Image cytometry reveals that cell cycle regulation minimizes pedigree depth in the C. elegans germline within constraints of rapid growth and reproduction

Permalink
https://escholarship.org/uc/item/22t1p55q

Author
Chiang, Michael

Publication Date
2015

License
CC BY 4.0

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA,
IRVINE

Image cytometry reveals that cell cycle regulation minimizes pedigree depth in the
C. elegans germline within constraints of rapid growth and reproduction

DISSECRATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Michael Chiang

Dissertation Committee:
Professor Olivier Cinquin, Chair
Professor Lee Bardwell
Professor Charless C. Fowlkes
Professor Arthur D. Lander
Professor Edward S. Monuki

2015
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF SUPPLEMENTAL DATA</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>ix</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>x</td>
</tr>
<tr>
<td>CURRICULUM VITAE</td>
<td>xi</td>
</tr>
<tr>
<td>ABSTRACT OF THE DISSERTATION</td>
<td>xii</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 1: Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 2: Characterizing the effect of cell cycle regulation on pedigree depth reduction</td>
<td>16</td>
</tr>
<tr>
<td>CHAPTER 3: Parismi — image cytometry software for the <em>C. elegans</em> germline</td>
<td>56</td>
</tr>
<tr>
<td>CHAPTER 4: Characterizing spatial cell cycle properties in the <em>C. elegans</em> germline using Parismi</td>
<td>103</td>
</tr>
<tr>
<td>CHAPTER 5: Conclusion</td>
<td>135</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>144</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Pedigree depth depends on the shape of a stem cell system's lineage tree</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Schematic diagram of a hypothetical stem cell / niche system where state of differentiation depends on proximity to the niche</td>
<td>13</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Cell cycle regulation can shape the developmental program of a stem cell system when differentiation state depends on proximity to the niche.</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>The C. elegans germline is a model system for stem cells and their niches</td>
<td>15</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>A unique ratio of stem cell cycling speed to transit amplifying cell cycling speed minimizes pedigree depth</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Simworm models are biologically realistic</td>
<td>36-37</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Cell cycle length is 3.4 hours for during early germline development</td>
<td>38</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Parismi is a pipeline customized to segment three-dimensional confocal images of the C. elegans germline</td>
<td>75</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Parismi cell detection is implemented using HOG features</td>
<td>76</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Parismi uses a top-layeredness metric to identify top layer cells</td>
<td>77</td>
</tr>
</tbody>
</table>
Figure 3.4  Benchmarking active contours, marker controlled watershed, and truncated voronoi segmentation 78

Figure 3.5  Parismi fluorescence quantification is accurate 79-80

Figure 3.6  Parismi cell row counter is accurate 81

Figure 3.7  G1, S, G2, and M-phase germ cells can be separated using DNA morphology 82

Figure 3.8  G1, S, G2, and M-phase germ cells have different spatial extents and number of connected components 83

Figure 3.9  The C. elegans germline arrests immediately after starvation 84

Figure 3.10  Parismi has applications to other model systems 85

Figure 3.11  Precision-recall curves benchmarking cell center detection 86

Figure 3.12  Relationship between Nanog and Cdx2 contents in preimplantation mouse embryos 87

Figure 3.13  Differences in cell phase indices along the medial-lateral axis of the mouse olfactory epithelium 88

Figure 4.1  Randomly chosen segmentations used in cell cycle fits 114

Figure 4.2  Spatial cell cycle phase indices at L4 and L4+1 115
Figure 4.3  EdU pulse-chase experiments reveal that cells in the DMMZ cycle more slowly than those in the MMZ

Figure 4.4  Best cell cycle fits of EdU pulse-chase data using DEMD and FLM metrics

Figure 4.5  Cell cycle fits of EdU pulse-chase data using DEMD and FLM metrics

Figure 4.6  Simple assumptions about pre-meiotic interphase recapitulates cell phase measurements

Figure 4.7  FLM histogram data over the entire mitotic zone

Figure 4.8  Cyclin E levels are graded across the DMMZ and MMZ, and are differentially dependent on cell cycle phase in the DMMZ and MMZ

Figure 4.9  EdU pulse-chase data at L4+3

Figure 4.10  Description of clock plot fitting procedure

Figure 4.11  Benchmarking clock plot fitting procedure

Figure 4.12  Continuous EdU labeling experiments

Figure 4.13  Clock plots across a range of genotypes, mating conditions, and ages
**LIST OF TABLES**

| Table 2.1 | A single self-fertilized gonadal arm produces 176 progeny over five days | 39 |
| Table 2.2 | Cells exiting the meiotic zone in Simworm simulations either undergo apoptosis or gametogenesis | 40 |
| Table 2.3 | A cell cycle gradient minimizes pedigree depth in Simworm models | 41 |
| Table 3.1 | Average precision of cell detection over a variety of experimental conditions | 89 |
| Table 3.2 | Average precision of cell detection over the whole gonad | 90 |
| Table 3.3 | Average overlap of cell segmentations across different experimental conditions | 91 |
| Table 3.4 | Benchmarking EdU quantification accuracy | 92 |
| Table 3.5 | Number of connected components in G1, S, G2, M phase cells | 93 |
| Table 3.6 | Statistics for number of connected components | 94 |
| Table 3.7 | Spatial extent of G1, S, G2, and M phase cells | 95 |
Table 3.8  Statistics for spatial extent of G1, S, G2, and M-phase cells  
Table 3.9  Sensitivity and specificity of cell phase classification  
Table 4.1  KS test for DNA content histograms at L4  
Table 4.2  KS test for DNA content histograms at L4+1  
Table 4.3  Chi-square test for fraction labeled mitoses histograms at L4  
Table 4.4  Chi-square test for fraction labeled mitoses histograms at L4+1  
Table 4.5  Results of two-compartment cell cycle fit  
Table 4.6  Lengths of G1, S, G2, M-phase at start of DMMZ and end of MMZ  
Table 4.7  Results of three-compartment cell cycle fit
### LIST OF SUPPLEMENTAL DATA

<table>
<thead>
<tr>
<th>Supplemental Data</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Full table of <em>Simworm</em> optimization results</td>
<td>42-48</td>
</tr>
<tr>
<td>2.2</td>
<td>MCMC statistics for <em>Simworm</em> optimizations</td>
<td>49-55</td>
</tr>
<tr>
<td>3.1</td>
<td>Mosaic of G1, S, G2, and M-phase DNA morphologies</td>
<td>98-102</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS AND DEFINITIONS

DEMD, DNA earth mover’s distance
DMMZ, distal-most mitotic zone
FLM, fraction labeled mitosis
L2, second larval stage
L4, fourth larval stage
L4+1, one day after mid-L4 (mid-L4 is hereafter abbreviated simply as L4)
L4+3, three days after L4
MZ, mitotic zone
MeZ, meiotic zone
MFT, M-phase fraction threshold
MMZ, medial mitotic zone
OE, olfactory epithelium
Pedigree depth, the number of divisions that separate a cell from its ancestral founder
PMZ, proximal mitotic zone
TA (cell), transit amplifying (cell)
TZ, transition zone
Acknowledgments

FUNDING SOURCES

Bioinformatics Training Program Grant T15 LM7443-10 2012

Developmental and Systems Biology Training Grant T32 HD060555-01. 2010
CURRICULUM VITAE

Michael Chiang

EDUCATION HISTORY

University of California, Irvine
Ph.D. in Biological Sciences 2015

California Institute of Technology
B.S. in Physics 2008

PUBLICATIONS (PEER REVIEWED)


AWARDS AND SCHOLARSHIPS

Center for Complex Biological Systems project competition winner, University of California, Irvine, $10,000 granted for project “Lipid accumulation in macrophages as an atherosclerosis marker”, $500 award for student 2012

Bioinformatics Training Program Grant T15 LM7443-10 2012

Developmental and Systems Biology Training Grant T32 HD060555-01. 2010
Abstract of the Dissertation

Image cytometry reveals that cell cycle regulation minimizes pedigree depth in the C. elegans germline within constraints of rapid growth and reproduction

by

Michael Chiang

Doctor of Philosophy

University of California, Irvine, 2015

Professor Olivier Cinquin, Chair

Minimizing the number of mutations accumulated by differentiated cells is a widely accepted performance objective of stem cell systems. One possible strategy to reduce mutation accumulation is to reduce pedigree depth (the number of cell divisions that separates a given cell from its ancestral founder) through cell cycle regulation. Briefly, a tissue can modulate the shape of its stem cell lineage tree via spatial regulation of cell cycle length. My thesis work focuses on (1) characterizing the theoretical cell cycle profile that minimizes mutation accumulation in the C. elegans germline (a model for stem cells and their niches) via computer simulations, (2) developing a computational approach to analyze spatial cell cycle data in microscopy images, and (3) using the aforementioned computational approach to characterize germline cell cycle properties.

We developed Simworm, a biologically realistic computational model of the germline. Using Simworm, we found that an approximate two-fold slowdown in stem cell cycle length optimizes pedigree depth, and this optimal profile is a result of balancing competing performance objectives of fast progeny production and growth with pedigree depth minimization.

In order to enable sensitive and high-throughput spatial studies of cell cycle kinetics, we developed Parismi which is a software package customized for image cytometry studies of the C. elegans germline. Parismi also has broader applications to other model systems such as HeLa cell culture, early mouse embryos, and the mouse olfactory epithelium.

To test our theoretical predictions about the existence of a two-fold cell cycle gradient minimizing pedigree depth, we used Parismi to characterize the substructure of the germline’s mitotic zone. We found the that the mitotic zone can be classified into three distinct sub-compartments with different cell
cycle properties. We also identify a spatial gradient of Cyclin E as a possible mechanism for regulation of cell cycle length. Finally, we characterize an intermittent cycling phenotype in older worms, and show that intermittent cycling is a result of sperm depletion. Altogether, our study shows that germline cell cycle properties are consistent with design to minimize pedigree depth and provides insights into understanding a fundamental mechanism through which tissues reduce mutation rate.
Chapter 1: Introduction

Section 1.1: What biological forces determine mutation rate?

The existence of sophisticated DNA proofreading/repair mechanisms and the evolution of sexual recombination suggests that processes that reduce the harmful effects of mutations are actively selected for [1][2]. Indeed, mutation accumulation experiments show the net effect of spontaneous mutation decreases fitness [3][4][5]. These results generalize to natural populations; in Drosophila, it is estimated that 70% of all amino acid polymorphisms are deleterious [6]. One might expect that natural selection minimizes mutation rate in order to maximize organismal fitness, but the existence of mutants with mutation rates lower than wild-type indicates otherwise [7][8]. If mutations are harmful to organismal fitness the vast majority of the time, then why are non-zero mutation rates observed across different species (12.9 * 10^{-9} in humans [9], 3.5 * 10^{-9} in D. melanogaster [10], 2.7 * 10^{-9} in C. elegans [11])?

There are three general forces that exert upward pressure on the wild-type mutation rate. The first is selection for an increased rate of beneficial mutations. Consider an allele that increases mutation rate, and suppose this increased mutation rate generates a rare beneficial mutation that is strongly selected for. Then, all alleles linked to the beneficial mutation are also selected for. This general phenomenon is known as genetic hitchhiking, and thus mutator alleles can hitchhike with the beneficial mutations that they generate [12]. Theoretical modeling suggests that hitchhiking may play a significant role in determining a non-zero optimal mutation rate in asexual populations [13][14], although such approaches assume the absence of horizontal gene transfer in asexual populations which is known to be untrue [15]. Certainly, in sexual populations the presence of recombination limits the role of genetic hitchhiking in setting the optimal mutation rate [16][17].

The second force involved in setting the lower bound on mutation rate is the “cost of fidelity”. Suppose there exists a tradeoff between fast DNA replication and high fidelity DNA replication. Here, the pleiotropic effects of decreased mutation rates (for example, slower rates of replication in RNA viruses [18]) introduce drawbacks that outweigh the benefits of low mutation rate. The cost of fidelity is widely considered to be an important factor in determining the optimal rate of mutation [19][20][21].

The third force that exerts upward pressure on mutation rate is genetic drift. A beneficial trait will not be fixed within a population if the magnitude of the selective advantage conferred by the trait cannot overcome drift. Thus, the observation of non-zero mutation rates may be an effect of selection
being unable to reduce the rate further. In a diploid organism, selection for a trait dominates over genetic drift if $|s| > 1/(2N_e)$ where $s$ is the selection coefficient of the trait and $N_e$ is the effective population size of the organism. The selection coefficient for anti-mutator traits is given by $s = 2 s_d \Delta U$ where $s_d$ is the average fitness cost of a single mutation, $\Delta U$ is the per generation change in deleterious mutation rate conferred by the anti-mutator trait, and the factor of two is a result of assuming that most mutations occur on chromosomes unlinked to the anti-mutator allele and thus retain association for an average of two generations [22]. In totality, selection dominates over drift for an anti-mutator trait if:

$$\Delta U > 1/(4 N_e s_d)$$

where we have assumed that the cost of enhanced replication fidelity is zero. Note that this expression for the anti-mutator selection coefficient assumes sexual reproduction; for asexual and selfing organisms, recombination plays a reduced role in unlinking mutations and the selection coefficient is given by $s = \Delta U$ and $s = \Delta U/2$, respectively [23]. Thus, it is straightforward to determine whether an anti-mutator trait can be selected for given $N_e$ and $s_d$. The effective population size $N_e$ can be inferred from the nucleotide diversity $\pi$, which itself can be estimated from sequencing data [24]; $\pi$ is expected to equal $4*N_e*\mu$, where $\mu$ is the mutation wild-type mutation rate. For multicellular organisms, values of $N_e$ typically range from $10^5$ in humans to $10^6$ in D. melanogaster (see [25] for a table of $N_e$ across different species). The average fitness effect per deleterious mutation $s_d$ can be calculated from mutation accumulation experiments using methods such as the Bateman-Mukai or Maximum Likelihood technique (an overview of these methods are given in [21]); typical estimates of $s_d$, are on the order of magnitude of $10^{-3}$ to $10^{-2}$ [26].

In summary, natural selection will reduce mutation rate as long as the cost of fidelity does not exceed the selective advantage conferred by decreased mutation rate, and the conferred selective advantage is sufficient to overcome genetic drift. It is possible that a higher mutation rate may be selected for due to genetic hitchhiking, but recombination limits the extent of this effect. It should be noted that all conclusions drawn above apply only to germline mutations. While somatic mutations are linked to negative phenotypes such as cancer and are of obvious importance to organism health, from a fitness perspective the soma only exists to propagate the germline and post-reproductive somatic failure is of little consequence to fitness (disregarding second order effects such as the grandmother hypothesis [27]). Mutation accumulation experiments, used to derive key parameters such as the decline in fitness per mutation and the deleterious mutation rate itself, provide no information about somatic mutation since
only germline mutations are transmitted across generations. Thus, our understanding of the forces that drive somatic mutation rates are primarily qualitative comparisons to that of germline mutation rates. First, genetic hitchhiking will play absolutely zero role in exerting upward pressure on somatic mutation rates since any beneficial mutations in somatic tissue will not be inherited and thus somatic-mutator alleles cannot hitchhike. Secondly, the cost of fidelity will still play an important role in setting a minimum mutation rate in somatic tissue; proper soma function is still crucial to germline function. Finally, selection for anti-somatic-mutator alleles is expected to be weak primarily due to a decreased average fitness effect per somatic mutation; by the time negative somatic phenotypes such as cancer arise, the organism might very well be in a post-reproductive state. In addition, note that in asexual and selfing taxa the non-inheritance of deleterious somatic mutations will also have a large effect in weakening selection for anti-mutator traits; in sexual taxa this effect is limited due to recombination. Altogether, given the disposability of the soma, it is unsurprising that mutation rates in somatic cell types are anywhere from 20 times [9] to 80 times [28] larger than that of germ cells. Nevertheless, the existence of DNA proofreading mechanisms within somatic cells suggests that reducing mutation rate is still an important performance objective.

Section 1.2: Strategies to reduce the negative fitness effects of mutation: a historical perspective

What strategies can be employed to reduce the negative fitness effects of mutation? One possibility is to reduce \( s_d \), the average fitness cost of per mutation; indeed, the existence of tumor suppressor genes suggests as much [29]. A second possibility is to reduce the mutation rate itself; on the single cell level, this can be done with DNA proofreading and repair mechanisms, and much work has been done in elucidating said mechanisms as well as quantifying their relative contributions to DNA replication fidelity [30][31][32].

The strategies listed above apply to single cells; more conceptually sophisticated mechanisms exist on the systems level. In his seminal 1975 paper [33], Cairns proposed three independent strategies for a stem cell system to reduce the fitness effects of mutation accumulation. The first strategy is the compartmentalization of stem cells into non-competing subsystems. For example, the intestinal epithelium is organized into crypts; stem cells proliferate at the base of these crypts and push neighboring cells out to the tip of the villus, where they terminally differentiate and are eventually shed into the gut. This crypt architecture may act as a cancer suppressor since it minimizes the spread of pre-cancerous mutants [34].
The idea that cancers often reshape their microenvironment in order to invade surrounding tissue is now a core concept in cancer biology [35][36][37][38].

The second strategy is the controversial “immortal strand” hypothesis. Briefly, it has been proposed that in some systems chromosomes are asymmetrically segregated during cell division so that the stem cell always retains a full set of parental DNA strands where the parental set of DNA strands will not accumulate mutations associated with DNA synthesis. Evidence for the immortal strand hypothesis has been reported in a variety of different tissues including the plant root tips [39], the intestinal crypt [40], mammary epithelial cells [41], neural stem cells [42], and muscle satellite cells [43]. In addition, the left-right dynein motor has been proposed as a putative mechanism for asymmetric chromosomal segregation [44]. On the other hand, there exists significant controversy over the correctness of previously mentioned findings. Many stem cell systems undergo symmetric divisions and daughter cells have identical fates [45][46]. In this case, asymmetric chromosomal segregation has little effect of reducing mutation accumulation, since the high-fidelity set of template DNA strands will either belong to a differentiated cell and not be replicated or there will exist a substantial population of low-fidelity non-template strands competing against the template strand. Other common critiques include the mis-identification of stem cell populations in previous immortal strand experiments [47][48]. In addition, the immortal strand is most likely not a universal mechanism; researchers have failed to find evidence for the immortal strand in hematopoietic stem cells, a system with readily available stem cell markers [49]. Altogether, consensus is lacking and the immortal strand hypothesis is still an active subject of study.

The final strategy (and the subject of this thesis) is perhaps the simplest strategy of all. Cairns suggested a developmental program of slow cycling stem cells and fast cycling transit amplifying cells could minimize the number of accumulated mutations. In this situation, mutations accumulated by fast cycling transit amplifying cells are quickly flushed out of the system while slow cycling stem cells periodically refresh the transit amplifying cell population with high-fidelity cells.

Section 1.3: Reducing mutation rate by number of cell divisions: a historical perspective

The number of cell divisions is known to play a large role in the etiology of cancer; two thirds of cancer risk is attributable to random mutations arising during cell division [50]. Thus, reducing the average number of divisions separating differentiated cells from their ancestral founders is a possible mechanism
to minimize the harmful effect of accumulated mutations. How can a stem cell system reduce the number of cell divisions that occur? The first solution is to reduce the number of differentiated cells produced by the system; all else being equal, a smaller number of cell divisions will occur in a tissue that produces three differentiated cells as opposed to one that produces six differentiated cells. This trivial solution is not considered further since producing a smaller number of cells would likely compromise tissue function. A second solution is to optimize the stem cell lineage tree so that the number of cell divisions are reduced while still producing the same number of differentiated cells (Figure 1.1).

The question of how to optimize lineage trees with respect to number of cell divisions has been explored in the seemingly disparate field of computer science. Binary trees (analogous to lineage trees) are data structures used to store and look up data; trees with low average pedigree depth are desirable in order to minimize lookup time. The balance $\beta$ is a metric that parameterizes the shape of the tree; balance is calculated by iterating through all tree nodes and finding the minimal ratio of branch weights [51]. A tree with low balance tends to have a high number of cell divisions (e.g., Figure 1.1A), while a tree with high balance tends to have a low number of cell divisions (e.g., Figure 1.1B). Adjusting the shape and hence balance of a stem cell tree is a potential mechanism to reduce mutation rate.

Do real stem cell systems shape their lineage trees to reduce mutation rate? Answering this question experimentally requires mapping cell lineage trees in real model systems; in the past, such lineaging efforts were conducted manually with great effort [52][53]. More recently, automated lineaging solutions have been developed [54] which show promise in streamlining fate mapping experiments. However, at the current time, lineaging data is still difficult to acquire and thus little has been published about how experimental stem cell systems shape their lineage trees to reduce mutation rate. There has been theoretical work done to characterize what lineage trees are optimal for minimizing cancer incidence rates in a hypothetical stem cell system [55]; unsurprisingly, this study showed that minimal cancer risk is obtained by using the fewest possible stem cell divisions subject to constraints imposed by tissue self-renewal. Overall, as of the writing of this thesis, the strategy of shaping lineage trees in order to reduce the number of cell divisions has not been thoroughly characterized.
Section 1.4: Mutation rate, pedigree depth, and cell cycle regulation are interrelated

Consider the lineage trees diagrammed in Figure 1.1; this tree forms a structure known in computer science as a binary tree, where in this case each cell has either 0 or 2 descendants. Define “pedigree depth” to be the number of cell divisions that separate a differentiated cell from its ancestral founder. In Figure 1.1, all three developmental programs produce eight differentiated cells but average pedigree depth differs between trees. Average pedigree depth is minimized when the binary tree is balanced, i.e. when all terminal nodes have the same depth [51]. However, the balanced tree strategy to minimize pedigree depth and thus replication-dependent mutation accumulation has significant drawbacks. This strategy produces a balanced pedigree tree by maintaining an expanding pool of progenitors in which all cells keep cycling at the same rate until the time the organ has reached its final set number of cells (Figure 1.1B). It precludes the differentiation of cells before that time, requires a large pool of progenitors, and is impractical for organs that must undergo self-renewal throughout life. Early cell differentiation and small progenitor pools are made possible by the naïve, alternative strategy that consists of maintaining a lineage of asymmetrically-dividing progenitors — but this comes at the cost of an unbalanced pedigree tree and thus increased pedigree depth (Figure 1.1A). A third strategy is possible that compromises between the two previous strategies: a population of long-lived, slow cycling progenitors divides asymmetrically to self-renew and to give rise to faster-cycling progenitors that only persist transiently before differentiating. This strategy, to which we refer as the pedigree depth quasi-minimization strategy hereafter, can lead to a highly balanced pedigree tree while allowing early production of differentiated cells and small progenitor pool size (Figure 1.1C).

How can a stem cell system exert control over its lineage tree in order to implement the pedigree depth quasi-minimization strategy? One possibility is molecular regulation of stem cell self-renewal and differentiation in order to control the relative numbers of stem cells and transit amplifying cells. A second not necessarily mutually exclusive possibility is regulating the relative cell cycle lengths of stem cells and transit amplifying cells. Many organs have a spatial structure with stem cells located in a niche and cells outside of the niche undergoing differentiation. This structure can provide for simple control of both the transition between the stem cell and differentiated states and the transition between slow and fast cycling states, if cells are displaced from the niche as a result of proliferation, and if the
transitions are controlled by distance to the niche. In such a system, a stem cell system can modulate the shape of its associated lineage tree through cell cycle regulation.

To better illustrate the idea described above, we propose a simple thought experiment—consider a one-dimensional stem cell system where stem cells divide symmetrically in the first \( n \) cell rows adjacent to the niche (hereafter referred to as mitotic zone) and cell movement is a simple function of mitotic pressure, i.e., cells are pushed proximally by distal division events and there is no active transport (Figure 1.2). If cells differentiate once they leave the mitotic zone, then only the first cell immediately neighboring the niche is functionally a stem cell; the other cells in the mitotic zone are functionally transit amplifying cells since they will eventually differentiate once they are pushed out of the mitotic zone. In this hypothetical system, the ratio of stem cell divisions versus transit amplifying cell divisions depends on the spatial profile of cell cycle lengths, and a number of different lineage trees can be generated via regulation of the spatial distribution of cell cycle lengths; a moderate cell cycle gradient maximizes balance, and hence minimizes average pedigree depth (Figure 1.3).

Section 1.5: The C. elegans germline is a model system for stem cells and their niches

A stem cell system can control the shape of its associated lineage tree if proliferation state depends on distance from the niche and cell movement is a simple function of mitotic pressure (Figure 1.3). Is there a model organism that possesses these properties? The C. elegans nematode is a small worm that reproduces primarily through hermaphroditic selfing. Young adult C. elegans hermaphrodites have two gonadal arms; each gonadal arm possesses a capped tube structure and holds approximately 1,000 germ cells patterned along the distal-proximal axis (Figure 1.4). Germline stem cells undergo mitotic divisions at the distal end of the gonad in the “mitotic zone” (MZ). The noted absence of oriented divisions in the mitotic zone [56] suggests that cell movement in the germline is primarily caused by mitotic pressure; cell divisions push more proximal cells out of the mitotic zone into the transition zone. Upon entering the transition zone, germ cells enter meiotic prophase and commit to eventual differentiation into sperm (pre-L4) or oocyte (post-L4); approximately 350 progeny are produced by a single selfed worm. Since any mutations accumulated by gametes are passed down to successive generations of worms, and because of its predominantly selfing mode of reproduction, reduction of mutation rate is an obvious performance objective of the C. elegans germline; a high mutation rate would
lead to rapid extinction of the species via Mueller’s ratchet since deleterious mutations cannot be expunged through sexual recombination [57][58]. Indeed, the C. elegans mutation rate is observed to be \(~3\times10^{-9}\) [11] or \(~10^{-8}\) [59] per nucleotide site per generation, slightly lower than the human rate [60][61][62][63][64]. The C. elegans gonad thus provides an amenable model system to ask how organs minimize the accumulation of mutations and what role cell cycle regulation plays in that minimization.

