Interconnected Genomic Landscapes of Sequence Variation, Meiotic Recombination, and Germline Chromatin in C. elegans

by

Zachary Daniel Bush

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Biology

Dissertation Committee:

Nadia Singh, Chair

Diana Libuda, Advisor

Bruce Bowerman, Core Member

Kryn Stankunas, Core Member

Brad Nolen, Institutional Representative

University of Oregon

Spring 2024

DISSERTATION ABSTRACT

Zachary Daniel Bush

Doctor of Philosophy in Biology

Title: Interconnected Genomic Landscapes of Sequence Variation, Meiotic Recombination, and Germline Chromatin in *C. elegans*

Meiosis is a specialized cell division used by sexually reproducing organisms to generate haploid gametes, such as sperm and eggs. During meiosis, cells must repair DNA damage and accurately segregate parental copies of each chromosome into daughter cells. Although there is potential for new DNA mutations and chromosome rearrangements in each meiotic division, meiotic cells preferentially use high-fidelity mechanisms of DNA repair such as crossovers to ensure faithful genome inheritance. Crossovers serve critical functions in repairing DNA damage and promote accurate chromosome segregation, but they also introduce genetic diversity in progeny. In the nematode *Caenorhabditis elegans,* like many species, there is sex-specific regulation of crossing over, but the mechanisms that lead to sexual dimorphisms in this process remain unclear. To investigate sex-specific regulation of crossing over, I leveraged the density of genetic variation in the Bristol and Hawaiian populations of *C. elegans* to generate high-resolution maps of crossovers in sperm and egg cells, respectively. In Chapter 2, I completed whole-genome assembly of the Bristol and Hawaiian strains of *C. elegans* and comprehensively detailed their genetic variation at multiple scales and complexities. I found while many genetic variants are small, such as single nucleotide polymorphisms (SNPs) and insertion/deletions (<50bp), most of the variation between these two populations is comprised of large (>50bp) sequence gains, losses, and rearrangements. Further, I demonstrate the role of specific

chromosome structures in influencing where SNPs, indels, and rearrangements accumulate in the genome. In Chapter 3, I defined genomic variations between different laboratory lineages of the Bristol and Hawaiian strains to demonstrate the degree of genetic drift and genomic structural variations accumulating in laboratory model organisms. In chapter 4, I developed a method that leverages the SNPs identified in Chapter 2 to map crossovers with sub-kilobase precision *C. elegans* sperm and eggs, respectively. I found that the crossover distribution and rate is sexually dimorphic, as well as demonstrating that the chromosomal structures associated with different states of germline gene expression are differentially associated with crossing over in developing eggs versus sperm. By determining the genomic features associated with crossover sites in each sex, I have illuminated the potential mechanisms that lead to sexually dimorphic distributions of crossing over. Taken together, the work in this dissertation fills critical gaps in our knowledge of how specific chromosome structures influence mechanisms that promote genomic integrity for inheritance by the next generation.

This dissertation includes previously unpublished co-authored material.

ACKNOWLEDGMENTS

I wish to express my sincerest thanks to all the biology educators and mentors throughout my life that have supported and encouraged my academic pursuits leading to this point. I must thank my first research mentor Dr. Glenn Rowe for his support throughout my undergraduate studies and encouragement to pursue my PhD. Above all, my greatest appreciation goes to Diana Libuda. I trusted her to support and challenge me to become the best scientist I'm capable of being, and she has done nothing less than been the best mentor I could have asked for throughout this dream of mine. Her kind guidance and devotion to excellent science have forever shaped the way I will advance my own career as well as how I will treat and support my peers in their own advancement. I thank my committee members, Nadia Singh, Jeff McKnight, Bruce Bowerman, Brad Nolen, and Kryn Stankunas for their constructive criticisms throughout the development of my projects. I want to thank the post-doctoral scholars in my lab, Nicole Kurhanewicz, Cori Cahoon, and Alice Naftaly whose constant support and friendship were vital to the completion of my studies. To my fellow students in the Libuda lab, Acadia, Jordan, and Hannah- to say I have valued our friendship is an incredible understatement and I thank you for keeping me sane in lab.

I also need to thank the members of PhD cohort, especially Elizabeth Vargas and Jordan Munroe for being my best friends. I thank Ethan Shaw and Murillo Rodrigues for being the best roommates and friends I could have asked for while living in Eugene. I also thank my partner Rose Al-Saadi for making the final phase of my PhD the happiest one. Finally, thanks to Nimbus for her aloof affection, endearing shenanigans, and constant meowing that warms my heart. Without the love and support of everyone mentioned and many more, I would not be here today.

DEDICATION

To my parents, Amy and Jason, and my stepparents, Patrick and Holly, for their endless love and support no matter how far away I am. Their dedication, patience, humility, and kindness live in me and carry me ever forward.

"My candle burns at both ends;

It will not last the night;

But ah, my foes, and oh, my friends—

It gives a lovely light!"

-Edna St. Vincent Millay, 1920

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Figure Page

CHAPTER 1: INTRODUCTION

DNA, a double-stranded polymer composed of four distinct nucleotide bases, is the biochemical source heredity for every living organism (Avery, MacLeod, and McCarty 1944; Hershey and Chase 1952). Genes are the fundamental units of heredity encoded within DNA. Each gene encodes the requisite information for the creation of functional RNAs and proteins needed for growth, survival, and reproduction (Crick 1958). The genome refers to the total quantity and number of distinct DNA molecules, called chromosomes, shared by every individual of a given species. While these individuals share highly similar genomes, the exact sequence of nucleotides in genes is variable and leads to diversity of life that is observed across all biological systems. Further, not only does sequence variation lead to functional diversity in genes, variation in the three dimensional organization of DNA also leads regulates genome function.

Chromatin is the state in which DNA is three dimensionally organized within cells. Nucleosomes, the fundamental units of chromatin, are an octamer of the four histone proteins H2A, H2B, H3, and H4 around which 147 base pairs of DNA is wrapped (Richmond and Davey 2003). Chemical modifications (*e.g.* methylation) to specific amino acids residues of the histones and higher-order interactions between modified nucleosomes can further change how densely the DNA is packaged (Ho et al. 2014; Rando and Winston 2012). Heterochromatin describes the state in which the DNA is densely packaged and often inaccessible to many DNA binding proteins. In contrast, euchromatin is the state of loosely packaged DNA where the sequence is accessible to other molecular and environmental factors. Notably, differences in chromatin state and DNA accessibility regulate many genomic processes including gene expression, the

frequency at which sequence variants arise, and the repair of DNA damage (Lawson, Liang, and Wang 2023; Rando and Winston 2012; Caridi et al. 2017; Chiolo et al. 2011; Ho et al. 2014; Janssen et al. 2016; Kouzarides 2007; Mikkelsen et al. 2007; Schuster-Böckler and Lehner 2012; Makova and Hardison 2015).

Given the significant impacts of variation in DNA sequence and chromatin structure on the integrity and function of genomes, it is critical to study how these properties are interrelated. To fully understand the functional consequences of sequence variation on genome function, we must understand all the types of sequence variations, their location, and which regions of the genome are most susceptible to change. To understand which regions of the genome are susceptible to the accumulation of variation, we must understand how differences in chromatin organization regulate the rise of different sequence variations for genome stability. Finally, to understand how chromatin organization regulates genome stability, we must understand how the mechanisms that promote genome integrity, like DNA repair, are influenced by different chromatin states. These three critical gaps in our knowledge of how the genome is maintained for faithful inheritance outline the core goals of this dissertation.

How and where do different sequence variants arise in the genome?

Sequence variations can affect genome function through a variety of mechanisms. Changes within the sequence of the coding region of genes can lead to several potential effects on the resulting protein. The range of these effects range from no noticeable changes to protein structure and function to complete loss of the protein or its function (Hartl 1996; Malinin et al. 2009). Even changes within introns, the non-coding sequence of genes normally spliced from

eukaryotic mRNAs, can affect the resulting mRNA and protein sequence. Introns are spliced from mRNA based on recognition of a DNA sequence motif, so sequence variations altering these splice sites in introns can severely impact the mRNA/protein products and contribute to the development of disease (Y. Lee and Rio 2015; Faustino and Cooper 2003). Further, DNA sequence variants can affect regions of the genome that are responsible for regulating the amount of RNA or protein produced from genes. These regulatory regions, such as promoters or enhancers, often rely on the recruitment of other proteins that recognize a highly specific DNA sequence (Schramm and Hernandez 2002; Szutorisz, Dillon, and Tora 2005; Z. Liu et al. 2014; Busby 1994; Fickett and Hatzigeorgiou 1997). The probability of incurring a deleterious effect to protein function is highly dependent on the size, location, and type of mutation (*e.g.* substitution versus deletion). Given these potential impacts to gene expression and genome function, it is critical to identify and precisely locate DNA sequence variation in all its forms.

DNA is subject to spontaneous mutations that can lead to single nucleotide variations (SNVs) or multi-nucleotide variations (Table 1.1). Mutations can be introduced from both environmental and intracellular sources including toxins, ionizing radiation, errors in DNA replication, and unrepaired DNA damage (Houston et al. 2018; Cortes-Bratti, Frisan, and Thelestam 2001; Ravanat and Douki 2016; Aquilina and Bignami 2001). Single nucleotide polymorphisms (SNPs) are the most common form of SNVs, and there are approximately 84.7 million SNPs in the 3.055 billion base pairs in the human genome (Haraksingh and Snyder 2013; The 1000 Genomes Project Consortium et al. 2015). SNPs result from the substitution of one nucleotide for another and are given the distinction of "SNPs" if they are present in more than 1% of individuals in a population (Table 1.1; Figure 1.1). Insertions or deletions (indels) under

50 nucleotides in length are the second most common sequence variant (The 1000 Genomes Project Consortium et al. 2015) (Table 1.1; Figure 1.1). Finally, the rarest and largest form of variants are termed structural variations (SVs), and their size ranges from 50 to hundreds or even millions of nucleotides (Table 1.1; Figure 1.1). SVs are commonly observed as large insertions or deletions, though they can often present as rearrangements such as duplications, inversions, or translocations (The 1000 Genomes Project Consortium et al. 2015; Sudmant et al. 2015). Importantly, any of these sequence variants (Table 1.1, Figure 1.1), whether they affect one or millions of bases, can have profoundly adaptive or sometimes lethal effects due to their ability to alter or eliminate the function of genes.

Table 1.1 Types and characteristics of genomic variation.

Variation Type		Context of change	Size
Single Nucleotide Polymorphisms (SNPs)		Substitution of one base for another	1 base pair
Short insertions/deletions (indels)		Gain or loss of sequence	1-50 base pairs
Structural Variants (SV)	**Insertion/Deletion**	Gain or loss of sequence	≥ 50 base pairs
	Inversion	Rearrangement to opposite orientation	
	Duplication	Gain of identical sequence copies	
	Translocation	Movement of sequence to a new region	
	Complex	Combination of SV characteristics (e.g. inverted translocation)	
Highly Divergent Regions (HDRs)		Dense accumulation of multiple variants in one region	≥ 50 base pairs

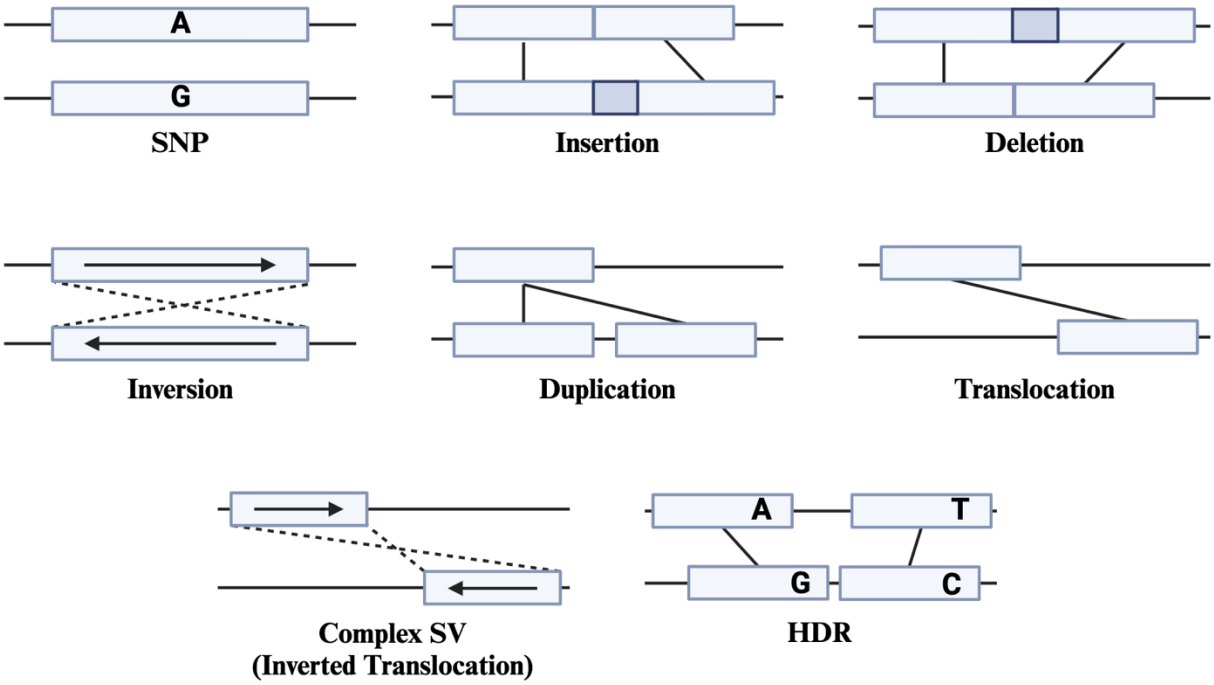


Figure 1.1 Types of genomic variation. For each variant schematic, parallel blocks represent a pairwise sequence alignment between two different genomes.

Due to their size and complexity, SVs have a high potential to disrupt genome function, but their accurate detection and characterization has been challenging. Detection of all sequence variants has historically relied on DNA sequencing and the alignment of sequencing reads (often only 100-200 base pairs in length) to a reference genome. Depending on the size and type of SV, the alignment of shorter sequencing reads either precludes their detection entirely or cannot distinguish between some types (Lesack et al. 2022a; Mahmoud et al. 2019; Goel et al. 2019; Nattestad and Schatz 2016) . An especially challenging source of genomic SVs to identify are transposable elements (TEs). TEs are highly repetitive sequences that also encode for a transposase protein enabling them to autonomously excise and reinsert into new genomic locations (Feschotte and Pritham 2007; Eide and Anderson 1985). TE mobility can create new

sequence variations through a variety of mechanisms. Their excision can lead to the generation of new SNPs or indels, and their insertion at new locations can be seen as translocations, duplications, and/or inversions (Wicker et al. 2016; Z. Zhang and Saier Jr. 2011; Geurts et al. 2006). Their highly repetitive nature and larger size also makes them incredibly challenge to map, and they have often been excluded from genomic analyses of SVs (Platt, Blanco-Berdugo, and Ray 2016; Tørresen et al. 2019; Mahmoud et al. 2019). Thus, in this dissertation I aim to address the need for alternative methods to sequencing read alignment required for comprehensive detection of all SVs and tracking of TEs.

Aside from the challenges in merely detecting all SNPs, indels, and SVs in a genome, an outstanding question in biology centers around how different chromatin states affect the rate at which these variants accumulate. The presence of DNA damage or errors in DNA synthesis must be resolved, and new sequence variants can arise if the DNA is inaccessible by other factors for restoration. While there is growing evidence for higher rates of SNPs and indels in heterochromatic regions (Makova and Hardison 2015; Schuster-Böckler and Lehner 2012), how differences in chromatin state contribute to the rise of large genomic SVs has long been hypothesized. Thus, I aim to integrate the genome-wide analyis of DNA variations and chromatin states to further our understanding of how hierarchical organization of the DNA regulates the maintenance of genome integrity for inheritance.

How do germ cells regulate faithful inheritance of the genome?

The maintenance of genome integrity during meiosis, the specialized cell division that produces germ cells such as sperm and eggs, is critical for faithful inheritance. Developing sperm

and eggs must repair DNA damage and accurately segregate chromosomes into daughter cells to protect progeny from new mutations and aneuploidy (Petronczki, Siomos, and Nasmyth 2003; Murray and Szostak 1985; Shirleen Roeder 1990). Notably, the processes of DNA repair and chromosome segregation in germ cells are coupled through the mechanisms that repair DNA double-strand breaks (DSBs) (Keeney 2008; Keeney, Giroux, and Kleckner 1997). DSBs are particularly deleterious because they result in the loss of multiple nucleotides, so this information must be restored to protect genome integrity and function. There are many error prone DSB repair mechanisms, such as non-homologous end joining (Chiruvella, Liang, and Wilson 2013; Rodgers and McVey 2016), that result in new mutations such as indels (Table 1.1, Figure 1.1). Error prone DSB repair mechanisms, however, they are strongly suppressed in developing germ cells (Zierhut et al. 2004; K. P. Kim et al. 2010; Lao and Hunter 2010; Schwacha and Kleckner 1997; Joyce et al. 2012). In developing sperm and eggs, the high-fidelity DSB repair mechanism of homologous recombination is preferentially used to faithfully restore the DNA sequence where DSBs occur (Szostak et al. 1983; Joyce et al. 2012). Homologous recombination is considered a high-fidelity repair mechanism because DSBs are processed in a way that allows them to engage with the homologous chromosome as a DNA template for repair (Figure 1.2). While homologous recombination does not introduce new mutations, it does introduce a reassortment of preexisting parental sequence variations into new recombinant chromosomes (Figure 1.2, green versus purple DNA sequence). Notably, completion of the homologous recombination program results in two basic outcomes: crossovers and noncrossovers (Schwacha and Kleckner 1995). Noncrossovers repair DNA on the damaged homologous chromosome by a transferal of sequence information from the opposing homolog in a “copy-and-paste” fashion.

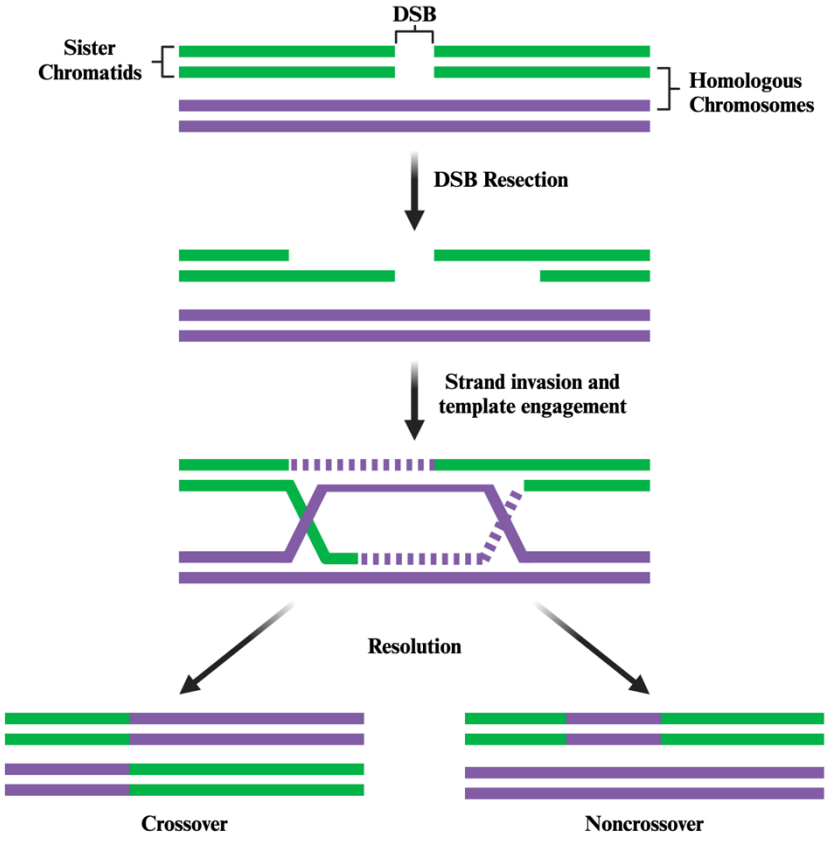


Figure 1.2. DNA double-strand break (DSB) repair by homologous recombination. DSBs are purposefully induced early in meiosis. During early stages of the DSB repair process, ends of the damaged homologous chromosome are resected. Following resection, a single strand of DNA invades the other homologous chromosome to use as a template for repair. Engagement of the homologous template and synthesis of new DNA results in the formation of a double Holliday junction. Double Holliday junctions can then be resolved into either crossover or noncrossover outcomes.

In contrast, crossovers are the result of reciprocal exchange of information between homologous chromosomes, and crossing over also creates a physical linkage between the homologs that ensures their accurate segregation during meiosis I (Murray and Szostak 1985; Shirleen Roeder 1990; Petronczki, Siomos, and Nasmyth 2003; Page and Hawley 2003). Notably, failure to recombine and form crossovers is one of the leading causes of aneuploidy, infertility, and birth defects in humans (Hassold and Hunt 2001). Thus, the study of homologous recombination, how

it is regulated, and whether this regulation is different in sperm versus egg cells is fundamental to improving our understanding of genome inheritance.

In the germ cells of most species, crossover formation is stringently regulated due to its necessity for genome integrity and accurate chromosome segregation. Crossover homeostasis ensures both the proper formation and quantity of crossovers occurring between each set of homologous chromosomes (Liangran Zhang, Liang, et al. 2014; Cole et al. 2012; Martini et al. 2006; Yokoo et al. 2012; Globus and Keeney 2012; Liangran Zhang, Wang, et al. 2014). Developing sperm and egg cells ensure the formation of at least one obligatory crossover per pair of homologous chromosomes through a mechanism termed crossover assurance (Gareth H. Jones and Franklin 2006). Too many crossovers, however, can be inhibitory to chromosome segregation, so germ cells have evolved mechanisms like crossover interference that restrict the formation of two crossovers near each other on the same chromosome (Sturtevant 1913; Muller 1916; G. H. Jones 1984; Hillers 2004; Meneely, Farago, and Kauffman 2002; Gerton et al. 2000). Thus, the genome-wide distribution of crossovers is non-random in species where interference is observable (Munz 1994; Strickland 1958; Otto and Payseur 2019; Berchowitz and Copenhaver 2010). Previous studies have demonstrated that while some meiosis-specific chromosome structures may be contributing to crossover interference (Libuda et al. 2013; Nabeshima, Villeneuve, and Hillers 2004), there are likely still many unidentified factors that regulate the rate and spatial distribution of crossovers.

