / \text{Red}$ . Because only utilization varies in time, it captures the dynamics of metabolic regulation:

$$\alpha(t) \propto \left( \frac{dV_{\text{worm}}}{dt}(t) \right) / \text{Red}. \quad (13)$$

# Vita

Joy Nyaanga is a first-generation Kenyan immigrant raised in Buffalo, NY. She attended John Carroll University for undergraduate study where she earned dual degrees in Cell & Molecular Biology and Chemistry. This experience ignited a fascination in the intersection of Biology and Chemistry. She followed this passion, earning a Masters at Princeton University where she investigated the interplay between RNA chemistry and cellular regulatory mechanisms. While at Princeton, a programming course introduced her to the application of quantitative methods in biological sciences and motivated her to search for a doctoral program committed to fostering this newfound interest. Joy went on to pursue a Ph.D. at Northwestern University in the Interdisciplinary Biological Sciences program. There she was a part of the NSF-Simons Center for Quantitative Biology, participating in research aimed at applying mathematical inquiry to the study of animal growth and development. During her doctoral research, Joy authored manuscripts spanning topics from developmental biology, genomics, and software development.