While many of the developmental regulators controlling the binary decision between self-renewal and differentiation have been well-established [65] [66] [67], the question of how the germline maps a gradient of Notch signaling to lineage progression remains unresolved. The MZ is known to be an inhomogeneous collection of cells; there is evidence that within the mitotic zone distal and proximal pools of cells differ with respect to commitment to differentiation [68]. How does a germ cell progress from a stem cell to become a transit amplifying cell before moving into a final differentiated state, and how does this progression fit within the genetic framework controlling the binary switch between mitosis and meiosis? There are two approaches to address this question; the first approach is to elucidate the gene network that regulates lineage progression. This task is difficult because no readily available markers to distinguish between stem cells and transit amplifying cells exist for the germline. The second approach is to sidestep the molecular definition of stem-ness in favor of a functional definition of stem-ness; if germline differentiation state is dependent on proximity to the niche and if cell movement is a simple function of mitotic pressure, then by definition only the distal-most germ cell is a stem cell and all other cells within the mitotic region are transit amplifying cells. By similarity to the thought experiment described in Figure 1.2, we expect spatial regulation of cell cycle length to play an important role in determining the balance, and hence average pedigree depth, of a stem cell system’s associated lineage tree.

Intriguingly, spatial measurements of mitotic index indicate that differences in cell cycle properties do exist along the distal-proximal axis of the germline [69]. Note that this study does not show that absolute differences in cell cycle length exist along the distal proximal axis—mitotic index measures only the ratio of M-phase length to total cell cycle length. Further work in characterizing cell cycle length within the mitotic zone is hampered by technical difficulties. Continuous BrdU labeling experiments fail to measure gross spatial differences in cell cycle properties and provide very different estimates of cell cycle length (~20 hours vs. ~8 hours, [56][70]). The absence of measured spatial differences may be due to cell movement; if the timescale of cell movement is shorter than its associated cell cycle length (i.e., fast
cycling cells are pushed out of the mitotic zone by slow cycling cells), then sophisticated methods are required to reverse the effects of cell movement in spatial cell cycle measurements.

One potential approach for measuring cell cycle speed within the germline is flow cytometry. Accurate measurements of cell cycle kinetics can be obtained by measuring the DNA content of a labeled cohort of cells through the cell cycle [71]. However, flow cytometry relies on tissues being dissociated prior to analysis which removes the spatial information present in the sample. On the other hand, imaging of undissociated tissues or organs can be readily performed with current technologies such as confocal microscopy, but such imaging does not immediately lead to cell-by-cell information without extensive analysis. This analysis, termed image cytometry, requires segmentation of individual cells in three dimensional images. The development of image cytometry tools specifically for segmenting and analyzing C. elegans germline microscopy images would aid in elucidating the link between cell cycle regulation and stem cell lineage progression.

Section 1.6: There exists a need for image cytometry tools customized for germline studies

Although image segmentation is a very active research area, few tools are readily applicable for cytometric studies of the germline. While commercial and public domain tools such as Volocity (PerkinElmer) and CellProfiler [72][73][74] can be applied to a wide range of images, they require extensive customization and tuning in order to achieve accurate segmentation of three-dimensional microscopy images. Techniques such as geodesic distance transform [75], gradient flow smoothing [76], and watershed transform have all been used successfully to perform 3D volumetric segmentation of cell nuclei [77][78][54]; however, these techniques typically require sample-specific post-processing [79] to correct segmentation errors. The need for extensive tuning has prevented widespread adoption of image cytometry for the germline and current studies of the germline generate three-dimensional cell segmentations by hand which is a time-consuming and error-prone task for human scientists [70][80].

Spatial cell cycle studies of the germline pose further challenges beyond image segmentation. For example, quantification of fluorescence intensity may be marred by attenuation artifacts. Specifically, images of tissues often contain attenuation artifacts with a positional dependence; these artifacts must be corrected for to prevent bias in spatial cell cycle studies. Furthermore, existing protocols for downstream analysis of image cytometry data is ill-suited for application to the germline. Consider the standard
protocol to measure cell cycle length by observing the DNA content of a labeled cohort of cells through the cell cycle [71]. This protocol only applies to homogenous cell cultures; in a tissue with a cell cycle gradient, additional analytical sophistication is necessary to account for cell movement within the tissue while the labeled cohort travels through the cell cycle. Issues such as these must be resolved and integrated into the image cytometry pipeline in order to encourage widespread adoption of cytometric studies within the C. elegans community. Tools that facilitate automated or semi-automated image segmentation and cell cycle analysis would thus be highly desirable and would enable worm biologists as well as the larger scientific community to analyze imaging data they routinely acquire in a much more powerful way.

Section 1.7: Thesis outline

The broad aim of this manuscript is to ask whether and how cell cycle regulation plays an important role in minimizing pedigree depth, and hence mutation accumulation in the C. elegans germline. To this end, the thesis is structured into five chapters:

In chapter one, a definition of pedigree depth is given. Through a simple thought experiment, we demonstrate that cell cycle regulation, pedigree depth, and mutation minimization are interconnected properties of a stem cell system where differentiation state is dependent on distance from the niche. We propose a technical approach, image cytometry, to address whether cell cycle regulation plays a role in pedigree depth minimization within the C. elegans germline.

In chapter two, we characterize the theoretical spatiotemporal cell cycle profile that minimizes pedigree depth in the C. elegans germline. We propose three different germline models of increasing complexity: (1) a naive model where average pedigree depth is minimized over 3,000 cells produced (the approximate number of mitotic divisions during the reproductive lifespan of the selfed worm), (2) a more complex model where gametogenesis of 44 spermatocytes and 176 oocytes (the reproductive output of a single selfed gonadal arm) is explicitly simulated, and (3) a biologically realistic model where mitotic zone properties are based on experimental measurements of the germline. We describe Simworm, an agent-based computational model that simulates cell lineages given input cell cycle behavior; Simworm is released as open-source software and is available at https://github.com/mc-simworm/simworm. Via computational modeling, we show that slower cycling stem cells reduce average pedigree depth at the
cost of slower cell production and progeny production. We show that when constrained by experimentally
derived cell production and progeny production schedules, an approximate two-fold difference in cell
cycle length along the distal-proximal axis of the germline minimizes pedigree depth, and that this optimal
profile is a result of the germline balancing competing performance objectives in producing gametes
quickly and producing gametes with low pedigree depth.

In chapter three, we describe the development of Parismi, image cytometry software that is
customized for cell cycle studies of the germline using three-dimensional confocal microscopy images.
We show that our software can accurately quantify the spatial position and fluorescence of single cells in
germline images. We further show the utility of our software by deriving two novel results about the
germline: (1) germ cell cycle phase can be classified based on morphological markers, and (2) germ cells
can stop at multiple points of the cell cycle (possibly spanning all of G1, S, and G2) in response to
nutritional starvation. We demonstrate the wide applicability of the software by applying Parismi to image
cytometry analyses of the mouse olfactory epithelium and mouse blastocyst. Parismi is open-source and
is available at https://github.com/cinquin/parismi.

In chapter four, we use Parismi to characterize spatial cell cycle properties in the C. elegans
germline. We show that the mitotic zone is composed of distinct sub-compartments with different cell
cycle properties, and demonstrate that distal cells cycle approximately 1.5-fold more slowly than proximal
cells at the L4 and L4+1 time points which is consistent with theoretical predictions; we further show that
this 1.5-fold gradient is selectable through evolution. In our investigation of the L4+3 time point, we find
that old germlines cycle intermittently, a finding not predicted by our computational simulations. Finally, we
correlate the experimentally measured cell cycle profile to a spatial gradient of cyclin E within the
germline.

In chapter five, we speculate how intermittent cycling might fit in the framework of pedigree depth
minimization. We consider other effects of pedigree depth minimization on fitness besides mutation
accumulation, about other explanations for slow stem cell cycling besides pedigree depth minimization,
and alternative strategies to minimize mutation accumulation. Finally, we speculate how our results might
apply to other model systems, and what role Cyclin E plays in regulating cell cycle length. Taken together,
this work shows that cell cycle regulation is an important component of pedigree depth reduction within
the germline and possibly in other systems.
Figure 1.1: Pedigree depth depends on the shape of a stem cell system’s lineage tree

Three different developmental programs for a stem cell system are given (A, B, C) all of which produce a total of eight differentiated cells (yellow). A balanced pedigree depth tree (B) minimizes average pedigree depth; a description of balance is given in Section 1.3. (A) Average pedigree depth is \((1+2+3+4+5+6+7+7)/8 = 4.375\). (B) Average pedigree depth is \((3+3+3+3+3+3+3+3)/8 = 3\). (C) Average pedigree depth is \((2+3+3+3+4+4+4+4)/8 = 3.375\).
Consider a stem cell / niche system where differentiation state of a cell is dependent on proximity to the niche. Here, cells inside the niche (stem cells in red, transit amplifying cells in blue) are maintained in a stem-like state. Mitotic pressure pushes cells out of the niche where they differentiate (yellow). This hypothetical stem cell system has several attractive design properties. First, tissue size / stem cell number is maintained in a robust manner. Second, there is steady production of differentiated cells. Finally, cell cycle regulation provides a means to modulate the lineage tree associated with this hypothetical stem cell system (see Figure 1.3). Note that this hypothetical stem cell system bears resemblance to the *C. elegans* germline (See Figure 1.4).
Figure 1.3: Cell cycle regulation can shape the developmental program of a stem cell system when differentiation state depends on proximity to the niche.

Consider a stem cell / niche system where differentiation state of a cell is dependent on proximity to the niche (see Figure 1.2 for further explanation). Lineage tree balance depends on the spatial distribution of cell cycle lengths. $\alpha$, stem cell cycle length (red) divided by transit amplifying cell cycle length (blue); $\beta$, balance of lineage tree; $p$, average pedigree depth of six differentiated cells. A moderate value of $\alpha$ maximizes the balance of the tree and minimizes average pedigree depth in this example. A definition of balance and a mathematical relationship between balance and average node depth is given in [51].
Figure 1.4: The *C. elegans* germline is a model system for stem cells and their niches

The germline is spatially patterned along its distal-proximal axis. Germ cells undergo mitotic divisions in the “mitotic zone” (MZ) at the distal end of the germline. Germ cells move proximally and enter meiotic prophase in the “transition zone”, where they commit to eventual differentiation into sperm or oocyte.
Chapter 2: Characterizing the effect of cell cycle regulation on pedigree depth reduction

Section 2.1: Contributions

The data presented in this chapter is published in the paper “Control of C. elegans germline stem cell cycling speed meets requirements of design to minimize mutation accumulation” (BMC Biology, 2015).

- Dr. Amanda Cinquin and Adrian Paz performed experiments.
- Michael Chiang developed Simworm.
- Prof. Olivier Cinquin provided software for sparse grid search optimization.
- Dr. Edward Meeds and Prof. Max Welling provided software for MCMC as well as guidance in running the analysis and interpreting its results.

Section 2.2: Introduction

The C. elegans germline is a model system for stem cells and their niches. Differentiation state within the germline is determined by proximity to the niche; germ cells within the mitotic zone (MZ) in the distal-most ~20 cell rows undergo mitosis while germ cells in the more proximal transition zone (TZ) enter the meiotic cell cycle. In Chapter 1, we observed that a cell cycle gradient can modulate the balance and hence pedigree depth of a stem cell system’s associated lineage tree (Figure 1.3). To place our observations on more quantitative footing, we explicitly model the simple stem cell system diagrammed in Figure 1.2 to derive average pedigree depth as a function of number of cells produced and the spatial cell cycle profile. Here, the spatial cell cycle profile is parameterized by $\alpha$, the ratio of the distal-most cell cycle length (Figure 1.2, red circle) to all other cell cycle lengths (Figure 1.2, blue circles).

We find that a unique value of $\alpha$ minimizes average pedigree depth for a given number of produced cells. In other words, minimal pedigree depth is achieved when $\alpha$ exists within a Goldilocks’s zone that is neither too high nor too low in magnitude (Figure 2.1). We observe that optimal $\alpha$ increases as the total number of cells to be produced increases. We also expect that optimal $\alpha$ depends on MZ geometry since MZ size is proportional to the number of rounds of symmetric division a distal cell can undergo before exiting and differentiating. However, MZ geometry is itself linked with the relative number of stem cells and transit amplifying cells in a phenomenological manner; $\alpha$ in this context is defined to be
the cell cycle speed of the distal-most cell divided by that of all other cells. Increasing MZ size changes the relative numbers of the distal-most cells versus all other cells. Thus, we do not consider MZ size in this simple set of simulations other than to say that MZ size has a direct effect on the relative numbers of stem cells and transit amplifying cells for a given $\alpha$.

How do the results described above apply to C. elegans germline? The effect of cell cycle regulation upon pedigree depth is difficult to assess in vivo. Although it is possible to arrest the cell cycle via depletion of cell cycle factors such as Cyclin E or CDK1, controlling the spatial distribution of cell cycle lengths is currently infeasible in the germline. In addition, no experimental protocols exist for lineage tracing of germ cells. Computational modeling provides an approach to sidestep these experimental difficulties. Previous studies with a theoretical emphasis have explored particular principles governing the ratio between the speed at which stem cells cycle and the speed at which their differentiating descendants cycle. For example, one study defined a performance objective as minimizing the chance of multiple mutational “hits” causing cancer, not considering speed of development, and assumed an intrinsic difference in mutation rates between stem cells and their differentiating descendants [55]; slower stem cell cycling was reported to be favored when the stem cells mutation rate was orders of magnitude lower than that for other cells. Another study focused on speed of development as a performance objective, not considering mutation accumulation, and found that the relative stem cell cycle speed should be high during a first phase of development before abruptly switching to a lower value, following the “bang-bang” principle of control theory [81]. Because both mutation minimization and speed of development are performance objectives relevant to biological systems, here we ask how the slow stem cell cycling principle outlined by [33] applies when considering these objectives jointly.

We developed an agent-based model of the C. elegans germline, hereafter referred to as Simworm. Simworm allows us to simulate the pedigree depth of individual germ cells and their dependence on the spatial distribution of cell cycle lengths under a wide variety of perturbations and assumptions. Beyond characterizing the relationship between cell cycle and pedigree depth, Simworm is also useful for fitting experimental cell cycle data; in particular, in Chapter 4, we use Simworm to correct for cell movement in spatial cell cycle fits and to characterize intermittent cycling across gonads. Simworm is implemented in C++ and released as open-source software at https://github.com/mc-simworm/simworm.
Section 2.3: Simworm models are biologically realistic

In the previous section, we observed that biological parameters such as number of differentiated cells produced have an effect on the optimal cell cycle profile that minimizes pedigree depth. It follows that Simworm must be biologically realistic if we are to derive meaningful predictions. In this section, we describe three Simworm models of increasing complexity that build toward a biologically realistic simulation of the C. elegans germline (Figure 2.2). An in-depth explanation of Simworm implementation details is given in the Sections 2.11 — 2.22.

All models share the same computational core. The mitotic zone is modeled as a two-dimensional lattice that wraps around itself to form a hollow cylinder mimicking the shape of the gonadal arm. Only one cell can occupy a lattice point at any given time. When a cell divides, geometrical constraints force cascades of cell movement until a single cell exits the mitotic zone (see Section 2.11). In addition, the spatiotemporal cell cycle profile in all models are parameterized by the same two spatial control points (distal and proximal) and four temporal control points (L2, L4, L4+1, and L4+3). Rationale for the selection of temporal control points is given in Section 2.15.

In the first and most basic class of Simworm model (hereafter referred to as a Class 1 model, Figure 2.2A), average pedigree depth is computed over the first 3,000 differentiated cells that exit the mitotic zone. Note that approximately 3,000 cell divisions occur during the reproductive lifespan (up to L4+3) of the selfed worm; this value was calculated by plugging the experimentally measured cell cycle length (Table 4.5) into our most biologically realistic (Class 3) Simworm model. The spatial cell cycle profile is assumed to be linear, and the proximal spatial control point is fixed to 2.8 hours without loss of generality (see Section 2.22). In addition, Class 1 models are typically optimized over different MZ geometries which stay constant over time.

In the second class of Simworm model (hereafter referred to as a Class 2 model, Figure 2.2B), gametogenesis is explicitly simulated and average pedigree depth is computed over the first 176 oocytes produced, where the number 176 is drawn from progeny counts (Table 2.1). In order to simulate gametogenesis, we simulated cellular events in the meiotic zone (MeZ) in addition to the mitotic zone. Cells that exit the MZ enter the MeZ, and geometrical constraints on the MeZ force cells to exit in first-in-first-out order. The first 44 cells that exit the MeZ undergo spermatogenesis, while subsequent exiting
cells either undergo apoptosis or oogenesis with some given probability (Table 2.2). An in-depth explanation of gametogenesis is given in Section 2.13 and Section 2.16. As with Class 1 models, the spatial cell cycle profile is assumed to be linear and Class 2 models are typically optimized over different MZ geometries which stay constant over time. However, the proximal control point is allowed to vary in Class 2 models.

In our final and most biologically realistic class of Simworm model (hereafter referred to as a Class 3 model), the mitotic zone is simulated in a biologically realistic manner. In particular, we simulated the following based on experimental data:

- Pre-meiotic interphase. Spatial measurements of mitotic phase index indicate that some fraction of MZ cells do not undergo proliferative divisions prior to entering mitosis [69]. Our own quantification of cell phase indices indicate that a sizable fraction of cells arrest in G2 after 15 cell rows at L4, and 11 cell rows at L4+1 (Figure 4.2). We explicitly modeled pre-meiotic arrest in Class 3 models; implementation details are given in Section 2.19.

- Piecewise linear cell cycle profile. Cell cycle fits indicate that the spatial cell cycle profile is flat in the proximal mitotic zone (see Section 4.6). Thus, we constrained the spatial cell cycle profile to be flat after 15 cell rows at L4 and 11 cell rows at L4+1.

Class 3 models also have a fixed MZ geometry based on experimental measurements (Section 2.18); note that this experimentally-derived geometry changes over time.

Section 2.4: Pedigree depth of all cells produced over the lifetime of the germline can be efficaciously minimized via cell cycle regulation

We first sought to establish whether the MZ’s tubular organization can efficaciously minimize pedigree depth when combined with a cell cycle gradient using Class 1 models (Figure 2.2A). The minimal average pedigree depth of the ~3,000 germ cells produced over the life time of a gonadal arm is \( \log_2(3,000) = 11.55 \). This minimal value can only be reached by keeping all cells in a cycling state until the time the population number reaches its final value; the body of a young adult C. elegans hermaphrodite could most likely not fit such a high number of germ cells. We thus asked whether average pedigree depth of differentiated cells can be minimized to a value close to its theoretical minimum even with an MZ of limited size. Our optimization procedure identified a set of parameters that minimized the average
pedigree depth of the first 3,000 differentiated cells output by the MZ, optimizing over MZ dimensions and a cell cycle profile with four degrees of temporal freedom. This minimal pedigree depth, achieved with an MZ comprising 359 cells, was 11.74 (Table 2.3, optimization 1; full optimization results are given in Supplemental Data 2.1); this is close to the theoretical minimum of 11.55.

We next asked whether the cell cycle gradient present along the distal-proximal axis plays a substantial role in minimizing pedigree depth in optimization 1. We performed a second optimization that was identical to the first except that cell cycle length was fixed both in space and in time, and constrained total MZ cell number to be no more than the optimum for optimization 1 (359 cells); the minimal pedigree depth was 13.94 (Table 2.3, optimization 2). Note that temporal freedom in cell cycle length does not make any difference to pedigree depth in the absence of spatial freedom, since a fixed number of cells are produced. The difference between optimizations 1 and 2 can thus be fully ascribed to the lack of spatial freedom in optimization 2. We next performed a third optimization that lifted the total MZ cell number constraint in optimization 2, and found that even without that constraint minimal pedigree depth was 12.05, still higher than when a cell cycle gradient is allowed (Table 2.3, optimization 3). Lastly, when changing optimization 2 to fix MZ cell number to 200 — the approximate number of actively-cycling cells in the MZ [70]— pedigree depth was 17.20 (Table 2.3, optimization 4), a high value that further underscores the importance of cell cycle gradients. Overall, germ cell pedigree depth can be efficaciously minimized by slow cycling of stem cells and differentiation of cells pushed away from the niche.

Section 2.5: Pedigree depth of gametes produced over the lifetime of the selfed germline can be efficaciously minimized via cell cycle regulation

We next focused on germ cells that become gametes utilizing Class 2 models (Figure 2.2B), because only they can transmit mutations to future generations. The majority of germ cells do not undergo gametogenesis but instead undergo apoptosis ([82]; apoptosis starts occurring at the same time gonads switch to oogenesis at the end of larval development). The production of the ~220 germ cells with a gamete fate is intertwined through time with the production of ~3,000 meiotic cells whose eventual fate is not gametogenesis. The minimal average pedigree depth of gametes is thus larger than the minimal average pedigree depth for 220 cells that are the only descendants of a primordial progenitor — i.e.
\[ \log_2(220) = 7.78 \] — and likely smaller than the minimal average pedigree depth for 3,000 cells — i.e. \[ \log_2(3,000) = 11.55. \]

Taking apoptosis and gametogenesis into account, we identified a minimal gamete pedigree depth of 9.62, achieved with an MZ that was 112 cell diameters long and 1 cell diameter wide (Table 2.3, optimization 5, which is set up in an identical way to optimization 1 except for the change to a Class 2 model instead of Class 1 model). This gamete pedigree depth, achieved with a steep cell cycle gradient that is further discussed below, is substantially lower than the theoretical minimum of 11.55 for the production of 3,000 cells. It is also lower than the minimal pedigree depth of the first 3,000 cells, because differentiated cells produced early in development (that have a lower pedigree depth) are more likely to become gametes than differentiated cells produced later. We next ran an optimization identical to optimization 5 except that it constrained MZ dimensions to 30*30 rows, closer to experimental MZ dimensions, which led to a minimal increase of optimal pedigree depth to 9.65 (Table 2.3, optimization 6).

The cell cycle gradient present along the distal-proximal axis plays a substantial role in minimizing pedigree depth: with a cell cycle length fixed in space and time, the minimal pedigree depth was 9.91 without constraints on MZ size (Table 2.3, optimization 7). Note that again temporal freedom in cell cycle length does not make any difference to pedigree depth in the absence of spatial freedom, because inputs to the simulations that are not optimized over are defined in terms of cell divisions rather than elapsed time (see section 2.22). Overall, gamete pedigree depth can be efficaciously minimized by slow stem cell cycling even when the size of the stem cell pool is constrained.

**Section 2.6: Rapid growth and progeny production is a competing performance objective against pedigree depth minimization**

Having established that the simple rules we used for control of germ cell cycling and differentiation make it possible to minimize gamete pedigree depth, we turned to the tradeoff between pedigree depth minimization and speed of reproduction. The cell cycle speed profiles reported above that minimize gamete pedigree depth (Table 2.3, optimization 5) would cause slower germ line development than is experimentally observed. Assuming that a minimal gonad size is required before oogenesis begins, for example because of the role of meiotic cells in streaming content to oocytes [83], a slower development rate delays the onset of reproduction. Using a matrix population model (see Section 2.24),
we computed that the slower reproductive schedule imparted by the optimal cell cycle profile derived from optimization 5 leads to a 6-fold increase in population doubling time compared to a flat cell cycle length profile fixed at 2.8 h (76 h vs. 12 h, respectively). This delay would lead to a fitness loss sufficiently large for natural selection to act upon since mutants with a developmental delay as small as 2.6 h can be outcompeted by wild-type [84].

To ask where the optimum between pedigree depth minimization and fast reproduction lies, we defined an aggregate fitness value that captures the effects of delays in the reproductive schedule due to slow stem cell cycling and of long-term mutational load stemming from replication-dependent mutations (see Section 2.25). Applying this fitness value to a Class 3 model, we found that the distribution of cell cycle speeds that maximizes this fitness is one where all cells cycle essentially as fast as permissible — which comes at the cost of sub-optimal germ cell pedigree depths (Table 2.3; optimization 8 and optimization 9). In other words, minimization of mutation accumulation essentially comes second to timely reproduction. Overall, within the context of the C. elegans germ line, the pressure for slow stem cell cycling to minimize pedigree depth is strong only as long as this slow cycling does not significantly delay the reproductive schedule.

Section 2.7: A ~two-fold slowdown in stem cell cycle length optimizes C. elegans germline mutation accumulation

We asked which MZ dimensions and cell cycle profile minimized pedigree depth while allowing for a speed of germline development and reproduction that were at least as high as determined experimentally by germ cell counts and rates of oocyte production (see Section 2.21). We ran a Class 2 optimization identical to optimization 6, except that we introduced development and reproduction rate constraints and fixed cell cycle length at larval stage L2 to its experimentally-determined value of 3.4 h. The latter change was important because a short L2 germ cell cycle is favored by optimizations, which get close to the 2.8 h limit (optimizations 5, 6, 8, 9); yet the germ cell cycle at that stage is longer (3.4 h, Figure 2.3), possibly because of physical constraints beyond the scope of our simulations (such as limited nutrient availability in early larvae whose intestine is substantially smaller than that of L4 larvae). This optimization minimized pedigree depth to a value of 9.72 (Table 2.3, optimization 10). For comparison, a flat profile produced progeny with average pedigree depth of 9.96; the advantage afforded by the cell
cycle gradient is thus ~0.2 fewer divisions in the germ cell lineage per generation, i.e. ~0.1 division per day given the C. elegans generation time of 2-3 days. Optimal MZ dimensions were 19 cells long by 12 cells wide (95% credible intervals: 17—22 x 10—13), and the optimal cell cycle gradient amplitude was ~two-fold (95% credible intervals for ratio: 1.81—2.11 at L4 and 1.34—2.62 at L4+1). Experimentally-determined dimensions are 19.3 cells along the long axis (n = 157 gonadal arms, 95% rank sum CI = 19.0 — 19.5) and 13.5 cells on average along the short axis (n = 157 gonadal arms, 95% rank sum CI = 13.1 — 13.7). Optimal dimensions derived from simulations are thus in remarkably close agreement with experimental measurements.

We asked whether the optimal cell cycle profile would hold if we used biologically-realistic assumptions about the sub-structure of the mitotic zone. To this end, we ran a simulation identical to optimization 10 except we applied a Class 3 model instead of a Class 2 model. Optimal cell cycle profiles are qualitatively similar whether they are derived using Class 3 or Class 2 models (Table 2.3, compare optimizations 10 and 11).