There are likely multiple layers of crossover regulation enacted at the chromosomal and cellular levels. At the chromosome level, specific chromatin states are known to regulate the initiation and completion of crossing over in species like yeast and humans. The position and

density of nucleosomes as well as specific euchromatic vs heterochromatic histone modifications are known to influence the rate and placement of crossovers in these species (Powers et al. 2016; Pan et al. 2011; Lascarez-Lagunas et al. 2023; F. Baudat et al. 2010; Parvanov, Petkov, and Paigen 2010). Additionally, biological sex is known to be yet another source of differences in the rate and placement of crossovers. Evidence from studies in plants, mollusks, arthropods, amphibians, reptiles, birds, and mammals all show that the frequency and spatial distribution of crossovers on each chromosome is sexually dimorphic in eggs versus sperm (Peterson and Payseur 2021; Sardell and Kirkpatrick 2020). While there is greater evidence for the individual contributions of chromatin and sex to crossover regulation, it remains unclear as to how sex differences in recombination are further regulated by sex differences in the distribution of chromatin states themselves. In this dissertation, I directly address this gap in knowledge of whether different chromatin states differentially regulate crossing over, and thus faithful chromosome segregation, in developing sperm versus eggs.

Methods of detecting sequence variation and DNA repair

The advent of modern DNA sequencing technologies has been pivotal in advancing our understanding of sequence variation, DNA repair, and genome inheritance. Whole genome sequencing enables the detection of SNPs and indels with high confidence by aligning short sequencing reads 100-150 base pairs in length to a reference genome sequence (Lappalainen et al. 2019; Li and Durbin 2009; Li 2018). Due to the large size and complex nature of many SVs, they are challenging to detect with sequencing reads of the same or shorter length than the variant itself (Mahmoud et al. 2019). Next-generation long-read sequencing technologies can

produce reads that are tens of thousands or sometimes millions of nucleotides in length, which is particularly useful for the *de novo* assembly of new reference genomes with unprecedented structural accuracy (Jain et al. 2018; B. Y. Kim et al. 2021; Yoshimura et al. 2019). The structural accuracy of reference genome assemblies built from longer reads is therefore essential for the detection of SVs, particularly through the use of technologies that analyze assembly-to-assembly alignments of whole genomes (Delcher et al. 1999; Mahmoud et al. 2019). Thus, high-quality reference genomes are a critical resource for biological research using the human genome or the genomes of model organisms used in many laboratories.

The accurate detection of sequence variants, namely SNPs, not only builds our understanding of genome function but can be further leveraged to detect DNA repair events such as crossovers. Depending on the density of SNPs between individuals, these SNPs can provide an extensive map of genetic markers that can be used to pinpoint the location a crossover in the developing sperm or egg genome (Rockman and Kruglyak 2009; Bernstein and Rockman 2016; Rowan et al. 2015; Kianian et al. 2018). Using whole genome sequencing, a reciprocal exchange between homologous chromosomes can be identified as a switch in the linear arrangement of unique SNP markers present in the genome of each parent (Figure 2, purple versus green sequence). Given the vital functions of crossovers, understanding where they occur across the genome of developing sperm and eggs is a critical first step in understanding the mechanisms that regulate recombination and ultimately chromosome segregation.

In these dissertation studies, I utilize *Caenorhabditis elegans* as an ideal model system to research the regulation of genomic variation and crossing over. The long-term geographical isolation and genetic divergence between the Bristol and Hawaiian isolates of *C. elegans* has led

to a great accumulation of all forms of genomic variation that is ripe for study (Brenner 1974; J Hodgkin and Doniach 1997; Thompson et al. 2015; D. Lee et al. 2021). Further, the incredible density of SNPs between these *C. elegans* also provides means to map and study crossovers with high resolution. *C. elegans* also produce sperm and eggs continuously through adulthood, which is particularly attractive for the study of sexual dimorphisms in meiosis. Further, this species is highly amenable to genomic techniques previously discussed as necessary for the rigorous study of mechanisms that maintain genome integrity and function for faithful inheritance.

Dissertation outline

This dissertation aims to improve our understanding of the complex nature of genomic variation and the mechanisms that regulate crossing over. I address proposed gaps in knowledge using *Caenorhabditis elegans* as an ideal model system for these studies. I combine the use of whole-genome sequencing, comparative genomics, and genetically divergent populations of *C. elegans* to comprehensively detect all forms of genomic variation. Further, I then use the same genetic variation detected between two populations to create high-resolution recombination maps of crossovers in *C. elegans* germ cells.

In chapter 2, we present the analyses comparing the genomes of two isolates of *C. elegans* from Bristol and Hawaii. Not only are there a plethora of different types of genetic variation, but the rate at which they accumulate in the genome can vary depending on the local sequence context and surrounding DNA organization. In this chapter, we use *de novo* assembly of long-read reference genomes to comprehensively detect SNPs, indels, SVs, and the movement of transposable DNA elements. Further, we examine which regions of the genome are uniquely

susceptible to the accumulation of different types of DNA variants by studying the association of different variant types with the global distribution of multiple chromatin states.

Chapter 3 extends our approach in the analyses of genomic variation to an examination of how much genomic variation exists between laboratory lineages of canonical wild type strains. Labs using *C. elegans* have passaged their strains for thousands of generations since the species was pioneered as a model organism in the late 1960s. Since then, there is growing evidence for widespread phenotypic variation between labs with regards to lifespan and reproduction. In these studies, we compared the reference genomes of the Bristol and Hawaiian lineages in our labs to recently published genomes from other labs. Our detection and characterization of the SNPs, indels, and SVs between lab lineages of *C. elegans* highlights the growing genomic divergence in laboratory model organisms underpinning phenotypic variance and differences in experimental outcomes.

In Chapter 4, I utilize the map of SNPs discussed in chapter 2 to detect recombination events in individual oocyte and spermatocyte genomes from *C. elegans* meioses. Sexual dimorphism is a ubiquitous feature of biology, and in this chapter, we explore sex differences in the maintenance of genome integrity through crossing over. We analyzed the spatial distribution and rates of crossing over at multiple genomic scales in each sex to identify regions of the genome with sex-based biases in crossover formation. We then further delve into sequence-level features of the DNA to understand whether intergenic or genic regions uniquely influence the crossover distribution. Lastly, we examine how sex-specific gene expression and specific chromatin states differentially influence crossing over in developing sperm versus eggs.

Taken together, this dissertation outlines the research into how the interrelated nature of genomic variation, chromosome structure, and biological sex ultimately promote the integrity and faithful inheritance of the genome for future generations.

Bridge to Chapter 2

In Chapter 1, I introduced the key concepts to understanding the structure of DNA and the hierarchical organization of the genome in our cells. I expanded on these fundamental ideas of genomic structure and organization by discussing how the genomic landscapes of sequence variation, chromatin, and recombination are interconnected. To more deeply understand how how the genome is faithfully maintained across generations, we need to understand whether specific sequences or structures in the genome influence the rise of different types of genomic variations. There are many approaches to identify and analyze genomic variation, including whole-genome sequencing, genome assembly, and simulation-based permutation tests for feature enrichment. In chapter 2, I present a comprehensive analysis of genomic variation between the genetically divergent Bristol and Hawaiian isolates of *Caenorhabditis elegans*. In this next chapter, I will demonstrate the role that DNA sequence context and higher-order chromatin structures play in shaping the landscape of genomic diversity and discuss perspectives on future research in this area.

CHAPTER 2: COMPREHENSIVE DETECTION OF STRUCTURAL VARIATION AND SEQUENCE DIVERGENCE IN WILD TYPE C. ELEGANS ISOLATES

(In review at *G3: Genes, Genomes, Genetics*)

Zachary D. Bush[1], Alice F. S. Naftaly[1], Devin Dinwiddie[1], Cora Albers[1], Kenneth J. Hillers[2], and Diana E. Libuda[1]*

[1] Institute of Molecular Biology, Department of Biology, University of Oregon, 1229 Franklin Blvd Eugene, OR 97403, USA

[2] Biological Sciences Department, California Polytechnic State University, San Luis Obispo, California, US

Author Contributions

Z.D.B. polished primary genome assemblies, assessed genome quality, annotated the genomes, and analyzed genomic variation and structures. A.F.S.N. helped conceive this study and developed protocols for DNA purification, short-read Illumina sequencing, and variant calling. D.D. developed protocols for long-read sequencing, devised strategies for genome assembly, assembled the contigs and primary assembly, and filled gaps. C.A. identified transposable elements and tracked copies containing SNPs between genome assemblies. K.J.H. helped conceive this study, develop protocols for DNA purification, and purified the DNA for the long-read PacBio sequencing. D.E.L. helped conceive this study, led discussions for the comparison of different genomes, and coordinated work. All authors contributed to the manuscript, with most of the writing by Z.D.B., A.S.F.N., and D.E.L.

Abstract

Genomic structural variations (SVs) and transposable elements (TEs) can be significant contributors to genome evolution, altered gene expression, and risk of genetic diseases. Recent advancements in long-read sequencing have greatly improved the quality of *de novo* genome assemblies and enhanced the detection of sequence variants at the scale of hundreds or thousands of bases. Comparisons between two diverged wild isolates of *Caenorhabditis elegans*, the Bristol and Hawaiian strains, have been widely utilized in the analysis of small genetic variations. Genetic drift, including SVs and rearrangements of repeated sequences such as TEs, can occur over time from long-term maintenance of wild type isolates within the laboratory. To comprehensively detect both large and small structural variations as well as TEs due to genetic drift, we generated *de novo* genome assemblies and annotations for each strain from our lab collection using both long- and short-read sequencing and compared our assemblies and annotations with that of other lab wild type strains. Within our lab assemblies, we annotate over 3.1Mb of sequence divergence between the Bristol and Hawaiian isolates: 246,298 homozygous SNPs, 73,789 homozygous small insertion-deletions (<50 bp), and 4,334 structural variations (>50 bp). We also define the location and movement of specific DNA TEs between N2 Bristol and CB4856 Hawaiian wild type isolates. Specifically, we find the N2 Bristol genome has 20.6% more TEs from the *Tc1/mariner* family than the CB4856 Hawaiian genome. Moreover, we identified Zator elements as the most abundant and mobile TE family in the genome. Using specific TE sequences with unique SNPs, we also identify 53 TEs that moved

intrachromosomally and 8 TEs that moved to new chromosomes between the N2 Bristol and CB4856 Hawaiian genomes. Further, we show an enrichment of variation in transposon sequences and silenced heterochromatic regions of germline chromosomes. demonstrating. Taken together, our studies define and illuminate specific regions of the genome, including large scale repetitive regions, that are more susceptible to accumulation of genetic variation and genome structure.

Introduction

The rise of genomic variation between individuals and genetic drift between populations underly a core process of evolution. Functional characterization of sequence variants guides our understanding of phenotypic variances within species while also being critical to identifying heritable disease-causing mutations (Haraksingh and Snyder 2013). Genomic variation has been reported at multiple scales, from single nucleotide polymorphisms (SNPs) to short insertions/deletions (indels) to much larger structural variants (SVs). SVs are defined as insertions, deletions, or chromosomal rearrangements at least 50 bp in length. SVs can cause loss-of-function mutations through large gene deletions or alter gene expression by disrupting spatial interactions between regulatory sequences (Stranger et al. 2007; Hurles, Dermitzakis, and Tyler-Smith 2008). Accurate detection of both sequence variants and chromosome rearrangements is critical for understanding how genomic variation may contribute to phenotypic plasticity in individuals and populations of the same species.

Transposable elements (TEs) are a class of repetitive DNA sequences capable of moving to new locations in the genome. TE mobility is another source of genomic structural variation

that can also alter gene expression (L. Girard and Freeling 1999; Slotkin and Martienssen 2007) and can drive rapid evolutionary changes within species (Van't Hof et al. 2016; Feschotte and Pritham 2007). Notably, transposons account for a significant fraction of the total DNA sequence in many eukaryotic species (Chalopin et al. 2015; Gilbert, Peccoud, and Cordaux 2021), which provides many opportunities for TE-driven structural rearrangements. The *Tc1/mariner* family of DNA transposons is one of the most abundant TEs across species (Eide and Anderson 1985; Plasterk, Izsvák, and Ivics 1999), and early studies in *C. elegans* found it to be one of the few mobile transposons observed under laboratory conditions (Fischer, Wienholds, and Plasterk 2003). Transposon mobility and repression is tightly regulated through multiple mechanisms including chromatin modification and RNA interference (Sijen and Plasterk 2003; H.-C. Lee et al. 2012), and naturally acquired mutations in the transposase coding sequence can also disrupt their mobility (Lohe, De Aguiar, and Hartl 1997). Despite their ubiquity and impact on genomic architecture, the comprehensive detection, annotation, and inclusion of TEs in comparative genomic analyses remains challenging due to their highly repetitive sequence nature. Many studies have incompletely characterized the genomic distribution of TEs because older, short-read based methods could not accurately map the full content and location of repetitive sequences (Goerner-Potvin and Bourque 2018). Further, programs that automatically detect TEs based on sequence homology and conserved sequence elements rely heavily on libraries of older reference sequences that may predate the discovery of TE fragments and newer TE families. As new families of transposable elements are discovered (Bao et al. 2009) along with new technology that aids their annotation and tracking (Riehl et al. 2022), determining the genomic

composition and mobility of new TEs will enable our understanding of their role in genome evolution and genome integrity.

Foundational research on genomic variation utilized next-generation short-read sequencing, long-read sequencing, and the direct comparison of reference genome assemblies to identify genomic variants (Mahmoud et al. 2019; Lappalainen et al. 2019). SNPs and indels, ranging in size from 1 bp to 50 bp, can be identified with high confidence using short sequencing reads that are 100-150 bp and have high read depths (Muzzey, Evans, and Lieber 2015). In contrast, SVs are challenging to annotate using short-read sequencing because the sequencing reads are often smaller than the size of an SV (Sudmant et al. 2015; Mahmoud et al. 2019; Lesack et al. 2022b). Similarly, the highly repetitive sequences of TEs present significant challenges to mapping and annotation with traditional short read sequencing methods. With the advent of higher quality long-read sequencing technologies which generate ~10kb-30kb reads with lower genomic coverage, the accurate annotation of large regions of genomic variation such as SVs and transposable elements has become easier (Sakamoto et al. 2021). Methods of genome assembly that leverage the strengths of both short- and long-read sequencing can provide more accurate reference sequences to fully address undiscovered genomic variations previously not detected by short-read sequencing alone. Notably, some new tools to identify SVs via assembly-to-assembly alignments (Delcher et al. 1999; Nattestad and Schatz 2016; Li 2018; Goel et al. 2019) are not constrained by read-length to identify SVs and depend on high-quality reference assemblies. Overall, a high-quality reference genome assembly using multiple sequencing platforms and tools can generate a more comprehensive, accurate genome that serves as a critical resource for genomic and genetic studies in any model organism.

To fully understand how SNPs, indels, and SVs contribute to genetic drift, it is important to know which regions of the genome are most likely to change over time. Not all regions of the genome accumulate sequence variants at the same rate. For instance, in non-coding DNA regions, base substitutions and short indels are frequently less deleterious than changes to exons (Ellegren, Smith, and Webster 2003). There is also evidence for enhanced rates of variation in tandemly repeating sequences (G. Zhang, Wang, and Andersen 2022), suggesting that local sequence context may be a major determinant for mutation accumulation. Further, there is increasing evidence that the structural organization of DNA in the context of different chromatin states plays a role in shaping the landscape of genetic variation in multiple species (Makova and Hardison 2015). Importantly, while regional differences in the rates that SNPs and indels accumulate have been studied, how structural variations and rearrangements contribute to variable mutation rates is not fully known. The development of methods that can identify both structural variations and larger regions of exceptional sequence divergence provide new opportunities to study how different regions of the genome accumulate different types and sizes of variants.

Caenorhabditis elegans was the first multicellular organism to have its genome fully sequenced (C. elegans Sequencing Consortium 1998) and has been exploited to pioneer many comparative genomic studies. To understand how genetic variation influences phenotypic differences and genomic processes within species, *C. elegans* researchers primarily utilize two highly diverged wild type strains estimated to have diverged 30,000-50,000 generations ago (Thomas et al. 2015): N2 (isolated in Bristol, England) and CB4856 (isolated in Maui, Hawaii) (Nicholas, Dougherty, and Hansen 1959; Sulston and Brenner 1974; J Hodgkin and Doniach

1997; Crombie et al. 2019). Earlier comparisons of the Bristol and Hawaiian lineages were critical for studying genetic variation, gene families, and evolution of genome structures (Koch et al. 2000; Wicks et al. 2001; Stewart et al. 2005; Maydan et al. 2010). The *C. elegans* genome, comprised of five autosomes and the *X* chromosome, displays a nonuniform distribution of sequence variation when comparing the genomes of wild isolates. Although a large amount of sequence divergence was previously found between the N2 Bristol and CB4856 Hawaiian lineages (Thompson et al. 2015; Andersen et al. 2012), the increased quality of reference genomes, sequencing technology, and variant detection methods renews the need for comprehensive identification of variations (in particular large structural variations) that previously went undetected in these *C. elegans* genomes.

The first CB4856 Hawaiian genome assembly was completed in 2015 by iteratively correcting the pre-existing N2 Bristol reference assembly (*C. elegans* Sequencing Consortium 1998) with short-read sequencing data (Thompson et al. 2015). This study identified 327,050 single-nucleotide polymorphisms (SNPs) and nearly 80,000 indels relative to N2; a marked increase relative to previous comparisons, which had identified 6,000-17,000 SNPs and small indels (Wicks et al. 2001; Swan et al. 2002) between N2 Bristol and CB4856 Hawaiian. Due to the size of the short-read sequences employed in the analysis, the iterative correction method used to assemble the CB4856 Hawaiian genome may not have detected all structural rearrangements and repetitive sequences. In 2019, the first *de novo* CB4856 Hawaiian assembly from Nanopore long-read sequencing extended the length of the Hawaiian genome by multiple megabases. By comparing this CB4856 Hawaiian assembly to the original N2 assembly, researchers in this study were able to characterize over 3,000 SVs (Kim et al. 2019). Thus,

combining long-read and short-read sequencing in *de novo* genome assembly not only extended the known length of the CB4856 Hawaiian genome, but broadened our understanding of how much genomic variation exists between these wild-type strains.

TEs are highly abundant in the *C. elegans* genome and are likely a source of genomic SVs due to their ability to autonomously excise and subsequently insert themselves into new locations. Early analyses of the *C. elegans* genome indicated that approximately 12-16% of the genome is comprised of transposable elements (TEs) (*C. elegans* Sequencing Consortium 1998; Bessereau 2006), with *Tc1/mariner* elements as one of the most widely studied DNA transposons that can be active in laboratory strains (Emmons et al. 1983; Liao, Rosenzweig, and Hirsh 1983). The Zator superfamily, a relatively new TE in its characterization, is a transposase that is distantly similar to those in the Tc1/*mariner* superfamily evolved from bacterial *IS630* transposons. Further, phylogenetic analyses suggest that Zators can be considered as a distinct family of eukaryotic TEs evolved from a bacterial *TP36*-like transposon (Bao et al. 2009). To our knowledge, movement of Zator elements and other recently identified TE families has not yet been analyzed in *C. elegans* laboratory strains. While transposable element distributions have been assessed in wild *C. elegans* strains using older reference genomes and Illumina short-read sequencing (Laricchia et al. 2017), the complete TE composition has not yet been reassessed in a *de novo* assembly built from long-read sequencing nor since new families of eukaryotic Class II transposons have been discovered (Bao et al. 2009). Notably, it remains unclear if these emerging families of DNA transposable elements comprise a significant proportion of the *C. elegans* genome.

To more accurately determine the extent of sequence and structural variation between N2 Bristol and CB4856 Hawaiian genomes, we generated two high-quality, long-read reference assemblies with the same assembly pipelines for both the N2 and CB4856 strains used in our laboratory. By leveraging recent technological advancements in sequencing and variant detection, we provide a comprehensive annotation of SNPs, indels, structural variations, and transposable elements between our lineages of the Bristol and Hawaiian strains. From our comprehensive mapping of TEs in our reference genomes, we report Zator elements to be the most abundant and mobile TE family in the *C. elegans* genome. Further, we find that variations are depleted from gene regulatory sequences such as promoters and enhancers in intergenic regions of the genome. Notably, we find that TE sequences and inaccessible, heterochromatic regions of the genome harbor an enrichment of all types of variants detailed in this study, particularly for structural variants and highly diverged regions. Taken together, our systematic and comprehensive analysis of genetic variation in these two wild isolates reveals how different chromatin environments and the extent at which TEs may uniquely contribute to genetic drift and the evolution of genome structure.

Results

De novo genome assembly using combined long and short-read sequencing

To perform systematic comparisons of multiple wild type genomes from different laboratory isogenic strains, we generated *de novo* assemblies of N2 Bristol and CB4856 Hawaiian. The N2 Bristol genome was assembled from PacBio long-reads with 136x coverage producing 121 contigs and a 100.4Mb genome (Figure 2.1A) The CB4856 Hawaiian genome

was generated from PacBio long-reads with 132x coverage from 169 contigs to give a 98.8Mb assembly (Figure 2.1B). These long-read assemblies were then supplemented with Illumina paired end short-reads with a sequencing depth of 540x and 628x for N2 Bristol and CB4856 Hawaiian respectively (Figure 2.1A-B).

To assess the quality of our reference genomes, we examined assembly-to-assembly alignments between our N2 Bristol and Hawaiian genomes and quantified the orthologous gene content for each assembly. Since the genomes of these two isolates are highly homologous, whole genome alignments should show a high proportion of aligned bases synteny when comparing our N2 Bristol to CB4856 Hawaiian assemblies. Indeed, 99.2% of bases across our N2 Bristol genome were aligned to our CB4856 Hawaiian genome assembly, and more than 92.2% of bases within alignments were syntenic (Table 2.1). Analysis of universal single-copy orthologs (Simão et al. 2015; Manni et al. 2021) in our *de novo* N2 Bristol and CB4856 Hawaiian genomes revealed greater than 98% completeness (Supplemental Figure S2.1) and validate that our assemblies are high quality.

De novo genome assemblies of the N2 Bristol and CB4856 Hawaiian isolates enhance detection of genomic variation

To detect any genomic variations that may have been missed in prior analyses with short-read sequencing-based reference genomes, we performed sequence alignments and comparisons with our reference genomes that were assembled using both short- and long-read sequencing technology and the same assembly method. First, we aligned our CB4856 Hawaiian short reads to our N2 Bristol assembly to quantify SNPs and indels. This analysis revealed a total of 337,584

SNPs and 94,503 indels across the genome, of which 246,298 SNPs and 73,789 indels are homozygous and not segregating in our populations of each strain (Table 2.1, Figure 2.1C-D). Similar to previous observations (Thompson et al. 2015), we note a few highly variable regions with a greater density of SNPs and short indels in the center regions of the autosomes, particularly on chromosomes *IV* and *V* (Figure 2.1C-D). We also find that many SNPs and indels overlapped with gene annotations, particularly on chromosomes *I* and *III* where more than 60% of SNPs and indels overlap with gene sequences (Figure 2.1E-F). Taken together, our detection of SNPs and indels indicate an abundance of sequence variation is unevenly distributed between genes and intergenic space on multiple chromosomes.

To identify large sequence variants and chromosome rearrangements, we used whole-genome alignments as opposed to the alignment of sequencing reads with our reference genomes (see Methods). This analysis identified a total of 4,364 structural variants, which are categorized as insertions, deletions, and other chromosomal rearrangements spanning at least 50 bp (Table 2.1). We also identified 1,174 Highly Divergent Regions (HDRs) (Goel et al. 2019) across the

---

Figure 2.1. Genomic distribution of SNPs and indels between the N2 Bristol and CB4856 Hawaiian genomes. (A) Line plots showing the average sequencing coverage in 100 kb bins across each chromosome in the N2 Bristol genome. **(B)** Line plots showing the average sequencing coverage in 100 kb bins across each chromosome in the CB4856 Hawaiian genome. For each plot in A and B, the coverage for Illumina short-read sequencing is shown in blue, and sequencing coverage for PacBio long-reads is shown in red. For both A and B, sequencing reads from the genome of each isolate were re-aligned to the completed genome assemblies, respectively. **(C)** Histograms depicting the distribution of CB4856 Hawaiian SNPs across each N2 Bristol chromosome in 100 kb bins. **(D)** Histograms of the distributions of CB4856 Hawaiian indels across each N2 Bristol chromosome in 100 kb bins. **(E)** The proportion of SNPs that overlap with remapped gene annotations versus intergenic regions in the N2 Bristol genome. **(F)** The proportion of indels that overlap with gene versus intergenic regions in the Bristol genome.

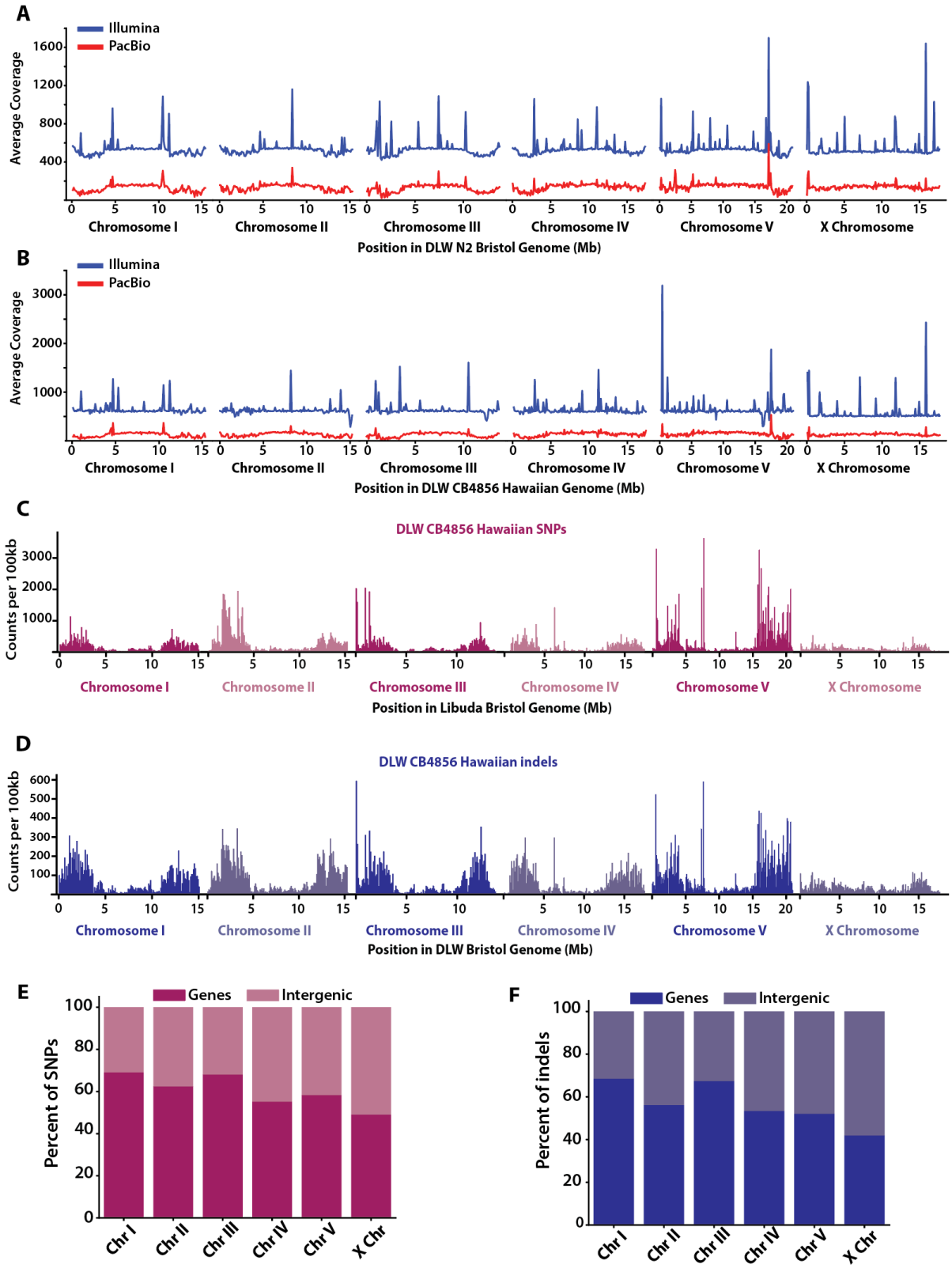

A
Illumina
PacBio
Average Coverage
1600
1200
800
400
Chromosome I
Chromosome II
Chromosome III
Chromosome IV
Chromosome V
X Chromosome
Position in DLW N2 Bristol Genome (Mb)
B
Illumina
PacBio
Average Coverage
3000
2000
1000
Chromosome I
Chromosome II
Chromosome III
Chromosome IV
Chromosome V
X Chromosome
Position in DLW CB4856 Hawaiian Genome (Mb)
C
DLW CB4856 Hawaiian SNPs
Counts per 100kb
3000
2000
1000
Chromosome I
Chromosome II
Chromosome III
Chromosome IV
Chromosome V
X Chromosome
Position in Libuda Bristol Genome (Mb)
D
DLW CB4856 Hawaiian indels
Counts per 100kb
600
500
400
300
200
100
Chromosome I
Chromosome II
Chromosome III
Chromosome IV
Chromosome V
X Chromosome
Position in DLW Bristol Genome (Mb)
E
Genes
Intergenic
Percent of SNPs
100
80
60
40
20
0
Chr I
Chr II
Chr III
Chr IV
Chr V
X Chr
F
Genes
Intergenic
Percent of indels
100
80
60
40
20
0
Chr I
Chr II
Chr III
Chr IV
Chr V
X Chr

Table 2.1. Comparisons between the N2 Bristol genome (this study) and CB4856 Hawaiian genome (this study)

	Chromosomes						Total
	I	*II*	*III*	*IV*	*V*	*X*	
N2 Bristol Chromosome Length (this study)	15,114,068	15,311,845	13,819,453	17,493,838	20,953,657	17,739,129	100,431,990
CB4856 Hawaiian Chromosome Length (this study)	15,045,644	15,257,363	13,206,755	17,183,882	20,547,529	17,584,915	98,826,088
N2 Bristol Bases Aligned	15,100,574	15,303,320	13,222,676	17,330,119	20,947,147	17,738,394	99,642,230 (99.21%)
% Syntenic Aligned Bases	93.31	88.56	90.61	95.42	87.04	98.73	92.23
SNPs*	30,394	48,365	29,881	30,497	87,300	19,861	246,298
Indels*	11,460	13,716	10,530	11,221	20,063	6,799	73,789 (275,442 bp)
SVs	863	808	649	619	925	470	4,334 (2,654,902 bp)
HDRs	185	270	165	138	356	60	1,174 (6,864,884 bp)

* All variants listed are only those for which the CB4856 Hawaiian genome was homozygous

genome. HDRs are defined here as regions of the genome over 50 bp in length that result in low-quality pairwise alignments due to the presence of multiple gaps within these alignments (Goel et al. 2019), and their classification is distinct from the hyper divergent regions previously characterized in the *C. elegans* genome (D. Lee et al. 2021). Overall, greater than 9.9% of the N2 Bristol genome (~10.0 Mb) displayed variation through SNPs, indels, SVs, and HDRs when compared to the CB4856 Hawaiian genome. SVs and HDRs represented only 1.3% and 0.3% of variant sites between N2 Bristol and CB4856 Hawaiian, respectively, but accounted for over 94% (9.5Mb) of sequence variation (Table 2.1). Including heterozygous variants, our short-read analysis detected 3% more SNPs and 18% more indels than previously discovered using short-

read assemblies of N2 Bristol and CB4856 Hawaiian (Thompson et al. 2015). Utilizing whole-genome alignment comparisons (Li 2018; Goel et al. 2019), we identified 985 more SV sites than previously reported (Nattestad and Schatz 2016; Kim et al. 2019). This increased sensitivity in variant site detection highlights the power of combining long-read and short-read sequencing to create accurate genome assemblies for comparative genomic studies.

Given an enhanced detection of variant sites between our N2 Bristol and CB4856 Hawaiian assemblies, we were interested whether the canonical "arms"-versus-"center" distribution is upheld in the larger variants we detected as well. Prior studies report have detailed punctuated regions of sequence divergence at the scale of SNPs and indels, (Thompson et al. 2015; Kim et al. 2019), so we expected a greater density of variation in the terminal thirds (the arm-like regions) of each chromosome for SVs and HDRs as well. Indeed, there is a significant concentration of SNPs, indels, SVs and HDRs in the arm-like regions relative to the center region of each chromosome (Supplemental Figure S2.2). Over 78% of all SNPs, indels, SVs, and HDRs are in the arm-like domains of each chromosome (Genome-wide averages: 75.12% of SNPs, 78.24% of indels, 71.39% of SVs, 90.77% of HDRs). To determine if the enrichment of SNPs, indels, and SVs in the chromosomal arm-like regions was significant, we compared the observed distribution of each variant category with random permutations of each category of variant (Heger et al. 2013). SNPs, indels, and HDRs on the autosomes were significantly enriched in the arm-like regions (SNPs: 1.36 to 1.77-fold enrichment; Indels: 1.47 to 1.84-fold enrichment; HDRs: 1.70 to 2.06-fold enrichment; $p < 0.001$ by hypergeometric test). SVs, however, were only significantly enriched on the arm-like regions of autosomes *I, III*, and *IV* (1.64 to 1.92-fold enrichment; $p<.001$ by hypergeometric test). The fold enrichment of all

variants on the arm-like regions of the *X* chromosome was slightly weaker, ranging from 1.23-1.64 (SNPs: 1.26-fold enrichment; Indels: 1.26-fold enrichment; SVs: 1.23-fold enrichment; HDRs: 1.64-fold enrichment; all p-values < 0.05 by hypergeometric test). Taken together, we find that all types and sizes of genomic variants share the same "arms"-vs-"center" distribution pattern across each chromosome.

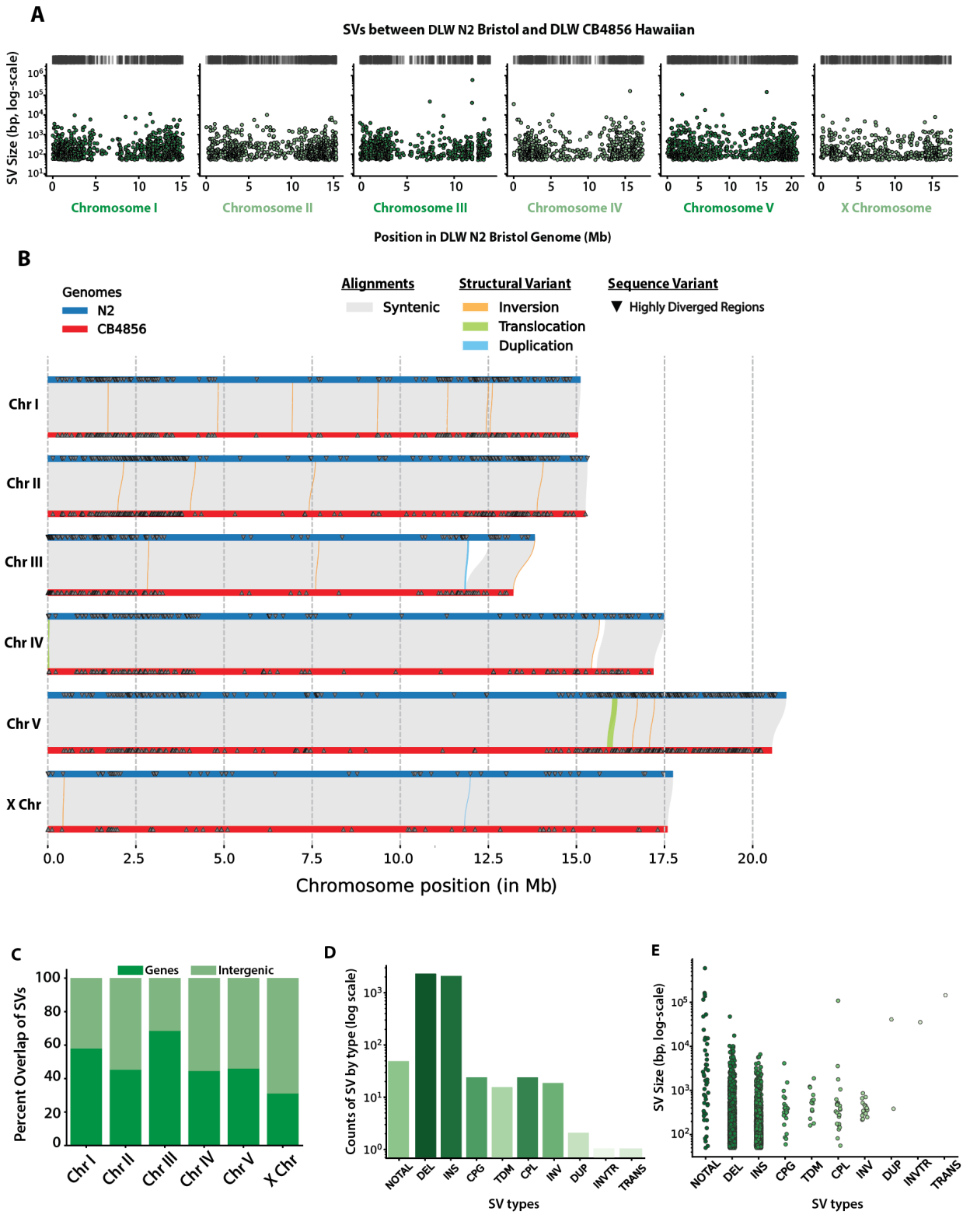
Millions of base pairs affected by genomic structural variation and highly divergent regions

We next analyzed which regions of the genome were most affected by structural variation. Broadly, the distribution of SVs and HDRs across each chromosome resembles the genomic distribution of SNPs and indels (Figure 2.2A-B). For both SVs and HDRs, we note that many of these sites reside on the arm-like domains of the autosomes, with the *X* chromosome displaying far fewer counts of structural variation and highly divergent regions (Figure 2.2A, Table 2.1). While most structural variants range in size from 50 bp-1 kb (Figure 2.2A and 2E), we did find multiple instances of SVs 10 kb or greater in size on each chromosome, with some approaching hundreds of kb in size. Due to the large sizes of structural variants, they have the potential to overlap with many coding sequences in the genome. Thus, we wanted to know whether SVs overlapped with coding regions at the same frequency as SNPs and indels. On the autosomes, 44.5-68.5% of SVs overlapped with gene regions compared 31.1% on the *X* chromosome (Figure 2.2C). Consistent with our previous analyses of SNPs and indels, SVs and HDRs are also present in greater numbers on the autosomes than the *X* chromosome. Taken together, the sequence of the *X* chromosome appears more stable across the long divergence times separating the N2 Bristol and CB4856 Hawaiian isolates.

To assess which types of structural variants are most prevalent in the genome, we detailed the number and size of each type of structural variant we identified (Figure 2.2D-E). Previous studies have identified many large insertions, deletions, and expansions/contractions of tandemly repeated regions (C. Kim et al. 2019; G. Zhang, Wang, and Andersen 2022), but less is known about the genome-wide prevalence of other types of SVs such as inversions, duplications, and translocations. Most SVs identified were large insertions and deletions, with relatively fewer SVs being characterized as duplications or rearrangements like inversions (Figure 2.2D). Among the SVs detected, we identified 47 non-alignable structures, 2 duplications, 18 inversions, and 2 translocations. Non-alignable regions (NOTALs) are highly diverged regions containing many repeats and low-complexity sequences that are inhibitory to whole-genome alignment. From our whole-genome alignments of the N2 Bristol and CB4856 Hawaiian genomes, the non-alignable regions between the two genomes comprise 1.39 Mb of sequence, ranged in size from 50-592 kb, and comprise <0.5% of coding genes in the Bristol genome. The SVs identified ranged in size

---

Figure 2.2. Genomic distribution and size of SVs between the N2 Bristol and CB4856 Hawaiian genomes. (A) Histograms depicting the distribution of SVs across each chromosome in 100kb bins. Black dashes above each histogram correspond to the genomic locations of SVs that are greater than 20kb in size. **(B)** Chromosome alignment plot depicting syntenic regions between N2 Bristol and CB4856 Hawaiian, structural variants, and highly divergent regions (HDRs). The width of lines showing SVs are proportional to their size. Only rearrangements 1kb or greater in size are shown. **(C)** Stacked bar plots showing the percentage of CB4856 Hawaiian SVs that overlap with intergenic and gene-coding regions of the N2 Bristol genome. **(D)** Bar plots showing the number of each type of SV identified. **(E)** Strip plots showing the log-scaled size distribution of SVs separated by type. For SV types: NOTAL = non-aligned regions, DEL =deletion, INS = insertion, CPG = copy gain in query genome, CPL = copy loss in query genome, TDM = tandem repeat region, INV = inversion, DUP = duplication, TRANS = translocation, and INVTR = inverted translocation. For D and E, different colors only correspond to the different types of SV identified.


A
SVs between DLW N2 Bristol and DLW CB4856 Hawaiian
SV Size (bp, log-scale)
Chromosome I
Chromosome II
Chromosome III
Chromosome IV
Chromosome V
X Chromosome
Position in DLW N2 Bristol Genome (Mb)
B
Genomes
N2
CB4856
Alignments
Syntenic
Structural Variant
Inversion
Translocation
Duplication
Sequence Variant
Highly Diverged Regions
Chr I
Chr II
Chr III
Chr IV
Chr V
X Chr
Chromosome position (in Mb)
C
Genes
Intergenic
Percent Overlap of SVs
D
Counts of SV by type (log scale)
NOTAL
DEL
INS
CPG
TDM
CPL
INV
DUP
INVTR
TRANS
SV types
E
SV Size (bp, log-scale)
SV types

from 50 bp to 592 kb (Figure 2.2E), and HDRs ranged from 50 bp to 199 kb (Supplemental Data). Insertions and deletions are the most common type of SV, and their size distribution includes a greater proportion of variants 50-200 base pairs. The proportion of duplications and other rearrangements at magnitudes of 1-100 kb, in contrast, is much higher (Figure 2.2E). We identified one 156 kb region in the N2 genome that was translocated upstream on the right end of CB4856 Hawaiian chromosome *V* (*V*:15,871,614-16,027,614 bp), while the other translocation (38 kb in size) was found to be inverted near a telomere of CB4856 Hawaiian chromosome *IV* (*IV*: 176:38,447 bp). The largest duplication was found on Hawaiian chromosome *III* (*III*: 11,819,363-11,860,261). Together, our analyses detected additional variant sites in N2 and CB4856 genomes, and further illuminates the contribution of large SVs and HDRs to genome evolution.

Minimal movement of DNA transposons between the N2 Bristol and CB4856 Hawaiian lineages

To identify and locate known TE sequences in our N2 Bristol and CB4856 Hawaiian assembled genomes, we used a TE identification pipeline that applies an ensemble of programs to find all known RNA and DNA TE families (Riehl et al. 2022). We found that approximately 14.7 and 14.3% of our N2 Bristol and CB4856 Hawaiian assemblies, respectively, are composed of TE sequences (Table 2.2). For both genome assemblies, the distribution of TEs was concentrated in the terminal third, arm-like regions of each chromosome (Figure 2.3A-B).

Table 2.2. Transposable Elements identified in the N2 Bristol genome (this study) vs CB4856 Hawaiian genome (this study).