Section 2.8: Alternative mechanisms for mutation accumulation and calculating fitness

We asked how the optimal cell cycle profile is affected by two putative mechanisms that alter the accrual of replication-dependent mutations and their distribution to daughter cells. We first considered a model according to which the accrual of mutations is inversely proportional to the speed of the cell cycle — e.g. because a slower cell cycle could leave more time for proof-reading and correction of replication errors. This leads to a cell cycle length gradient that is marginally steeper than when mutation accrual is independent of cell cycle speed (Table 2.3, optimization 12).

We next considered an “immortal strand” model [85], in which stem cells located in row 1 do not retain mutations caused by DNA replication, which are instead passed on to daughter cells. We found that under immortal strand assumptions, the optimal cell cycle profile is inverted (Table 2.3, optimization 13); it is advantageous for the distal-most cells to cycle as quickly as possible relative to more proximal cells. Intuitively, this is because mutations are flushed out of the tissue by stem cell cycling.
Section 2.9: Pedigree depth minimization is insensitive to the L4+3 control point

The optimal pedigree depth using a Class 3 model with cell production and progeny production constraints is 9.85 (Table 2.3, optimization 11); we asked how sensitive this result is to the L4+3 control point. We ran pedigree depth simulations using the optimal cell cycle profile found in optimization 11, except we set the cell cycle profile to be constant past L4+1. We found that this modification has a minimal effect on pedigree depth (pedigree depth = 9.86; for comparison, the pedigree depth for a simulation with a flat profile is 10.26). Thus, pedigree depth minimization is insensitive to the L4+3 control point. Given that very few oocytes are produced past L4+3, this is not wholly unexpected.

Section 2.10: Conclusion

The competing performance objectives of mutation minimization and timely reproduction allows us to make qualitative and quantitative predictions about the germline cell cycle profile. First, starting from L4 cells in the distal MZ should cycle more slowly than cells in the proximal MZ; this difference should persist in adulthood. Second, the cell cycle speed difference between distal and proximal MZs should be on the order of two-fold. Finally, pedigree depth is relatively insensitive to the cell cycle profile at L4+3 compared to L4 and L4+1. A natural question is whether our theoretical predictions are borne out in actual experiment. We set out to quantify cell cycle speed along the MZ distal-proximal axis in worms at the L4 and young adult stage; this aim led to the creation of Parismi, image cytometry software that enables spatial cell cycle studies of the germline. The development of Parismi is detailed in the next chapter.
Supplemental Text

Section 2.11: Simworm — rules for cell movement within the mitotic zone

The mitotic zone (MZ) is modeled as a two-dimensional lattice with a long axis corresponding to the distal-proximal axis of the gonadal arm and a short axis that wraps around itself to form a hollow cylinder mimicking the shape of the gonadal arm. Only one cell can occupy a lattice point at any given time. When a cell divides, one daughter remains at the same location and one daughter needs to find a new position. If an empty lattice point exists in the same row the division occurred, cells in the row are pushed across the short axis so that the nearest empty point in the row is filled. Otherwise, if the next cell row has an empty position the daughter cell is pushed forward to that row, and cells within the new row are displaced as necessary so that the empty position is occupied. If both the row in which the division occurred and the next row are full, the daughter is either pushed forwards to the next row or sideways in the same row with equal probability and thereby displaces another cell. The same movement rules are then iteratively applied to this displaced cell and other cells that are subsequently displaced, until either an empty point is filled in the lattice or a cell is pushed out. The randomness in simulated cell movement is inspired from the randomness observed in the orientation of cell division planes [56].

In Class 3 models (Figure 2.2C), the shape of the MZ lattice changes with worm age to match experimental behavior. For rows that see their capacity diminish, thus requiring cell rearrangement, the same movement rules described above are applied.

Section 2.12: Simworm — rules for cell division within the mitotic zone

Cells can only divide in mitotic zone. At each cell birth during the simulation, a time of next cell division is computed independently for the two daughters by sampling from a uniform distribution whose mean is determined by the piecewise linear function described above and whose width is 1% of cell cycle length. For Class 3 models (Figure 2.2B) where pre-meiotic arrest is explicitly simulated, the time at which the cell will enter G2 is computed based experimentally-determined cell cycle phase indices (see Section 2.20). Both time of division and time of G2-M transition are entered into a priority queue that keeps track of the next event to take place in the simulation. If the cell is pushed forward before it has divided, the time to next division is scaled using the ratio of cycle lengths between the new row and the old row. Time
in the simulation moves forward by retrieving the next simulation event from the priority queue each time the previous event — either cell division or G2-M transition — has been processed.

Section 2.13: Simworm — rules for gametogenesis

The following section only applies to Class 2 and 3 models (Figure 2.2B,C) and not Class 1 models (Figure 2.2A). Cells that exit the MZ enter the meiotic region (MeZ) which is modeled as a one-dimensional lattice. Once the MeZ is filled past capacity, cells exit in first-in-first-out order. The first 44 cells that exit the MeZ become primary spermatocytes (Table 2.1), while the rest of the cells either undergo apoptosis or oogenesis with some time-dependent probability (Table 2.2). An explanation of how these probabilities are derived is given in Section 2.17.

Pedigree depth of progeny is calculated from sperm-oocyte pairs. Note that spermatocytes contributed a four-fold weighted contribution to average pedigree depth since each primary spermatocyte produces four spermatids.

Section 2.14: Simworm — initialization and termination conditions

Simworm is initially seeded with a single progenitor cell at the distal end of the MZ for all simulations described in this chapter. Each cell keeps a record of the number of divisions that link it to the progenitor cell — i.e. of its pedigree depth. Simworm terminates when 3,000 cells are produced for Class 1 Models. When gametogenesis is simulated (Class 2 and Class 3 Models), Simworm terminates once 176 oocytes are produced.

For fits to experimental cell cycle data (see Chapter 4), gonadal arms are seeded with a population of cells whose initial “age” within the cell cycle is taken from an exponential distribution [86], and whose overall cell phase distribution matched experimentally-determined cell cycle phase indices. Simulations were pre-run for a period of two hours in simulation time, at which points cells in S phase are marked as “EdU-positive”. In this case, Simworm terminates once an appropriate chase time is simulated.

Section 2.15: Simworm parameter selection — Timing of temporal control points

Developmental stages L2, L4, L4+1, and L4+3 were chosen as representative temporal control points for all Simworm models. Rationale for this choice follows:
The start of germ cell division is the biological equivalent of initializing Simworm. Biologically, this occurs sometime during L1/L2 in the *C. elegans* germline; for convenience, we assume that germ cells start dividing at L2.

- L4 is a representative time point for the start of oocyte production since the worm starts laying eggs shortly after L4. In addition, L4 is easy to stage experimentally, an important detail when validating Simworm predictions with experimental measurements. Note that L4 here actually means mid-L4, but is truncated to reduce verbosity (see List of Abbreviations).

- L4+1 is a stage that the worm community often uses to study the adult germline, and thus is chosen as a representative time point.

- L4+3 is a representative time point for the end of progeny production because there is a drastic decrease in progeny production rate in selfed worms.

Developmental timings were defined in cell divisions as follows:

- 0 cell divisions elapsed at L2.
- \( \sim 400 \) cell divisions elapsed at L4.
- \( \sim 1200 \) cell divisions elapsed at L4+1.
- \( \sim 2400 \) cell divisions elapsed at L4+3.

**Section 2.16: Simworm parameter selection — MeZ geometry**

The following section only applies to Class 2 and 3 models (Figure 2.2B,C) and not Class 1 (Figure 2.2A). No oogenesis occurs prior to L4 and so we cannot assume that the MeZ is completely filled. Thus, we cannot infer the size of the MeZ based on total cell counts at L4. Instead, we calculate the size of the MeZ at L4 so that egg-laying begins at the appropriate time.

- The total length of L4 is 56 - 43.5 = 12.5 hours and the time to last molt to egg laying is 65 - 56 = 9h (WormAtlas). We measured the time from L2 to L4 to be 30 hours. Thus, the time from L2 to egg laying is (time from L2 to L4) + 0.5 * length of L4 + (time from last molt to egg laying) = 30 + 0.5 * 12.5 + 9 = 45.25. We adjusted the size of the meiotic region at mid-L4 so that the first cell that exits the meiotic region at the appropriate time. The appropriate size of the meiotic region is 500 cells at 400 cell divisions. Note that a MeZ size of 500 cells does not mean that the MeZ is filled to capacity; MeZ lattice points can be empty.
Since oogenesis occurs by L4+1 and L4+3 days, the MeZ and the MZ are assumed to be completely filled at these time points. Thus, the number of cells in the MeZ should equal the total number of cells minus the number of cells in the MZ.

- There are 253 cells in the MZ at L4+1 (Section 2.18). Based on our own measurements, there are 1002 cells total in the germline at L4+1. Thus, there are **749 cells** in the MeZ at **1200 cell divisions**.
- There are 201 cells in the MZ at L4+3 (Section 2.18). Based on our own measurements, there are 1278 cells total in the germline at L4+3. Thus, there are **1077 cells** in the MeZ at **2400 cell divisions**.

Meiotic region geometries at intermediate time points were inferred via linear interpolation, and assumed to stay constant pre-L4 and post-L4+3.

Section 2.17: Simworm parameter selection — gametogenesis and apoptosis probabilities

The following section only applies to Class 2 and 3 models (Figure 2.2B,C) and not Class 1 (Figure 2.2A). The first 44 cells that exit the MeZ become primary spermatocytes. Note that each primary spermatocyte eventually produces four spermatids. Otherwise, cells either undergo oogenesis or apoptosis when they exit the MeZ. Apoptosis probabilities were calculated by plugging in the experimentally measured cell cycle profile (Table 4.5) into a Class 3 model, and finding the apoptosis probability that would generate a simulated reproductive schedule that matches experimental measurements. These values are given in Table 2.2.

Section 2.18: Simworm parameter selection — MZ geometry

The following section only applies to Class 3 models (Figure 2.2C) and not Class 1 or 2 models (Figure 2.2A). The MZ lattice is assumed to be filled at L4, L4+1, and L4+3. Image segmentation was used to count the number of cells per row.

- **[5,5,6,7,7,8,8,9,9,10,10,10,10,11,11,11,11,11,10,10] at L4**
- **[6,7,8,10,10,11,13,14,15,16,16,16,17,17,16,16,15,0,0,0,0] at L4+1**
- **[9,10,13,15,16,17,19,19,20,21,21,21,12,0,0,0,0,0,0,0,0] at L4+3**

Geometry at intermediate time points was inferred via linear interpolation. Geometry was assumed to stay constant pre-L4 and post-L4+3.
### Section 2.19: Simworm parameter selection — premeiotic arrest

The following section only applies to Class 3 models (Figure 2.2C) and not Class 1 or 2 models (Figure 2.2A). To model pre-meiotic arrest in the PMZ, we assume cells arrest at the G2-M checkpoint if the simulated mitotic index exceeds that of experimentally measured indices in the PMZ. This is hereafter referred to as the M-phase fraction threshold (MFT). We assumed no pre-meiotic cells exist prior to L4. Thus, the MFT is:

- \([1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1] \text{ prior to L4}\)

At L4, there is a sharp increase in G2-index at 15 cell rows (Figure 4.2). This is consistent with cells entering pre-meiotic G2 and arresting. We assumed that no pre-meiotic cells exist in the first 15 cell rows. After row 15, the MFT is the experimentally measured M-phase index (Figure 4.2):

- \([1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0.040,0.036,0.035,0.037,0.017,0.0096,0.013,0.0] \text{ at L4}\)

At L4+1, there is a sharp increase in G2-index starting at 11 cell rows (Figure 4.2). As before, we assumed no pre-meiotic cells exist in the first 11 cell rows. After row 11, the MFT is the experimentally measured M-phase index (Figure 4.2):

- \([1,1,1,1,1,1,1,1,1,1,1,1,0.024,0.027,0.022,0.023,0.013,0.011,0.0079,0.0043,0,0,0,0] \text{ at L4+1}\)

M-phase fraction thresholds at intermediate time points were inferred via linear interpolation. At L4+3, it was difficult to resolve a distinct PMZ (likely due to intermittent cycling). In the absence of information, the MFT was assumed to stay constant post-L4+1. This model of pre-meiotic arrest shows excellent agreement with experimental data (Figure 4.6).

### Section 2.20: Simworm parameter selection — Timing of G2/M transition

The following section only applies to Class 3 models (Figure 2.2C) and not Class 1 or 2 models (Figure 2.2A). The timing of the G2-M transition is used to compute the dynamics of pre-meiotic entry. Based on experimental data (Figure 3), the cell phase index distribution at L4 is:

- **G1** = \([0.11,0.09,0.12,0.16,0.09,0.10,0.13,0.05,0.07,0.10,0.09,0.06,0.09,0.07,0.08,0.08,0.08,0.08,0.08,0.08,0.08]\)

- **S** = \([0.39,0.52,0.52,0.55,0.60,0.64,0.61,0.71,0.63,0.63,0.64,0.69,0.66,0.67,0.70,0.70,0.70,0.70,0.70,0.70]\)

- **G2** = \([0.46,0.37,0.32,0.25,0.27,0.23,0.21,0.19,0.26,0.23,0.22,0.19,0.20,0.21,0.18,0.18,0.18,0.18,0.18,0.18,0.18,0.18]\)
• $M = [0.04, 0.02, 0.03, 0.05, 0.04, 0.03, 0.05, 0.04, 0.04, 0.05, 0.06, 0.05, 0.04, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05]$

Similarly, at L4+1:

• $G1 = [0.04, 0.09, 0.09, 0.09, 0.06, 0.08, 0.10, 0.11, 0.11, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08, 0.08]$

• $S = [0.65, 0.57, 0.64, 0.65, 0.72, 0.69, 0.68, 0.69, 0.67, 0.71, 0.72, 0.72, 0.72, 0.72, 0.72, 0.72, 0.72, 0.72, 0.72, 0.72, 0.72, 0.72]$

• $G2 = [0.29, 0.30, 0.24, 0.23, 0.19, 0.20, 0.19, 0.18, 0.20, 0.19, 0.17, 0.17, 0.17, 0.13, 0.13, 0.13, 0.13, 0.13, 0.13, 0.13, 0.13, 0.13]$

• $M = [0.02, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.02, 0.02, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03, 0.03]$

At L4+3, it was difficult to resolve spatial cell phase indices (likely due to intermittent cycling). In the absence of information, cell phase indices were assumed to stay constant post-L4+1 and pre-L4. Note we assume that the cell cycle phase distribution of non-premeiotic cells remains constant past row 15, 11 at L4 and L4+1.

The timing of the G2-M transition (and all other cell phase transitions) can be directly computed from cell phase index data. Assuming an exponentially-decreasing cell age distribution $f$ such that $f(0) = 2^{-f(1)}$ [86], we computed the relative length of G1 by solving $F(x) = p_{G1}$, where $F(x) = 2 - 2^{1-x}$ is the fraction of cells younger than $x$ according to the exponential age distribution, and $p_{G1}$ is the observed G1 phase index (and so forth for subsequent phases).

Section 2.21: Simworm parameter selection — Cell production constraints and progeny production constraints

The minimum cell cycle length that we observed experimentally was 2.8 hours (Table 4.5). In optimizations 10—13, we assumed that a lower limit on cell cycle length exists and that pedigree depth optimization must satisfy cell production constraints and progeny production constraints.

We first determined the the timing (hours) of temporal control points L2, L4, L4+1, and L4+3 relative to the start of germ cell division. We measured the time from the first germ cell division to L4 to be approximately 30 hours. Assuming that germ cell divisions begin at L2, we thus have hourly developmental timings of 30 hours from L2 to L4, 54 hours from L2 to L4 +1 day, and 102 hours from L2 to L4+3. Combined with developmental timings from Section 2.15, we derive cell production constraints of
400 cell divisions by 30 hours, 1200 cell divisions by 54 hours, and 2400 cell divisions by 102 hours. The same process was used to derive progeny production constraints given by Table 2.1.

Section 2.22: Simworm — non-dimensionality

The timing of L2, L4, L4+1, and L4+3 is recorded in cell divisions elapsed (which is non-dimensional) rather than hours elapsed. We used non-dimensional units in order to decouple spatial variation in cell cycle length from temporal variation in cell cycle length; the pedigree depth of a germline simulation with a cell cycle profile magnitude of 1 hour is the same as that with a cell cycle profile magnitude of 1000 hours. Thus, only relative differences in spatial cell cycle length matter when computing pedigree depth; pedigree depth is entirely parameterized by $\alpha$ and an infinite number of cell cycle profiles can yield optimal $\alpha$.

Section 2.23: Pedigree depth optimization via self-refining grid search and Markov Chain Monte Carlo

Simworm allows us to calculate pedigree depth statistics for a given cell cycle profile. Finding the profile that minimizes pedigree depth is an altogether different problem known as optimization. Pedigree depth optimization is non-trivial due to the dimensionality of the spatiotemporal cell cycle profile space and the stochastic nature of our simulations (due to randomness in cell movement and in cell cycling). Asking what cell cycle profile minimizes mutation accumulation thus requires minimizing a stochastic, high-dimensional objective function and deriving a range of parameters that perform reasonably well around that minimum. Compounding the difficulty of the problem, many of our simulations are performed under constraints on speed of development and reproduction that are subject to the same stochastic fluctuations.

We took a two-step approach to identify parameter sets that minimize mutation accumulation. First, we performed a grid search for parameters that met constraints on average and that minimized the empirical average of the objective function, sampled at least 450 times at each point. To optimize performance given the relatively high number of dimensions in our grid searches, we used a grid that dynamically self-refined around the parameter regions in which the objective function was lowest. We used custom-written software that used Java Remote Method Invocation to distribute jobs to $\sim$1,000

31
single-threaded workers provided by a cluster of 64-core nodes, and dynamically adjusted the grid using aggregated results. Using this setup, a six-dimensional optimization takes ~1–2 days to complete.

As a second step, we used Markov Chain Monte Carlo [87] to establish a posterior distribution on the parameters that did at least as well as the best parameter identified by grid search. Each chain was initiated using that parameter. Burn-in was calculated post-hoc so that the autocorrelation of all output parameters decayed to at least 1/e. Each iteration computed an empirical average for the objective function and for constraints using 450 samples; proposed moves were rejected if the empirical average of the objective function was ε higher than the grid-search optimum or if constraints were violated with corresponding functions ε higher than for the grid-search optimum (ε = 0.1 for cell production constraints, ε = 1.0 for fecundity constraints, and ε = 0.01 for the fitness metric unless otherwise specified in Supplemental Data 2.1). Each chain ran for 10,000 iterations. In order to establish posterior distributions, we thresholded samples along the chain path to keep those that met constraints and did at least as well as the starting point in terms of minimizing the objective function (note that some posterior distributions only contain a large number of repeats of the same point). After thresholding, each optimization had at least 50 samples used to construct the posterior distributions and 95% credible intervals. We used a parallelized version of our simulations for fast computation of empirical averages at each point, and used custom Python software to drive the process (details of that software will be reported elsewhere); each chain takes ~3 days to complete on a 64-core computer. Detailed results are shown in Supplemental Data 2.2.

**Section 2.24: Computation of population growth rate**

We also used our simulations of germ cell cycling and differentiation to quantify the impact of slow germ cell cycling on the overall population growth rate. For a given profile of cell cycle lengths along the MZ, we recorded the times at which cells destined to become oocytes were pushed out of the MeZ. Since fertilization occurs concomitantly with ovulation, this defined progeny birth times. We computed an average reproductive schedule based on 450 simulation runs, and used that schedule to define a transition matrix whose dominant eigenvalue yielded average population growth rate [88]. Since Simworm is initialized at L2, we add a 26 hour delay to account for the time between egg-laying at L2 (WormAtlas).
Section 2.25: Fitness function for joint optimization of mutation rate and growth rate

We formalize the trade-off between delay in reproductive schedule due to slow germ cell cycling and long-term mutational load stemming from replication dependent mutations using a population genetics approach. Consider the selection coefficient $s_{ΔU}$ for a trait that changes deleterious mutation rate $ΔU$, and the selection coefficient $s_{Δm}$ for a trait that changes population growth rate by $Δm$. Then, assuming independence of the effects on generation rate and mutation accumulation, the aggregate selection coefficient

$$s = s_{ΔU} + s_{Δm}$$

The selection coefficient $s_{ΔU}$ for a trait that changes deleterious mutation rate $ΔU$ is given in given

$$s_{ΔU} = -ΔU/2$$ [89][19][23]. Now, consider a wild-type population of worms with gonads that produce progeny with pedigree depth $p_0$, and a mutant population with gonads that produce progeny with pedigree depth $p$. Then, assuming that all mutations are replication dependent and each cell division has an equal probability of incurring a mutation, the change in deleterious mutation rate is $ΔU = U (p - p_0) / p_0$. Thus, the selection coefficient for a trait that changes pedigree depth from $p_0$ to $p$ is:

$$s_{ΔU} = -U * (p - p_0)/(2*p_0)$$

The selection coefficient $s_{Δm}$ for a trait that changes population growth rate from $m$ to $m_0$ is given by Equation 0:

$$ln(1 + s_{Δm}) = ln(f) * (m-m_0)/m_0$$

The above equation is derived below. Suppose we have a wild-type population of worms and a mutant population of worms. By definition, the selection coefficient $s_{Δm}$ for the mutant trait is given by Equation 1:

$$N(τ)/N_0(τ) = N(0)/N_0(0) * (1 + s_{Δm})^{τ}$$

where $τ$ is the number of discrete generations that elapse. Now, suppose that our worm populations exhibit continuous growth with overlapping generations. Then, the worm number of the wild-type and mutant population $N$ and $N_0$ is given by:

$$N(t) = N(0) * exp(m^*t) , N_0(t) = N_0(0) * exp(m_0^*t)$$

From here, it is straightforward to derive Equation 2:

$$N(t)/N_0(t) = N(0)/N_0(0) * exp((m - m_0) * t)$$

Combining Equations 1 and 2, we derive Equation 3:

$$(1 + s_{Δm})^{τ} = exp((m - m_0) * t)$$
Can we find an expression for \( r \) in terms of \( t \), i.e., can we find the time \( t \) it takes for \( r \) generations to elapse in a continuous population? Then, we have:

\[
f^\tau = \exp(m_0 \cdot t)
\]

where \( f \) is the number of progeny produced by a single individual in one generation. From here, it is straightforward to derive Equation 4:

\[
\tau = m_0 \cdot t / \ln(f)
\]

Substituting Equation 4 into Equation 3, it is straightforward to derive the original Equation 0:

\[
\ln(1 + s_{\Delta m}) = \ln(f) \cdot (m - m_0)/m_0
\]

Note that Equation 0 matches that of [90] without the simplifying assumptions of \( f=2 \) (bacteria produce two daughter cells per generation) and \( s_{\Delta m} = \ln(1 + s_{\Delta m}) \), i.e., \( s_{\Delta m} \ll 1 \). In our simulations, we simulate a single germline producing 176 progeny, and thus \( f = 176 \). \( m \) can be calculated \( m = \ln(r) \), where \( r \) is the dominant eigenvalue of the population transition matrix resulting from the simulated reproductive schedule.

Finally, taking our equations for \( s_{\Delta U}, s_{\Delta m} \), and substituting into \( s = s_{\Delta U} + s_{\Delta m} \), we have the aggregate selection coefficient for a trait that changes pedigree depth from \( p_0 \) to \( p \) while simultaneously changing growth rate from \( m \) to \( m_0 \)

\[
s = -U \cdot (p - p_0)/(2*p_0) + \exp(\ln(176) \cdot (m-m_0)/m_0) - 1
\]
Figure 2.1: A unique ratio of stem cell cycling speed to transit amplifying cell cycling speed minimizes pedigree depth

Simulations of the hypothetical stem cell system diagrammed in Figure 1.2 reveal that pedigree depth (p) is minimized when the ratio of stem cell cycling speed to transit amplifying cell cycling speed is neither too high nor too low. Here, \( \alpha \) is stem cell cycle length divided by transit amplifying cell cycle length. The optimal value of \( \alpha \) increases as the number of cells to be produced increases.
Simworm models fall under three broad classes. In all 3 classes of simulation, the mitotic zone (MZ) was modeled as a lattice. Cell divisions (red) cause cascades of cell movement (green arrows) that push a single proximal cell out of the MZ (purple). (A) In Class 1 models, the MZ was modeled as a rectangular lattice. The average pedigree depth over the first 3,000 cells that exit the MZ is assayed as a function of the cell cycle profile. (B) Gametogenesis was explicitly modeled in Class 2 models. The meiotic zone (MeZ) is modeled as a one-dimensional lattice. Cells are pushed from the MZ into the MeZ; once in the MeZ, cells no longer divide (cyan circles). MeZ cells are pushed proximally until they reach the proximal end of the MeZ, at which time they exit in FIFO order. The first 44 cells that exit the MeZ become spermatocytes and the rest either undergo oogenesis or apoptosis. The average pedigree depth of the first 176 progeny is calculated as a function of cell cycle profile, where the pedigree depth of one progeny is the average pedigree depth of its constituent oocyte and sperm. (C) A biologically realistic MZ is simulated in Class 3 models. Pre-meiotic arrest is modeled in the PMZ (black circles) and MZ geometry is drawn from experimental measurements; note that MZ geometry is not an optimization parameter in Class 3 models. In addition, the shape of the cell cycle profile is constrained to be flat in the proximal mitotic zone (PMZ) in accordance with experimental measurements (Section 4.6).
Figure 2.2: *Simworm* models are biologically realistic (2/2)

A

Long cell cycle length

Cell cycle profile

Short cell cycle length

Cell exits MZ

Mitotic zone (MZ)

distal - proximal

B

Long cell cycle length

Cell cycle profile

Short cell cycle length

Cell pushed into MeZ

Meiotic zone (MeZ)

Apoptosis or oogenesis or spermatogenesis

C

Long cell cycle length

Flat cell cycle profile in PMZ

Short cell cycle length

No optimization over shape of MZ

Apoptosis or oogenesis or spermatogenesis

Meiotic zone (MeZ)
Figure 2.3: Cell cycle length is 3.4 hours for early germline development

Time course of larval germ cell proliferation at its onset. A fit assuming exponential growth gave a cell cycle length of 3.4 h for early germ line development.
Table 2.1: A single self-fertilized gonadal arm produces 176 progeny over five days

The numbers are derived from the average reproductive schedule measured from n = 19 worms by counting progeny laid at days 1 to 7 of adulthood, and dividing the total of number of progeny by 8 for spermatocytes (since 4 spermatocytes give rise to 4 sperm, and since there are 2 gonadal arms per worm), and the daily counts by 2 for oocytes.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Time window</th>
<th>Gametes produced</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 to L4</td>
<td>30h</td>
<td>44 primary spermatocytes</td>
</tr>
<tr>
<td>L4 to L4 + 1</td>
<td>24h</td>
<td>23 oocytes</td>
</tr>
<tr>
<td>L4 + 1 to L4 + 2</td>
<td>24h</td>
<td>83 oocytes</td>
</tr>
<tr>
<td>L4 + 2 to L4 + 3</td>
<td>24h</td>
<td>64 oocytes</td>
</tr>
<tr>
<td>L4 + 3 to L4 + 4</td>
<td>24h</td>
<td>5 oocytes</td>
</tr>
<tr>
<td>L4 + 4 to L4 + 5</td>
<td>24h</td>
<td>1 oocyte</td>
</tr>
</tbody>
</table>
Table 2.2: Cells exiting the meiotic zone in Simworm simulations either undergo apoptosis or gametogenesis

Probabilities are given as a function of the sequence index of cells leaving the MZ (cells leave the MZ and Meiotic Zone in the same order), and computed from cell cycle and germ cell count data as described in Section 2.17. Primary spermatocytes do not undergo apoptosis.