	N2 Bristol	CB4856 Hawaiian
Class I Transposable Elements (Retrotransposons)	710 (2,688,730 bp)	776 (2,522,357 bp)
Gypsy	557 (2,195,895 bp)	592 (2,031,038 bp)
Copia	134 (472,195 bp)	161 (465,785 bp)
SINE	9 (2,146 bp)	9 (1,945 bp)
ERV	7 (8,280 bp)	6 (7,569 bp)
LINE	3 (10,214 bp)	8 (16,038 bp)
Class I intrachromosomal transpositions*	0	
Class I interchromosomal transpositions*	0	
Class II Transposable Elements (DNA transposons)	17,682 (12,055,357 bp)	17,310 (11,606,010 bp)
Tc1/Mariner	1870 (1,298,386 bp)	1,550 (1,131,443 bp)
hAT	3,999 (3,988,461 bp)	3,818 (3,725,667 bp)
CMC	1,679 (3,138,647 bp)	2,011 (3,260,455 bp)
Zator	9,159 (3,009,341 bp)	8,980 (2,907,391 bp)
Novosib	46 (12,060 bp)	28 (12,088 bp)
Helitron	39 (368,980 bp)	43 (329,238 bp)
Sola	821 (226,645 bp)	699 (196,797 bp)
MITE	69 (12,837 bp)	181 (42,931 bp)
Class II intrachromosomal movement*	53	
Class II interchromosomal movement*	8	

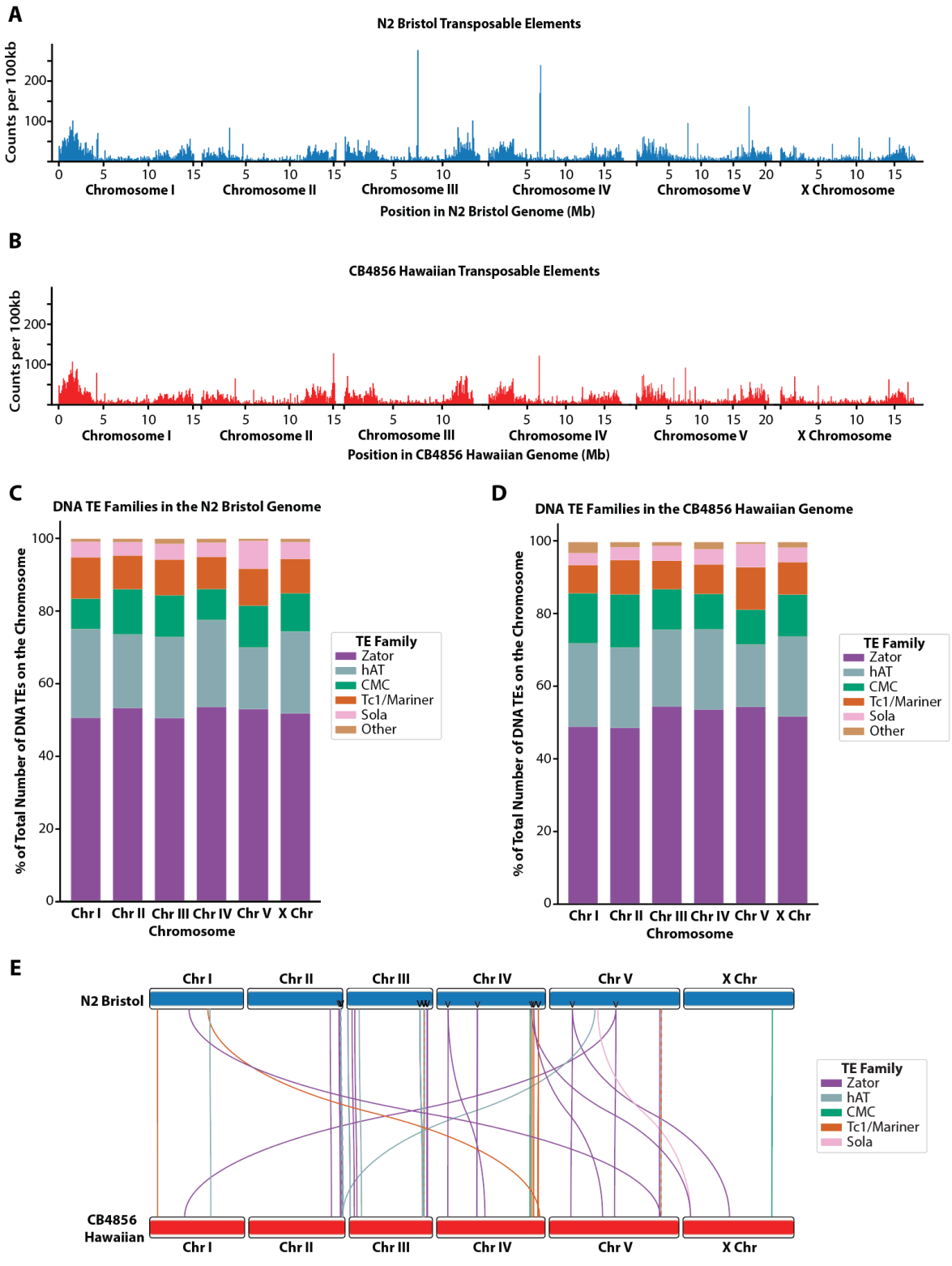
***** All TE sequences with predicted transpositions are relative to the N2 Bristol genome.

Class II DNA TEs represented 96% of all TEs identified in each genome, and Zator elements are 52% of these Class II DNA TEs present in each genome (Supplemental Table 2.1, Figure 2.3C-D). To our knowledge, movement of Zator elements and other recently identified TE families has not yet been analyzed in *C. elegans* laboratory strains. Further, we also found that our N2 Bristol genome has 20.6% more TEs from the *Tc1/mariner* family than our CB4856 Hawaiian genome assembly (Table 2.2).

Since the N2 Bristol and CB4856 Hawaiian lineages were geographically isolated for thousands of generations, we sought to utilize our new TE annotation set to identify individual transposition events that occurred over the course of divergence between the two strains. Using whole-genome alignments and the SNPs we previously defined between these two lineages, we identified specific TE sequences with unique polymorphisms that enables individual transposons to be tracked between the N2 Bristol and CB4856 Hawaiian genome assemblies. Of the 18,392 total transposable elements identified in the N2 Bristol genome, 9,377 TEs were uniquely identifiable by sequence polymorphisms. Among all N2 Bristol TEs with SNPs, only 1,535 elements were detectable in the CB4856 Hawaiian genome. While the vast majority of TEs were found to have not moved within either genome, we did track 53 Class II DNA TEs to new locations on the same chromosome and eight TEs that appeared on different chromosomes in the CB4856 Hawaiian genome (Figure 2.3E, Table 2.3). Specifically, we detected five Zator elements and one each of *Tc1/mariner*, Sola, and hAT elements on different chromosomes

---

Figure 2.3. Genomic distributions of transposable elements in the DLW N2 Bristol and DLW CB4856 Hawaiian genomes. Histograms depicting the distributions of transposable elements across the DLW N2 Bristol genome in 100kb bins. **B**) Histograms depicting the distributions of transposable elements across the DLW CB4856 Hawaiian genome in 100kb bins. **C,D)** Stacked bar plot depicting the percent of total DNA transposable elements on DLW N2 Bristol (C) and DLW CB4856 Hawaiian (D) chromosomes accounted for by specific DNA transposon families. For TE families: CMC= CACTA, Mirage and Chapaev families; hAT = hobo and Activator families; Other = MITE, Novosib and Helitron families. **E**) Ideogram depicting the locations of individual DNA transposable elements that moved between the DLW N2 Bristol genome and the DLW CB4856 Hawaiian genome. DLW N2 Bristol chromosomes are represented by the blue boxes on the top, and DLW CB4856 Hawaiian chromosomes by the red boxes on the bottom. Each line represents an individual transposable element sequence, traced from its position on the DLW N2 Bristol genome to its unique position on the DLW CB4856 Hawaiian genome. Transposable elements predicted to have translocated are colored according to transposon class. Arrow heads across the Bristol N2 chromosomes indicate DNA TEs where duplicated copies are found in the Hawaiian CB4856 genome.


A
N2 Bristol Transposable Elements
Counts per 100kb
Chromosome I
Chromosome II
Chromosome III
Chromosome IV
Chromosome V
X Chromosome
Position in N2 Bristol Genome (Mb)
B
CB4856 Hawaiian Transposable Elements
Counts per 100kb
Chromosome I
Chromosome II
Chromosome III
Chromosome IV
Chromosome V
X Chromosome
Position in CB4856 Hawaiian Genome (Mb)
C
DNA TE Families in the N2 Bristol Genome
% of Total Number of DNA TEs on the Chromosome
Chr I
Chr II
Chr III
Chr IV
Chr V
X Chr
Chromosome
TE Family
Zator
hAT
CMC
Tc1/Mariner
Sola
Other
D
DNA TE Families in the CB4856 Hawaiian Genome
% of Total Number of DNA TEs on the Chromosome
Chr I
Chr II
Chr III
Chr IV
Chr V
X Chr
Chromosome
TE Family
Zator
hAT
CMC
Tc1/Mariner
Sola
Other
E
Chr I
Chr II
Chr III
Chr IV
Chr V
X Chr
N2 Bristol
CB4856 Hawaiian
TE Family
Zator
hAT
CMC
Tc1/Mariner
Sola

between the two lineages. In this analysis, we note that 23 unique copies of Class II DNA TEs in the N2 Bristol genome that had duplicated copies in the CB4856 Hawaiian genome (Figure 2.3E, arrowheads). While we were able to identify transposition events relative to the N2 Bristol genome, we cannot accurately infer the history of each CB4856 Hawaiian copy to determine which resulted from transposition versus duplication. Overall, the landscape of transposable elements remains largely unchanged across the history of divergence between the N2 Bristol and CB4856 Hawaiian lineages.

Table 2.3. Intra- and interchromosomal movement of TEs by family.

Class II DNA TE Family	Intrachromosomal movements (8)	Interchromosomal movements (53)
Zator	5	26
hAT	1	14
CMC	0	2
Tc1/Mariner	1	11
Sola	1	0

Variants are strongly depleted from the coding sequences in genes and gene regulatory regions

While much of the sequence variation is enriched in intergenic sequences of the arm-like regions, we wanted to determine whether this variation differentially affected genic versus intergenic and/or regulatory sequences across the genome. Based on our remapped annotations (see Methods, LiftOff (Shumate and Salzberg 2021)), approximately 61.8% of the N2 Bristol

genome is comprised of gene sequences, with exons and introns representing 28.6% and33.2% of the genome, respectively. Thus, we would expect corresponding proportions of each variant type to overlap within each annotation if variant sites were uniformly distributed across the genome. While we did find that many SNPs and indels do overlap with gene sequences, our more detailed enrichment analyses find that these are likely occurring within intronic sequences (Figure 2.4A and Supplemental Figure S2.3). For all variant types, we find significant depletions in the CDS and exon sequences of each gene. Short indels under 50 bp, display the strongest depletion from CDS and exon sequences on each chromosome (Log_2(fold) values -1.56 to -3.12). Further, we find that SVs display a much greater depletion in the 5' and 3' untranslated regions relative to CDS/exons (Log2(fold) values -.99 to -3.15; all p values < 0.05 except 5' UTRs on chromosome *II*). Taken together, we find that much of the sequence variants and structural variations overlapping with genes reside within the introns where their effects on the overall function of genes are likely minimized.

Since each variant type displayed significant depletions in coding regions of the genome, we then assessed whether regulatory and other intergenic sequences outside of genes were enriched for genomic variation. SNPs and indels displayed modest depletions within regulatory sequences such as transcription factor binding sites, promoters, and enhancers (Figure 2.4B). SVs, however, were notably depleted at transcription factor binding sequences on every chromosome (Log2(fold) values -0.56 to -1.93). SVs were also significantly depleted from promoters on chromosomes *I* and *V* and depleted from enhancers on *I, II, IV,* and *V* (Figure 2.4B

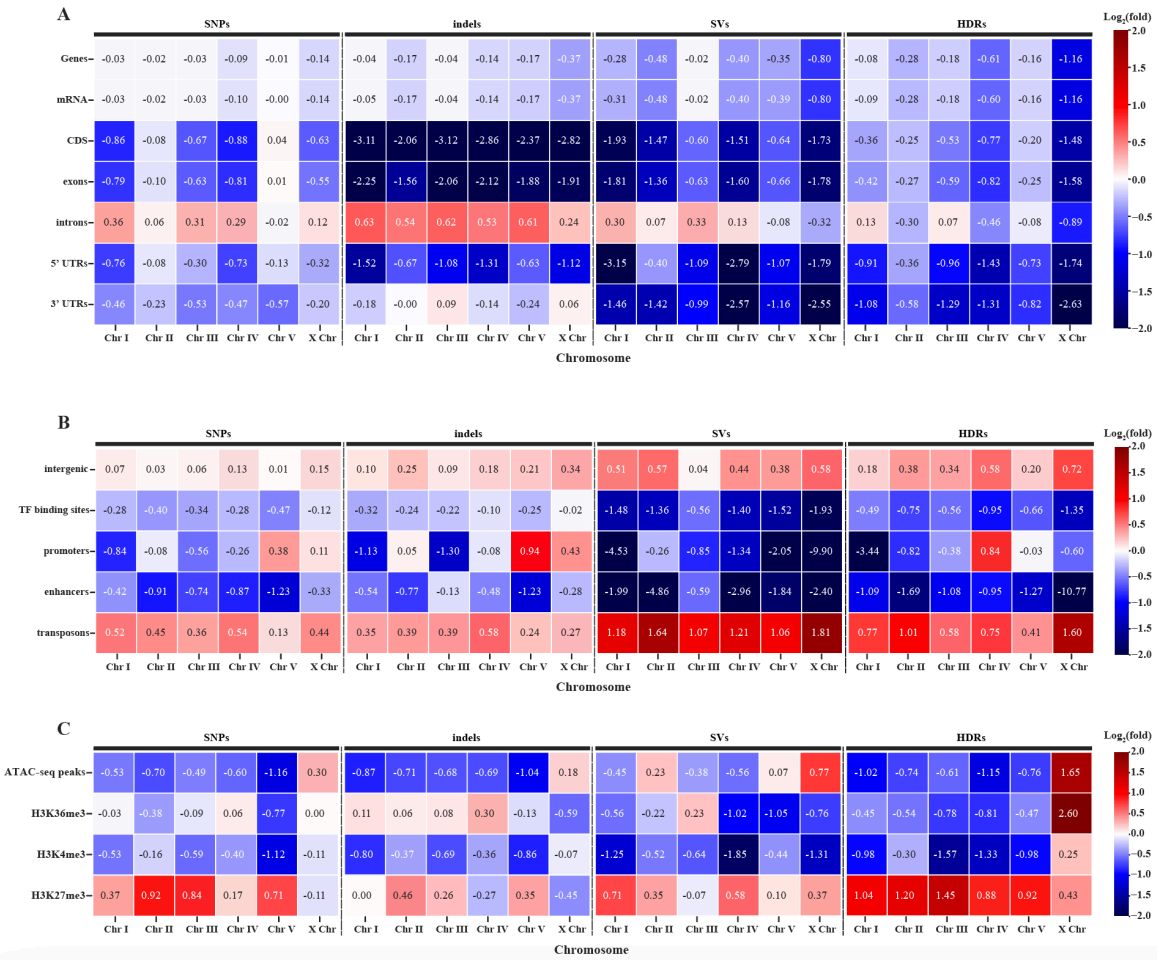


Figure 2.4. Genomic variants are depleted from coding regions and enriched in TEs and heterochromatin. A) Heatmap showing the Log2(fold) enrichment or depletion of each variant type in genes and sub-gene annotations. B) Heatmap showing the Log2(fold) enrichment or depletion of each variant type in intergenic regions, gene regulatory sequences, and transposons. C) Heatmap showing the Log2(fold) enrichment or depletion of each variant type in different chromatin profiles. ATAC-seq and ChIP-seq data were taken from germline-specific datasets and calculated with MACS3 (see methods).

and Supplemental Figure S2.3). Thus, many of the intergenic SNPs, indels, SVs, and HDRs may be accumulating in different regions outside of genes and their regulatory sequences.

Genomic variation is enriched in TEs and silenced regions of the genome

To determine whether another source of intergenic variation may be occurring via TEs, we measured the overlap of our variant dataset with our transposon annotations. We find a consistent enrichment of each variant type in transposable element sequences. SNPs and indels were relatively modest in their enrichment in these regions (Log2(fold) values 0.13 to 0.58, p-values < 0.05), whereas SVs were markedly more enriched in TEs (Log2(fold) values 1.06 to 1.81, p-values < 0.05). SVs account for most of the base pairs affected by variation between the N2 Bristol and CB4856 Hawaiian genomes, and approximately 32% of the base pairs in SVs reside within TE sequences. Overall, large genomic variants are depleted from regulatory sequences where gene function may be disrupted, and TE sequences are the most prone to changes in genome structure.

Chromatin, the hierarchical organization of DNA around histone proteins, can adopt DNA-accessible or dense inaccessible conformations to regulate gene/TE silencing versus gene expression (Ho et al. 2014; Lawson, Liang, and Wang 2023; Rando and Winston 2012). To assess whether specific higher-order chromatin structures influence where variants accumulate in the genome, we measured the association of each variant type with accessible or inaccessible chromatin regions. Chromatin can broadly be classified into accessible, transcriptionally active euchromatin or inaccessible or silenced heterochromatic regions. When we measured the association of each variant type with accessible regions of chromatin determined by ATAC-seq (Serizay et al. 2020), we find that SNPs and indels are depleted from genomic regions with accessible chromatin on each autosome (Log2(fold) values -0.49 to -1.16, p-values < 0.05; Figure 2.4C). HDRs displayed similar patterns of depletion, whereas SVs were displayed more

modest depletions in accessible chromatin. In contrast to the patterns observed on the autosomes, SVs and HDRs were enriched in ATAC-seq peaks on the *X* chromosome (Log2(fold) values 0.77 for SVs and 1.65 for HDRs, p-values ≤ 0.05). In summary, we find that chromatin accessibility negatively correlates with the presence of sequence and genomic structural variation on the autosomes.

Specific chemical modifications to histone proteins, particularly the methylation of histone H3 lysine residues, are used to induce the accessible versus inaccessible chromatin states. To test whether specific chromatin modifications contribute to the depletion of variants in accessible chromatin on the autosomes, we measured how often variants overlapped with peaks of canonical heterochromatic and euchromatic modifications. H3K4 trimethylation and H3K36 trimethylation are euchromatic modifications associated with transcribed genes in the germline (C. L. Liu et al. 2005; Pokholok et al. 2005; Mikkelsen et al. 2007; Rando and Winston 2012; Ho et al. 2014). In contrast, chromatin modifications like H3K9me3 and H3K27me3 mark heterochromatic and thus transcriptionally inactive regions like transposons, silenced genes, and intergenic regions, which all have lower DNA accessibility (Ho et al. 2014; X. Zhang et al. 2007; Jamieson et al. 2013; Connolly, Smith, and Freitag 2013; Basenko et al. 2015; Lewis 2017). SNPs and indels displayed the greatest depletion from H3K4me3-assciated regions (Figure 2.4C) which often mark the promoter sequences of genes, consistent with our association data in promoters (Figure 2.4B). SVs and HDRs displayed similar patterns of enrichment in H3K36me3 regions, and much stronger depletions in H3K4me3 peaks compared with SNPs and indels (Figure 2.4C). Strikingly, we find a significant enrichment of all variant types in H3K27me3 marked heterochromatin. SNPs on chromosomes *II, III*, and *V* (Log2(fold) values 0.71 to 0.92, p-

values < 0.05) and HDRs on every chromosome (Log2(fold) values 0.43 to 1.45, p-values < 0.05) displayed the greatest enrichments in these silenced regions relative to indels and SVs. For each variant type, a large fraction of the base pairs affected by these sequences overlap with H3K27me3 heterochromatin (42.7% of HDR base pairs, 34.0% of SNPs, 32.1% of SV base pairs, and 26.0% of indel base pairs). Taken together we find that inaccessible heterochromatic regions of the genome, which include TE sequences, appear to accumulate genomic variation more than coding, euchromatic regions with accessible chromatin.

Discussion

Detection and characterization of sequence variation between individuals or across species is fundamental to our functional understanding of genomic elements and consequences of variation. Since the first draft of the *C. elegans* genome was released in 1998, the highly divergent strains N2 Bristol and CB4856 Hawaiian have been used extensively for comparative genomics studies(C. elegans Sequencing Consortium 1998; Koch et al. 2000; Wicks et al. 2001; Maydan et al. 2010; Andersen et al. 2012; D. Lee et al. 2021). The combined usage of short and long read sequencing to assemble genomes and to compare them has both increased the quality of our reference genomes as well as enhanced the genome-wide detection of sequence variants, new genes, and new genomic regions (Yoshimura et al. 2019; C. Kim et al. 2019; B. Y. Kim et al. 2021; Sarsani et al. 2019). In this study, we generate *de novo* assemblies for the N2 Bristol and CB4856 Hawaiian *C. elegans* isolates from our lab lineage using short-read and long-read sequencing. Our examination of the many types of genomic variants that arise in these diverged strains demonstrate the role that specific sequences and their associated chromatin structures

have in shaping the evolution of genome structure across large timescales. Further, these genomes will serve as additional tools for future comparative genomics studies, especially in the functional characterization of structural variations identified through whole-genome alignments.