<table>
<thead>
<tr>
<th>Cell sequence index</th>
<th>Stage of Meiotic Zone exit</th>
<th>Apoptosis probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 210</td>
<td>L3 - L4+1</td>
<td>68%</td>
</tr>
<tr>
<td>211 - 682</td>
<td>L4+1 - L4 + 2</td>
<td>83%</td>
</tr>
<tr>
<td>683 - 1150</td>
<td>L4 + 2 - L4+3</td>
<td>86%</td>
</tr>
<tr>
<td>1151 - 1478</td>
<td>L4+3 - L4 + 4</td>
<td>98%</td>
</tr>
<tr>
<td>1479 - Inf</td>
<td>L4 + 4 - Inf</td>
<td>99%</td>
</tr>
</tbody>
</table>
Table 2.3: A cell cycle gradient minimizes pedigree depth in *Simworm* models

Each row shows results for one kind of problem, defined by the objective function to optimize (PD: Pedigree Depth), the parameters that are free to evolve within bounds during the optimization procedure (which can include MZ dimensions and distal cell cycle lengths at various stages), and other characteristics grouped under “Optimization setup”. Results shown are derived using grid search; \( \alpha \): ratio of cell cycle lengths between the distal end of the DMMZ and proximal end of the MMZ (slower distal cell cycle for \( \alpha > 1 \)). See Supplemental Data 2.1 for an extended version of this table, including credible intervals derived from MCMC.

<table>
<thead>
<tr>
<th>Optimization</th>
<th>Objective function</th>
<th>Free parameters</th>
<th>Optimization setup</th>
<th>Optimization results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Notes on optimizations:
1: Optimization 8 was run with deleterious mutation rate \( U = 0.03 \), and optimization 9 with \( U = 0.48 \)
2: Optimization 12 assumes a cell cycle speed dependent mutation rate
3: Optimization 13 assumes that distal cells preserve an immortal DNA strand
Supplemental Data 2.1: Full table of *Simworm* optimization results (1/7)

Truncated results are given in Table 2.3.

<table>
<thead>
<tr>
<th>Optimization ID #</th>
<th>Model Class (see Figure 2.2)</th>
<th>Metric</th>
<th>Fecundity and cell production constraints</th>
<th>Uses experimental geometry</th>
<th>MCMC metric ε</th>
<th>MCMC cell production constraint ε</th>
<th>MCMC fecundity constraint ε</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>P.D. first 3000 cells</td>
<td>No</td>
<td>No</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>P.D. first 3000 cells</td>
<td>No</td>
<td>No</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>P.D. first 3000 cells</td>
<td>No</td>
<td>No</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>P.D. first 3000 cells</td>
<td>No</td>
<td>No</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>P.D. gametes</td>
<td>No</td>
<td>No</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>P.D. gametes</td>
<td>No</td>
<td>No</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>P.D. gametes</td>
<td>No</td>
<td>No</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>Aggregate fitness</td>
<td>No</td>
<td>Yes</td>
<td>0.00</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>Aggregate fitness</td>
<td>No</td>
<td>Yes</td>
<td>0.00</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>P.D. gametes</td>
<td>Yes</td>
<td>No</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>P.D. gametes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.01</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>P.D. gametes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.01</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>P.D. gametes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.01</td>
<td>0.10</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Supplemental Data 2.1: Full table of *Simworm* optimization results (2/7)

<table>
<thead>
<tr>
<th>Optimization ID #</th>
<th>Max MZ width/height</th>
<th>Max MZ cell number</th>
<th>Number temporal control points</th>
<th>Pre-meiotic arrest within MZ</th>
<th>Special notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2000</td>
<td>2000</td>
<td>4</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2000</td>
<td>359</td>
<td>1</td>
<td>No</td>
<td>1 with flat profile, max 359 cells</td>
</tr>
<tr>
<td>3</td>
<td>2000</td>
<td>2000</td>
<td>1</td>
<td>No</td>
<td>1 with flat profile</td>
</tr>
<tr>
<td>4</td>
<td>2000</td>
<td>200</td>
<td>1</td>
<td>No</td>
<td>1 with flat profile, max 200 cells</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>2000</td>
<td>4</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>2000</td>
<td>4</td>
<td>No</td>
<td>5 with MZ 30 x 30</td>
</tr>
<tr>
<td>7</td>
<td>2000</td>
<td>2000</td>
<td>1</td>
<td>No</td>
<td>5 with forced spatially flat profile</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>Yes</td>
<td>8 with U=0.48</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>500</td>
<td>3</td>
<td>No</td>
<td>6 with cell production and fecundity constraints</td>
</tr>
<tr>
<td>11</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>Yes</td>
<td>Experimental, cone-like geometry</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>Yes</td>
<td>12 with cell cycle speed dependent mutation rate</td>
</tr>
<tr>
<td>13</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>Yes</td>
<td>12 with immortal strand</td>
</tr>
</tbody>
</table>

43
Supplemental Data 2.1: Full table of *Simworm* optimization results (3/7)

<table>
<thead>
<tr>
<th>Optimization ID #</th>
<th>Optimized metric (posterior mean)</th>
<th>Optimized metric (posterior CI)</th>
<th>Optimized MZ width</th>
<th>Optimal MZ width posterior mean</th>
<th>Optimal MZ width CI</th>
<th>Optimal MZ height</th>
<th>Optimal MZ height posterior mean</th>
<th>Optimal MZ height CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.74</td>
<td>11.74</td>
<td>11.74 - 11.74</td>
<td>359</td>
<td>351.32</td>
<td>319 - 383</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>13.94</td>
<td>13.94</td>
<td>13.94 - 13.94</td>
<td>1</td>
<td>1.00</td>
<td>1 - 1</td>
<td>359</td>
<td>359.00</td>
</tr>
<tr>
<td>3</td>
<td>12.05</td>
<td>12.05</td>
<td>12.05 - 12.05</td>
<td>1523</td>
<td>1528.88</td>
<td>1455 - 1610</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>17.20</td>
<td>17.19</td>
<td>17.19 - 17.19</td>
<td>1</td>
<td>1.00</td>
<td>1 - 1</td>
<td>200</td>
<td>200.00</td>
</tr>
<tr>
<td>5</td>
<td>9.62</td>
<td>9.61</td>
<td>9.59 - 9.62</td>
<td>112</td>
<td>67.51</td>
<td>51 - 122</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>9.91</td>
<td>9.89</td>
<td>9.85 - 9.90</td>
<td>1</td>
<td>31.56</td>
<td>1 - 199</td>
<td>216</td>
<td>159.82</td>
</tr>
<tr>
<td>8</td>
<td>-0.0054</td>
<td>-0.0062</td>
<td>-0.0078 - -0.0057</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>0.0005</td>
<td>-0.0011</td>
<td>-0.0035 - 0.0001</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>9.85</td>
<td>9.84</td>
<td>9.81 - 9.85</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>8.88</td>
<td>8.87</td>
<td>8.87 - 8.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>8.85</td>
<td>8.84</td>
<td>8.84 - 8.85</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Optimization ID #</td>
<td>L2-D</td>
<td>L2-D posterior mean</td>
<td>L2-D CI</td>
<td>L2-P</td>
<td>L2-P posterior mean</td>
<td>L2-P CI</td>
<td>α</td>
<td>α L2 posterior mean</td>
</tr>
<tr>
<td>-------------------</td>
<td>------</td>
<td>---------------------</td>
<td>---------</td>
<td>------</td>
<td>---------------------</td>
<td>---------</td>
<td>---</td>
<td>---------------------</td>
</tr>
<tr>
<td>1</td>
<td>1893.15</td>
<td>1571.17</td>
<td>1435.22 - 1773.36</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>676.13</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>2.80</td>
<td>7.49</td>
<td>2.80 - 20.41</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>2.68</td>
</tr>
<tr>
<td>6</td>
<td>2.80</td>
<td>4.93</td>
<td>2.80 - 9.09</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>1.76</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>2.80</td>
<td>2.80</td>
<td>2.80 - 2.80</td>
<td>2.80</td>
<td>2.80</td>
<td>2.80 - 2.80</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>2.80</td>
<td>2.80</td>
<td>2.80 - 2.80</td>
<td>2.80</td>
<td>2.80</td>
<td>2.80 - 2.80</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>3.40*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Supplemental Data 2.1: Full table of *Simworm* optimization results (5/7)

<table>
<thead>
<tr>
<th>Optimization ID #</th>
<th>L4-D posterior mean</th>
<th>L4-D CI</th>
<th>L4-P posterior mean</th>
<th>L4-P CI</th>
<th>α L4 posterior mean</th>
<th>α L4 CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1985.02</td>
<td>1610.86</td>
<td>1206.51 - 1975.93</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>708.94</td>
<td>575.31</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>135.61</td>
<td>865.85</td>
<td>86.51 - 1437.09</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48.43</td>
<td>309.23</td>
</tr>
<tr>
<td>6</td>
<td>9.79</td>
<td>45.76</td>
<td>9.79 - 85.57</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.50</td>
<td>16.34</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>2.80</td>
<td>2.84</td>
<td>2.80 - 2.86</td>
<td>2.80</td>
<td>2.80 - 2.80</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.80 - 2.80</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>2.80</td>
<td>2.82</td>
<td>2.80 - 2.89</td>
<td>2.80</td>
<td>2.80 - 2.80</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.80 - 2.80</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>6.00</td>
<td>5.64</td>
<td>5.38 - 5.97</td>
<td>2.80</td>
<td>2.94</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.81 - 3.16</td>
<td>2.14</td>
</tr>
<tr>
<td>11</td>
<td>6.45</td>
<td>6.23</td>
<td>5.82 - 6.57</td>
<td>2.80</td>
<td>2.86</td>
<td>2.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.81 - 2.98</td>
<td>2.30</td>
</tr>
<tr>
<td>12</td>
<td>6.55</td>
<td>6.57</td>
<td>6.56 - 6.58</td>
<td>2.80</td>
<td>2.81</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.81 - 2.81</td>
<td>2.34</td>
</tr>
<tr>
<td>13</td>
<td>2.80</td>
<td>2.83</td>
<td>2.82 - 2.83</td>
<td>4.42</td>
<td>4.42</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.42 - 4.47</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.63 - 0.64</td>
<td>0.63</td>
</tr>
</tbody>
</table>
### Supplemental Data 2.1: Full table of *Simworm* optimization results (6/7)

<table>
<thead>
<tr>
<th>Optimization ID #</th>
<th>L4+1-D posterior mean</th>
<th>L4+1-D CI</th>
<th>L4+1-P posterior mean</th>
<th>L4+1-P CI</th>
<th>α L4+1 posterior mean</th>
<th>α L4+1 CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>553.53</td>
<td>698.02</td>
<td>473.52 - 964.70</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>253.95</td>
<td>126.00</td>
<td>17.79 - 319.90</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>3.80</td>
<td>9.05</td>
<td>3.18 - 26.99</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>2.86</td>
<td>2.81</td>
<td>2.81 - 2.86</td>
<td>2.80</td>
<td>2.80</td>
<td>1.02</td>
</tr>
<tr>
<td>9</td>
<td>2.82</td>
<td>2.83</td>
<td>2.82 - 2.85</td>
<td>2.80</td>
<td>2.80</td>
<td>1.01</td>
</tr>
<tr>
<td>10</td>
<td>5.42</td>
<td>6.32</td>
<td>5.64 - 7.40</td>
<td>2.80</td>
<td>3.44</td>
<td>1.94</td>
</tr>
<tr>
<td>11</td>
<td>4.00</td>
<td>8.14</td>
<td>5.54 - 9.42</td>
<td>2.80</td>
<td>2.96</td>
<td>1.43</td>
</tr>
<tr>
<td>12</td>
<td>7.81</td>
<td>7.14</td>
<td>4.92 - 7.55</td>
<td>2.80</td>
<td>2.90</td>
<td>2.79</td>
</tr>
<tr>
<td>13</td>
<td>2.80</td>
<td>2.85</td>
<td>2.80 - 2.87</td>
<td>3.95</td>
<td>3.75</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Supplemental Data 2.1: Full table of *Simworm* optimization results (7/7)

<table>
<thead>
<tr>
<th>Optimization ID #</th>
<th>L4+3-D</th>
<th>L4+3-D posterior mean</th>
<th>L4+3-D CI</th>
<th>L4+3-P</th>
<th>L4+3-P posterior mean</th>
<th>L4+3-P CI</th>
<th>α L4+3 posterior mean</th>
<th>α L4+3 CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.80</td>
<td>2.86</td>
<td>2.80 - 2.98</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>1321.45</td>
<td>1058.60</td>
<td>476.16 - 1703.07</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
<td>471.95</td>
<td>378.07</td>
</tr>
<tr>
<td>6</td>
<td>149.84</td>
<td>333.06</td>
<td>85.70 - 686.38</td>
<td>2.80*</td>
<td>-</td>
<td>-</td>
<td>53.52</td>
<td>118.95</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>2.85</td>
<td>2.99</td>
<td>2.83 - 3.29</td>
<td>2.80</td>
<td>2.82</td>
<td>2.80 - 2.82</td>
<td>1.02</td>
<td>1.06</td>
</tr>
<tr>
<td>9</td>
<td>2.86</td>
<td>2.93</td>
<td>2.81 - 3.37</td>
<td>2.80</td>
<td>2.82</td>
<td>2.80 - 2.86</td>
<td>1.02</td>
<td>1.04</td>
</tr>
<tr>
<td>10</td>
<td>7.24</td>
<td>9.81</td>
<td>8.53 - 11.05</td>
<td>3.18</td>
<td>4.45</td>
<td>3.05 - 6.33</td>
<td>2.28</td>
<td>2.31</td>
</tr>
<tr>
<td>11</td>
<td>8.31</td>
<td>5.07</td>
<td>3.25 - 8.41</td>
<td>2.80</td>
<td>3.62</td>
<td>2.89 - 4.43</td>
<td>2.97</td>
<td>1.49</td>
</tr>
<tr>
<td>12</td>
<td>7.76</td>
<td>7.11</td>
<td>6.42 - 8.60</td>
<td>2.80</td>
<td>2.91</td>
<td>2.82 - 3.52</td>
<td>2.77</td>
<td>2.45</td>
</tr>
<tr>
<td>13</td>
<td>2.80</td>
<td>2.94</td>
<td>2.92 - 2.96</td>
<td>2.91</td>
<td>4.66</td>
<td>3.91 - 5.06</td>
<td>0.96</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Supplemental Data 2.2: MCMC statistics for Simworm optimizations (1/7)

A (optimization 1)

B (optimization 2)
Supplemental Data 2.2: MCMC statistics for *Simworm* optimizations (2/7)

C (optimization 3)

D (optimization 4)
Supplemental Data 2.2: MCMC statistics for *Simworm* optimizations (4/7)

**G (optimization 7)**

**H (optimization 8)**
Supplemental Data 2.2: MCMC statistics for Simworm optimizations (5/7)

I (optimization 9)

J (optimization 10)
Supplemental Data 2.2: MCMC statistics for Simworm optimizations (6/7)

K (optimization 11)

L (optimization 12)
Supplemental Data 2.2: MCMC statistics for Simworm optimizations (7/7)

M (optimization 13)

![Graphs showing MCMC statistics for Simworm optimizations](image-url)
Chapter 3: Parismi - image cytometry software

for the C. elegans germline

Section 3.1: Contributions

The data presented in this chapter has been submitted for publication.

• Michael Chiang and Prof. Olivier Cinquin designed and implemented active contours and downstream analyses
• Sam Hallman and Prof. Charless C. Fowlkes designed and implemented automatic cell detection and validation benchmarks
• Dr. Amanda Cinquin, Nabora Reyes de Mochel, Adrian Paz, and Dr. Shimako Kawauchi designed and performed experiments and interpreted the results with supervision from Prof. Anne L. Calof, Prof. Ken Cho, and Prof. Olivier Cinquin.

Section 3.2: Introduction

In Chapter 2, we found that a ~two-fold difference in cell cycle length minimizes pedigree depth in computational simulations. We want to ask whether a two-fold difference in cell cycle length actually exists in a real germline. Image cytometry may provide a sensitive approach to measuring spatial differences in cell cycle speed. Although image segmentation is a very active research area, cytometric studies of the germline still rely on hand-segmenting three-dimensional confocal images, a time-consuming and error prone task for the human researcher [70][80].

A number of general image segmentation tools exist that are specifically targeted at biological applications, including both open source [74][91][92][93][94][95] and commercial software (e.g., Imaris, Bitplane or Volocity, PerkinElmer). For more extensive surveys, see [96][97][98]. Despite rapid development (see cell tracking benchmark competition [99]), the problem of automatically producing high-quality 3D segmentations of cells in general images remains unsolved due to the wide variation in appearance across different tissue and cell types, labeling procedures and imaging methods. Rather than tuning existing pipelines or developing custom segmentation algorithms that might improve performance on images of particular cell types, we decided to design a pipeline that maximizes the utility of the most accurate but most expensive resource in image segmentation: time spent by experts providing ab initio annotations or correcting computer-derived segmentations. This pipeline aims to provide automation of
repetitive tasks for which there is no need for user input (such as applying image transformations like blurring with pre-determined parameters or segmenting out the region around a putative cell location), and to allow the user to focus on the tasks that provide the highest added value.

We designed our pipeline Parismi (Pipeline for Automated oR Interactive SegMentation of Images) around a simple, two-step idea. Cells are first detected, and these detections are then used to seed a segmentation algorithm. Detection can be performed manually (using a 3D browsing interface similar to VANO [95]) or by a machine learning algorithm trained from a set of manual annotations used to bootstrap the procedure (we chose a machine learning procedure, similar to [100], as opposed to ad-hoc processing of the fluorescence signal [101][102][103], to facilitate reuse across sample types that vary in nuclear morphology and imaging conditions). The output of the machine learning algorithm can be reviewed and corrected by the same interface. As the set of segmented cells that have received manual curation expands, the machine learning algorithm can be re-trained from these segmentations, providing for iterative improvements in the quality of the automatic detection step. This approach is loosely conceptually similar to "semi-supervised learning" [104] and "active learning" [100][105], although our current implementation is not fully interactive in that sense.

As the second step of our segmentation procedure, we use “active contours” (implemented following [106]), which are closed surfaces that are initialized from the detected center point and grow smoothly outwards in three dimensions until they encounter the putative cell boundary (suggested by membrane staining) or until they collide with surfaces corresponding to neighboring cells. The surface evolution is governed by both membrane staining (also referred to as the guide image) and by the curvature of the surface itself; penalizing high local curvatures helps the surface maintain a roughly spherical shape, which provides robustness, e.g., to noise in the guide image. In the case of stains that are not limited to the periphery of the structure being segmented, such as DAPI or Hoechst stains for nucleus segmentation, pre-processing of the image can be used to produce a guide image that outlines boundaries, so that active contours can still be applied.

The use of active contours has a long history in cell segmentation and has proven to be a robust approach for identifying cellular and nuclear volumes in three dimensions [106][107][108]. Other approaches to segmentation such as geodesic distance transform [75], gradient flow smoothing [76], and watershed transform have also been used successfully to perform 3D volumetric segmentation of cell
nuclei in specific sample preparations [77][54]; however, these techniques often require post-processing [79] to correct segmentation errors. In particular, segmenting densely labeled whole cells (rather than nuclei) requires high-quality membrane staining to achieve sufficient local contrast [109]. Our choice of seeded segmentation and active contours avoids difficulties that arise in purely segmentation-based approaches, particularly in our C. elegans gonadal arm data, where the spatial distribution of DNA towards the periphery of the nucleus results in gaps that can be larger than the separations between neighboring nuclei and that hence cannot be easily resolved by local smoothing.

In summary, Parismi is composed of four broad components: (1) an interface to manually annotate cells in 3D images; (2) an automatic cell detector that can be trained from manual annotations; (3) an active contour implementation that produces cell segmentations, from which fluorescence content and position in the tissue of individual cells are computed (Figure 1); and (4) a number of plugins for segmentation analysis. Overall our procedure is similar to previous reports in that it relies on machine learning for segmentation of biological images (see ilastik [100] or Trainable Weka [94]) but distinct in that it provides for full automation of repetitive steps, which has allowed us to segment hundreds of thousands of cells, and in that it relies on active contours instead of thresholding, watershed or identification of boundary pixels, which we have found provides for more robustness when cells are tightly packed and/or not perfectly separated by a clean boundary signal. Code for Parismi is open-source and is available at https://github.com/cinquin/parismi. Datasets supporting results of this article are available for download at http://cinquin.org.uk/Parismi_dataset.tgz. A schematic overview of Parismi is given in Figure 3.1.

Section 3.3: Cell detection in Parismi is based on DNA-stained images via SVM with HOG features

We trained an automatic cell detector that predicts cell center locations from DNA-stained image stacks by classifying each sub-window of the stack as either containing a cell center or not. Positive example sub-windows were specified by hand-clicking cell centers, and negative examples automatically extracted from locations farther than one cell radius from all positive labels.

We extracted image features from the training set using 2D windows taken along the xy and xz planes running through detection centers. From the xy slice we computed two features: average pixel brightness of the detection window, and a histogram of oriented gradients (HOG features) computed over a grid of non-overlapping sub-windows ([110], Figure 3.2). Within each sub-window, the image gradient
was estimated at each pixel and binned into one of 18 orientations. These histograms were normalized and the normalization factor along with the normalization of neighboring bins were stored as additional features. The same HOG features were computed for the xz slice. Since there was no a priori favored image orientation, the features were symmetrized left-right and top-bottom for positive training examples. Overall, 2,147 cells across 6 gonadal arms were used for training.

We trained a support vector machine (SVM) classifier to distinguish positive detections from negative detections in the training set. A given feature vector corresponds to a cell center if \( w^T v > \tau \), where \( w \) is a weighted vector learned by the SVM, \( v \) is a feature vector, and \( \tau \) is a tunable parameter. A lower value of \( \tau \) yields more cell detections at the cost of more false positives. Since there are millions of negative detections in our training set, we used an iterative training algorithm to decrease run time and memory requirements. This iterative algorithm trains the detector with a subset of negative examples, searches for additional high-scoring negative examples, adds them to the training set, and re-trains the classifier. This iterative approach of hard negative mining is mathematically equivalent to performing training on the set of negative detections. To make the training algorithm robust to potential errors in the localization of cell centers by human labelers, we performed latent estimation of the “true” cell center for each positive training example [111]. Briefly, once a detector had been trained, we ran the detector on the positive training data and re-estimated the center of each cell as the maxima of the detector response within a small radius of the original ground-truth detection. The detector was then re-trained with these updated set of positive locations.

The final step in automatic detection applies the SVM classifier to all sub windows in a given DNA stained image. To accurately handle natural variation in cell sizes, this detection process was carried out multiple times on scaled versions of the original image stack (scale factor ranging from 0.7-1.5 for MZ stacks). Since a given cell may produce multiple positive detections in slightly offset sub-windows, we suppressed detections which overlapped with any higher-scoring detection within one cell radius.

Section 3.4: Preprocessing guide images is necessary to generate good segmentation masks

Membrane staining in the germline is often noisy, and thus we generate an enhanced “guide image” for active contours by preprocessing membrane images. This preprocessing consists of four steps: first, we removed low frequency noise from the membrane image by “sharpening” the image. We
normalized pixels to the average pixel value in a surrounding 2D sliding square parallel to the xy-plane; we chose the size of the sliding square to be the average cell diameter; this normalization had the additional effect of evening fluorescence intensities across the z-axis. Second, we removed sharp discontinuities and high frequency noise by blurring the sharpened image. We set the standard deviation of the Gaussian kernel to the membrane thickness. This choice of standard deviation allowed the blurring filter to remove small image artifacts without compromising the membrane signal. Third, we enhanced “sheet-like” structures using a principal curvature approach [112]. For each pixel in the blurred image \( I \), we calculated the Hessian matrix \( H \):

\[
H(x, y, z) = \begin{bmatrix}
I_{xx} & I_{xy} & I_{xz} \\
I_{yx} & I_{yy} & I_{yz} \\
I_{zx} & I_{zy} & I_{zz}
\end{bmatrix}
\]

with ordered eigenvalues \( d_1(x,y,z) \), \( d_2(x,y,z) \), \( d_3(x,y,z) \). Then, we computed the intermediate image \( I' \):

\[
I'(x, y, z) = -d_1 e^{-(\frac{d_2}{2\sigma_1})^2} e^{-(\frac{d_3}{2\sigma_1})^2}
\]

Note that blob-like structures (where \( |d_1| \gg 1 \), \( |d_2| \ll 1 \), and \( |d_3| \ll 1 \)) and line-like structures (where \( |d_1| \gg 0 \), \( |d_2| \gg 0 \), and \( |d_3| \gg 0 \)) are attenuated in \( I' \), while sheet-like structures (where \( |d_1| \gg 0 \), \( |d_2| \ll 0 \), and \( |d_3| \ll 0 \)) are enhanced. Thus membrane signal, which is sheet-like in three-dimensions, in enhanced in intermediate image \( I' \).