Highly variable arm-like domains on *C. elegans* chromosomes

The arm-like regions of *C. elegans* chromosomes exhibit a striking degree of variation that is highly correlated with large domains of increased recombination, which is a pattern observed in many species (Andersen et al. 2012; D. Lee et al. 2021; Kern and Hahn 2018; Rockman and Kruglyak 2009). In *C. elegans,* these divergent autosomal arm-like domains coincide with a disproportionate fraction of newer, rapidly evolving genes as compared to the center regions of each chromosome, which house highly conserved essential genes (C. elegans Sequencing Consortium 1998; Kamath et al. 2003). The development of new tools to detect larger structural variations through alignment of assemblies or long sequencing reads has revealed many SVs on the chromosomal arm-like domains (Mahmoud et al. 2019; C. Kim et al. 2019). The fact that SVs are enriched in the arm-like regions, which also display elevated levels of recombination, is notable given the fact that large structural variants such as inversion are typically inhibitory to recombination (Miller, Cook, and Hawley 2019). The arm-like regions of *C. elegans* chromosomes are enriched for many repetitive elements, including transposable elements, tandem repeats, and low complexity repeat sequences (C. elegans Sequencing Consortium 1998; Surzycki and Belknap 2000). The presence of many SVs in the arm-like regions could be due to errors in double-strand DNA break repair and heterologous

recombination in regions adjacent to highly repetitive sequences, thereby causing chromosomal rearrangements. Similar rearrangement events are known to contribute to many human genomic disorders like Prader-Willy Syndrome or Charcot-Marie-Tooth disease (Carvalho and Lupski 2016; Stankiewicz and Lupski 2010). Future investigations assessing the occurrence of SVs adjacent to highly repetitive regions and sites of homologous recombination will be invaluable in understanding how differences in genomic organization arise between divergent lineages of *C. elegans*.

An emergent role of new TE superfamilies in genomic structural variation

With regard to genomic rearrangements and their impact on genome function, renewed attention must be given to the contribution of transposable elements and their mobility within and between chromosomes. While Sola and Zator elements are relatively recent in theirdiscovery within *C. elegans* and other eukaryotic genomes (Bao et al. 2009; Riehl et al. 2022), our data suggests there may be many active TE copies in these families, particularly Zator elements. Historically, much attention has been given to the impact of *Tc1/Mariner* transposition on genomic architecture, but the contribution of Zator elements to changes in genome structure and gene regulation merits further future investigation. Our analysis of TE mobility only examines two endpoints across the long period of divergence between the Bristol and Hawaiian lineages. It remains unclear, however, whether many of these newly characterized TEs remain active and whether they contribute to the growing catalog phenotypic differences displayed between laboratory lineages of Bristol and Hawaiian *C. elegans*.

The role of chromatin in shaping patterns of genomic variation

Our data shows that many SVs are enriched in silenced, heterochromatic regions of the *C. elegans* genome, which supports the growing body of evidence for local variations in mutation rates. New sequence variants and structural variations are the foundation of genome evolution. Short sequence variants, SNPs and indels, are the most common type of genomic variant, and studies have found that mutation rates for these short sequence variants are not equal across the genome (Wolfe, Sharp, and Li 1989; Hodgkinson and Eyre-Walker 2011; Hardison et al. 2003). Chromatin modifications and chromatin accessibility likely play a significant role in how variants accumulate in structurally distinct compartments of the genome (Makova and Hardison 2015). For instance, differences in chromatin organization within somatic cancer cells can shape local heterogeneities in the mutation rate (Schuster-Böckler and Lehner 2012). Germline variants have much more profound impacts than somatic variants in shaping the genetic variation of individuals between populations, and they produce lasting effects on the evolution of genome structures within species (Yu et al. 2024). Our results suggest germline chromatin states may be influence genome evolution, thereby serving as the foundation for future research to understand how distinct chromatin states directly affect the specific rates at which SNPs, indels, or SVs accumulate across the genome.

How might chromatin states and sequence context dictate where variations appear? Base substitutions and short indels that occur naturally in regions of open chromatin may be more likely to be repaired. DNA repair proteins and other factors that monitor sequences mismatches are more likely to eliminate new mutations and variants if the region is physically accessible (Prendergast et al. 2007).Further, we found an abundance of variant sites nested within TE

sequences, which also reside in heterochromatic domains. In the case of TEs, one hypothesis for this enrichment is that the accumulation of variants in in TEs could be protective of genome integrity. TE insertion into genes or regulatory sequences can be highly disruptive, and there are multiple mechanisms to silence their mobility to protect genome integrity (Lawson, Liang, and Wang 2023). For example, an enhanced mutation rate in TEs provides an alternative method to silence their movement since mutations could disrupt TE excision (Lohe, De Aguiar, and Hartl 1997). TE sequences, therefore, could be more likely to accumulateboth small sequence variants as well as structural variations. Taken together, the landscape of chromatin states that silence specific genes or TEs is likely contributing to patterns of sequence divergence and changes in genome structure.

Conclusion

In summary, our *de novo* generation of long-read genome assemblies for both N2 Bristol and CB4856 Hawaiian isolates has revealed a large extent of genomic variation and potential mechanisms that lead to these changes. Our findings suggest that heterochromatic regions of the genome are uniquely susceptible to give rise to many small and large genomic variants. These silenced regions of the genome are likely strong contributors to the large-scale differences in genome structure between individuals and across populations and may underpin a greater proportion of the phenotypic plasticity and genetic variation than previously appreciated. In conclusion, we generated and characterized new reference genomes of different wild-type isolates of *C. elegans* to comprehensively detect sequence variations, which revealed how

different chromatin states and the extent in which TEs contribute to the evolution of genome structure.

Methods

C. elegans culture and sucrose floatation

The N2 Bristol and CB4856 Hawaiian strains of *C. elegans* were grown at 20°C on standard NGM agar plates seeded with the OP50 strain of *E. coli* as a food source. To minimize bacterial contamination in downstream gDNA sample preps, we performed sucrose floatation on pooled populations of each isolate. Worms were washed from plates with 8mL cold M9 buffer and transferred to 15mL glass centrifuge tubes using a glass Pasteur pipette. Collected worms were centrifuged at 3000rpm at 4°C and washed in 4mL of fresh M9 twice. To separate worms from bacteria and other debris, 4mL of 60% sucrose solution was added to 4mL of M9 buffer and worms and vortexed briefly. The mixture was then spun at 5000 rpm at 4°C for 5 minutes. Using a glass pipette, the floating layer of worms were transferred to a new glass centrifuge tube on ice and brought up to 4mL in fresh M9. Worms were then incubated at room temp for 30 minutes and gently vortexed every 5 minutes. Worms were washed three times in equal volume of fresh M9 were performed before storing collected worms in M9 at 20°C before genomic DNA (gDNA) extraction.

Long-read and short-read sequencing

Genomic DNA was extracted from worms using the Qiagen DNeasy Blood and Tissue Kit. Sequencing was performed on pooled populations of N2 and CB4856 after reducing

bacterial contamination by sucrose float for each strain. For PacBio long-read sequencing, library preparation was performed on pooled populations of worms for each isolate by the University of Oregon's Genomics and Cell Characterization Core Facility and sequenced on the Sequel II system. Raw PacBio subreads were then refined into circular consensus sequencing (CCS) and assessed for read accuracy (Supplemental Figure 2.1). CCS reads were then used for all further applications. For Illumina short-read sequencing, library preparation was performed on pooled populations of worms for each isolate by the University of Oregon's Genomics and Cell Characterization Core Facility. The short-read libraries were then sequenced on an Illumina HiSeq4000 (2 x 150 bp).

Long-read genome assembly and short-read refinement

PacBio long-reads were aligned to the E. coli genome using BWA (Li and Durbin 2009) (version 0.7.17), and reads that aligned to the bacterial genome were removed. *De novo* genome assembly was performed for N2 Bristol and CB4856 Hawaiian using Canu (Koren et al. 2017) (version 1.7). To refine the long-read assemblies, short reads from each isolate were aligned to their respective long-read assembly using BWA-MEM (version 0.7.17). Aligned reads in SAM format were sorted and converted to BAM format using SAMtools (Li et al. 2009). Using Picard (https://broadinstitute.github.io/picard/), read groups were added via AddOrReplaceReadGroups, and duplicate reads were filtered using MarkDuplicates. Some bases may have been inaccurately called due to lower sequencing coverage, larger error rate in PacBio sequencing, or predominating alleles present in the population of each isolate that could be revealed by greater sequencing depth afforded by Illumina sequencing. GATK's HaplotypeCaller (McKenna et al.

2010) and Freebayes (Garrison and Marth 2012) were utilized to generate VCF files representing potentially inaccurate sites in each initial assembly. Coverage thresholds were manually determined using the Integrative Genomics Viewer (IGV) for each assembly(Robinson et al. 2011). Sites were filtered according to manual values using VCFtools (Danecek et al. 2011; Danecek and McCarthy 2017)). Error correction was performed on single-nucleotide alleles using BCFtools *consensus* (Danecek and McCarthy 2017) and alternate indel alleles. After filtering potential sites by sequencing depth thresholds determined for each chromosome, this left 4,237 and 36,145 corrections for the N2 Bristol and CB4856 Hawaiian genomes, respectively. Of these sites, less than 0.7% were unable to be resolved, and all of these were short indels comprising less than 0.001% of each genome.

Assessing genome assembly completeness

To determine whether any assembly artifacts remained after polishing, sequencing reads from each genome were then re-aligned to their respective, completed genome assembly. No significant losses in read alignment or coverage were observed. To further assess the quality and completeness of our N2 Bristol and CB4856 Hawaiian assemblies, we used BUSCO (Simão et al. 2015; Manni et al. 2021). BUSCO was run in a Docker container (https://busco.ezlab.org/busco_userguide.html) in genome mode. For each assembly, the quality and presence of expected orthologous genes was checked against the nematoda and metazoan lineage databases.

SNP and indel Calling in N2 and CB4856 assemblies

Illumina short reads from the N2 Bristol and CB4856 Hawaiian genome were trimmed using Trimmomatic (Bolger, Lohse, and Usadel 2014) to remove adapter and barcode sequences. The trimmed CB4856 reads were then aligned to the N2 Bristol reference genome using BWA-MEM so that SNPs and indels present between N2 Bristol and CB4856 Hawaiian could be identified. All resulting variant positions comparing our N2 Bristol and CB4856 Hawaiian genomes are in relation to the N2 Bristol assembly. Aligned reads in SAM format were then sorted using SAMtools (Li et al. 2009) and converted to BAM files. Using Picard read groups were added via AddOrReplaceReadGroups, and duplicate reads were filtered using MarkDuplicates as described above. BAM files with filtered duplicate reads were used to call variants using a combination of GATK HaplotypeCaller, Freebayes, and BCFtools. The three resulting VCF files containing SNPs and indels were then concatenated, further filtered for duplicate sites and low-quality variants, and sorted using BCFtools. SNPs with QUAL scores of 30 or greater, a minimum of 10 variant reads, and a minimum of 30 total, high-quality reads were retained.

Calling Structural Variants using whole-genome alignments

All assembly-to-assembly alignments were performed using Minimap2 (Li 2018). SyRI (Goel et al. 2019) was then used to parse the resulting SAM files and call structural variants and highly divergent regions (Structural rearrangements were plotted with the aid of Plotsr within the SyRI package. “NOTAL” or non-alignable regions in each genome were retained as SVs. To acquire NOTAL regions in each query genome, the Minimap2 alignment was repeated with the

original reference and query genomes swapped. The sizes of HDRs depicted in Tables 1-3 are sizes relative to the reference genome in each comparison (*i.e.* N2 Bristol in Table 2.1). HDRs called by SyRI represent regions of pairwise alignment between reference genome assemblies where there are multiple gaps present at the same corresponding locus in each genome. These could be due to the presence of multiple SNPs, indels, or SVs of varying complity that each genome acquired independently at the same site. These are distinct from "hyper divergent regions" as seen in prior studies (Thompson et al. 2015; D. Lee et al. 2021) that use different approaches (e.g. sliding window analyses) to measure the local density of variants in kilobase scale regions. All SV and HDR calls from SyRI are available in Supplementary File 1.

Converting gene annotations between assemblies

We converted gene annotations from the N2 reference assembly (cel235) to our N2 Bristol and CB4856 Hawaiian assemblies. The gene annotations for the WBcel235 genome assembly were downloaded in GFF3 format from Ensembl (http://ftp.ensembl.org/pub/release-105/gff3/caenorhabditis_elegans/). Unlike previously established tools that require pre-generated chain files (James et al. 2003), Liftoff (Shumate and Salzberg 2021) can accurately remap gene annotations onto newly generated assemblies using Minimap2 assembly-to-assembly alignments. Rather than aligning whole genomes, Liftoff aligns only regions listed in the annotation files so that genes may be remapped even if there are large structural variations between two genomes. The Liftoff program was then used to remap annotations between the WBcel235 assembly onto each new genome assembly for N2 Bristol and CB4856 Hawaiian.

Testing the association of variant sites in gene annotations and chromatin profiles

For each chromosome, to determine whether SNPs or indels were enriched within gene annotations, fold enrichment analyses were performed using the genomic association tester (GAT) (Heger et al. 2013) tool (https://github.com/AndreasHeger/gat.git). The observed enrichment of each variant type in gene annotations was compared to overlaps in simulated distributions SNPs or indels. Simulated distributions were created using 20,000 iterations whereby each variant type was randomly and uniformly distributed across each chromosome. SNPs and indel distributions were compared against intergenic, gene, intron, exon, and UTR annotations. Comparing the observed enrichment to the simulated distributions, statistical significance was assigned to the observed fold enrichment with p-values calculated from a hypergeometric test calculated within GAT.Per-chromosome BED files for SNP intervals were created from their original VCF using AWK. Per-chromosome BED files for indel intervals were calculated using a custom script. The GFF3 formatted annotations generated via liftoff were then broken down by chromosome, gene, exon, and UTR regions. Because intron regions were not explicitly written into each GFF3 file, they were calculated using BEDtools (Quinlan and Hall 2010). First, a joint BED file containing the UTR and exon regions were made using awk and sorted first by chromosome then by position. Using BEDtools these intervals were combined, and intronic regions were calculated by finding regions in gene intervals not covered by either UTR or exons. Intergenic spaces on each chromosome were calculated with the gene BED files and chromosome sizes as inputs. Germline ChIP-seq (NCBI BioProject PRJNA475794) and ATAC-seq (NCBI Gene Expression Omnibus accession GSE141213) data were aligned to the N2 genome and peaks were called in MACS3 (Feng et al. 2012) using parameters as previously

described (Tabuchi et al. 2018; Serizay et al. 2020). Germline-specific genomics datasets were chosen as mutations and variations that arise in the germline are more likely to persist through development and fix in larger populations (Yu et al. 2024).

Transposable Element Identification and Tracking

The TransposonUltimate pipeline was run for both our N2 Bristol and CB4856 Hawaiian genome assemblies. MUST and SINE finder were run independently and integrated into the filtering steps of the pipeline manually. Additionally, we added LTR retriever to the TE identification ensemble to supplement LTR harvest and LTR finder. TE sequences that overlapped with SNPs were identified using BEDtools. CB4856 Hawaiian SNPs were applied to corresponding N2 Bristol TE sequences, and these sequences were cross-referenced with the original TransposonUltimate output for CB4856 Hawaiian for matches. Unique polymorphic TE sequences found in both genomes were then assessed for translocation events by examining genomic start coordinates in each genome. Utilizing whole-genome alignments for each chromosome, TEs were predicted to have moved if starting coordinates for each TE pair were did not correspond to relative changes in coordinates due to alignment.

Data Availability Statement

The PacBio long-read and the Illumina short-read data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA907379. Code for our transposon annotation and tracking method can

be found on github (https://github.com/libudalab/transposon-tracking).All custom scripts are available upon request. Strains are available upon request.

Acknowledgements

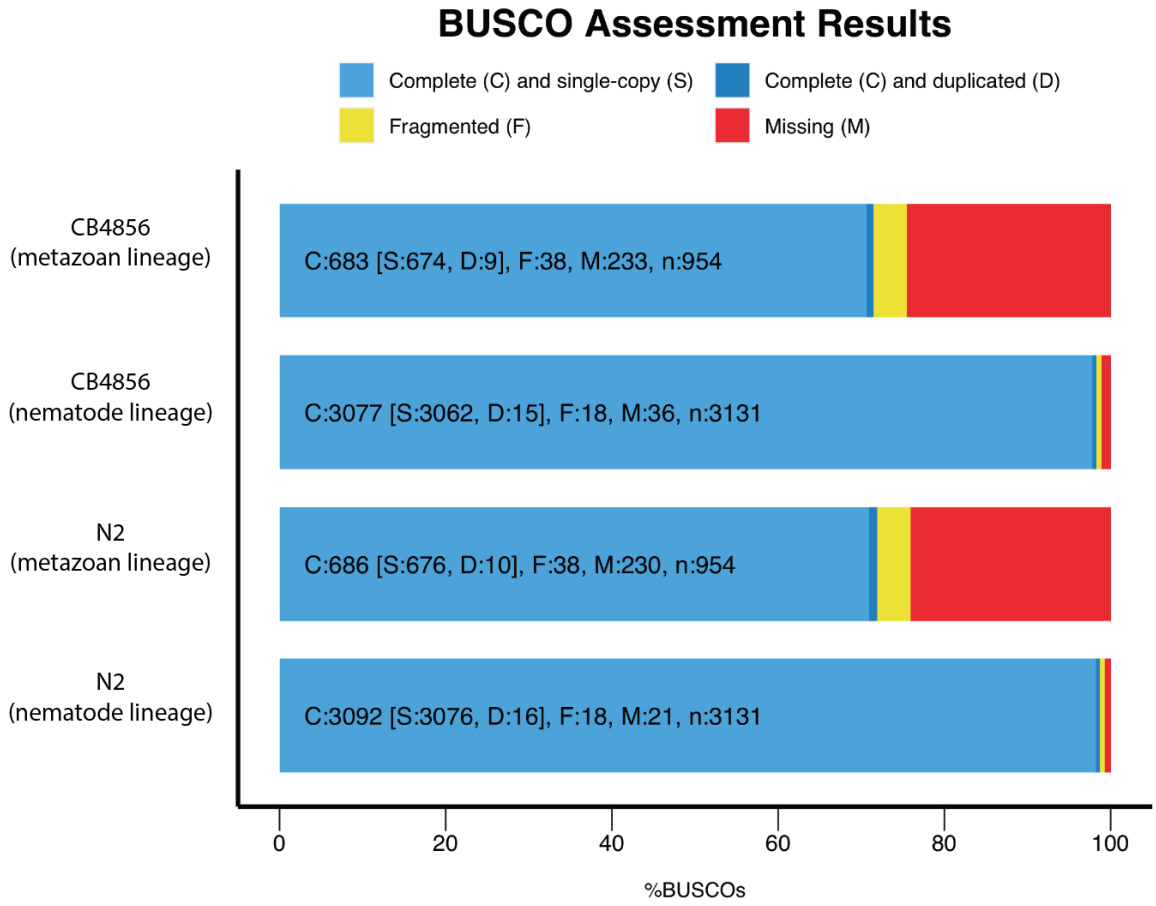
We thank N. Kurhanewicz, C. Cahoon, E. Toraason, and J. Conery in the Libuda Lab for thoughtful discussion and comments on the manuscript. We are grateful to the University of Oregon's Genomics and Cell Characterization Core Facility for sample prep and sequencing of our N2 Bristol and CB4856 Hawaiian strains. Over the course of many years beyond the time of this study, thanks for inspiration from Walter Rogers and invaluable guidance from Dr. Glenn C. Rowe given to Z.D.B.

Funding

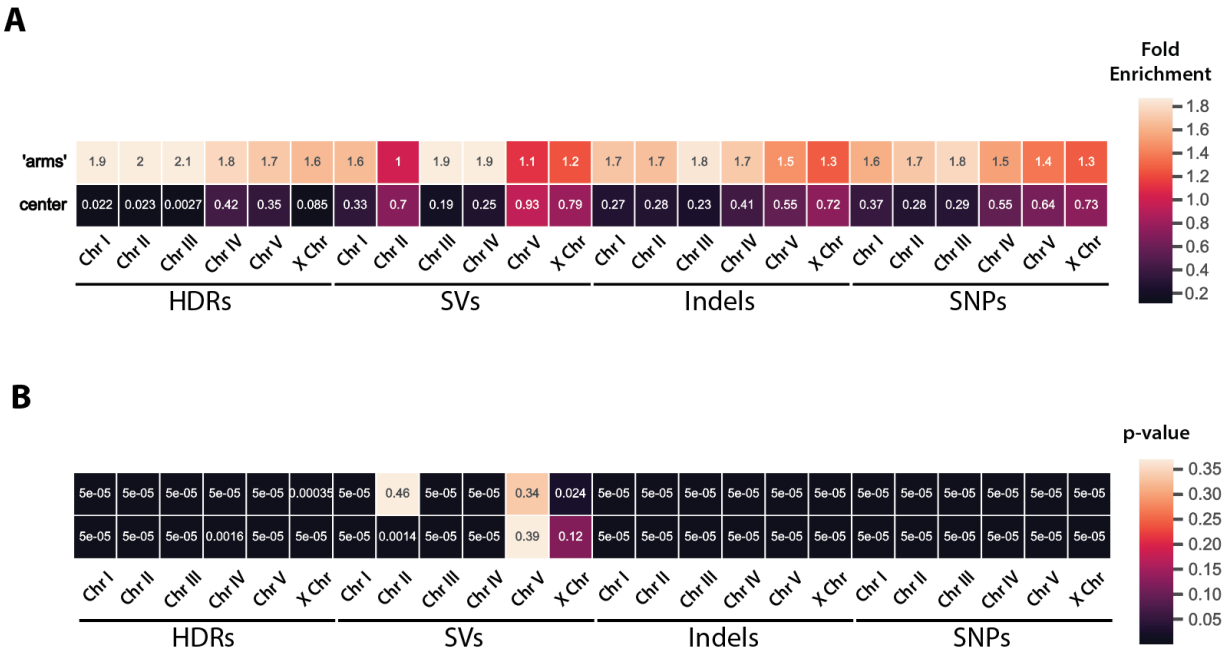
This work was supported by the National Institutes of Health T32HD007348 to Z.D.B and National Institutes of Health R35GM128890 and University of Oregon start-up funds to D.E.L.