Finally, we generated the final preprocessed image by removing sharp discontinuities through blurring, with the standard deviation of the Gaussian kernel again set to the membrane thickness. The end result is an guide image where cellular boundaries are clearly demarcated by enhanced membrane signal. An example preprocessed guide image is given in Figure 3.1A.

Section 3.5: Segmentation in Parismi is implemented using active contour segmentation of cellular volumes

Implicit active contours, first described in [113], have been used extensively in biological image analysis [106]. Active contours are model-based and work well even in the presence of a poor guide image. Since \( C. elegans \) germ cells are roughly spherical and uniform in size, active contours are a good choice for segmenting them. Consider the partial differential equation:
\[
\frac{d\phi}{dt} = g \|\nabla \phi\| (1 - c_1 \epsilon) + c_2 \nabla g \cdot \nabla \phi
\]

used to update active contours. Here, \( g \) is the inverted preprocessed guide image, \( \phi \) is a higher dimensional function that embeds segmentation mask composed of points where \( \phi(x,y,z) < 0 \), and \( \epsilon \) is the curvature:

\[
\epsilon = \nabla \cdot \left( \frac{\nabla \cdot \phi}{\|\nabla \phi\|} \right)
\]

used to enforce smooth segmentation borders. \( \phi \) is initialized at cell detection points and then active contours are run in two steps. The first step of active contours is conservative; masks stop short of boundaries in the guide image. This is achieved with \( c_2 < 0 \) so that contours are pushed backwards as they approach inner edges. The second step of active contours refines the masks so that they stop on the boundaries of the guide image. This is achieved with \( c_2 > 0 \) so that the contour is pulled forward as it approaches an inner edge, then pushed back as it approaches an outer edge. In order to prevent overlapping segmentation masks, we set \( d\phi(x,y,z)/dt = 0 \) when two different masks “collide”.

In order to decrease run-time, we ran individual \( \phi \) corresponding to individual cells in cropped sub-windows. In addition, we computed updates to \( \phi \) using the narrow band level set method [114].

**Section 3.6: Quantification of top-layeredness allows Parismi to correct for z-attenuation artifacts**

Germlines imaged in a confocal microscope display z-attenuation, where cells deep in the organ display weaker fluorescence due to light absorption by intervening tissue. Correction of z-attenuation is non-trivial because the germline has a tapered geometry such that distance to the top layer of cells does not have a simple dependence on \( z \). In order to resolve the problem of z-attenuation, we sought to identify cells on the top layer of the gonadal arms which have direct line of sight to the microscope objective and thus exhibit minimal attenuation along the z-axis. We defined the top-layeredness \( \theta \) of a cell to be the cross-sectional area of its segmentation mask that is unobscured in the z-projection over all cell segmentation masks (Figure 3.3). Thus, \( \theta = 0 \) means that a cell is completely obscured from the path of the microscope objective, while \( \theta = 1 \) means that a cell has direct line of sight to the microscope objective. In practice, \( \theta > 0.1 \) was a good threshold to select cells on the top layer of the gonadal arm.
An alternative method would have been to select “stack top” cells based on z position within the top $n^{th}$ percentile of z positions, where $n$ can be adjusted on a stack-by-stack basis so that each stack contributes a given number of cells. Because attenuation is stronger when light travels through tissue than when it travels through immersion medium, and because z variation throughout MZs was overall small (6 µm between MZ rows 1 and 10, $n = 18$), the top layer metric provided more accurate fluorescence quantification than the stack top metric (Figure 3.3I-J; z position percentile adjusted so that both metrics selected the same number of cells per MZ, to ensure a fair comparison).

Section 3.7: A normalization procedure is necessary for accurate DNA content quantification.

The naive way to calculate cellular DNA content is to simply sum pixels in a given segmentation mask. However, DNA fluorescence displays appreciable fluorescence attenuation along the distal-proximal axis and the z-axis of the gonadal arm; this could introduce bias into spatial cell cycle studies. We correct for these artifacts when quantifying DNA content using the following normalization procedure:

1. We filter segmented cells to only keep those in the top layer using the “top-layer metric” described in Section 3.6; this minimizes artifactual variations in DNA content due to fluorescence attenuation along the z axis.

2. For each segmentation mask, we computed the raw DNA content (sum over all DNA pixel values inside the mask) and the 95% DNA intensity percentile (95% percentile of DNA pixel values inside the mask).

3. We fit a cubic spline to the empirical distribution of 95% DNA intensity percentiles as a function of geodesic distances, on a gonad-by-gonad basis. We normalized raw DNA contents against this spline to derive spline-normalized DNA contents. This step reduces potential bias from fluorescence attenuation along the distal-proximal axis of the gonads.

4. Cellular data was binned by spatial position. The 10th and 85th percentile of spline-normalized DNA contents in each bin was normalized to 2c and 4c DNA content, respectively. A bin size of four cell rows was used. This step allows us to aggregate data across germ lines by assuming each spatial bin contains the same proportion of G1/S/G2/M-phase cells across germ lines.

In practice, this normalization scheme produced reasonable DNA contents and benchmarking statistics are given in Section 3.12.
Section 3.8: A normalization procedure is necessary for accurate EdU content quantification.

The naive way to quantify cellular fluorescence would be to simply sum pixels in a given segmentation mask. However, EdU staining is quite variable within the germline and care must be exercised to avoid errors when calculating binary EdU contents. We correct for noisy EdU staining when quantifying fluorescence using the following normalization procedure:

1. We first apply a median filter and then threshold the image. All pixel values less than $t_1$ are set to $t_1$, and all pixel values greater than $t_2$ are set to $t_2$, where $t_1$ and $t_2$ are determined on an image-by-image basis. The resulting image was then scaled so the minimum, maximum pixel value was 0, 1, respectively.

2. For each segmentation mask, we summed all normalized EdU values of pixels inside the mask. We then normalized the 10th and 85th percentiles of cellular EdU contents in a given gonadal arm to 0 and 1, respectively. This step allows us to aggregate fluorescence contents across gonads.

3. We classify cells as EdU-positive or EdU-negative by applying a manually set threshold. In practice, a threshold of 0.2 works well.

In practice, this normalization scheme produces reasonable EdU contents and benchmarking statistics are given in Section 3.12.

Section 3.9: Parismi computes position in cell rows

The C. elegans community measures germ cell position in two different ways: (1) physical distance along the axis of the germline (see [69] for an example of measuring distance in microns), or (2) counting cell rows to the distal end of the germline (see [68] for an example of measuring distance in cell rows). Thus, we designed Parismi to also compute cellular spatial position in two ways- “geodesic distance” and “cell row distance”.

In order to calculate geodesic distance, we fit a principal curve to cell detection points using the algorithm detailed in [115]. We then computed the distance of each cell to the distal end of the gonadal arm along the principal curve. Parismi contains functionality to determine the distal end of the germline automatically based on shape, or the user can manually specify the distal end.
In order to calculate cell row distance, we generated a connectivity map between germ cells based on touching segmentation masks. Consider two cell segmentations with surface areas $A_1$ and $A_2$, and a combined touching area of $A_t$. These two segmentation masks are considered to be touching if $A_t / (A_1 + A_2) > \delta$ where $\delta$ is a tunable parameter; in practice, $\delta = 0.01$ produced reasonable cell row measurements. Then, we then computed the minimal path of a given cell to the distal end of the gonadal arm via Dijkstra’s algorithm [116].

Section 3.10: Parismi cell detection is accurate

A cell detection is considered a true positive if it is within 1.5 µm (approximately one cell radius) of a manually-annotated cell center and there are no other detections closer to the manually annotated center. Otherwise, a cell detection is considered a false positive. A manually-annotated center with no cell detections within a 1.5 µm radius is considered a false negative. Our automatic cell detector contains a tunable threshold $\tau$; for high thresholds, the detector returns only a few detections and naturally achieves high precision (few false positives) at the expense of low recall (many false negatives). To summarize detector performance in a manner independent of $\tau$, we computed precision and recall at all thresholds and report the average precision (AP), the area under the precision-recall curve.

We trained our detector on one experimental dataset composed of thirteen MZ image stacks, then applied our classifier across twelve independent experimental samples composed of worms of different genotypes, stages of development, and feeding or mating treatments (Table 3.1). In these twelve samples, the detector achieved near perfect accuracy in the MZ with average AP = 98.7±1.8%. Visual inspection showed that most errors were associated with condensation of DNA during M phase. We also measured detector performance in seven whole gonadal arm images. Detection performance decreased when evaluated on the whole gonad as opposed to its MZ subset, with average AP = 90.6±9.4% (Table 3.2); this is likely due to wider range of nuclear morphologies in the proximal germ line. The range of nuclear morphologies in the gonadal arms is substantially more varied than typically seen in other tissues or organs. Thus, the relatively high AP over the whole gonadal arm suggests that the automated nuclear detection is robust and generalizes well.

To quantify the amount of training data needed for good detector performance, we also trained the detector on varying sized subsets of the thirteen training images. The detector was trained on each
subset and then evaluated on the test dataset. On average, a detector trained with only a single MZ image achieved average AP = 95.3±4.4%. Average precision quickly reached a plateau, reaching 99.5±5.3% at eight stacks (Figure 3.2C). Altogether, these results demonstrate that our automatic cell detector is remarkably accurate in the MZ, while being robust to different experimental conditions such as genotype, developmental stage, and replicate variability. In addition, training the detector does not require an inordinate amount of labeled training examples: detector performance plateaus at eight training images.

Altogether, these results demonstrate that our automatic cell detector is remarkably accurate in the MZ, while being robust to different experimental conditions such as genotype, developmental stage, and replicate variability. In addition, training the detector does not require an inordinate amount of labeled training examples: detector performance plateaus at eight training images.

Section 3.11: Parismi image segmentation is accurate

In order to quantify segmentation accuracy, we scored the overlap between an automatically-produced segmentation mask and a ground-truth segmentation mask, computed as the ratio of the volume of the intersection of the two specified regions to the volume of their union. This ratio, which has a maximum of 1 when the masks are identical, penalizes segments returned by the algorithm that are too small or too large. To aggregate accuracy over a whole collection of segmented cells, we first computed an optimal one-to-one matching between the machine and human masks that maximized the overlap between matching masks, and then calculated the Jaccard index (hereafter referred to as the average overlap AO) averaged across all matched masks. We hand-constructed segmentations to serve as "ground truth" using Fiji’s “segmentation editor” [94]. Since hand-segmentation of three-dimensional images is an arduous and time-consuming task, we performed this validation on three image stacks consisting of 856 cells.

We evaluated segmentation accuracy comparing our implementation of active contours against the more classical method of marker-controlled watershed [117][118] and a simple baseline method termed “truncated Voronoi” segmentation, which assumes constant radius and non-overlapping cells (see Section 3.18 and 3.19). Marker-controlled watershed (AO = 0.53), truncated Voronoi (AO = 0.61), and active contours (AO = 0.62) performed similarly under ideal conditions with perfectly centered
segmentation seeds and clean membrane images (Figure 3.4, Table 3.3). Since experimental conditions are often less than ideal, we also characterized segmentation accuracy in the presence of a membrane guide image that was artificially degraded to mimic suboptimal staining (see Section 3.20). We found that the segmentation accuracy for marker-controlled watershed decreased drastically when the membrane image was degraded (AO = 0.13, 75.5% relative decrease) while active contours were minimally affected (AO = 0.57, 7.1% relative decrease). Truncated voronoi does not use the membrane signal and hence is unaffected.

To measure the influence of imperfectly localized cell detections, we computed segmentations from marker locations offset by uniform spherical noise of 0.5 µm in radius; this resulted in a 3.3% relative decrease in AO for active contour segmentation and a larger 8.6% relative decrease in AO for truncated Voronoi. This noise level roughly matches the statistics of automatic detections, which had an average distance of 0.5 µm from the “true center” (calculated from manually-constructed segmentations). Similarly, if offset noise was increased to 1 µm, the AO of active contour segmentation decreased by 9.6%, while truncated Voronoi AO decreased by 28.7%, i.e. nearly three times as much.

Altogether, our results demonstrate that marker-controlled watershed is not robust to poor guide image quality and that active contours provide more accurate estimates of cell volume than truncated Voronoi which assumes constant sized cells and doesn’t utilize the guide image (Figure 3.4). Thus, active contours are the most appropriate segmentation method for our images.

Section 3.12: Parismi fluorescence quantification is accurate

We quantified DNA fluorescence intensities in cells from gonads that were pulsed-fixed with the thymidine analog EdU, which helps identify the phase of the replication cycle in which cells reside at the time of the pulse. We tested accuracy of our estimates of DNA content using the facts that (1) cells at the G1 phase of their cycle have not initiated DNA replication and should thus have minimal content, and (2) that cells in the G2 and M phases have finished replication and should thus have maximal content. As expected, DNA content histograms of cells that were EdU-negative, indicating that they were not replicating their DNA at the time of the pulse-fix and were thus in G1 or in G2/M, displayed characteristic peaks at 2C (minimal content) and 4C (maximal content) with a coefficient of variation of ~20% (Figure 3.5B, left). Conversely, cells that were EdU-positive, indicating they were replicating their DNA at the time
of the pulse-fix, had intermediate DNA contents. To further test our procedure, we quantified the DNA content of manually annotated M-phase cells. Single M-phase cells were either annotated as one or two separate cells based on its sub-phase (Figure 3.5A); as expected, DNA content histograms of annotated M-phase cells displayed characteristic peaks at 2C and 4C (Figure 3.5B, right). To rule out spatial bias in DNA quantification, we verified that 2C/4C peaks remained well-separated along the distal-proximal axis of the gonadal arms (Figure 3.5C). Finally, we verified that our classification of cells as EdU-positive or EdU-negative was accurate; we found specificity and sensitivity of 85% and 88%, respectively, using manual annotations as "ground truth" (Table 3.4). Altogether, these results demonstrate that our image analysis pipeline provides accurate fluorescence quantification.

Section 3.13: Parismi cell row position quantification is accurate

We compared measurements of MZ length performed using our cell row counter against manual measurements. The average percent deviation of automatic cell row measurements from manual measurements was 9.4%, and we observed strong correlation between measurements ($R^2 = 0.72$; Figure 3.6). Thus, our cell row counter is accurate.

Section 3.14: Morphological predictors of cell cycle phase exist for C. elegans germ cells

One advantage that image cytometry holds over more conventional methods of single cell analyses is retention of morphological information. As an example application of Parismi, we asked whether DNA morphology changes with phase of the cell cycle. Although M-phase and its sub-phases have characteristic morphologies that can be classified using machine learning approaches [119][120], the other phases G1, S, and G2 superficially appear similar to one another. Previous studies have suggested that cell images carry information enabling computer analysis to distinguish between cell types [121], and in particular that a relationship exists between chromatin texture and cell cycle phase [122] [123].

Our observations made using mosaics of segmented and classified cells suggested that the DNA of G1 phase cells tends to have a punctate morphology with approximately 5-6 puncta per cell; S phase DNA has a smooth morphology without readily visible puncta; and G2 phase DNA has a punctate morphology with a variable number of puncta (See Figure 3.7 for representative images, Supplemental
Data 3.1 for full mosaic of all classified cells). In addition, the area covered by DNA in G2 phase cells appeared larger than that in S-phase cells, which in turn appeared larger than that in G1-phase cells. In order to place these observations on a more quantitative footing, we chose a 2D slice in the middle of each cell, thresholded the DNA image from each of those slices using Otsu’s method [124], and measured the total number of connected components (i.e., the number of spots) in segmented DNA as well as the area (i.e., the spatial extent) of the segmented DNA. We found that the average number of spots in G1 or G2 phase nuclei is larger than that observed in S-phase nuclei (p < 1e-12, Bonferroni corrected rank-sum test; see Figure 3.8C, Table 3.5, Table 3.6) and that the average spatial extent of DNA fluorescence in G1 phase nuclei is smaller than that in S-phase nuclei, which in turn is smaller than that in G2-phase nuclei (p < 2e-14, Bonferroni corrected rank-sum test; see Figure 3.8B, Table 3.7, Table 3.8).

Having established that morphological differences exist between the nuclei of G1-, S-, and G2-phase cells, we asked whether these differences could provide a basis for cell phase classification. We cropped 2D cell segmentations to a maximum size of 34x34 pixels and extracted the following features:

1. Number of connected components of the thresholded foreground mask. For each segmented cell, we generated a foreground mask via Otsu thresholding of the DNA channel. Then, we counted the number of connected components via Matlab’s `regionprops` command.

2. Number of pixels composing the thresholded foreground mask. For each segmented cell, we generated a foreground mask via Otsu thresholding of the DNA channel. Then, we counted the number of pixels inside the foreground mask.

3. Center-surround Haar-like features. Haar-like features are simple convolution masks that can be used to detect puncta [125]. Let \( r = (\alpha, \beta, u, v) \) parameterize a rectangle such that \( r \) is composed of all points bounded by \( \alpha \leq x \leq \alpha + u \) and \( \beta \leq y \leq \beta + v \). Then a Haar-like feature can be specified by two rectangles \( r_1, r_2 \) where \( r_1 \) encompasses \( r_2 \). In our application of Haar-like features, we used “center-surround” features such that \( r_1, r_2 \) are squares and \( r_2 \) is positioned in the center of \( r_1 \). In addition, we set the maximum size of \( r_1 \) to be eight pixels wide; altogether, there are 5,239 possible Haar-like features that are eight or less pixels wide in a 34x34 image frame. The image response to a Haar-like feature is the average pixel value within \( r_2 \) minus the average pixel value within \( r_1 \) not in \( r_2 \).
We used these features to train SVM classifiers (see Section 3.21); mean sensitivity and specificity exceeded 0.66 for all classifiers (Table 3.9).

Thus, segmentation of individual cells shows that DNA morphology carries substantial information about cell cycle phase. A crucial advantage of using this morphology information is that it is acquired as a matter of course in most imaging experiments, and does not require fluorescent transgene expression or live imaging that facilitate cell cycle phase identification [126][127][128] but limit the kind of tissue that can be imaged, the strains that can be used, and the number of imaging channels that are available for readouts unrelated to the cell cycle. Future work will focus on improving classifier performance, using an extended set of features and more powerful classification techniques.

Section 3.15: Cell cycle pauses upon starvation in the C. elegans germ line

As a second application of Parismi, we asked how C. elegans germ cells respond to worm starvation, which is expected to occur frequently in the wild [129]. Although the germ line is known to undergo dramatic cell death or regeneration upon changes in nutritional status [130], and larval germ cells are known to arrest in G2 in starved larvae [131][132], the kinetics of cell cycle response to food removal remain uncharacterized in adults. To ask whether cells stop at a particular point of the cell cycle, we tracked cell cycle progression of labeled and unlabelled cells in germ lines pulsed with EdU and chased over a five-day starvation period. We observed little to no cell cycle progression (Figure 3.9) in two independent experimental repeats that included a total of 20,022 MZ cells in 73 gonadal arms. This unexpected result suggests that there are a large number of points along their cycle at which cells can pause in response to food removal, at least in adult nematodes.

Section 3.16: Parismi can be applied to other model systems

Parismi is designed to analyze C. elegans germline images but can be applied to other model systems. To ask how Parismi generalizes to other systems, we turned to datasets from cultured human cells (from the HeLa line), early mouse embryos, and the mouse olfactory placode. First, we asked whether Parismi’s active contour implementation generalizes to images from these systems, whose appearance is substantially different from that of C. elegans gonads. We found that with some adjustments to the image pre-processing steps applied prior to running active contours (see Section
3.22), cells from these three sources can be segmented reasonably well (Figure 3.10). We also evaluated automatic detection accuracy in early mouse embryos. We found detection performance to be comparable to that for worm germ cells; the detector’s AP was 0.968, and the precision-recall curve shows that, with an appropriate threshold, we detect more than 80% of cells without a single false positive (Figure 3.11).

We next asked whether we could derive new insights in regulation of cell differentiation in pre-implantation mouse embryos. We quantified expression of Cdx2 and Nanog, two antagonistic regulators involved in the early differentiation decision made in the embryo that guides the establishment of an embryonic stem cell-like population [133][134]. Despite being antagonistic, these factors are co-expressed early during embryonic development ([134]; Figure 3.12A–C); such paradoxical early co-expression is also true of a number of other antagonistic gene groups that mediate cell fate decisions, and models have been proposed to account for initial co-expression or concomitant upregulation of antagonistic factors [135][136][137]. To further assess the relationship between Cdx2 and Nanog in different cell sub-populations in the pre-implantation mouse embryo, we used a recently published dataset in which we had annotated cells as being either on the “interior” of the embryo or on the “periphery” of the embryo and quantified the levels of Cdx2 and Nanog in each subpopulation [138]. As expected, cells on the periphery of the embryo had higher Cdx2 content and lower Nanog content than cells on the interior (Figure 3.12D). Interestingly, we observed that despite the Nanog/Cdx2 antagonism both inner and outer cell populations show a significant positive correlation between Cdx2 and Nanog expression levels at the ~29 cell stage (Figure 3.12C). Surprisingly, however, the positive correlation was removed specifically in the outer cells upon chemical inhibition of BMP signaling, which we recently showed to be active in early embryonic development (Figure 3.12E; [138]). These results suggest that BMP signaling may play a context-dependent role in the regulatory interactions between Nanog and Cdx2 or their upstream controls; this intriguing context-dependence would have been obscured had we evaluated only average expression levels across all cell sub-populations. These findings further emphasize the utility of segmentation methods in understanding complex regulatory networks that underlie cell differentiation.

Finally, we asked whether Parismi could allow us to make new findings on cell cycle behavior during embryonic development, using mouse olfactory placode as a model system. The olfactory placode is a thickened region of head ectoderm that invaginates into developing head mesenchyme to form the
olfactory mucosa, a highly-branched mucosa whose epithelial lining contains the primary sensory neurons that subserve the sense of smell [139]. Questions remain about the forces that drive the early phases of invagination in this and other ectodermal placodes of the head. A number of possible mechanisms have been proposed for such morphogenetic changes, including local modulation of cell proliferation rates [140]. We used Parismi to ask whether we could detect different cell cycle behavior in sub-regions of olfactory placodes, which we called center and outer “rings” of each placode based on known patterns of gene expression and cell differentiation state [141][139]. We processed and imaged placodes from the developing heads of E9.5 mouse embryos, and categorized cells as G1, S, G2 (based on DNA content and EdU content), or M-phase (using manual annotations based on DNA morphology). S-phase index was higher in the outer rings than in the center rings of these placodes (p < 0.0062, categorical chi-square; Figure 3.13). There was also a significant change of overall cell cycle phase distribution (p = 0.037, categorical chi-square test). The increased S-phase index of outer ring placode cells is particularly interesting because this is the region in which Sox2 and Fgf8 expressing stem cells of the early olfactory epithelium are found in highest number [141], suggesting that proliferation of these stem cells may be particularly important for driving early morphogenesis of the olfactory epithelium.

Section 3.17: Conclusions

Although many image segmentation software packages are available exist, none are readily applicable to the *C. elegans* germline. In this chapter, we describe the development of Parismi, a image segmentation tool specialized for germline studies. We use Parismi to derive two novel results about the germline: we demonstrate that (1) sub-phases of interphase can be classified morphologically, and (2) cells within the adult germline can stop at multiple points of the cell cycle (possibly spanning all of G1, S, and G2) in response to starvation. In addition, Parismi was used to derive results described in Chapter 4. In order to demonstrate broader application of Parismi, we successfully adapted the software tool a number of different sample types— HeLa cell culture, early mouse embryo, and mouse olfactory epithelium— showing that our software makes it possible to answer different kinds of biological questions in a variety of different model systems. We hope that Parismi is of use to the *C. elegans* germline researchers as well as the wider scientific community.
Supplemental Text

Section 3.18: Truncated Voronoi

For a set of cell detections \( p = \{ p_1, p_2, ..., p_n \} \), consider the following: (1) The Voronoi diagram \( V \) consisting of Voronoi polygons so that \( V = \{ v_1, v_2, ..., v_n \} \), and (2) the set of points \( c_i \) within radius \( r \) of point \( p_i \) so that \( C = \{ c_1, c_2, ..., c_n \} \). We computed the truncated Voronoi segmentation \( S \) as segmentation masks \( s_i \) that consist of the intersection of points between \( v_i \) and \( c_i \)

\[ S = \{ s_1, s_2, ..., s_n \} , s_i = v_i \cap c_i \]

Note that the result of truncated Voronoi only depends on input cell detections, and not on the preprocessed guide image.

Section 3.19: Marker-controlled watershed

A detailed description of marker-controlled watershed can be found in [118]. In brief, let \( I \) be the guide image used to define segmentation boundaries and let \( p \) be a set of cell detections. We imposed local minima at \( I(p) \) via morphological reconstruction. Then, we ran a watershed transform on \( I \) in order to generate segmentation masks. We thresholded segmentation masks based on size in order to remove gross mis-segmentations.

Section 3.20: Artificial degradation of preprocessed guide images

A degraded preprocessed guide image was generated for segmentation benchmarking using the following procedure:

1. Let \( I \) denote the original preprocessed membrane image. An image kernel \( K \) was generated by cropping a 108 x 108 x 40 section of \( I \).
2. The kernel was inverted and thresholded so that no pixel values fell above 0.85. Then, the kernel was scaled so that all pixel values fall between \([0,1]\).
3. The kernel was tiled to form a new image the same size as \( I \).
4. The degraded preprocessed image was generated by multiplying \( I \) and \( K \).

Section 3.21: Training SVMs for cell phase classification
We trained and ran SVMs using Matlab’s `svmtrain` and `svmclassify` commands. We trained SVMs using the following parameters:

1. Maximum number of iterations = 150,000
2. Tolerance = 1e-7 for the G1 classifier, 1e-8 for the S classifier, 1e-8 for the G2 classifier, 1e-8 for the M classifier.
3. Box constraint = 1e-5 for the G1 classifier, 1e-3 for the S classifier, 1e-2 for the G2 classifier, 1e-3 for the M classifier.

When training and running SVMs, we split datasets randomly into equally-sized, non-intersecting training and testing subsets. In order to generate classifier statistics, we repeatedly resampled training and testing subsets.

Section 3.22: Adapting Parismi to other model systems

The following changes were made when adapting Parismi to model systems other than the C. elegans germline (mouse olfactory epithelium, early mouse embryo, and HeLa cell culture):

- Cell centers in images of the mouse OE, mouse embryo, and HeLa cells were curated manually.
- Preprocessing of mouse pre-implantation embryo guide images. We used the DNA channel as a guide image for segmentation. We preprocessed the images as follows. First, we thresholded the DNA image \( D \) using an adaptive algorithm. Let \( p = \{x,y,z\} \) be a given cell detection point, and let \( \{p_i\} \) be a set of cell detections. For a small window around a given \( p_i \), we calculated the mean DNA pixel value \( m(x_i, y_i, z_i) \). Then, we fit coefficients \( c_1, c_2, c_3, c_4 \) to the model \( m(x, y, z) = \exp(c_1 z) (c_2 x + c_3 y + c_4) \). We calculated an adaptive threshold \( t(x, y, z) \):

\[
t(x, y, z) = ke^{c_1 z} (c_2 x + c_3 y + c_4)
\]

and applied this threshold to the DNA image. We chose \( k \) heuristically; \( k = 1/3 \) worked well in practice.

Second, we removed high frequency noise from the thresholded image through median filtering. Third, we inverted the image. Finally, we removed sharp discontinuities via blurring, yielding the final preprocessed guide image for active contours.
- Preprocessing of HeLa cell images. We pre-processed HeLa cell DNA images in the same fashion as mouse embryo images, with the difference that we thresholded using Otsu’s method [124].
• Quantification of DNA content in the mouse olfactory epithelium. We corrected for fluorescence fluctuation along the long, medial-lateral axis of the OE (which corresponded to the x-axis in our images) by fitting a second order polynomial to the 90% percentile pixel intensity in each x-slice, then normalizing against this polynomial. We corrected for z-attenuation by fitting a first order polynomial to the 90% percentile pixel intensity in the middle 25 z-slices of each image, then normalizing against this polynomial. We used the middle 25 z-slices because cells were evenly distributed in this region. We measured DNA content for each cell by summing the normalized DNA fluorescence intensity inside the segmentation mask. DNA content was normalized based on M-phase cell annotations.

• Quantification of fluorescence content in mouse pre-implantation images. Fluorescence levels were computed by taking the sum of pixel intensities within a given segmentation mask and dividing by the volume of the mask.
Figure 3.1: Parismi is a pipeline customized to segment three-dimensional confocal images of the C. elegans germline

Parismi is composed of four components. (A) Preprocessing a membrane image to create a guide image using a principal curvature approach. (B) Detecting cell centers using HOG features. (C) Segmenting cellular volumes by running active contours on the preprocessed guide image initialized from cell detections. (D) Quantifying cellular position and fluorescence content using cell segmentation masks.
Figure 3.2: Parismi cell detection is implemented using HOG features

(A) HOG weights learned from germ cells. The left two panels display the component of the weight vector learned from xy sections, while the right two panels correspond to xz sections. (B) Variation of training samples. The left panel corresponds to the 20 highest-scoring training samples and the right panel corresponds to the 20 lowest-scoring training samples. (C) Eight curated MZs are sufficient for accurate training of the cell detector.
Figure 3.3: Parismi uses a top-layeredness metric to identify top layer cells

(A) "Top-layeredness" $\theta$ is measured as the fraction of surface area that is visible looking down the z-axis.

(B) Gonadal arm images where all (red) and top-layer only (purple) cells have been segmented. In practice, $\theta > 0.1$ is a suitable threshold for top-layer cells.
Figure 3.4: Benchmarking active contours, marker controlled watershed, and truncated voronoi segmentation

(A) Active contours, marker controlled watershed, and truncated Voronoi all perform well with a high quality guide image and accurately localized cell centers. (B) Marker controlled watershed performs poorly when guide image is low quality. (C) Active contours are more robust to detection noise than truncated Voronoi. (D) Overlay of guide image (white signal) with hand-constructed segmentation (green, top row) or automatic segmentation (red, bottom row). Average overlap is 74%. Seven z-slices are shown, which cover the cell visible in the center of each slice.
Figure 3.5: Parismi fluorescence quantification is accurate (1/2)

(A—B) Distinct G1 (blue) and G2 (red) peaks are observed as expected in DNA content histograms of EdU-negative cells for gonadal arms labeled without a chase at L4, both in the region used for DEMD fits (panel A) and in pairs of rows all along the MZ distal-proximal axis (panel B); rows were paired for visualization purposes. The y axis of histograms shows cell frequency. Overlap between red and blue histograms is due to thresholds being chosen on a row-by-row basis, and some thresholds falling in-between bin boundaries. (C—D) DNA contents of M-phase cells show two peaks of DNA content at 2C and 4C (color based on same thresholding as in panels A, B); this is as expected, because mitotic cells are segmented either as one cell with 4C content (see panel I), or as two separate cells with 2C content if the cell has sufficiently progressed through cytokinesis (see panel J). (E—H) Same as panels A—D for the L4+1 stage.

(I—J) Comparison of DNA quantification results derived using “top layer” or “stack top” metrics (see Section 3.5) to select cells in which fluorescence signal attenuation is minimal. The top layer metric leads to better-defined and sharper G1 and G2/M DNA content peaks for EdU-negative cells (I; graph based on 48 MZs) and M-phase cells (J; graph based on 157 MZs). (K—L) Example segmentations (green outline) of a mitotic cell before cytokinesis has initiated (panel K; arrow), and of a mitotic cell in which cytokinesis is completing, leading to segmentation of two nuclei (panel L; arrows).
Figure 3.5: Parisimi fluorescence quantification is accurate (2/2)
Figure 3.6: Parismi cell row counter is accurate

(A) Overlay of microscope image (white signal, derived from the DNA stain DAPI) with segmentations color-coded by cell row position from the distal end (left), as computed automatically using our counter.

(B) Size of the MZ scored manually vs size computed through the automatic counter. A small amount of noise (0.5 cell rows) was added in order to aid visualization of overlapping data. Positions scored manually and automatically are in close agreement; average percent deviation is 9.4%. Diagonal shown for reference in red.
Figure 3.7: G1, S, G2, and M-phase germ cells can be separated using DNA morphology

Cell cycle phase was assayed by quantifying EdU and DNA content in EdU pulse experiments using Parismi. See Supplemental Data 3.1 for the full mosaic of all classified cells. (A) G1-phase cells are small and have 5-6 puncta. (B) S-phase cells are large and have a smooth morphology. (C) G2-phase cells are large and have a punctate morphology. (D) M-phase cells are blobby.
Figure 3.8: G1, S, G2, and M-phase germ cells have different spatial extents and number of connected components

(A) Thresholding of raw cell images via Otsu's method to partition DNA signal into foreground and background pixels. (B) Spatial extent of DNA fluorescence, assayed by counting the number of foreground pixels in thresholded images. On average, G1 is smaller than S is smaller than G2. (C) Spottiness of DNA morphology, assayed by counting the number of connected components in thresholded images. On average, G1 and G2 phase cells have more connected components than S phase cells, which have more connected components than M phase cells. See Tables 3.5, 3.6 for statistical comparisons.
Figure 3.9: The *C. elegans* germline stops cycling immediately after starvation

(A) DNA content histograms of EdU- and EdU+ cells remain constant over the course of 5 days when adults are starved, indicating that little to no cell cycle progression occurred from the onset of starvation.

(B) For comparison, there is clear cell cycle progression over 5 hours in well-fed worms.
Figure 3.10: Parismi has applications to other model systems

(A) Example 3D segmentations of a *C. elegans* gonadal arm. (B) A mouse early embryo. (C) Cultured HeLa cells. (D) Mouse olfactory epithelium. Red, DNA channel; green, segmentation mask boundary.
Figure 3.11: Precision-recall curves benchmarking cell center detection

(A-B) Precision-recall tradeoff as the detection threshold is varied for C. elegans MZ cells (A) and mouse embryo cells (B). Based on these curves, an automatic detection threshold can be chosen that yields high precision (e.g. $P = 0.98$); automatic detection using this threshold can be followed by manual curation using Parismi's annotation tool to add cells missed by the detector.
Figure 3.12: Relationship between Nanog and Cdx2 contents in preimplantation mouse embryos

(A-C) Mouse embryo stained for DNA (A) Cdx2 (B) and Nanog (C; white arrows point to the same two cells in each panel). (D) Control embryos cultured for 12 h (28.5±6.8 cells per embryo, n = 8) show a positive relationship between Nanog and Cdx2 (interior linear fit slope 0.13, 95% CI = [0.08, 0.18]; periphery linear fit slope 0.11, 95% CI = [0.05, 0.17]). (E) Embryos cultured for 24 h with BMP signaling inhibitor LDN (29.4±10.2 cells per embryo, n = 12) do not show a positive relationship between Nanog and Cdx2 content (interior linear fit slope 0.07, 95% CI = [0.03, 0.10]; periphery linear fit slope 0.01, 95% CI = [-0.03, 0.05]).
Figure 3.13: Differences in cell phase indices along the medial-lateral axis of the mouse olfactory epithelium

(A) Left: A transverse vibratome section of an E9.5 embryo including the two olfactory placodes (yellow box). Right: High magnification image of the placodes stained for DNA (cyan) and EdU (green). Placodes were divided into “center” and “outer ring” regions for analysis. (B) DNA content histograms of EdU+ (red) and EdU− (blue) cells in center and outer ring regions. (C) During cell detection, M−phase cells in the process of cytokinesis were segmented as two separate cells and were annotated as being a "half cell". Conversely, M−phase cells before cytokinesis were annotated as being a “full cell”. (D) DNA content histogram of "half" (magenta) and "full" (green) cells. Histograms are appropriately centered at 2C and 4C.
Table 3.1: Average precision of cell detection over a variety of experimental conditions

Overall average precision across twelve replicates is 98.7±1.8%. Unless otherwise stated, worms were not mated and were fixed at L4+1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average precision (AP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fog-2 starved 5 days</td>
<td>1.0000</td>
</tr>
<tr>
<td>N2 #1</td>
<td>1.0000</td>
</tr>
<tr>
<td>cye-1/+</td>
<td>0.9997</td>
</tr>
<tr>
<td>fog-2</td>
<td>0.9962</td>
</tr>
<tr>
<td>fog-2 mated</td>
<td>0.9956</td>
</tr>
<tr>
<td>fog-1</td>
<td>0.9939</td>
</tr>
<tr>
<td>N2 L4+3 #1</td>
<td>0.9897</td>
</tr>
<tr>
<td>L4 N2</td>
<td>0.9878</td>
</tr>
<tr>
<td>N2 L4+3 #2</td>
<td>0.9875</td>
</tr>
<tr>
<td>N2 #2</td>
<td>0.9869</td>
</tr>
<tr>
<td>gld-1::gfp</td>
<td>0.9654</td>
</tr>
<tr>
<td>inx-22; fog-2</td>
<td>0.9391</td>
</tr>
</tbody>
</table>
Table 3.2: Average precision of cell detection over the whole gonad

Overall average precision is 90.6±9.4%.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Average precision (AP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2 gonad #12</td>
<td>0.9822</td>
</tr>
<tr>
<td>N2 gonad #8</td>
<td>0.9670</td>
</tr>
<tr>
<td>N2 gonad #1</td>
<td>0.9394</td>
</tr>
<tr>
<td>N2 gonad #9</td>
<td>0.9392</td>
</tr>
<tr>
<td>N2 gonad #11</td>
<td>0.9271</td>
</tr>
<tr>
<td>N2 gonad #17</td>
<td>0.8812</td>
</tr>
<tr>
<td>N2 gonad #1</td>
<td>0.7056</td>
</tr>
</tbody>
</table>
Table 3.3: Average overlap of cell segmentations across different experimental conditions

Active contours performs better than marker-controlled watershed when membrane signal is poor quality, and active contours performs better than truncated Voronoi when cell detections are not perfectly centered.

<table>
<thead>
<tr>
<th>Segmentation method</th>
<th>Guide image quality</th>
<th>Noise (µm)</th>
<th>Average overlap (AO)</th>
<th>AO Confidence interval (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active contour</td>
<td>High</td>
<td>0</td>
<td>0.62</td>
<td>0.61 - 0.62</td>
</tr>
<tr>
<td>Watershed</td>
<td>High</td>
<td>0</td>
<td>0.53</td>
<td>0.52 - 0.54</td>
</tr>
<tr>
<td>Voronoi</td>
<td>High</td>
<td>0</td>
<td>0.61</td>
<td>0.61 - 0.62</td>
</tr>
<tr>
<td>Active contour</td>
<td>Low</td>
<td>0</td>
<td>0.57</td>
<td>0.57 - 0.58</td>
</tr>
<tr>
<td>Watershed</td>
<td>Low</td>
<td>0</td>
<td>0.13</td>
<td>0.12 - 0.14</td>
</tr>
<tr>
<td>Voronoi</td>
<td>Low</td>
<td>0</td>
<td>0.61</td>
<td>0.61 - 0.62</td>
</tr>
<tr>
<td>Active contour</td>
<td>High</td>
<td>1</td>
<td>0.60</td>
<td>0.59 - 0.60</td>
</tr>
<tr>
<td>Watershed</td>
<td>High</td>
<td>1</td>
<td>0.48</td>
<td>0.47 - 0.50</td>
</tr>
<tr>
<td>Voronoi</td>
<td>High</td>
<td>1</td>
<td>0.57</td>
<td>0.57 - 0.58</td>
</tr>
<tr>
<td>Active contour</td>
<td>High</td>
<td>1</td>
<td>0.51</td>
<td>0.50 - 0.52</td>
</tr>
<tr>
<td>Watershed</td>
<td>High</td>
<td>1</td>
<td>0.33</td>
<td>0.31 - 0.34</td>
</tr>
<tr>
<td>Voronoi</td>
<td>High</td>
<td>1</td>
<td>0.45</td>
<td>0.44 - 0.45</td>
</tr>
</tbody>
</table>
Table 3.4: Benchmarking EdU quantification accuracy

Samples 1, 2, and 3 are representative of the range of EdU staining quality observed in all samples.

<table>
<thead>
<tr>
<th>MZ sample</th>
<th>True positives</th>
<th>True negatives</th>
<th>False positives</th>
<th>False negatives</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Processed cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>76</td>
<td>21</td>
<td>9</td>
<td>2</td>
<td>0.974</td>
<td>0.700</td>
<td>108</td>
</tr>
<tr>
<td>2</td>
<td>66</td>
<td>47</td>
<td>6</td>
<td>20</td>
<td>0.767</td>
<td>0.887</td>
<td>139</td>
</tr>
<tr>
<td>3</td>
<td>68</td>
<td>20</td>
<td>1</td>
<td>7</td>
<td>0.907</td>
<td>0.952</td>
<td>96</td>
</tr>
<tr>
<td>Aggregate</td>
<td>210</td>
<td>88</td>
<td>16</td>
<td>29</td>
<td>0.879</td>
<td>0.846</td>
<td>343</td>
</tr>
</tbody>
</table>
Table 3.5: Number of connected components in G1, S, G2, M phase cells

The number of connected components is calculated by thresholding DNA fluorescence signal and counting the number of connected components.

<table>
<thead>
<tr>
<th></th>
<th>Number of components</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>3.54</td>
<td>1.00-7.20</td>
</tr>
<tr>
<td>S</td>
<td>2.71</td>
<td>1.00-7.00</td>
</tr>
<tr>
<td>G2</td>
<td>3.67</td>
<td>1.00-8.00</td>
</tr>
<tr>
<td>M</td>
<td>1.90</td>
<td>1.00-5.00</td>
</tr>
</tbody>
</table>
Table 3.6: Statistics for number of connected components

Table shows pairwise rank sum p-values with Bonferroni correction.

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>S</th>
<th>G2</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>4.00E+00</td>
<td>9.70E-13</td>
<td>1.90E+00</td>
<td>4.40E-47</td>
</tr>
<tr>
<td>S</td>
<td>-</td>
<td>4.00E+00</td>
<td>1.80E-33</td>
<td>1.60E-51</td>
</tr>
<tr>
<td>G2</td>
<td>-</td>
<td>-</td>
<td>4.00E+00</td>
<td>7.60E-100</td>
</tr>
<tr>
<td>M</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.00E+00</td>
</tr>
</tbody>
</table>
Table 3.7: Spatial extent of G1, S, G2, and M phase cells

The spatial extent is calculated by thresholding DNA fluorescence signal and computing the area of the foreground mask.

<table>
<thead>
<tr>
<th></th>
<th>Spatial extent of DNA signal ((\mu m^2))</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>3.31</td>
<td>2.14 - 4.83</td>
</tr>
<tr>
<td>S</td>
<td>3.92</td>
<td>2.61 - 5.30</td>
</tr>
<tr>
<td>G2</td>
<td>4.17</td>
<td>2.87 - 5.70</td>
</tr>
<tr>
<td>M</td>
<td>3.96</td>
<td>2.10 - 6.20</td>
</tr>
</tbody>
</table>
Table 3.8: Statistics for spatial extent of G1, S, G2, and M-phase cells

Table shows pairwise rank sum p-values with Bonferroni correction.

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>S</th>
<th>G2</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>4.00E+00</td>
<td>3.10E-35</td>
<td>3.50E-48</td>
<td>1.10E-20</td>
</tr>
<tr>
<td>S</td>
<td>-</td>
<td>4.00E+00</td>
<td>1.40E-14</td>
<td>3.90E+00</td>
</tr>
<tr>
<td>G2</td>
<td>-</td>
<td>-</td>
<td>4.00E+00</td>
<td>2.40E-07</td>
</tr>
<tr>
<td>M</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4E+00</td>
</tr>
</tbody>
</table>
Table 3.9: Sensitivity and specificity of cell phase classification

All binary classifiers have sensitivity and specificity above 66%.

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Sensitivity 95% CI</th>
<th>Specificity</th>
<th>Specificity 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G1</strong></td>
<td>0.66</td>
<td>0.57 - 0.75</td>
<td>0.67</td>
<td>0.61 - 0.72</td>
</tr>
<tr>
<td><strong>S</strong></td>
<td>0.73</td>
<td>0.69 - 0.76</td>
<td>0.67</td>
<td>0.63 - 0.70</td>
</tr>
<tr>
<td><strong>G2</strong></td>
<td>0.76</td>
<td>0.70 - 0.81</td>
<td>0.74</td>
<td>0.71 - 0.76</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>0.71</td>
<td>0.68 - 0.75</td>
<td>0.84</td>
<td>0.82 - 0.87</td>
</tr>
</tbody>
</table>
Supplemental Data 3.1: Mosaic of G1, S, G2, and M-phase DNA morphologies (1/5)

M-phase cells and interphase cells (G1, S, G2) were classified based on manual annotations. Interphase cells were sub-classified into G1, S, G2 fractions based on DNA and EdU quantification in EdU pulse experiments. Note that this figure covers five pages.
Supplemental Data 3.1: Mosaic of G1, S, G2, and M-phase DNA morphologies (2/5)
Supplemental Data 3.1: Mosaic of G1, S, G2, and M-phase DNA morphologies (3/5)

S continued
Supplemental Data 3.1: Mosaic of G1, S, G2, and M-phase DNA morphologies (4/5)
Supplemental Data 3.1: Mosaic of G1, S, G2, and M-phase DNA morphologies (5/5)
Chapter 4: Characterizing spatial cell cycle properties
in the C. elegans germline using Parismi

Section 4.1: Contributions

The data presented in this chapter is published in the paper “Control of C. elegans germline stem cell cycling speed meets requirements of design to minimize mutation accumulation” (BMC Biology, 2015).

- Michael Chiang developed computational tools and applied them with help with data curation from Christopher A. Price.
- Dr. Amanda Cinquin performed experiments.

Section 4.2: Introduction

A number of experimental studies have addressed cell cycle properties of stem cells in various contexts. In vertebrates, although stem cells are thought to often reside in a quiescent state, many organs maintain stem cell populations that cycle fast (for example, [142]). Such fast-cycling populations appear to be supported by “reserve” populations that cycle less frequently and that are for example mobilized upon injury [143][144]. Multiple stem cell subpopulations can thus exist in the same organ; since their discovery is often prompted by the use of new markers or combinations of markers, more are likely to be discovered in the future. These multiple subpopulations, whose properties and relative contributions to tissue homeostasis can be difficult to measure over extended periods of time and often generate debate [145][146], make it challenging to test quantitatively whether cell cycle control follows the pedigree depth minimization principle. Such a test is thus best performed in an experimental model system where the contribution of all proliferating cells can be readily assayed.

In Chapter 2, we showed that an approximate two-fold difference in cell cycle length along the distal-proximal axis of the germline minimizes pedigree depth. While it was previously observed that M-phase index differs along the distal-proximal axis of the gonad [69], cell phase indices cannot be used to directly estimate total cell cycle length; cell phase indices only tell us about the relative lengths of G1, S, G2, and M to each other. Measurements of total cell cycle length using continuous BrdU labeling experiments failed to measure any difference along the distal proximal axis [56]; the absence of significant differences are likely a result of the low resolution of continuous labeling experiments. In order
to make more precise measurements of cell cycle properties, we developed Parismi, software that quantifies fluorescence content and spatial position of germ cells in microscopy images of intact gonads. In this chapter, we utilize Parismi to measure cell cycle properties in the germline and compare with theoretical measurements given in Chapter 2; a randomly chosen subset of Parismi germline image segmentations are given in Figure 4.1. We show that the mitotic region is composed of three sub-compartments with different cell cycle properties, and that the spatial profile of cell cycle length across these three sub-compartments is consistent with the theoretical predictions of Chapter 2 at the L4 and L4+1 time points. We further characterize an intermittent cycling phenotype at L4+3, and suggest that intermittent cycling may be an independent strategy to minimizes pedigree depth in an uncertain mating environment. Altogether, our results provide experimental evidence that cell cycle regulation contributes to mutation minimization in the *C. elegans* germline.

Section 4.3: The germline is composed of three sub-compartments with distinct cell cycle properties

To measure cell phase indices we performed pulse-fix labeling using the thymidine analogue EdU which incorporates into cells in S-phase. We first measured the distribution of cell cycle phases in 48 L4+1 gonadal arms containing 12,997 segmented MZ cells. These cells were first annotated as either being in M-phase or in interphase based on DNA morphology, then interphase cells were sub-classified into G1, S, and G2 fractions based on cellular DNA and EdU content derived using Parismi (see Chapter 3 for a detailed description of Parismi). We found that the first eleven cell rows of the germline, containing on average 142 cells, was composed of 8.6% G1, 68.0% S, 20.7% G2, and 2.7% M-phase cells. This is similar to previously reported results in the literature (2% G1, 57% S, 39% G2, 2% M; [70]). We further conducted the same analysis on 27 L4 germlines containing 5,947 cells; we found that the first fifteen cell rows of the germline, containing on average 125 cells, was composed of 9.1% G1, 62.7% S, 23.9% G2, and 4.3% M-phase cells.

Next, we asked whether spatial variation in cell cycle properties is present within the germline. We again calculated cell cycle phase indices, this time binning results along each cell row of the MZ (Figure 4.2). We observed a decrease in G2 index in the first eight cell rows of L4 gonadal arms, and a decrease in G2 index in the first six cell rows of L4+1 gonadal arms. In each case, the index of the first cell row was significantly higher than that of the last cell row (p < 0.01, categorical chi-square test). This
finding is of interest because it shows distinctive cell-cycle behavior of germ cells located at the distal end of the MZ at the G2 cell cycle phase — the length of this phase is known to be actively regulated in other systems [147]. In addition, we measured an increase in G2 index at fifteen cell rows in L4 germlines, and eleven cell rows at L4+1 germlines. This increase in G2 index corresponds with a decrease in mitotic index and most likely corresponds germ cell entry into pre-meiotic interphase [69].

Thus, based on our observations of G2 index, the germline can be divided into two spatial compartments: the distal mitotic zone (DMZ) containing cells that undergo proliferative divisions, and the proximal mitotic zone (PMZ) containing an increasing fraction of cells in pre-meiotic interphase. In addition, the DMZ can be further sub-divided into two sub-compartments: the distal-most mitotic zone (DMMZ) containing a high G2 index and a medial mitotic zone (MMZ) containing a lower G2 index. In L4 gonadal arms, the DMMZ consists of cell rows 1-8, the MMZ consists of cell rows 9-15, and the PMZ is composed of cell rows 16-23 (where 23 is the average length of the MZ at L4). In L4+1 gonadal arms, the DMMZ is composed of cell rows 1-6, the MMZ is composed of cell rows 7-11, and the PMZ is composed of cell rows 12-19 (where 19 is the average length of the MZ at L4+1).

Section 4.4: Cells in the DMMZ cycle more slowly than cells in the MMZ at L4 and L4+1.

Cell cycle length cannot be deduced from cell cycle phase indices alone, since indices only contain information about the relative lengths of sub-phases and not total cell cycle length. To make direct observations about germline cell cycle speed, we performed EdU pulse-chase experiments to observe progression of the labeled and unlabeled fractions through the cell cycle over time [71]. Comparisons of DNA content for EdU-negative and EdU-positive populations in the DMMZ and MMZ regions revealed differences consistent with the MMZ cycling more quickly than the DMMZ in both L4 and L4+1 gonadal arms (Figure 4.3A, 4.3C; p < 0.02 for 9 populations at a total of 5 time points; K-S tests with Bonferroni correction shown in Table 4.1, Table 4.2). To confirm this result we analyzed the data in an independent way by scoring the fraction of labeled mitoses [148]. We also found significant differences compatible with faster cycling of the MMZ (Figure 4.3B, 4.3D; p < 0.02 for 6 populations at 5 time points; categorical Chi-square tests with Bonferroni correction; Table 4.3, Table 4.4). Thus, two independent measurements show that cells in the MMZ cycle more quickly than the DMMZ at L4 and L4+1.
Section 4.5: Cells in the DMMZ cycle ~1.5 times slower than cells in the MMZ

Calculating cell cycle length in the germline based on image cytometry data is non-trivial due to cell movement. During the span of an EdU pulse-chase experiment, cells move from the DMMZ into the MMZ which minimizes the apparent differences between these regions; the DMMZ and MMZ thus cannot be analyzed independently. In order to make quantitative estimates of cell cycle length, we fit simulated data generated using Simworm (see Chapter 2) to experimental data using two independent metrics: the DNA Earth Mover’s Distance metric (DEMD) and the Fraction Labeled Mitoses metric (FLM). Algorithmic details about DEMD and FLM are given in Section 4.12 and 4.13.