Conflict of Interest

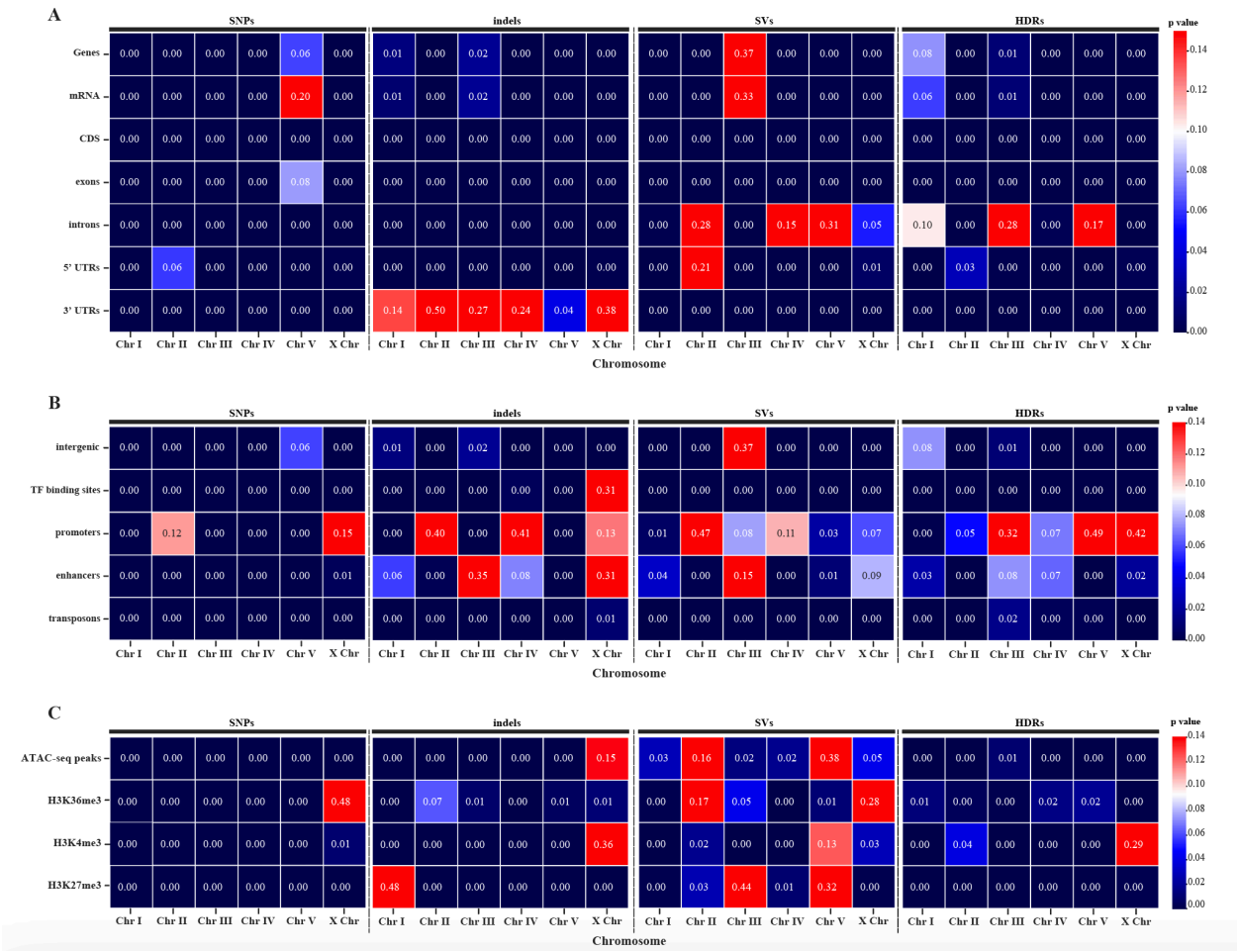
The authors declare no conflicts of interest.



Supplemental Figure S2.1. BUSCO analysis of the N2 Bristol and CB4856 Hawaiian genome assemblies. The presence of orthologous genes from metazoan and nematode lineages are shown for each genome assembly. Each orthologous gene analyzed is depicted as either Complete (C, blues), Fragmented (F, yellow), or Missing (M, red). Complete orthologs are then further categorized as single-copy (S, light blue) or duplicated (D, dark blue).



Supplemental Figure S2.2. Test of association results analyzing the overlap of CB4856 Hawaiian SNPs, indels, and SVs with arm-like versus center chromosome domains. A) Heatmap showing the fold enrichment of each variant type within each region for each chromosome. B) Heatmap of p-values associated with corresponding fold enrichments shown in panel A calculated by the hypergeometric test.



Supplemental Figure S2.3. Statistical significance of variants enriched/depleted in sequence annotations and chromatin profiles. A) Heatmap showing the p-values corresponding to enrichment or depletion of each variant type in genes and sub-gene annotations. B) Heatmap showing the p-values corresponding to enrichment or depletion of each variant type in intergenic regions, gene regulatory sequences, and transposons. C) Heatmap showing the p-values corresponding to enrichment or depletion of each variant type in different chromatin profiles. All p-values listed as 0.0 are < 0.001.

Bridge to Chapter 3

In chapter 2, I demonstrated the effectiveness of using *de novo* genome assembly and comparisons between reference genomes assembled via the same methods in identifying the global distribution of genomic variation. By leveraging the use of genetically divergent populations within the same species, our studies show that the majority of variant sites between diverged populations are SNPs and indels, but SVs (though much fewer in number) affect the most base pairs by far. Further, we demonstrate a role for chromatin accessibility, modulated by specific chemical modifications to histones, in regulating the accumulation of these large genomic SVs and other variants. These studies demonstrate how specific chromosome structures affect the accumulation of variation over tens of thousands of years of geographical isolation and genetic divergence. While this work gives us insight into how natural populations undergo genetic drift, we can also learn something about the mechanisms that regulate genome function by comparing the genomes of different laboratory lineages of wild type model organisms. Labs passaging many generations of model organisms have shown both phenotypic divergence and genomic divergence at smaller scales, but the genomic divergence has yet to be explored at a genome-wide scale. In chapter 3, I explore the extent to which laboratory domestication and cultivation of Bristol and Hawaiian *C. elegans* leads to the accumulated SNPs, indels, and SVs in different genomic regions.

CHAPTER 3: ACCUMULATED GENOMIC VARIATION BETWEEN LABORATORY LINEAGES OF WILD TYPE *C. ELEGANS*

Introduction

DNA mutations are a source of genetic variation and the prerequisite for many evolutionary processes acting on populations. In the human genome, the rate and prevalence of single nucleotide polymorphisms (SNPs) is the highest, followed by small insertions/deletions (indels, < 50bp) and structural variants (SVs, ≥ 50bp), respectively (The 1000 Genomes Project Consortium 2012; Nesta, Tafur, and Beck 2021). Mutations, whether they affect one, thousands, or millions of base pairs, can have profound impacts on the development, health, and survival of individuals (Stankiewicz and Lupski 2010; Gagliano et al. 2019). Combining modern DNA sequencing technologies with the use of model organisms has greatly enhanced our detection and understanding of how genomic variation in the contexts of both developmental and evolutionary biology.

Model organisms used in labs are also subject to the same processes of mutation and evolution. Genetic drift and repeated selection for desirable phenotypes in model organisms can result in lab-to-lab variation in the genetic background of wild type strains. Studies in bacteria, yeast, plants, and vertebrates demonstrate that laboratory cultivation of model organisms leads to the accumulation of genomic variation with functional consequences on processes like metabolism and reproduction (Yalcin et al. 2004; Guryev et al. 2006; Alonso-Blanco et al. 2003; Bentsink et al. 2006; Daranlapujade et al. 2003; Bradley et al. 2016). Notably, some inbred lines of mice have significant differences in their mutation rate (Chebib et al. 2021; Uchimura et al.

2015). Thus, studies into how genetic variation between different lineages of laboratory wild types affects the practice and interpretation of research is required.

Much of the existing work characterizing genetic variation between wild type strains focuses on the accumulation of SNPs and indels (Daranlapujade et al. 2003; Carreto et al. 2008; Y. Wang et al. 2018; Guryev et al. 2006). In contrast, the extent and impact of large genomic SVs between the genomes of laboratory lineages of wild type models remains underappreciated. Current long-read sequencing technologies have greatly aided the detection of SVs in humans and revealed that SVs between human populations affect many more base pairs than SNPs and indels (Hurles, Dermitzakis, and Tyler-Smith 2008; Sudmant et al. 2015; Sedlazeck et al. 2018). Large genomic SVs could pose disruptions to daily research practices if, for example, thousands of bases have been deleted or rearranged in an area targeted by PCR or CRISPR. Further, SVs are known to disrupt gene expression and lead to disease phenotypes such as cancer (Hurles, Dermitzakis, and Tyler-Smith 2008; Stankiewicz and Lupski 2010; Sakamoto et al. 2021). Knowing that disruptive SVs are likely accumulating in laboratory model organisms requires a more comprehensive analysis of genomic variation between laboratory model systems.

Caenorhabditis elegans is an excellent model organism to study the accumulation of mutations in laboratory model organisms. Most *C. elegans* researchers use the N2 Bristol isolate as the canonical wild type since it was first isolate established and extensively characterized in a laboratory setting by Sydney Brenner in the 1970s (Brenner 1974; Sulston and Brenner 1974). As a point of comparison to N2, many labs also use the CB4856 Hawaiian isolate to study genomic variation, recombination, and evolution. (J Hodgkin and Doniach 1997; Koch et al. 2000; Wicks et al. 2001; Rockman and Kruglyak 2009; Thompson et al. 2015; Crombie et al.

2019). While the remarkable sequence diversity between the N2 Bristol and CB4856 Hawaiian wild types has been the focus of *C. elegans* genomics studies, recent work on germline mutation rates, however, suggest that considerable genetic variation has also accumulated within the N2 Bristol strain since laboratory domestication (Denver et al. 2009). The germline rate of mutation accumulation is 2.7×10^{-9} mutations per site per generation (Denver et al. 2009) and the generation time is approximately three days. Depending on how often labs return to frozen stocks, each laboratory N2 lineage alone may have accumulated up to ~1,500 single nucleotide mutations since the 1970s and nearly 790 potential mutations since the first genome was published in 1998 (C. elegans Sequencing Consortium 1998). Notably, this predicted variation does not include the accumulation of multi-nucleotide indels and genomic SVs. While isolated examples of genomic variation in N2 between labs has been documented, a modern genome wide analysis has yet to reveal the full extent of genetic drift between labs.

The genomes of N2 Bristol and CB4856 Hawaiian that are currently in individual labs likely carry considerable genomic variation relative to other labs isolates. Previous studies using earlier genome assemblies have identified many segmental duplications between lab lineages of *C. elegans* wild type strains (Vergara et al. 2009) as well as duplications ranging in size from 200bp to 108kb that affect as many as 26 genes (Vergara et al. 2009). This variation may underpin phenotypic variation as well as previous work that has shown the lifespans of laboratory N2 Bristol isolates varies between 12-17 days (Gems and Riddle 2000). Further, prior research suggest that inter-lab genetic variation in wild type backgrounds also contributes to differences in reproduction, feeding, sensory signaling, and social behaviors with likely impacts on experimental outcomes (Sterken et al. 2015; Andersen et al. 2014; Duveau and Félix 2012;

Weber et al. 2010; McGrath et al. 2009). Given the growing evidence for phenotypic divergence between canonical wild type *C. elegans,* it is increasingly important to understand the underlying genomic changes that lead to these differences. Thus, high-quality lab-specific reference genomes are an important tool to understand how genetics influences the phenotypes and processes studied by different laboratory groups.

Recently, several labs produced *de novo* genome assemblies for their lineages of the Bristol and Hawaiian isolates using a combination of Illumina, PacBio, and Oxford Nanopore platforms (Yoshimura et al. 2019; Kim et al. 2019). Compared to the previous short-read based assemblies of N2 Bristol, the new assembly of N2 Bristol, called VC2010, identified 53 more predicted genes, 1.8Mb of additional sequence, and eliminated 98% of existing gaps in the N2 Bristol genome. Thus, the overall structure of the VC2010 Bristol genome very likely better represents the genome of Bristol *C. elegans* currently used in laboratories worldwide (Yoshimura et al. 2019). In 2019, the first *de novo* CB4856 Hawaiian assembly from long-read sequencing extended the length of the Hawaiian genome, and was further able to characterize over 3,000 previously uncharacterized SVs (Kim et al. 2019). Having multiple long-read assemblies of the Bristol and Hawaiian isolates from different labs provides the first opportunity for a modern analysis of not only SNPs and indels, but previously uncharacterized SVs as well.

To determine the extent of genetic variation between our laboratory lineages of N2 Bristol and CB4856 Hawaiian, we compared the Bristol and Hawaiian genome assemblies from our lab (discussed in Chapter 2) to the VC2010 Bristol (Yoshimura et al. 2019) and Kim CB4856 Hawaiian (Kim et al. 2019) genomes, respectively. We identified SNPs, indels, and SVs unique to the wild type strains in each lab. Between the different lab lineages of the Bristol strain, we

identify over 2.01 megabases of genomic variation. When comparing lab lineages of the CB4856 Hawaiian isolate, we identify over 6.92 megabases of sequence variation. Notably, more than 99% of the total genomic variation between lineages of each isolate is due to SVs. We find that SNPs, indels, and SVs were enriched in intergenic regions of the *C. elegans* genome, suggesting that variations in regulatory sequences and other non-coding regions may underlie the phenotypic variances previously observed between laboratory strains. Taken together, our systematic analysis of genetic variation between natural and laboratory wild type isolates highlights the impact of large structural variants and other chromosomal rearrangements accumulating in the genomes of laboratory model organisms.

Results

Genome divergence between laboratory lineages of the Bristol isolate

To assess how much genetic variation may exist between lab lineages of the most utilized wild-type strain N2 Bristol, we compared our N2 Bristol genome to VC2010 Bristol. Both Bristol assemblies were constructed from PacBio long reads, so we expect much of the structure of these genomes to be highly similar within alignable regions. An examination of synteny, colinear stretches of highly homologous DNA, indicates how well conserved the sequence and overall structure is between to genomes. We expected that examining whole-genome alignments to previously validated long-read assemblies would reveal a large degree of synteny and similarity, and whole genome alignments between the two N2 Bristol lineages revealed that 99.9% of bases were alignable and 99.3% of bases were syntenic (Table 3.1). To detect short sequence variations, we used Illumina short read alignments and identified 1,162 homozygous

SNPs and 1,528 homozygous indels. Further examination of our whole genome alignments revealed approximately 2.01Mb of genomic SVs. In total, over 2.07Mb were affected by SNPs, indels and SVs, with 99.7% of this sequence divergence due to structural variations (Table 3.1). Highly divergent regions (HDRs) identified between laboratory strains represent regional clusters of variants that create multiple gaps in pairwise alignments within regions of synteny (Goel et al. 2019). These can be due to the presence of multiple SNPs, indels, or SVs acquired independently by each genome at the same corresponding locus. While highly divergent regions have been observed (and defined differently) between wild populations of *C. elegans* (D. Lee et al. 2021; Thompson et al. 2015), we were also able to identify over 404kb of sequence as HDRs between these two laboratory Bristol lineages (Table 3.1). Overall, while the majority of DNA sequence is conserved between laboratory lineages of the N2 Bristol isolate, approximately 2% of the genome differs by large insertions, deletions, and rearrangements over 50 base pairs in length between the two lab lineages.

We then wanted to assess the global distribution of SNPs, indels, and SVs on each chromosome between the Bristol genome assemblies for the two lab lineages. SNPs, indels, and SVs in the Bristol genome are present in high numbers on the terminal megabases of each chromosome (Figure 3.1A-C). We did, however, find multiple 500kb regions with highly elevated SNP counts in the centers of chromosomes *IV, V*, and the *X* chromosome (Figure 3.1A). In contrast, only one 500kb region in the center of chromosome *V* showed a much higher accumulation of indels (Figure 3.1B). We identified more than 100 SVs on each chromosome except on chromosome *III* (Table 3.1). SVs appeared to cluster on the terminal thirds of

Table 3.1 Comparisons between the DLW N2 Bristol genome and VC2010 Bristol genome

	Chromosomes						Total
	I	***II***	***III***	***IV***	***V***	***X***	
DLW N2 Bristol Chromosome Length (this study)	15,114,068	15,311,845	13,819,453	17,493,838	20,953,657	17,739,129	100,431,990 bp
VC2010 Bristol Chromosome Length (Yoshimura et al., 2019)	15,331,301	15,525,148	14,108,536	17,759,200	21,243,235	18,110,855	102,078,275 bp
DLW N2 Bristol Bases Aligned	15,108,942	15,310,622	13,819,294	17,492,076	20,852,291	17,738,432	100,321,657 bp (99.89%)
% Syntenic Aligned Bases	98.83	99.83	99.47	99.12	99.05	99.50	99.28
SNPs*	169	124	164	209	280	216	1,162
Indels*	150	261	210	262	378	267	1,528 (3,465 bp)
SVs	113	134	83	228	175	164	897 (2,010,282 bp)
HDRs	8	14	10	21	24	11	88 (406,737 bp)

* All variants listed are only those for which the VC2010 Bristol genome was homozygous

each chromosome, where there are multiple instances of SVs 1-10kb in size (Figure 3.1C). Among these SVs, we identified two inverted duplications (5.4kb and 12.9kb on chromosomes *III* and *V*, respectively) and 39 simple inversions. Four of these inversions are over 29kb in size and account for 11.6% of all structural variation between our N2 Bristol and the VC2010 Bristol genomes. In summary, the distribution of variation on each chromosome is largely in the terminal domains, which resembles the broad pattern of variation observed when comparing Bristol and Hawaiian genomes (Chapter 2, Figure 2.1 and Supplemental Figure S2.2). There are, however, multiple punctuated regions of substantial sequence and structural variation in the gene dense centers of some chromosomes.

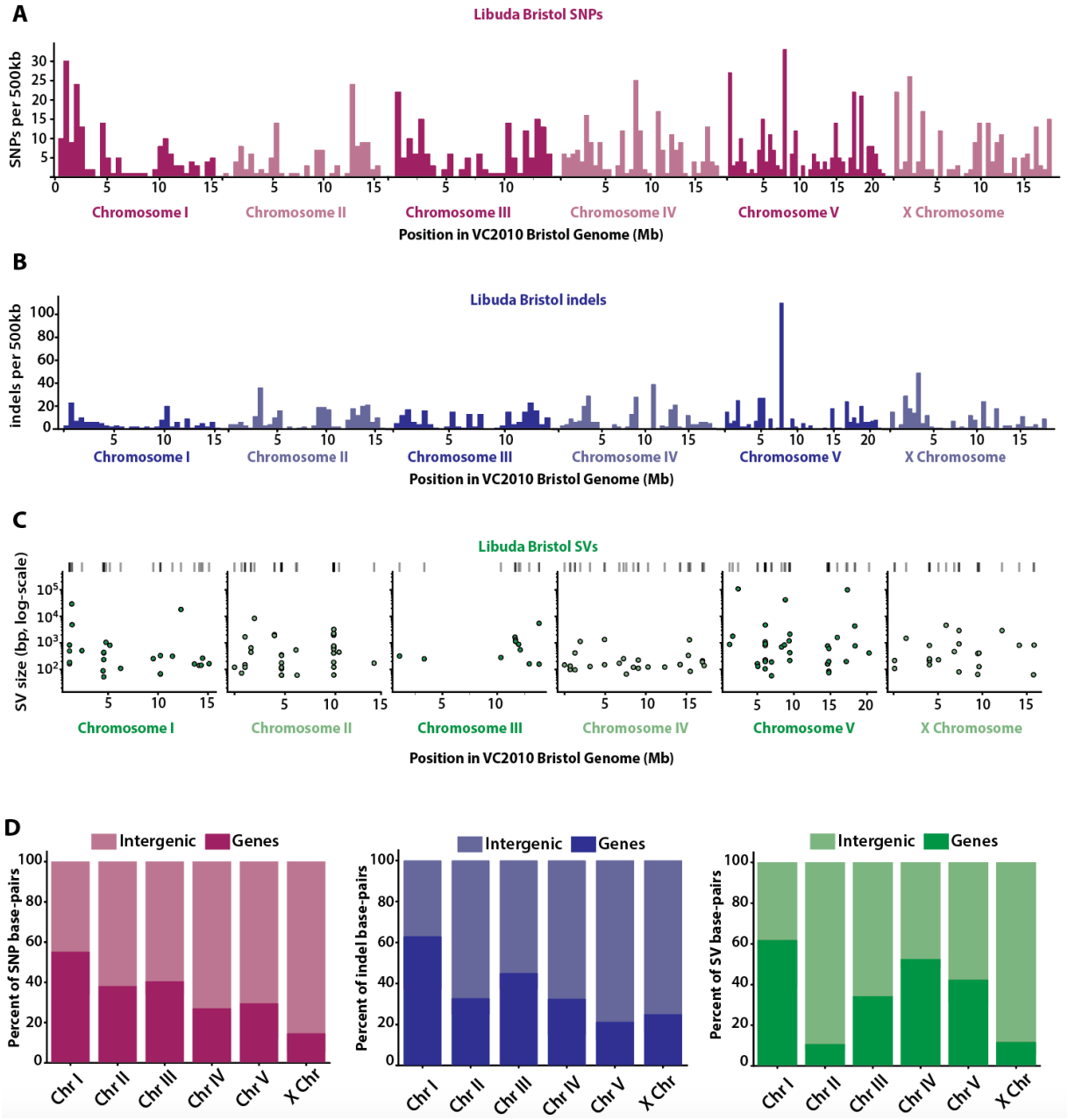


Figure 3.1. Genomic variation between the DLW N2 Bristol genome and the VC2010 Bristol genome. (A-B) Histograms depicting the distribution of DLW N2 Bristol SNPs and indels across each VC2010 Bristol chromosome in 500kb bins. All SNPs and indels represented include those that are both homozygous and heterozygous in the Libuda Bristol population. **(C)** Scatterplots showing the genomic position of SVs with the log-scaled size of each SV on the y-axis. Ticks above axes further indicate the genomic coordinates of each SV. **(D)** The proportions base pairs affected by DLW N2 Bristol SNPs, indels, and SVs that overlap with intergenic regions versus genes of the VC2010 Bristol genome.