We found that the best fit simulation data as determined via DEMD and FLM provides a close fit to experimental data (Figure 4.4), supporting the validity of our fitting procedure. At L4, the distal-most cells in the DMMZ have a cell cycle length of 4.23 hours and the proximal-most cells in the MMZ have a cell cycle length of 2.83 hours. At L4+1, the distal-most cells in the DMMZ have a cell cycle length of 6.71 hours and the proximal-most cells in the MMZ have a cell cycle length of 4.43 hours. Overall, the average ratio of cell cycle speeds between the distal DMMZ and proximal MMZ was 1.50 (95% bootstrapped CI: 1.26 — 1.67) and 1.53 (95% bootstrapped CI: 1.20 — 1.90) at the L4 and L4+1 stages, respectively (Figure 4.5, Table 4.5). Importantly, this result is supported by two independent analysis techniques: one based on the Fraction of Labeled Mitoses (FLM), which has been used before without distinguishing between subpopulations along the distal proximal axis [70], and the new technique we report based on DNA content histograms (DNA Earth Mover’s distance, DEMD) that makes use of all cells instead of only rare M-phase cells.

To ask how cell cycle length is regulated across the distal-proximal axis of the gonad, we computed the estimated distribution of cycle lengths based on our best-fit simulations (Figure 4.4). The length of G2 showed a clear reduction along the distal-proximal axis (71% and 61% decrease between rows 1 and 15 at L4, and rows 1 and 11 at L4 + 1 day, respectively; p < 0.05), while the other phases did not (Table 4.6). We thus conclude that distal-most cells cycle more slowly for the most part because they spend more time in G2.
Section 4.6: The cell cycle profile is spatially flat in PMZ

Cell cycle fits in the PMZ contain one extra degree of complexity; the PMZ contains cells in pre-meiotic interphase that do not divide. Thus, cell phase indices in the PMZ are not accurate readouts of relative lengths of G1, S, G2, and M. To illustrate, there is a sharp increase in G2 index in the PMZ; this sharp increase could be due to an increase in the relative length of G2 in dividing cells, or it could be due to G2 arrest prior to entry into meiotic prophase in pre-meiotic cells, or it could be a combination of the two. We asked whether G2 arrest by itself was sufficient to explain the increased G2 index in the PMZ. Let \( m(r) \) be the experimentally measured mitotic index at row \( r \); we modified our germline simulations so that cells undergoing the G2-M transition would arrest in G2 if the simulated mitotic index in row \( r \) exceeded \( m(r) \). This simple assumption was sufficient to recapitulate experimentally measured cell phase indices presuming the relative lengths of G1, S, G2 and M-phase in proliferative cells stay constant in the PMZ (Figure 4.6).

Next, we made direct measurements of absolute cell cycle length within the PMZ. Since the cell cycle length of an arrested pre-meiotic interphase cell is not well-defined, we focused on PMZ cells that undergo proliferative divisions and thus are captured by the FLM metric. FLM-based cell cycle fits of the PMZ indicate a flat cell cycle profile for cells that have not left the mitotic cycle; the cell cycle length breakpoints in the DMMZ, MMZ, PMZ are 4.26 hours, 2.86 hours, 2.99 hours at L4 and 6.49, 4.43, 4.23 at L4+1, respectively (Figure 4.7, Table 4.7). We note that three-compartment FLM fits across the DMMZ, MMZ, and PMZ are consistent with two-compartment fits across the DMMZ and MMZ only (Section 4.5), lending credence to our fitting algorithm. We also note that the cell cycle profile is essentially flat in the PMZ.

Section 4.7: A Cyclin E gradient exists in the distal MZ that does not depend on cell cycle phase

To begin identifying mechanisms potentially responsible for slower stem cell cycling in the C. elegans germ line, we quantified the spatial expression profile of the cell cycle regulator CYE-1. We focused on this regulator because it is expressed in the MZ and is required for germ cell cycling [149] [150] and because of its intriguing regulation: it is repressed by the proximal, differentiation-promoting factor GLD-1 [151][152], but its transcript is also bound by the repressor FBF-1 [153] that acts to promote the stem cell fate distally.
Nuclear CYE-1 expression was computed by summing pixels in a 0.4µm x 0.4µm x 1µm box centered on the nucleus and follows a biphasic gradient within the MZ at L4 + 1 day, with a peak at row 9 (Figure 4.8A, 4.8B). A gradient of CYE-1 thus spans the distal MZ (rows 1—11), in which we showed that a cell cycle gradient exists. The difference between the DMMZ and MMZ is modest (11%) but statistically significant (p < 1.0E-14; Wilcoxon rank sum test). Average nuclear CYE-1 levels thus correlate positively with cell cycle speed.

Since in most cell types CYE-1 levels oscillate with cell cycle phase, we asked whether lower CYE-1 levels in distal-most cells could be explained by their longer G2 phase. We first ascertained whether in the C. elegans germline CYE-1 expression levels oscillate with cell cycle phase. We quantified CYE-1 contents in rows 1—11, and found that cells at the beginning of the cycle indeed express moderately higher CYE-1 (see Figure 4.8C-E), but that this cell cycle phase dependence of CYE-1 levels is for the most part contributed by the MMZ and not the DMMZ (Figure 4.8F-G). We next asked whether the CYE-1 gradient we observed along the distal-proximal axis was predominantly contributed by cells at a specific phase of the cycle, but found no difference in overall CYE-1 profiles when considering only cells at the beginning or at the end of the cycle as defined by DNA content (Figure 4.8H-I). We conclude that CYE-1 expression levels are regulated in a way that is partly independent of cell cycle phase. Although more direct evidence awaits further study, this is consistent with CYE-1 playing a causative role in changes in cell cycle length along the distal-proximal axis.

Section 4.8: L4+3 gonadal arms display variability cell cycle properties

We attempted to characterize cell cycle properties in L4+3 gonadal arms using the same image cytometry procedure applied to L4 and L4+1 germlines. We found that DNA content histograms and fraction labeled mitoses histograms equilibrate over time (Figure 4.9); at later time points, there is little temporal change in fraction labeled mitoses histograms or DNA content histograms, and the DNA content histograms of EdU-positive and EdU-negative cells display high similarity. In order to place these qualitative observations on more quantitative footing, we developed “clock plots” as a method to quantify cell cycle variability on the gonad-by-gonad level (Figure 4.10, see Section 4.14).

We applied clock plot fitting to EdU pulse chase data at L4, L4+1, and L4+3 (Figure 4.10). As expected, DNA content histograms at L4 and L4+1 displayed minimal germline-level variability; germlines
were essentially identical to one another with respect to cell cycle properties at these time points. However, gonadal arms at L4+3 displayed high variability, indicating that cell cycle properties differed significantly between germlines in older worms.

Section 4.9: Germline-level cell cycle variability in L4+3 gonadal arms is due to intermittent cycling

There are two possible causes of germline-level variability in cell cycle properties: (1) each individual germline cycles continuously at a different speed from one another, or (2) each germline cycles intermittently, and start/stop times differ between germlines. In order to distinguish between these two possibilities, we performed continuous EdU labeling experiments. If all germlines cycle continuously at different rates, then germlines will quickly display some amount of EdU labeling. On the other hand, if germlines cycle intermittently and at any given time some fraction of germlines are arrested, then some fraction of germlines will not display EdU labeling for extended periods of time.

Continuous EdU labeling experiments show that all L4+1 germlines contain some amount of EdU labeling by two hours (Figure 4.12A). On the other hand, a sizable fraction of L4+3 germlines remain unlabeled at three hours. Thus, L4+3 clock plot heterogeneity is most likely due to intermittent cycling, and not continuously cycling germlines with different rates of cell cycle. In addition, the time that cells remain arrested during intermittent cycling is approximately three hours.

Section 4.10: Intermittent cycling in L4+3 germlines is caused by sperm depletion

Progeny production in selfed worms drops to minimal levels at L4+3; this decrease is due to sperm availability and not some fundamental limit in reproductive lifespan [154]. We asked whether the intermittent cycling phenotype we observed in older worms was caused by the absence of sperm. We performed EdU pulse-chase experiments in mated and selfed L4+3 germlines, and performed the same clock plot analysis described in previous section. We found that if L4+3 worms are pre-mated between L4 and L4+1, then clock plot analysis shows germlines cycle continuously at L4+3 (Figure 4.13A).

Next, we asked whether an intermittent cycling could be induced via sperm depletion. To address this question, we turned to feminized hermaphrodites — worms that are turned into females through mutation of genes such as fog-1 and fog-2. [155][156][157]. These females do not produce self-sperm, but form an otherwise fully-developed reproductive system in which oocyte maturation and growth by
cytoplasmic streaming are substantially reduced [158][83]. Feminized worms can bear progeny only after mating with males, whose sperm trigger oocyte maturation and fertilization. In conjunction with continuous EdU-labeling experiments (Figure 4.12B), clock plot analyses on fog-1 and fog-2 worms indicate that feminized germlines display the same intermittent cycling phenotype as old germlines (Figure 4.13B). In addition, mating fog-2 worms induces continual germline cycling (Figure 4.13C).

Finally, we asked whether oocyte fertilization is required for continuous cycling. Knockdown of spe-8 preserves stimulation of oocyte maturation by self-sperm that are incapable of fertilizing the oocytes [159] [160]. We found that spe-8 mutants cycle continuously (Figure 4.12B, Figure 4.13C), indicating that oocyte fertilization is not required for continuous cell cycling. We further asked whether oocyte maturation in the absence of sperm is sufficient to induce continuous cycling. Knockdown of inx-22 results in precocious oocyte maturation in feminized gonads even in the absence of the sperm signal [158][83][161], although the oocyte maturation rate is lower than that of wild-type. We found that inx-22; fog-2 mutants cycle intermittently (Figure 4.12B, Figure 4.13B) and thus the low rate of oocyte maturation found in inx-22; fog-2 is insufficient to induce continuous cycling.

Section 4.11: Conclusion

Using Parismi, we have shown that the mitotic zone of the germline can be partitioned into three sub-compartments with distinct cell cycle properties: the DMMZ, the MMZ, and the PMZ. In addition, we have shown that a ~1.5-fold difference in cell cycle speed exists between the DMMZ and the MMZ at L4 and L4+1; this is qualitatively consistent with theoretical predictions that a ~2-fold difference in cell cycle speed will minimize pedigree depth. Finally, we show that intermittent cycling occurs at L4+3, and is due to depletion of sperm. Although this intermittent cycling phenotype is not predicted by our computational model, it does not contradict the theoretical findings presented in Chapter 2. The presence of intermittent cycling is equivalent to introducing random “pauses” into Simworm, and has no effect on average pedigree depth in selfed worms.

What sort of pedigree depth reduction can we expect from slow cycling stem cells, and is this reduction evolutionarily selectable? Without a cell cycle gradient, our most biologically realistic germline simulation (optimization 11, Table 2.3) produces progeny with an average pedigree depth of 10.26. Plugging in our experimentally measured cell cycle gradient produces progeny with an average pedigree depth
depth of 9.99; this is reasonably close to the absolute minimum of 9.85 (optimization 11, Table 2.3). The advantage afforded by a cell cycle gradient is thus a 2.6% decrease in pedigree depth (and presumably mutational load) compared to that of a flat profile.

Is a 2.6% decrease in mutation rate selectable by evolution? Natural selection dominates over genetic drift if |s| > 1/(2N_e), where s is the selection coefficient and N_e is the effective population size. For selfing species, the selection coefficient for a trait that changes the deleterious mutation rate by ΔU is s = -ΔU/2 [23][89][19]. Putting everything together, we expect a 2.6% decrease in mutation rate to be selected for if:

$$0.026 \times U \times N_e > 1$$

The effective population size N_e is estimated to be anywhere from 9,600 [162] to 90,000 [163]. Estimates for the genomic deleterious mutation rate U are similarly variable, ranging a full order of magnitude from 0.03 [3] to 0.48 [164]. Nevertheless, even the most conservative estimates of deleterious mutation rate U = 0.03 and effective population size N_e = 9,600 yields:

$$0.026 \times 0.03 \times 9600 = 7.48 > 1$$

Thus, the ~1.5-fold difference in cell cycle speed we measured in the germline is expected to be selectable through evolution.
Supplemental Text

Section 4.12: DNA Earth Mover’s Distance (DEMD) — a metric for fitting DNA content histograms for EdU labeling experiments

Consider a series of EdU pulse-chase experiments across $T$ different chase times. Suppose we quantify DNA content, EdU content, and spatial compartment $C$ for each individual germ cell in our EdU pulse-chase experiments. It is then straightforward to generate a set of $T \times C \times 2$ DNA content histograms, where cells are partitioned based on chase time $T$, spatial position $C$, and EdU content (labeled or unlabeled). Define DEMD histograms as the set of histograms

$$g = \{g_1, ..., g_{T \times C \times 2}\}, \ h = \{h_1, ..., h_{T \times C \times 2}\}$$

Consider two sets of DEMD histograms $g$ and $h$. Define the EMD distance $d_{DEMD}$ between $g$ and $h$:

$$d_{DEMD}(g, h) = \sum_i n(g_i) n(h_i) d_{CEMD}(g_i, h_i)$$

where $n(.)$ gives the number of cells in a histogram and $d_{CEMD}$ is the Circular Earth Mover’s Distance [165]. Now, suppose $g$ is drawn from experimental data and $h(v)$ is drawn from simulations with cell cycle profile $v$. The goal of DEMD-based cell cycle fits is to perform the following minimization: $v_{DEMD} = \arg \min_v d_{DEMD}(g, h(v))$

Section 4.13: Fraction Labeled Mitoses (FLM) — a metric for fitting the fraction of labeled mitoses in EdU labeling experiments

Consider a series of EdU pulse-chase experiments across $T$ different chase times. Suppose we quantify cell phase, EdU content, and spatial compartment $C$ for each individual germ cell in our EdU pulse-chase experiments. It is then straightforward to generate a $T \times C$ matrix that records the percentage of M-phase cells at chase time $T$ and spatial position $C$ that are EdU-positive. Define this $T \times C$ matrix as the FLM matrix $p = \{e_t, c\}$. Consider two FLM matrices $g$ and $h$. Define the FLM distance $d_{FLM}$ between $g$ and $h$:

$$d_{FLM}(g, h) = \sum_i n(g_i) n(h_i) (g_i - h_i)^2$$

where $n(.)$ gives the total number of M-phase cells used to compute the percentage. Now, suppose that $g$ is drawn from experimental data and $h(v)$ is drawn from simulations with cell cycle profile $v$. The goal of FLM-based cell cycle fits is to perform the following minimization:

$$v_{FLM} = \arg \min_v d_{FLM}(g, h(v))$$
In our cell cycle studies, we found the optimal profiles $v_{FLM}$ and $v_{DEMD}$ via a grid search implemented in Matlab. Confidence intervals on $v_{DEMD}$ and $v_{FLM}$ were calculated via bootstrapping [166]. We performed bootstrapping in a way that each sample maintained the same number of gonadal arms at each chase time. Specifically, suppose we use an experimental dataset $z$ composed of $N$ gonadal arms. Suppose $z$ is partitioned into $T$ subgroups based on chase time:

$$z = \{z_1, z_2, ..., z_T\}, \quad z_i = \{g_1, g_2, ..., g_{n(i)}\},$$

where $g_i$ is a gonadal arm and where $n(1) + n(2) + ... + n(T) = N$. A bootstrap distribution for $v_{EMD}$, $v_{FLM}$ was derived by resampling each $z_i$ independently and re-running the grid search minimization.

Section 4.14: Clock plots — a method to quantify germline-level variability in cell cycle properties

We simulated germlines with uniform cell cycle lengths using Simworm (see Chapter 2), and recorded the simulated DNA content histograms at 0%, 5%, 10%, ..., 90%, and 95% cell cycle completion. Then, for each DNA content histogram corresponding to each individual gonadal arm, we found the best fitting simulated DNA content histogram and associated cell cycle percent completion using the DEMD metric (see Section 4.12). Percent completion was plotted on the unit circle for each germline, generating a spread of germline progression across successive chase times. A visual representation of clock plot fitting is given in Figure 4.10. In order to quantify spread, we plotted the resultant vector after adding up individual germline percent completions; if percent completions are evenly spread out across the unit circle, then the resultant vector has a magnitude of zero. Likewise, if percent completions all fall within the same bin on the unit circle, then the resultant vector has a magnitude of one. Thus, a short resultant vector length indicates high germline-level variability in cell cycle properties, and vice versa.

In order to validate our clock plot fitting procedure, we plotted the best, median, and worst fitting germline DNA content histogram at representative percent completion for our L4+3 dataset (Figure 4.11); visually, the clock plot fits appear reasonable.
Figure 4.1: Randomly chosen segmentations used in cell cycle fits

50 randomly chosen segmentations used in cell cycle fits. Green, DNA channel; red, segmentation mask boundary.
Figure 4.2: Spatial cell cycle phase indices at L4 and L4+1

Blue, G1 index; green, S-phase index; red, G2-index; cyan, mitotic index. 95% confidence intervals (thin lines) computed via bootstrapping. (A) Cell cycle phase indices at L4. The mitotic region (23 cell rows) can be partitioned into three distinct sub-compartments: the distal-most mitotic zone (DMMZ) consisting of cell rows 1-8 and decreasing G2 index, the medial mitotic zone (MMZ) consisting of cell rows 9-15 and flat G2 index, and the proximal mitotic zone (PMZ) consisting of cell rows 16-23 and increasing G2 index. (B) Cell cycle phase indices at L4+1. The mitotic region (23 cell rows) can be partitioned into three distinct sub-compartments: the distal-most mitotic zone (DMMZ) consisting of cell rows 1-6 and decreasing G2 index, the medial mitotic zone (MMZ) consisting of cell rows 7-11 and flat G2 index, and the proximal mitotic zone (PMZ) consisting of cell rows 12-19 and increasing G2 index.
Figure 4.3: EdU pulse-chase experiments reveal that cells in the DMMZ cycle more slowly than those in the MMZ

(A, C) DNA content histograms of EdU-positive (blue) and EdU-negative (red) cells in the DMMZ and MMZ at L4 and L4+1, respectively. It is visually apparent that cells in the MMZ cycle more quickly than those of the DMMZ; more EdU-positive cells have divided and entered G1 in the MMZ than in the DMMZ (black arrows). (B, D) Fraction labeled mitoses histograms in the DMMZ and MMZ at L4 and L4+1, respectively. It is visually apparent that cells in the MMZ cycle more quickly than those of the DMMZ; at the one hour time point (L4) and the two hour time point (L4+1), more mitotic cells have EdU label in the MMZ than in the DMMZ.
Figure 4.4: Best cell cycle fits of EdU pulse-chase data using DEMD and FLM metrics.

(A, B) DEMD fit of DNA content histograms of EdU-negative and EdU-positive cells, respectively at L4. Black, simulated data; red and blue, experimental data. (C) FLM fit of fraction labeled mitoses histograms at L4. (D, E) DEMD fit of DNA content histograms of EdU-negative and EdU-positive cells, respectively at L4+1. Black, simulated data; red and blue, experimental data. (F) FLM fit of fraction labeled mitoses histograms at L4.
Figure 4.5: Cell cycle fits of EdU pulse-chase data using DEMD and FLM metrics

(A, B) FLM (blue) and DEMD (green) fits at L4 and L4+1. Each dot represents one bootstrap run; a small amount of uniform rectangular noise (0.5 hours) was added to aid visualization of bootstrap runs. 95% confidence ellipses given. (C, D) Relative contributions of G1, S, G2, and M-phase to total cell cycle length.
Figure 4.6: Simple assumptions about pre-meiotic interphase recapitulates cell phase measurements

The following assumptions were made: proliferative cell phase indices stay constant in the PMZ, cells enter pre-meiotic interphase to satisfy experimentally measured M-phase index, and that the cell cycle properties of pre-meiotic and proliferative cells are identical save for G2 arrest. (A) Assumed cell phase indices of proliferative cells where cell phase indices stay flat in the PMZ. (B) Simulated combined (pre-meiotic + proliferative) cell phase indices (dotted line) versus experimentally measured combined cell phase indices (solid line). There is good fit between simulated and experimentally measured combined cell phase indices.

![Graph A](image1)

![Graph B](image2)
Figure 4.7: FLM histogram data over the entire mitotic zone

Fraction labeled mitoses histograms in the distal-most mitotic zone (DMMZ), medial mitotic zone (MMZ), and proximal mitotic zone (PMZ). Note that fraction labeled mitoses histograms only include proliferative cells, while DNA content histograms include both proliferative cells and pre-meiotic cells. Cell cycle fit results are given in Table 4.7. (A) Fraction labeled mitoses histograms at L4. (B) Fraction labeled mitoses histograms at L4+1.
Figure 4.8: Cyclin E levels are graded across the DMMZ and MMZ, and are differentially dependent on cell cycle phase in the DMMZ and MMZ (1/2)

(A) Example of CYE-1 staining pattern in a gonadal arm at L4+1 (color-coded using ImageJ’s “Fire” lookup table). CYE-1 levels appear to start low in the distal region, rise, and then fall in the proximal region. (B) Quantification of nuclear CYE-1 levels using 7,508 cells segmented from 30 gonadal arms. Each dot represents a cell; the red line is the average at each cell row, with a 95% bootstrapped confidence interval. (C—D) Cells with typical G1 morphology (arrows in C) have higher CYE-1 content than their neighbors (D; arrows point to same G1 cells as in C). (E) Scatterplot of nuclear CYE-1 content vs DNA content, showing that cells with lower DNA content — i.e. early in the cell cycle — have moderately higher levels of CYE-1 than cells with higher DNA content. Density colored via “jet” lookup table (red: high density, blue: low density). (F—G) Variation of CYE-1 content with cell cycle phase is lesser for cells in the DMMZ (F; virtually flat trend line) than in the MMZ (G; steeper trend line). The difference between DMMZ and MMZ is statistically significant (95% bootstrapped CI for difference in slopes of first component of trend lines: 0.024—0.38, n = 50,000 replicates). Arrows show two clusters at low and high DNA content. (H—I) Quantification of nuclear CYE-1 profile as in (A), but considering only cells with low (H) or high (I) DNA content.
Figure 4.8: Cyclin E levels are graded across the DMMZ and MMZ, and are differentially dependent on cell cycle phase in the DMMZ and MMZ (2/2)
Figure 4.9: EdU pulse-chase data at L4+3

EdU pulse-chase data “stabilizes” over time at L4+3 in a way consistent with intermittent cycling. (A, B, C) EdU pulse chase data at L4, L4+1, and L4+3, respectively. L4 and L4+1 data is given for comparison purposes. Top, DNA content histograms of EdU-positive (blue) and EdU-negative (red) cells; middle, fraction labeled mitoses histogram; bottom, clock plot. Description of clock plot is given in Figure 4.10.
Figure 4.10: Description of clock plot fitting procedure

DNA content histograms from individual germlines are scored for progression through the cell cycle. Progression through the cell cycle is plotted on the unit circle (with 0% progression at twelve o'clock and clockwise increasing progression). The resultant vector gives a measure of synchrony of cell cycle progression over a collection of germlines. If all germlines cycle continuously at the same speed, then the resultant vector has a length of one. If there is variability in cell cycle progression, then the resultant vector will have a length less than one.
Figure 4.11: Benchmarking clock plot fitting procedure

The best fit (minimum EMD), average fit (median EMD), and worst fit (maximum EMD) DNA content histograms of EdU-positive (blue) and EdU-negative (red) cell compared to simulated data (top) at 0%, 25%, 50%, and 75% completion of the cell cycle.
Figure 4.12: Continuous EdU labeling experiments

(A) All L4+1 germlines have some amount of EdU-labeling by two hours. Conversely, a high fraction of L4+3 germlines remain unlabeled at three hours. (B) fog-1, fog-2, and to a lesser extent inx-22; fog-2 worms have a sizable fraction of unlabeled germlines at two hours. Conversely, spe-8 and N2 worms have zero unlabeled germlines at two hours. All worms in panel B are young adult worms (L4+1).
Figure 4.13: Clock plots across a range of genotypes, mating conditions, and ages

Unless otherwise specified, worms are L4+1, selfed, and wild-type. Mating was performed in the interval between L4 and L4+1. (A) L4+3 selfed (blue) and L4+3 mated (red). (B) fog-2 (blue), fog-1 (red), inx-22; fog-2 (green). (C) fog-2 (blue), fog-2 mated (red), spe-8 (green). Note that spe-8 data only covers 0, 3, and 6 hour time points.
Table 4.1: KS test for DNA content histograms at L4

P-values for Kolmogorov-Smirnov (KS) tests of difference in DNA content histograms for cells in DMMZ and MMZ at L4 (Bonferroni correction applied); see Figure 4.3A.

<table>
<thead>
<tr>
<th>Chase time</th>
<th>0h</th>
<th>1h</th>
<th>2h</th>
<th>3h</th>
<th>4h</th>
<th>5h</th>
<th>6h</th>
<th>8h</th>
</tr>
</thead>
<tbody>
<tr>
<td>L4 EdU-negative</td>
<td>2.39E+00</td>
<td>8.25E+00</td>
<td>4.96E-06</td>
<td>8.53E-03</td>
<td>2.61E-05</td>
<td>1.23E+00</td>
<td>4.54E-08</td>
<td>9.43E-03</td>
</tr>
<tr>
<td>L4 EdU-positive</td>
<td>5.19E+00</td>
<td>6.15E+00</td>
<td>5.17E-09</td>
<td>7.01E+00</td>
<td>1.69E+00</td>
<td>1.19E+01</td>
<td>2.56E-08</td>
<td>1.18E+00</td>
</tr>
</tbody>
</table>
Table 4.2: KS test for DNA content histograms at L4+1

P-values for KS tests of difference in DNA content histograms for cells in DMMZ and MMZ at L4+1 (Bonferroni correction applied); see Figure 4.3C.

<table>
<thead>
<tr>
<th>Chase time</th>
<th>0h</th>
<th>2h</th>
<th>3h</th>
<th>4h</th>
<th>5h</th>
<th>6h</th>
<th>8h</th>
</tr>
</thead>
<tbody>
<tr>
<td>L4+1 EdU-negative</td>
<td>8.5E+00</td>
<td>8.8E-01</td>
<td>4.7E-05</td>
<td>8.2E-02</td>
<td>2.0E+00</td>
<td>6.3E-02</td>
<td>1.1E+01</td>
</tr>
<tr>
<td>L4+1 EdU-positive</td>
<td>1.9E+00</td>
<td>2.2E-01</td>
<td>4.5E-12</td>
<td>1.1E-02</td>
<td>3.3E+00</td>
<td>3.9E+00</td>
<td>3.9E+00</td>
</tr>
</tbody>
</table>
Table 4.3: Chi-square test for fraction labeled mitoses histograms at L4

P-values for categorical Chi-square tests of difference in FLMs for cells in DMMZ and MMZ at L4 (Bonferroni correction applied); see Figure 4.3B.

<table>
<thead>
<tr>
<th>Chase</th>
<th>0h</th>
<th>1h</th>
<th>2h</th>
<th>3h</th>
<th>4h</th>
<th>5h</th>
<th>6h</th>
<th>8h</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>n/a</td>
<td>3.8E-03</td>
<td>n/a</td>
<td>6.6E-03</td>
<td>6.4E-01</td>
<td>5.0E-01</td>
<td>1.6E+00</td>
<td>2.0E-05</td>
</tr>
</tbody>
</table>
Table 4.4: Chi-square test for fraction labeled mitoses histograms at L4+1

P-values for categorical Chi-square tests of difference in FLMs for cells in DMMZ and MMZ at L4+1 (Bonferroni correction applied); see Figure 4.3D.

<table>
<thead>
<tr>
<th>Chase</th>
<th>0h</th>
<th>2h</th>
<th>3h</th>
<th>4h</th>
<th>5h</th>
<th>6h</th>
<th>8h</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>2.5E+00</td>
<td>2.6E-03</td>
<td>2.1E+00</td>
<td>3.8E+00</td>
<td>1.4E-02</td>
<td>2.0E-01</td>
<td>3.3E-04</td>
</tr>
</tbody>
</table>
Table 4.5: Results of two-compartment cell cycle fit

Fits were conducted over the DMMZ and MMZ. The cell cycle profile was parameterized linearly. Both DEMD and FLM cell cycle fits show that $\alpha \approx 1.5$. Each DEMD and FLM bootstrap run was weighted equally when calculating the aggregate fitting metric.

<table>
<thead>
<tr>
<th>Fitting metric</th>
<th>Stage</th>
<th>DMMZ cell cycle length (h)</th>
<th>DMMZ cell cycle length CI</th>
<th>MMZ cell cycle length (h)</th>
<th>MMZ cell cycle length CI</th>
<th>$\alpha$</th>
<th>$\alpha$ CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEMD</td>
<td>L4</td>
<td>4.30</td>
<td>4.07 - 4.62</td>
<td>2.76</td>
<td>2.28 - 2.83</td>
<td>1.56</td>
<td>1.44 - 2.03</td>
</tr>
<tr>
<td>FLM</td>
<td>L4</td>
<td>4.16</td>
<td>3.66 - 4.62</td>
<td>2.91</td>
<td>2.62 - 3.10</td>
<td>1.44</td>
<td>1.13 - 1.62</td>
</tr>
<tr>
<td>Aggregate</td>
<td>L4</td>
<td>4.23</td>
<td>4.07 - 4.62</td>
<td>2.83</td>
<td>2.55 - 3.10</td>
<td>1.50</td>
<td>1.26 - 1.67</td>
</tr>
<tr>
<td>DEMD</td>
<td>L4+1</td>
<td>6.82</td>
<td>6.10 - 7.14</td>
<td>4.45</td>
<td>4.03 - 4.90</td>
<td>1.55</td>
<td>1.25 - 1.73</td>
</tr>
<tr>
<td>FLM</td>
<td>L4+1</td>
<td>6.60</td>
<td>6.10 - 7.66</td>
<td>4.42</td>
<td>3.86 - 5.07</td>
<td>1.50</td>
<td>1.20 - 1.90</td>
</tr>
<tr>
<td>Aggregate</td>
<td>L4+1</td>
<td>6.71</td>
<td>6.10 - 7.66</td>
<td>4.43</td>
<td>3.86 - 5.07</td>
<td>1.53</td>
<td>1.20 - 1.90</td>
</tr>
</tbody>
</table>
Table 4.6: Lengths of G1, S, G2, M-phase at start of DMMZ and end of MMZ

Confidence intervals for length of G1, S, G2, M-phase in Figure 4.5B,C were calculated via bootstrapping.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Stage</th>
<th>cell cycle length row 1 (h)</th>
<th>CI (row 1)</th>
<th>cell cycle length row 15 or 11 (h)</th>
<th>CI (row 15 or 11)</th>
<th>Percent change</th>
<th>Overlapping credible intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>L4</td>
<td>0.33</td>
<td>0.13 - 0.59</td>
<td>0.16</td>
<td>0.07 - 0.26</td>
<td>0.53</td>
<td>no</td>
</tr>
<tr>
<td>G1</td>
<td>L4+1</td>
<td>0.19</td>
<td>0.08 - 0.31</td>
<td>0.27</td>
<td>0.15 - 0.40</td>
<td>-0.41</td>
<td>no</td>
</tr>
<tr>
<td>S</td>
<td>L4</td>
<td>1.43</td>
<td>0.86 - 2.04</td>
<td>1.83</td>
<td>1.52 - 2.14</td>
<td>-0.28</td>
<td>no</td>
</tr>
<tr>
<td>S</td>
<td>L4+1</td>
<td>3.90</td>
<td>3.30 - 4.53</td>
<td>3.04</td>
<td>2.74 - 3.33</td>
<td>0.22</td>
<td>no</td>
</tr>
<tr>
<td>G2</td>
<td>L4</td>
<td>2.24</td>
<td>1.65 - 2.82</td>
<td>0.65</td>
<td>0.35 - 0.96</td>
<td>0.71</td>
<td>yes</td>
</tr>
<tr>
<td>G2</td>
<td>L4+1</td>
<td>2.44</td>
<td>1.79 - 3.01</td>
<td>0.96</td>
<td>0.71 - 1.21</td>
<td>0.61</td>
<td>yes</td>
</tr>
<tr>
<td>M</td>
<td>L4</td>
<td>0.23</td>
<td>0.12 - 0.34</td>
<td>0.19</td>
<td>0.15 - 0.24</td>
<td>0.16</td>
<td>no</td>
</tr>
<tr>
<td>M</td>
<td>L4+1</td>
<td>0.18</td>
<td>0.09 - 0.28</td>
<td>0.17</td>
<td>0.13 - 0.21</td>
<td>0.06</td>
<td>no</td>
</tr>
</tbody>
</table>
Table 4.7: Results of three-compartment cell cycle fit

Fits were conducted over the DMMZ, MMZ, and PMZ. Only the FLM fitting metric was used because the PMZ contains pre-meiotic cells, complicating DEMD analysis.

<table>
<thead>
<tr>
<th>Fitting metric</th>
<th>Stage</th>
<th>DMMZ cell cycle length (h)</th>
<th>DMMZ cell cycle length CI</th>
<th>MMZ cell cycle length (h)</th>
<th>MMZ cell cycle length CI</th>
<th>PMZ cell cycle length (h)</th>
<th>PMZ cell cycle length CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLM</td>
<td>L4</td>
<td>4.26</td>
<td>3.68 - 4.32</td>
<td>2.86</td>
<td>2.84 - 3.05</td>
<td>2.99</td>
<td>2.21 - 3.16</td>
</tr>
<tr>
<td>FLM</td>
<td>L4+1</td>
<td>6.49</td>
<td>6.16 - 6.95</td>
<td>4.43</td>
<td>4.05 - 5.11</td>
<td>4.23</td>
<td>3.26 - 5.37</td>
</tr>
</tbody>
</table>
Chapter 5: Conclusion

Section 5.1: General conclusions

While the link between cell divisions and mutation accumulation has been long recognized [50][167], current studies have not addressed whether previously proposed strategies to reduce number of cell divisions [33] are actually implemented in real stem cell systems. The work presented in this thesis describes the development of tools (Simworm and Parismi) that enable detailed and high-throughput cell cycle studies of the C. elegans germline, and quantitatively tests whether cell cycle control follows a pedigree depth minimization principle. We find that computational modeling predicts that an approximate two-fold difference in cell cycle length minimizes pedigree depth while still maintaining fast growth and reproduction. Using Parismi, we test whether a cell cycle gradient is found within the germline and find a 1.5-fold difference in distal/proximal cell cycle length exists, consistent with theoretical measurements. This experimentally measured profile is selectable through evolution (Section 4.11). We further show that there exists a gradient of Cyclin E expression that correlates with cell cycle speed, and also observe an intermittent cycling phenotype at L4+3 not predicted by simulations.

Section 5.2: Intermittent cycling and its relationship to pedigree depth

We observed that germlines cycle intermittently at L4+3. This finding is not predicted by computational modeling although it does not contradict theoretical predictions either; intermittent cycling is equivalent to adding random pauses to computational simulations and has no effect on pedigree depth. How does this finding fit within the broader framework of pedigree depth minimization? There are two naturally-occurring C. elegans sexes: males and hermaphrodites. Hermaphrodites can either self-fertilize with the ~300 stored self-sperm they produce during development, or be cross-fertilized with male sperm transferred during mating that allows brood sizes of up to 1,200 [84]. Since males are rare in the wild [162], we speculate that intermittent cycling is a bet-hedging strategy to reduce pedigree depth while maintaining the ability to reproduce quickly if and when a mating event occurs.

Bet-hedging is a well-established strategy followed by unicellular and multicellular organisms to avoid or spread risks in the face of uncertain environmental conditions [168][169], and phenotypic heterogeneity is a mechanism by which bet hedging can be implemented [170]. We speculate that under unfavorable mating conditions C. elegans populations hedge their bets by maintaining a population of
individuals that are primed for reproduction at the cost of faster senescence, and a population of individuals whose gonads are dormant and thus increment pedigree depth more slowly. Stochasticity in the behavior of individual gonads would then have two roles. First, it would allow individuals to modulate the average rate of germ cell cycling. Second, and more importantly, stochasticity is a parsimonious mechanism to develop a broad distribution of effective reproductive senescence rates in the population without those rates being pre-assigned to each individual.

Section 5.3: Minimizing pedigree depth may increase the quality of maternal factors in oocyte development

Population genetics provides a formalism to relate mutation accumulation to fitness [89], and thus it is altogether natural to frame strategies for pedigree depth minimization with respect to mutation rate. What other effects does pedigree depth have on fitness besides reducing mutation accumulation? The germline forms a syncytium and cytoplasmic streaming is known to transport RNAs and proteins made by pachytene nuclei to growing oocytes [83]. It is reasonable to suggest that pachytene nuclei that have undergone a lower number of total divisions produce higher quality maternal factors. Suppose for each progeny production event we average the pedigree depths of all meiotic zone cells, then average over all progeny. Then, without a cell cycle gradient, our most biologically realistic germline simulation (optimization 11, Table 2.3) has an average meiotic zone pedigree depth of 16.9 during progeny production. On the other hand, plugging in our experimentally measured cell cycle gradient reduces average meiotic zone pedigree depth to 15.2, a 10.1% decrease. This is substantially higher than the 2.6% decrease measured for pedigree depth of progeny (Section 4.10), and suggests that pedigree depth minimization has other additional effects in addition to reducing mutation accumulation in gametes.

Section 5.4: Minimizing pedigree depth may increase reproductive lifespan

An important goal of aging research is not just to extend lifespan — which in C. elegans can be simply achieved by a pause in developmental and reproductive activities in the “dauer” state [171]— but to do so in a way that increases “healthspan” without diminishing organ activity. To this end, it is critical to understand whether aging is driven by organ activity or whether it is a function of chronological age. In addition to reducing mutation rate, pedigree depth minimization strategies may play an important role in
extending the healthspan of the organ. In context to the germline, healthspan is analogous to “reproductive lifespan” which is the period of adulthood over which C. elegans hermaphrodites can bear progeny [172].

Previous studies report that aging individuals lose “reproductive capacity” — the maximum brood size an individual is capable of producing from a given point in time until cessation of reproduction — as a function of chronological age rather than reproductive activity [154]. However, in our own unpublished data, we observed that cell cycle arrest by hydroxyurea treatment or starvation delays reproductive aging. This indicates that reducing pedigree depth may be beneficial to organ healthspan, and also suggests a role for intermittent cycling in reducing the amount of needless cell cycling prior to a random mating event.

Section 5.5: Potential alternative explanations for slow stem cell cycling

There exist alternative explanations for the presence of a cell cycle length gradient besides pedigree depth minimization. For example, changes in cell cycle speed could be a side-effect of cells progressing through differentiation, or could even be part of the mechanism that promotes differentiation [173]. But the change commonly observed in the course of differentiation is a lengthening of the cell cycle, in contrast to the shortening of the cell cycle that we observed in C. elegans germ cells initiating differentiation.

It is also possible that a slower cell cycle allows for more efficient DNA repair, a lower DNA replication error rate, or lower metabolic demands on the cell that minimize production of DNA-damaging free radical species. Indeed, such slower cycling could be a requirement for the lower stem cell mutation rate posited in some models [55]. Data are lacking to use these ideas to extract quantitative predictions on the relationship between the extent of cell cycle lengthening and a reduction in mutation rate. We showed that our quantitative predictions of cell cycle length ratios were largely unchanged by the additional assumption that mutation rate is inversely proportional to cell cycle length, and that the pedigree depth quasi-minimization strategy is still effective at further reducing mutation accumulation. Since our quantitative predictions match experimental data closely, the pedigree depth quasi-minimization strategy is a strong candidate to explain how the speed of stem cell cycling was tuned by evolution.
Section 5.6: Other strategies to minimize mutation accumulation

We note that there are a number of strategies other than cell cycle control to minimize mutation accumulation. Another potential strategy is asymmetric segregation of “immortal” strands of DNA by stem cells [33]. By retaining the unreplicated DNA strands at each division, stem cells could segregate replication errors to their differentiating descendants and thus suppress the accumulation of mutations in the stem cell compartment. This strategy has been proposed to apply in different contexts to all chromosomes [174], some chromosomes [175], or not at all [49]. How does the pedigree depth quasi-minimization strategy interact with the immortal strand strategy, which does not rely on control of cell cycle length? Our results show that if this strategy were followed by the *C. elegans* germ line the cell cycle length profiles should be very different from those we observed experimentally: stem cells, which would not accumulate mutations, should cycle quickly (see also [55]). For organs that rely on a large pool of stem cells, if an immortal strand strategy applies slow cycling of cells at the top of the lineage hierarchy would be beneficial as the stem cell pool expands during development [33]; but once the stem cell compartment is fully developed stem cells would cycle quickly.

An independent strategy to minimize the accumulation of mutations, whether they were incurred from errors in DNA replication or not, is for cells that accrued mutations to senesce [176] or undergo apoptosis [177][178]. In the *C. elegans* germline, extensive apoptosis occurs in older adults. While this apoptosis could be explained by the elimination of nurse cells [82] or the need to reduce competition between developing germ cells [179], it appears that apoptosis could preferentially eliminate damaged cells in certain contexts [180][181]. This idea could be further explored in the future with tools to estimate the mutational load in populations of cells before and after they have been purged of apoptotic cells.

Section 5.7: Control of cell cycle length to minimize pedigree depth

Our study identified two cell cycle phases that show substantial variation in their duration. S phase is shorter during larval development than in adulthood, and G2 is longer in distal cells than in proximal cells both during larval development and in adulthood. Lengthening of G2 in preference to other cell cycle phases is consistent with mutation minimization, as replicated chromosomes offer the possibility of error-free damage repair with homologous recombination using the sister chromatid [182]. Regulation of G2 length has been reported in other contexts [147]. Why S phase would be lengthened as well as G2
when germ lines transition to the adult stage is less clear. We speculate that longer S phase could be less error prone because it allows more time for error-free repair before trans-lesion synthesis occurs [183]; S phase could be shorter during larval development because the benefits of faster development outweigh the costs of decreased DNA replication fidelity — which would be consistent with our findings and those of [18].

What role does Cyclin E1 play in control of cell cycle length? Our data contribute two new observations that expand understanding of that role. First, we extend previous reports that Cyclin E1 is expressed throughout the cell cycle [150][70][151][184] by showing with finer quantification that Cyclin E1 expression levels do change with cell cycle phase (albeit in a dampened manner compared to other cell types). Interestingly, a similar finding has been made in mammalian embryonic stem cells using APC activity as a readout [185], extending earlier reports highlighting the lack of robust oscillations of cell cycle regulators in these cells [186]. Second, and more importantly, we show that Cyclin E1 levels are graded along the distal-proximal axis of the *C. elegans* germ line in a way that is not solely dependent on changes in the lengths of cell cycle phases. This suggests that CYE-1 could play an upstream role in controlling overall cell cycle length, which is also compatible with the complex regulation of Cyclin E by the mitosis promoting factors FBF-1/2 and the meiosis-promoting factor GLD-1.

A role of CYE-1 in regulating cell cycle length along the distal-proximal axis could appear at first sight surprising: Cyclin E is better known for its role in driving G1 progression [187], but a minimal fraction of cells are in G1 along the distal proximal axis — even in the very proximal MZ, where Cyclin E1 levels drop significantly — and it is G2 whose length is modulated along that axis. A role of CYE-1 in regulating the length of G2 is possible given that Cdk2 is known to play a role in progression through S phase and to M phase [187]. This Cdk2 role is thought to normally rely on complex formation with Cyclin A2 [187], but continued expression of Cyclin E1 past G1 in cycling MZ cells could allow activity of a Cyclin E/Cdk2 complex past G1. Although in the C. *elegans* germline CYE-1 is the cell cycle regulator whose interplay with differentiation regulators is best documented [70][151][184], B-type cyclins could also play an important role in control of cell cycle length as they are also potential targets of both FBF-1/2 and GLD-1 [188][153][189]. Overall, it appears that there is a complex interplay between the cell cycle machinery and regulators of differentiation. The design principle highlighted in this study provides one potential reason for the need for fine cell cycle control as cells proceed through differentiation.
Section 5.8: Competition assays as an alternative methods to assay fitness

Our computational model of the germline assumes that it is desirable to minimize the mean number of mutations accumulated by progeny. This metric is desirable for two reasons: first, it is common in the mutation accumulation literature to consider the mean number of mutations per generation [22] and mean decrease in fitness per generation [3]. Second, it provides a straight-forward and non-contrived method of linking mutation rate to fitness [23][89][19]. However, it may be important to consider the distribution of mutation number when calculating population fitness. Consider a hypothetical germline that produces 175 progeny with zero pedigree depth and a single progeny with a very large pedigree depth. Then, the average pedigree depth over all progeny is then arbitrarily large; however, a single progeny with a very large pedigree depth most likely has a negligible effect on population fitness due to the presence of its siblings. Thus, the average fitness of individuals is not the same as the fitness of the population. Even in the case of individual progeny, the fitness effects resulting from new mutations does not follow a simple linear relationship, and the distribution of fitness effects is known to be complex and possibly multi-modal [190].

Can we find a more appropriate metric for assaying fitness? One way to approach this problem is to derive a more accurate quantitative metric that accounts for population fitness rather than the average fitness across individuals. Such a metric requires detailed knowledge of the natural habitat of *C. elegans* outside the confines of the petri dish. While efforts have been made to understand the ecology of *C. elegans* [129], there is still much to be learned about its life history and it is infeasible to create a biologically realistic fitness metric for a *C. elegans* population found in nature.

A second possibility is to approach the problem using an experimental approach. Rather than defining a possibly contrived fitness metric computationally, we can directly compare the fitness of cell cycle mutants and wild-type worms using competition assays. The biggest challenge in conducting this experiment is generating cell cycle mutants that have different cell cycle gradients in the mitotic zone of the germline. We have shown that a gradient of Cyclin E expression in the mitotic zone correlates with cell cycle speed; it is reasonable to hypothesize that differential regulation of Cyclin E can generate different cell cycle gradients. How do we approach this experimentally? While RNAi experiments are easy to do in the germline, the effect of knocking down cell cycle regulators is usually arrest rather than generating a different gradient shapes. Many genes are regulated post-transcriptionally in the germline.
using 3' UTRs [191]. This provides a potential mechanism to alter the spatial expression profile of CYE-1 in the distal mitotic zone. For example, the mRNA expression profile for a cyclin E transgene with the \textit{fbf-1} 3' UTR is expected to be the same as that of \textit{fbf-1} rather than endogenous cyclin E mRNA. This may be sufficient to generate a different cell cycle profile than that of wild-type and thus enable competition assay experiments.

While the Petri dish may be an imperfect representation of the worm’s natural ecology, competition assays represent a possibility to sidestep the problems associated with defining a possibly non-biological fitness metric. Given a transgenic worm with altered cell cycle properties, approximately how many generations must elapse before mutation accumulation has a noticeable impact? Suppose the transgenic worm is less fit than control, and we require that the transgenic worm population halves with respect to control before we are satisfied with our competition assay. Define \( s \) the strength of selection against the transgenic worm, where \( s>0 \) since the transgenic worm is less fit than control (by definition, the selection coefficient for control is 0). Let \( n \) be the number of generations that elapse before the transgenic worm population halves with respect to control. Then, \((1-s)^n = 1/2\), which can be simplified to

\[
n = \frac{-\log(2)}{\log(1-\Delta U/2)}
\]

where we substituted the equation for the selection coefficient for mutator traits in a selfing organism \( s = \Delta U/2 \) [23][89][19]. In Section 4.11, we saw that the wild-type cell cycle profile reduces mutation rate by 2.6% from that of a flat profile (i.e., \( \Delta U/U = 0.026 \)); assuming the transgene alters mutation rate by the same order magnitude (i.e., \( \Delta U/U = 0.02 \)), then the equation for \( n \) becomes:

\[
n = \frac{-\log(2)}{\log(1-0.01*U)}
\]

With estimates of \( U \) ranging between 0.48 [164] and 0.03 [3], we expect that anywhere from \( n=144 \) to \( n=2310 \) generations must elapse before the population of transgenic worms halves with respect to control. If it takes one week per worm generation, this competition assay can be completed in anywhere from 2.8 years to 44.3 years (one would hope for the shorter time frame).

Section 5.9: Extension to other organs

The pedigree depth quasi-minimization strategy extends to other tissues. In the following, we consider three differences between the \textit{C. elegans} gonad and other self-renewing organs that are relevant to pedigree depth quasi-minimization. First, a difference with many vertebrate organs is speed of development. While small developmental delays are expected to have a strong, deleterious effect on
fitness in an organism with a short life cycle and a “boom-bust” lifestyle such as *C. elegans*, they are likely to have a smaller impact on organisms with a longer life cycle. Such organisms are thus expected to favor low mutation accumulation over high speed of development at least to some extent, since pedigree depth quasi-minimization will come at a lessened cost. Notably however, it has been proposed that the development of mouse intestinal crypts is designed to minimize the time to formation of a mature crypt [81]. This strongly suggests that the tradeoff we have investigated between mutation minimization and speed of development is of broad relevance to animals other than *C. elegans*.

Second, a large difference lies in the number of cells to be produced over an individual’s lifetime — with a *C. elegans* gonadal arm producing ~3,000 cells and a human testis or hematopoietic system over $10^{12}$ [192] and $10^{15}$ [193][194], respectively. Because of these differences, the pedigree depth quasi-minimization strategy predicts that stem cells in vertebrates should have a slower cycling speed relative to their differentiating descendants than in *C. elegans*. While in many contexts the contribution of various stem cell populations remain to be established, the presence of sporadically cycling “reserve” populations [195] is consistent with this idea. The pedigree depth quasi-minimization strategy similarly predicts a negative correlation between stem cell cycling speed and number of cells to be produced over a lifetime; this correlation holds true when comparing hematopoiesis in a number of mammalian species.

Third, different organs may have different optimal distributions of mutations in the cells that they produce. In the context of somatic tissues, an important expected benefit of mutation minimization is reduction of cancer frequency. Since multiple “hits” are thought to be required for malignant transformation [196], it might be advantageous for a tissue to minimize the number of cells that carry two or more mutations [55][197], even if that came at the cost of an overall increase in mutation frequency. But mutator mutations likely play a significant role in tumorigenesis [198], and control of stem cell lineage might be better used to minimize the frequency at which the first mutator mutation occurs, since the carcinogenic effects of such a mutation might be difficult to counteract. In the context of the germ line, the performance objective assumed in the present study was minimization of the average number of mutations in progeny. The mutation frequency in *C. elegans* is low (~0.3 — 1 new mutation per progeny [11][59], suggesting that the problem of multiple mutations per progeny might not be of practical relevance — quantification of mutation distributions in progeny from old hermaphrodites could confirm this or provide data to guide modifications to the performance objective. Overall, the pedigree depth quasi-
minimization strategy is of broad relevance but would gain from being fine-tuned once the combined effects of multiple mutations carried by the same cell are better understood.
References


C. Tomasetti and B. Vogelstein, “Variation in cancer risk among tissues can be explained by the number of stem cell divisions,” *Science*, vol. 347, no. 6217, pp. 78-81, 2015.


discriminatively trained part-based models,” IEEE Transactions on Pattern Analysis and Machine


based on Hamilton-Jacobi formulations,” Journal of Computational Physics, vol. 79, no. 1, pp. 12-49,


1, no. 1, pp. 269-271, 1959.

simulations,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, pp. 583-598,

morphology and its applications to image processing, pp. 69-76, 1994.


[122] F. Giroud, C. Gauvain, D. Seigneurin, and V. von Hagen, “Chromatin texture changes related to