Given that the accumulation of SNPs, indels, and SVs has the potential to disrupt the expression of genes, we next assessed how much sequence divergence overlapped with gene versus intergenic regions. For SNPs, indels, and SVs, most of the base pairs affected by these variants (55%, 63%, and 61%, respectively) overlapped with gene annotations on chromosome *I*. SVs on chromosome *IV* also overlapped with approximately 52% of genes (Figure 3.1D). For SNPs and indels, relatively fewer base pairs affected by these variants overlapped with genes (SNPs: 14.3-40.2%; indels: 21.2-45.2%) on the other autosomes and the *X* chromosome. Chromosomes *III* and the *X* chromosome had remarkably fewer SVs compared to *I, II, IV,* and *V*. Only 10.4% and 11.5% of base pairs affected by SVs on chromosomes *II* and the *X* chromosome overlapped with gene sequences. Taken together, much of the genetic variation between lab lineages of the Bristol isolate is intergenic except for chromosome *I.*

Intergenic enrichment of Bristol genomic variation on the chromosome arms

To determine if SNPs, indels, and SVs are also enriched in the terminal domains of each chromosome, we performed permutation and enrichment analyses of variants in these regions. When comparing the Bristol and Hawaiian wild isolates to each other, much of the genetic variation is enriched in the "arm" like domains of each chromosome (Thompson et al. 2015; C. Kim et al. 2019; D. Lee et al. 2021; Andersen et al. 2012). Given the shorter divergence time between laboratory lineages of the Bristol wild type, we wanted to confirm whether these lineages accumulated variation in the same pattern. Our broad findings demonstrate that SNPs, indels, and SVs are enriched on the arms of each chromosome (Figure 3.2A-B). We find SNPs were 1.2 to 1.5-fold enriched in the arm-like regions of chromosomes *I, II, III*, and *V* (p-values

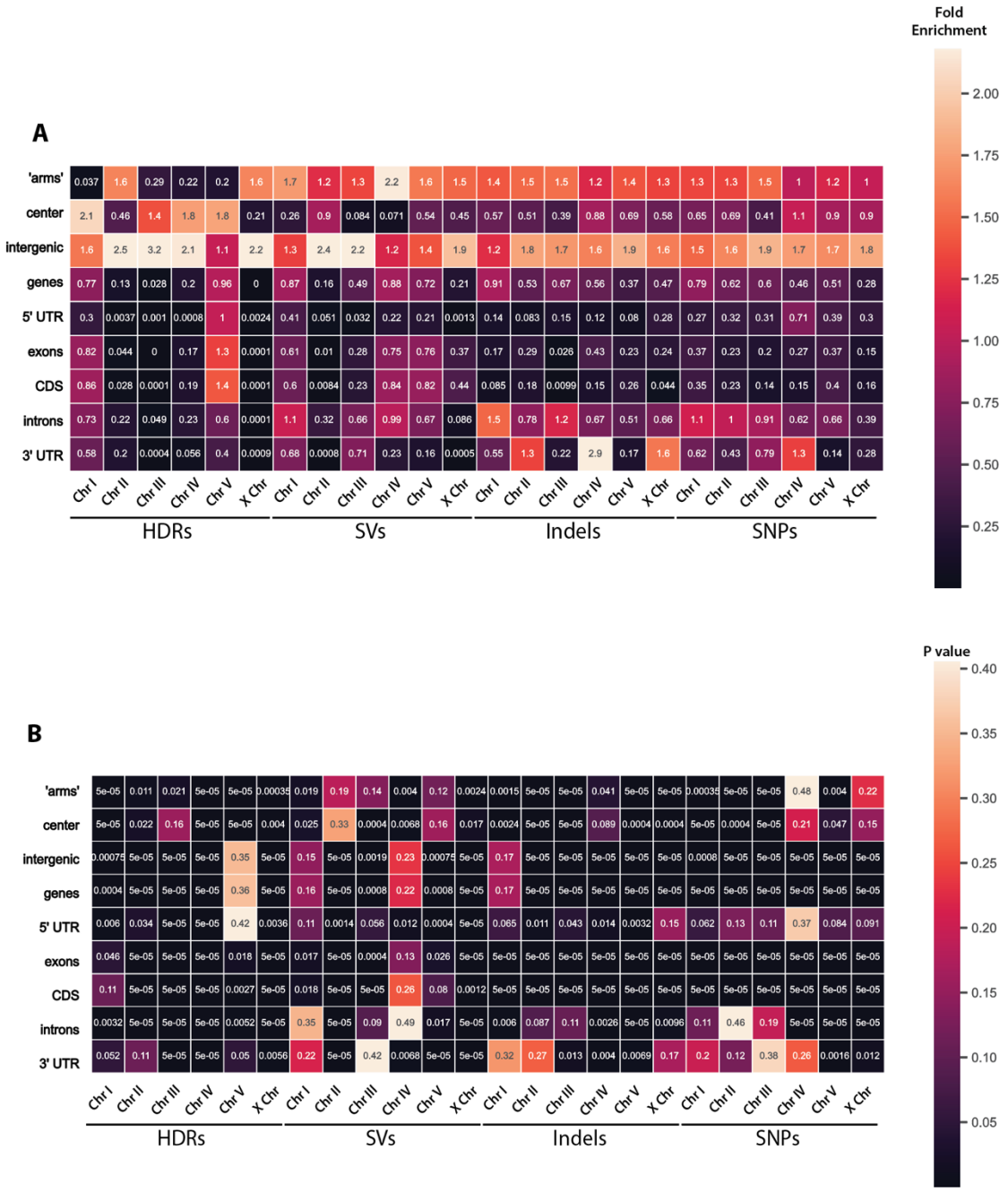


Figure 3.2. Intergenic enrichment of N2 Bristol variants. GAT interval-association test results analyzing the overlap of DLW N2 Bristol SNPs, indels, and SVs with remapped VC2010 Bristol genome annotations. A) Heatmap showing the fold enrichment of each variant type within gene annotations for each chromosome. B) Heatmap of p-values associated with corresponding fold enrichments shown in panel A calculated by the hypergeometric test.

<0.01). Indels, however, were concentrated in the arm-like regions of every chromosome with fold enrichments ranging from 1.2 to 1.5 (p-values < 0.05). While SVs were 1.2 to 2.2-fold enriched (p-values < 0.05) in the arm-like regions of each chromosome, this enrichment was only significantly higher than expected by null distributions on chromosomes *I, IV,* and the *X* chromosome (Figure 3.2A-B). Further, highly divergent regions between Bristol lineages were 1.6-fold enriched on the arm-like regions of chromosome *II* and the *X* chromosome (p-values < 0.05), while displaying significant 1.8 to 2.1-fold enrichments in the center regions of chromosomes *I, IV,* and *V* (p-values < 0.01). Thus, while much of the variation is in the arms of each chromosome, and examination of HDRS, which represent high density clusters of variation, are enriched in the gene dense centers of some chromosomes.

Finally, we also wanted to confirm whether the variant sites we detected between lab lineages of the Bristol isolate were depleted from genic regions. Between the two Bristol lineages, SNPs and indels were all significantly depleted from genes and sub-gene annotations, particularly exons/CDS annotations (SNPs: 0.14 to 0.37-fold, indels: 0.01 to 0.43-fold; p-values < 0.001) on every chromosome. SVs were significantly depleted from exons on every chromosome except *IV* (0.01 to 0.76-fold, p-values < 0.05), on chromosomes (Figure 3.2A-B). Although HDRs were enriched in intergenic regions on chromosomes *I, II, III,* and *IV* (1.1 to 3.2-fold, p-values < 0.001), a more careful examination of HDRs on chromosome *V* showed they are moderately enriched in exons and CDS annotations (1.3 and 1.4-fold, respectively; p-values < 0.05). Taken together, the 2.07 megabases of genomic variation between laboratory Bristol lineages is largely concentrated in non-coding intergenic regions of each chromosome.

Exceptionally dense genomic variation between CB4856 Hawaiian lineages

We then wanted to comprehensively analyze the genomic variation between lineages of the CB4856 Hawaiian wild isolate, which was discovered in 1978 (J Hodgkin and Doniach 1997) and used for comparative genetic studies. Due to differences in usage and generations passaged in labs, there may be differences in the magnitude of divergence relative to comparisons of the N2 Bristol lineage. Analysis of our CB4856 Hawaiian genome versus the Kim CB4856 Hawaiian genome showed that 96.1% of bases were alignable, with 92.3% of bases in syntenic alignments (Table 3.2). This proportion of syntenic alignments is approximately 7% lower than when comparing N2 Bristol genomes (Table 3.1, Table 3.2). Chromosomes *IV* and *V* display the lowest amount of synteny between laboratory lineages, with only 89.28% and 89.24% synteny, respectively. Thus, the genomes of the CB4856 Hawaiian strains present in labs likely harbor a greater degree of genomic variation than N2 Bristol lineages.

Our CB4856 Hawaiian lineage compared to the Kim CB4856 Hawaiian assembly revealed a greater number of base pairs affected by variation than comparisons between laboratory lineages of N2 Bristol. Small sequence variants were present at lower amounts across the whole genome, as we identified 541 homozygous SNPs and 1,298 homozygous indels (Table 3.2). SVs, however, were much more prevalent between Hawaiian genomes in contrast to N2 Bristol lineages. Our whole genome alignments allowed us to detect 2,070 structural variants that affect approximately 6.92 megabases of the genome (Table 3.2). Notably, approximately 66% of this structural variation is due to unique regions in either genome that cannot be aligned to each other (hereafter abbreviated as “NOTALs”). NOTAL regions are highly divergent with many

Table 3.2 Comparisons between the DLW CB4856 Hawaiian genome and Kim CB4856 Hawaiian genome

	Chromosomes						Total
	I	*II*	*III*	*IV*	*V*	*X*	
DLW CB4856 Hawaiian Chromosome Length (this study)	15,045,644	15,257,363	13,206,755	17,183,882	20,547,529	17,584,915	98,826,088
Kim CB4856 Hawaiian Chromosome Length (Kim et al., 2019)	15,528,896	15,813,191	14,110,336	17,985,219	21,389,866	18,073,349	102,900,857
DLW CB4856 Hawaiian Bases Aligned	14,620,886	14,680,704	12,451,582	16,482,840	19,427,050	17,371,269	95,034,331 (96.16%)
% Syntenic Aligned Bases	94.91	92.38	93.01	89.28	89.24	96.10	92.32
SNPs*	60	52	71	108	135	115	541
Indels*	240	190	175	242	238	213	1,298 (2,157 bp)
SVs	148	274	194	626	660	168	2,070 (6,923,335 bp)
HDRs	19	70	25	100	144	12	370 (3,327,407 bp)

* All variants listed are only those for which the Kim CB4856 Hawaiian genome was homozygous

gaps in pairwise alignments that often contain many tandem repeats and low-complexity sequences problematic for sequence alignment (Tørresen et al. 2019). Thus, roughly 2.31 megabases of genomic SVs between laboratory lineages of the Hawaiian strain can be attributed to large deletions, insertions, and rearrangements that can be accurately mapped by pairwise alignments. This magnitude of base pairs affected by SVs is similar what we observed in Bristol lineages (2.31 Mb vs 2.01 Mb, Table 3.1 vs Table 3.2). Further, over 3.3 Mb of variation in each Hawaiian genome can be categorized as HDRs, indicating many more punctuated regions of dense sequence divergence. Taken together, we identified more total variation between genome assemblies of the Hawaiian isolate than genomes of the Bristol isolate.

We then analyzed the genomic distribution of SNPs, indels, and SVs on each chromosome between the two CB4856 Hawaiian lineages. Our analysis found that SNPs, indels, and SVs tend to accumulate on the terminal domains of each chromosome (Figure 3.3A-B). Similar to comparisons between N2 Bristol lineages, we note multiple 500kb regions in the centers of chromosome *IV* and the *X* chromosome with elevated levels of small sequence variants (Figure 3.1A-B, Figure 3.3A-B). SVs in CB4856 Hawaiian are much more prevalent on chromosomes *IV* and *V* relative to other chromosomes (Table 3.2, Figure 3.3C), and they appear to follow a similar distribution pattern as SNPs and indels. Further, we find a greater proportion of SVs in the Hawaiian genome that are ten to hundreds of kilobases in size relative to SVs in the Bristol genome (Figure 3.3C). When we assessed how often SNPs overlapped with genes, we noted a markedly low count of these variants in coding regions compared to observations in the N2 Bristol lineages (Figure 3.3D). Only 1.9-8.7% of SNPs overlapped with gene annotations in the CB4856 Hawaiian lineages. In contrast, indels were present at greater levels in genes, which ranged from 16.7% on the *X* chromosome up to 37.7% on chromosome *I.* Notably, SVs displayed the highest overlap with gene annotations. Similar to indels, base pairs affected by SVs were least overlapped with genes on the *X* chromosome (23.9%). Most bases encompassed by SVs on chromosomes *I, III,* and *IV*, however, did overlap with coding regions (67.3%, 55.2%, and 50.7%, respectively). In summary, much of the genomic variation between laboratory lineages of the CB4856 Hawaiian isolate is due to structural variation, in which many overlap with genes on multiple autosomes.

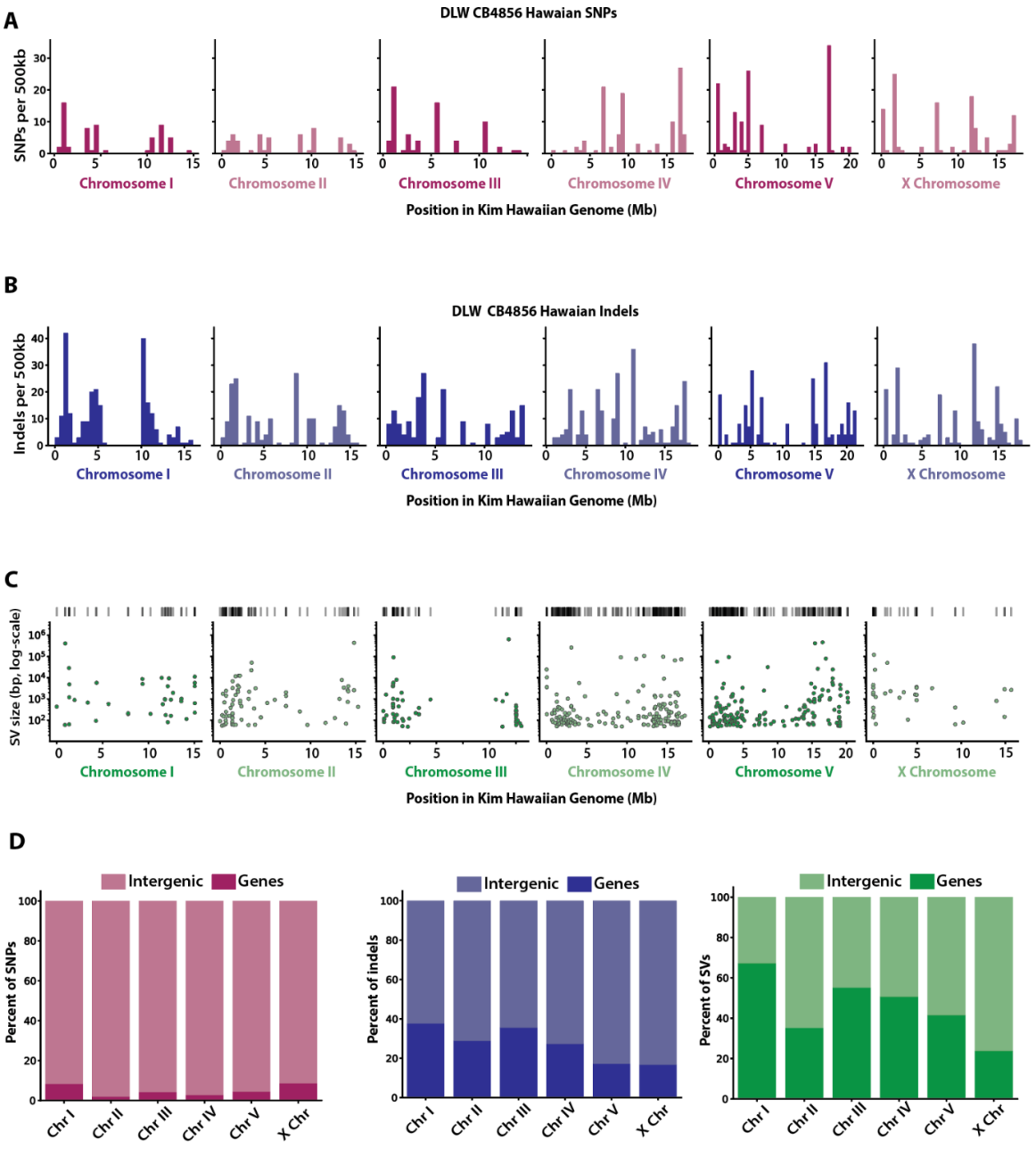


Figure 3.3. **Genomic variation between the DLW CB4856 Hawaiian genome and the Kim CB4856 Hawaiian genome.** (A-B) Histograms depicting the distribution of SNPs and indels across each Kim CB4856 Hawaiian chromosome in 500kb bins. (C) Scatterplots showing the genomic position of SVs with the log-scaled size of each SV on the y-axis. Ticks above axes further indicate the genomic coordinates of each SV. (D) The proportions of DLW CB4856 Hawaiian SNPs, indels, and SVs that overlap with intergenic versus gene-coding regions of the Kim CB4856 Hawaiian genome.

Depletion of CB4856 Hawaiian genomic variants in genic regions

We next examined the genomic distribution and enrichment of variant sites in gene annotations of the two CB4856 Hawaiian genomes to see if the patterns of enrichment were like the N2 Bristol lineages. All variant types on most chromosomes were enriched for variant sites in the arm-like regions and intergenic sequences, with a few exceptions (Figure 3.4). SNPs were only 0.84-fold depleted in the arm-like regions of the *X* chromosome, and indels were 0.81-fold depleted in the arm-like regions of chromosome *IV* (p-values < 0.05). In contrast, N2 lineages SNPs were found to be randomly associated (1.0-fold) and indels are 1.2-fold enriched on the arms of the *X* chromosome and chromosome *IV,* respectively. In CB4856 Hawaiian lineages, SVs were 0.64-fold depleted in the arm-like regions of chromosome *II*, and HDRs were 0.12-fold depleted in the arm-like regions of chromosome *I* (p-values < 0.001). In contrast to Hawaiian, N2 lineages showed strong depletions of HDRs on the arm-like regions (0.037 to 0.29-fold) on all chromosomes except *II*.

We then examined the enrichment of all variants in intergenic versus gene sequences between the two CB4856 genomes. Similar to N2 lineages, CB4856 Hawaiian SNPs and indels showed significantly high enrichment in intergenic regions on all chromosomes (1.7 to 2.8-fold, p-values < 0.001). SVs displayed a more moderate 1.2 to 1.7-fold enrichment in the intergenic regions of all chromosomes except chromosome *I* (all p-values < 0.05). HDRs were similarly 1.2 to 1.7-fold enriched in the intergenic regions of chromosomes *II, III, IV, V* and the *X* chromosome (all p-values < 0.05). Thus, laboratory lineages of the CB4856 Hawaiian isolate follow similar patterns of genetic variation in intergenic regions of the chromosome arms as seen in comparisons of N2 Bristol lineages.

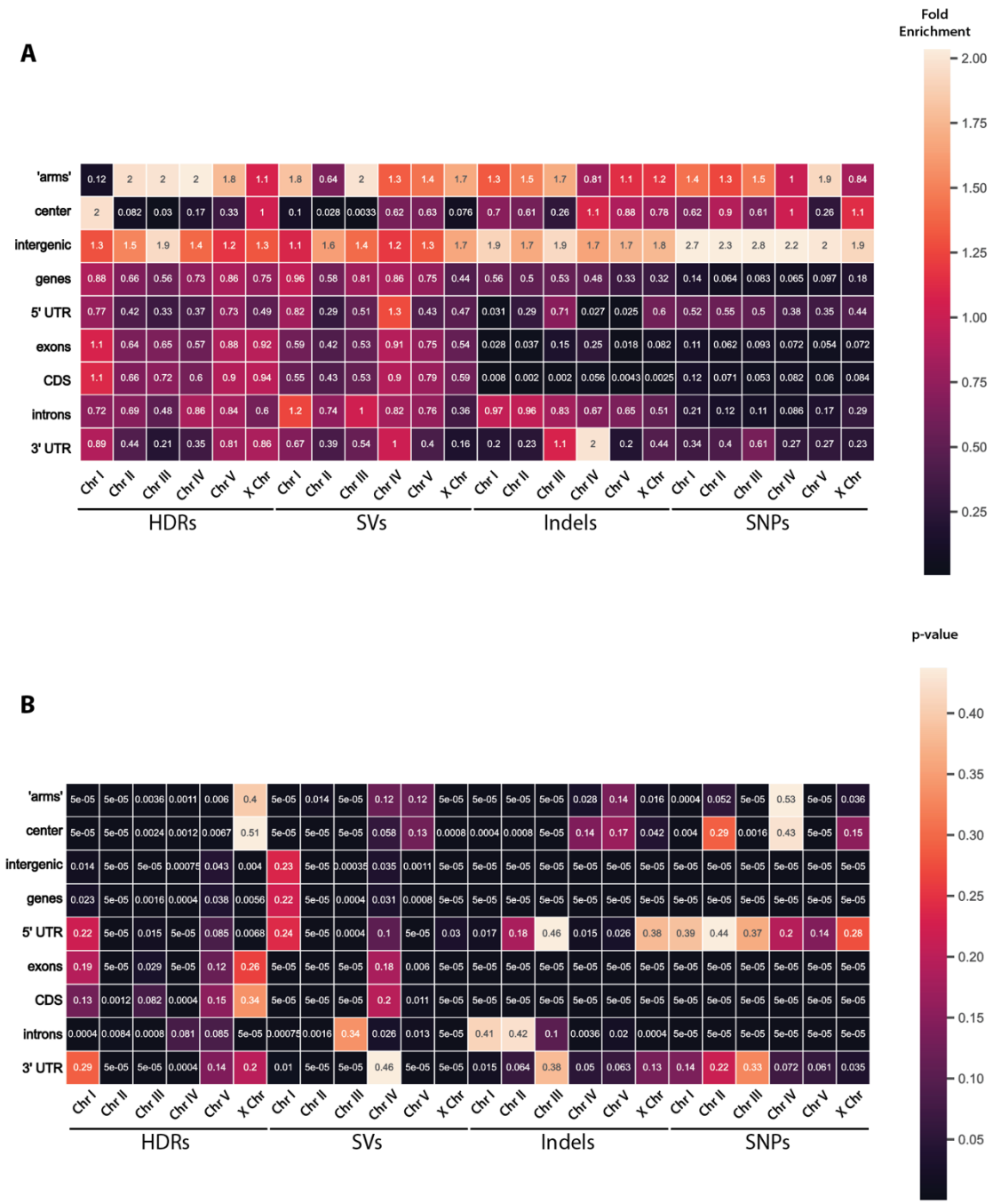


Figure 3.4. Non-coding variation dominates in the CB4856 Hawaiian genome. GAT interval-association test results analyzing the overlap of DLW CB4856 Hawaiian SNPs, indels, and SVs with remapped Kim CB4856 Hawaiian genome annotations. A) Heatmap showing the fold enrichment of each variant type within gene annotations for each chromosome. B) Heatmap of p-values associated with corresponding fold enrichments shown in panel A calculated by the hypergeometric test.

To test whether the high degree of overlap between SVs and genes on chromosomes *I*, *III*, and *IV* in the Hawaiian lineages, we tested for enrichment in sub-gene annotations. On chromosome *I,* SVs were significantly depleted from untranslated regions, exons and CDS (Figure 3.4A). Instead, SVs were enriched 1.2-fold higher than expected in intronic sequences (p-value < 0.001). On chromosome *III*, the overlap of SVs with genes is also largely observed in introns, although this is not higher than expected by random chance (Figure 3.4A). On chromosome *IV*, we noted a moderate 1.3-fold enrichment of SVs in the 5' untranslated regions of genes, though this is statistically insignificant in our sample size (p-value = 0.1). Overall, while we see that SVs and HDRs are depleted from genes and sub-gene annotations our analyses of Bristol lineages and Hawaiian lineages, we do note that the proportion of strong depletions (0 to 0.2-fold) is much higher in Bristol lineages (Figure 3.2A versus Figure 3.4A). In conclusion, analysis of the genetic variation between respective lab lineages of the CB4856 Hawaiian isolate revealed a striking amount of variation often present in intergenic regions and non-coding introns of genes.

Discussion

Our examination of the inter-lab genetic drift among wild-type strains suggests that laboratory domestication of multiple *C. elegans* isolates has led to the accumulation of substantial genomic variation. Much of the variation we identified lies within non-coding regions, but future investigation is required to determine whether any of the variants discovered in non-coding regions or the few that reside within coding regions lead to changes in gene function. Further, we find existence of large structural variations that could impact genomic-

based analyses and other experimental practices such as primer design. Overall, our work demonstrates the impact of long-term laboratory cultivation of strains in different laboratories on the genome sequence of established wildtype isolates. Lastly, the growing number of wild type genomes will serve as additional tools for future comparative genomics studies, especially in the functional characterization of structural variations identified through whole-genome alignments.

Genomic divergence of laboratory wild type lineages

Earlier studies uncovering phenotypic and genetic variations between lab wild-type *C. elegans* strains indicated that there are likely many underlying large-scale genomic differences (Denver et al. 2009; Vergara et al. 2009; Gems and Riddle 2000). Here we identify numerous SNPs, indels, SVs, and HDRs between different lab lineages of each wild isolate. The total amount of genomic variation is at levels higher than predicted by earlier mutation accumulation studies (Denver et al. 2009). Much of this variation, however, is due to SVs and HDRs, which have only recently become a detailed subject of study (Thompson et al. 2015; Kim et al. 2019; Lee et al. 2021). Our genome assemblies of the N2 Bristol and CB4856 Hawaiian strains corroborate prior results indicating that genomic variation is enriched in the distal arm-like regions of chromosomes between these natural isolates (Thompson et al. 2015; D. Lee et al. 2021). Evolutionary genomic analysis has shown that recombination in the arm-like regions of each chromosome and balancing selection likely have shaped this landscape of sequence divergence across the 30,000-50,000 generations these strains have been geographically isolated (Thomas et al. 2015; Kern and Hahn 2018). In contrast to comparisons between Bristol and Hawaiian genomes, we find that the distribution of variant sites across the arm-like regions

versus center domains between lab lineages is not as striking or consistent across each chromosome. This result could indicate that in relatively short timescales (~3,000-5,800 generations), selection for the accumulation of mutations in the arm-like regions, particularly in noncoding regions, is not sufficient to consistently eliminate sequence divergence away from the gene-dense chromosome centers. Further, we found that SNPs, indels, and structural variations were highly enriched in intergenic regions when comparing the genomes of laboratory strains. Although many of the sequence variants we identified are not directly disrupting coding sequences, it remains possible that genetic drift in these regions is altering the function of intergenic regulatory sequences such as promoters and enhancers. Thus, the accumulation of disruptive genomic changes within regulatory regions in the gene-dense centers of chromosomes may underpin many of the phenotypic differences observed in laboratory wild-type strains, such as variance in lifespan (Gems and Riddle 2000).

Potential impacts of accumulating structural variations

We detected many structural variants ranging from one to hundreds of kilobases in size. Although these were often in intergenic regions of the genome, this does not preclude any possible impacts on the regulation of gene expression. Other than loss of entire genes, SVs of this size can be particularly disruptive to gene expression by impairing long range interactions between regulatory sequences such as promoters and enhancers (Stranger et al. 2007; Hurles, Dermitzakis, and Tyler-Smith 2008). Further, the expression of eukaryotic genes relies on the splicing of introns out of pre-mRNAs which is mediated by the spliceosome at specific recognition sites (Y. Lee and Rio 2015). Large insertions, deletions, and rearrangements within

intronic sequences of genes can potentially disrupt pre-mRNA splicing and intronic copy number variations have been associated with variable gene expression in populations (Rigau et al. 2019). SVs within the introns of genes have also been shown to lead to the emergence of duplicated genes and give rise to functionally distinct paralogs (Xu et al. 2012). Therefore, the accumulation and prevalence of non-coding SVs cannot be ignored as they may lead to significant impacts on gene function and evolution between laboratory model systems.

Experimental practices and genomic analyses could be improved by resequencing the genomes of labs' wild-type strains or utilizing strains with recently published, accurate genome assemblies. Aside from phenotypic consequences, the accumulation of undetected indels and SVs could be inhibitory to basic molecular biology techniques such as PCR or CRISPR if the target sequence is missing, disrupted, or rearranged. Further, large gains and losses of sequence, which may include entire genes, would be inhibitory to DNA sequencing workflows where the alignment of sequencing reads is necessary for downstream analyses. Researchers should ideally be using strains or lines of their species with a reference genome that accurately reflects the genotype of their model system. Though whole genome sequencing and genome assembly is a costly option to resolve challenges in research stemming from genetic drift of model organisms, there are alternatives and better practices in the maintenance of inbred lines. Our data presents a strong argument for labs utilizing *C. elegans* in their research to use a lineage with a recently published genome or frequently return to cryogenically preserved stocks of their wild type strains.

Are Hawaiian *C. elegans* lineages exceptional in their variation?

We find that the total genomic variation in CB4856 Hawaiian lineages is slightly higher than our observations comparing N2 Bristol lineages. Although both strains passaged in labs have underappreciated amounts of variation, largely due to SVs, the frequency of use for each strain must be considered. By far, the N2 Bristol strain is largely the standard wild type strain for laboratory research, and has been in use since the late 1960s and 1970s (Brenner 1974). We expect that the CB4856 Hawaiian isolate, however, has not been maintained or passaged in *C. elegans* labs worldwide at the same rate as the N2 Bristol isolate. This could lead to expectations of lower frequencies of sequence variants between laboratory lineages of CB4856 Hawaiian. Despite this, CB4856 Hawaiian *C. elegans* are a more "social" species with higher male mating frequencies in the population in contrast to N2, which has lower frequencies of mating with males (Wegewitz, Schulenburg, and Streit 2008). This increased mating frequency could lead to increased heterozygosity and higher sequence divergence between Hawaiian lineages compared to N2. Further, the *Caenorhabditis* Genetics Center, an international repository and distributor of strains, returned to its 1995 working stock of CB4856 due to many labs reporting phenotypic abnormalities into 2013 and 2014. Depending on how long each lab has passaged their lineage of the N2 and CB4856 strains, our account of SNPs and indels could be reasonably explained by findings in previous mutation accumulation studies (Denver et al. 2009). The rate of base substitutions and indels in the germline do not account for SVs, though studies of human genomes support the lower counts of SVs observed in our *C. elegans* genomes (Nesta, Tafur, and Beck 2021). Why then, does there appear to be so much more genomic structural variation between lab lineages of the Hawaiian isolate? One possible source is that the Hawaiian genome

experienced expansion of tandemly repeated regions (which include rDNA and regions of tandemly repeated TE sequences), which are known to mutate at much higher frequencies in yeast and humans (Fan and Chu 2007). Analysis of tandem repeat expansion between human individuals have shown that some tandemly repeated regions can vary in size from approximately 159.8-441.8kb (Gondo et al. 1998). Another source of genomic structural variation could simply be due to variability and accuracy afforded by different genome assembly methods. Genomes assembled with Nanopore reads have successfully spanned and faithfully represented tandemly repeat regions that can increase the genome length by hundreds of thousands or of millions of base pairs, as is the case with *C. elegans* VC2010 and the recently completed human genome (Yoshimura et al. 2019; Nurk et al. 2022). Thus, to accurately determine the extent to which SVs account for elevated genomic variation in Hawaiian *C. elegans,* or between any two genomes, future studies using genomes assembled from the same sequencing technology will be a necessity.

Conclusion

Finally, the generation of multiple independent *de novo* genome assemblies for both N2 Bristol and CB4856 Hawaiian isolates provides an excellent system to study genetic drift between laboratory model organisms. Additionally, identification and functional characterization of polymorphic sites and structural variations present between lab lineages of N2 Bristol and CB4856 Hawaiian may provide new insights into how pronounced phenotypic differences in the lifespan, feeding behavior, and reproductive fitness arise in modern lab-derived strains (Gems and Riddle 2000; Zhao et al. 2018). Future studies utilizing identical sequencing technologies

and genome assembly methods for their comparisons will further illuminate the extent of genomic diversity between labs and allow for functional characterization of large genomic rearrangements. Overall, our findings here provide both evidence and a platform for future comparative genomic studies to improve our understanding of how mutation accumulation impacts the practice and interpretation of model organism research.

Methods

N2 Bristol and CB4856 Hawaiian genome assemblies

The Libuda N2 Bristol and CB4856 Hawaiian genomes used were previously assembled *de novo* from PacBio long reads and corrected with Illumina short reads as described in Chapter I. The VC2010 Bristol genome (European Nucleotide Archive: PRJEB28388) was previously assembled *de novo* from PacBio and Oxford Nanopore long read sequencing technologies and polished with Illumina and PacBio reads (Yoshimura et al. 2019). The Kim CB4856 Hawaiian genome (NCBI BioProject PRJNA523481) was previously assembled *de novo* from PacBio sequencing reads and iteratively polished with both PacBio and Illumina reads (C. Kim et al. 2019). For subsequent variant analyses, the VC2010 Bristol and Kim CB4856 Hawaiian genomes were used as the reference sequences.

SNP and indel variant calling in Bristol and Hawaiian genomes

To call short sequence variants, Illumina short reads from the Libuda N2 Bristol and CB4856 Hawaiian genomes were aligned to the VC2010 Bristol genome and Kim CB4856 Hawaiian genome genomes, respectively. Sequencing adapters and barcodes were trimmed from

raw reads using Trimmomatic (Bolger, Lohse, and Usadel 2014). The trimmed reads were then aligned to Libuda reference genomes using BWA-MEM (Li and Durbin 2009). Aligned reads in SAM format where coordinate sorted using SAMtools (Li et al. 2009) and converted to BAM files. Picard was then used to assign read groups via AddOrReplaceReadGroups, and duplicate reads were filtered using MarkDuplicates ("Picard Toolkit" 2019). BAM files with filtered reads were used to call SNP and indel variants using GATK HaplotypeCaller (McKenna et al. 2010), Freebayes (Garrison and Marth 2012), and BCFtools (Danecek and McCarthy 2017). The VCF files produced from each individual program were concatenated and further filtered for duplicate sites and low-quality variant calls using BCFtools. SNPs with phred-scaled QUAL scores of at least 30, a minimum of 10 variant reads, and a minimum of 30 total, high-quality reads were retained.

Whole-genome alignment and calling genomic structural variants

All assembly-to-assembly alignments were performed using Minimap2 (Li 2018) and SyRI (Goel et al. 2019) was then used to parse SAM files to call SVs and highly divergent regions. To acquire NOTAL regions in query genomes, Minimap2 alignments were repeated with Libuda N2 Bristol and CB4856 Hawaiian genomes as the reference sequences. When comparing our CB4856 Hawaiian genome to the Kim CB4856 Hawaiian genome, 89% of the size difference in assemblies can be accounted for in the net sequence gained from Kim HDRs and unique NOTAL structures. NOTAL structures and gap-adjacent sequences in the Kim Hawaiian genome are 1.5 to 1.6-fold enriched for low complexity sequences (e.g. "homopolymer runs" of at least 4 consecutive identical bases) and repeat sequences. These regions and sequence

contexts are challenging for genome assembly and likely contribute to the megabase-scale difference in assembly sizes. Further differences can be ascribed to the much greater read length afforded by Nanopore sequencing, which can be especially helpful in assembling tandemly repeated regions spanned by ultra-long reads (Jain et al. 2018).

Converting gene annotations between assemblies

I converted the coordinates of gene annotations from the original N2 reference assembly (cel235) to corresponding regions in the VC2010 Bristol and Kim CB4856 Hawaiian genomes. The gene annotations for the WBcel235 assembly were downloaded in GFF3 format from Ensembl (http://ftp.ensembl.org/pub/release-105/gff3/caenorhabditis_elegans/). Assembly-to-assembly alignments via Minimap2 and Liftoff (Li 2018; Shumate and Salzberg 2021) were used to remap annotations as done in Chapter I.

Assessing enrichment or depletion of variants in gene annotations

SNPs, indels, SVs, and HDRs were tested for their degree of association in gene annotations to determine if lab lineages of each strain were accumulating genetic variants in coding versus noncoding sequences. Log2(fold) values were computed using the Genomic Association Tester (GAT) tool as described in Chapter I. Briefly, the observed overlap of each variant type with annotations was calculated independently and then compared to the mean overlap of simulated null distributions of each variant type. Simulated distributions were made from 20,000 iterations of a random uniform distribution across each chromosome. Statistical significance of each fold difference was determined via hypergeometric tests. BED files for SNP,

indel, SV, and HDR intervals were made by exporting pandas data frames of each VCF in python as a tab-separated file. BED files for gene annotations on each chromosome were made by separating the gene, mRNA, exon, CDS, and UTR regions from remapped annotations in GFF3 format. Intron and intergenic regions not explicitly written into the original GFF3 file were calculated using BEDtools (Quinlan and Hall 2010) as described in Chapter I. GAT simulations and statistical tests were then performed within the VC2010 Bristol and Kim CB4856 Hawaiian genome assemblies on each chromosome.

Bridge to chapter 3

Here, I outlined how laboratory passaging of multiple strains of *C. elegans* has led to unexpected amounts of genomic divergence in different lineages of canonical wild types. Genomic SVs account for millions of base pairs of variation between lab strains, and these likely underpin variance in behavior, metabolism, lifespan, and reproduction between labs. Until now, I have explored the patterns of sequence and genomic structural variation followed with perspectives on how specific chromatin structures influence the rise of these variations. In chapter 4, I explore this further by examining how similar chromatin structures influence mechanisms of DNA repair like homologous recombination in the germline. Further, I examine how sperm versus egg cells are differentially influenced by specific germline chromatin structures to give rise to sexual dimorphisms in the spatial distribution and rate of crossing over despite the shared goal of meioses in securing an obligatory crossover for accurate chromosome segregation.

CHAPTER 4: SEXUALLY DIMORPHIC CROSSOVER LANDSCAPES ARE ASSOCIATED WITH GERMLINE CHROMATIN STATES IN *C. ELEGANS*

Introduction

Developing gametes, such as spermatocytes and oocytes, repair DNA double strand breaks (DSBs) with homologous recombination to form crossovers. Crossover formation physically links the homologs and facilitates faithful chromosome segregation during meiotic cell division (Petronczki, Siomos, and Nasmyth 2003; Page and Hawley 2003). A failure to induce or establish crossovers can lead to aneuploid gametes, infertility, and developmental disorders (Hassold and Hunt 2001; S. Wang et al. 2019).

To ensure faithful inheritance of the genome, crossover formation is stringently regulated in germ cells of many species. The formation of at least one crossover between each pair of homologous chromosomes is ensured by a process called crossover homeostasis (Liangran Zhang, Liang, et al. 2014; Cole et al. 2012; Martini et al. 2006; Yokoo et al. 2012; Globus and Keeney 2012; Liangran Zhang, Wang, et al. 2014). In many species, the distribution of crossovers across the genome is non-random. The formation of one crossover can inhibit the formation of subsequent nearby crossovers on the same chromosome through a phenomena known as crossover interference (Sturtevant 1913; Muller 1916; G. H. Jones 1984; Hillers 2004; Meneely, Farago, and Kauffman 2002; Lloyd 2023; Gerton et al. 2000). Analyses of crossover distributions in many species reveals that the strength of crossover interference varies greatly between species and between sexes within species (Otto and Payseur 2019; Berchowitz and Copenhaver 2010). In species like *S. pombe* and *A. nidulans,* crossovers form with little to no

interference (Munz 1994; Strickland 1958), whereas some species like *C. elegans* have nearly complete interference and each set of paired homologs form exactly one crossover event (Yokoo et al. 2012; Machovina et al. 2016). Previous studies have indicated that while chromosome structures may be involved in crossover interference, there are likely additional multi-tiered mechanisms that regulate the rate and placement of crossovers and remain to be uncovered.

The rate and distribution of crossovers across the genome, known as the recombination landscape, often differ between oogenesis to spermatogenesis in multiple model systems. Evidence from studies in plants, mollusks, arthropods, amphibians, reptiles, birds, and mammals all demonstrate that the distribution of crossover events during spermatogenesis is slightly elevated in sub-telomeric regions in contrast to the more centrally located crossovers in oogenesis (Sardell and Kirkpatrick 2020). Further, the crossover number per chromosome pair is also sexually dimorphic, with oogenesis having a higher crossover rate than spermatogenesis in many species (Sardell and Kirkpatrick 2020). Notably, there is increasing evidence that epigenetic modifications can influence sex differences in the recombination landscape. Mammalian oocytes, but not spermatocytes, undergo global DNA demethylation (Seisenberger et al. 2012; Smith et al. 2012), and DNA methylation has been shown to promote the initiation of recombination (Brick et al. 2018). DNA methylation occurs primarily at CpG nucleotides, which are enriched in sub-telomeric regions coincident with known male biases for crossing over (Bird 1986; Arndt, Hwa, and Petrov 2005; Lister et al. 2009; Bernardi 2000; Sardell and Kirkpatrick 2020). Additionally, studies in mice and plants have shown that chromatin modifications like H3K4me3 can differentially influence where recombination is initiated in each sex (Brick et al. 2018; Kianian et al. 2018). Thus, while sex differences in recombination with respect to

epigenetic modifications may not completely explain sexually dimorphic recombination landscapes, it is highly likely they play a substantial role.

Transcriptionally active regions, also known as euchromatin, are known to have more physically accessible DNA, which may promote initiation, processing, and maturation of crossover recombination. Local chromatin structure and nucleosome positioning around promoter regions can influence the formation of programmed DSBs, the initiating event for homologous recombination, by the highly conserved, topoisomerase-like protein SPO11 (Keeney, Giroux, and Kleckner 1997; Keeney 2008; Frédéric Baudat et al. 2000; Grelon 2001; Dernburg et al. 1998; Lange et al. 2016). In mice and humans, the histone methyltransferase PRDM9 binds specific DNA motifs and creates "hotspots" for crossover formation (F. Baudat et al. 2010; Myers et al. 2010; Parvanov, Petkov, and Paigen 2010; Powers et al. 2016). In species that lack PRDM9-mediated hotspots, such as budding yeast, DSBs and crossovers are enriched in physically accessible euchromatic regions, such as nucleosome-depleted sequences and gene promoters (Pan et al. 2011). In species that lack hotspots entirely, such as the nematode *Caenorhabditis elegans,* other chromatin modifications like H3K9 methylation are still known to play a role in shaping the recombination landscape in oocytes (Lascarez-Lagunas et al. 2023), and crossovers are preferentially positioned in multi-megabase domains at the terminal thirds of each chromosome (Barnes et al. 1995; Rockman and Kruglyak 2009).

The crossover landscape is sexually dimorphic in *C. elegans.* Crossover assessment at the resolution of multiple megabases via genetic assays have shown sex-specific differences in crossover frequencies on the arms of some autosomes (Lim, Stine, and Yanowitz 2008; Meneely et al. 2012; Meneely, Farago, and Kauffman 2002; Wagner et al. 2010). In *C. elegans* oocytes,

crossover homeostasis and interference is incredibly robust, where paired homologs receive only one crossover (Yokoo et al. 2012; Machovina et al. 2016). In contrast, the incidence of double crossover events is slightly elevated in spermatogenesis across multiple genetic intervals (Gabdank and Fire 2014; Henzel et al. 2011; Jonathan Hodgkin, Horvitz, and Brenner 1979; Lim, Stine, and Yanowitz 2008; Meneely, Farago, and Kauffman 2002; Zetka and Rose 1995), meaning that the strength of interference is likely sexually dimorphic and weaker in males. Notably, a few studies did detect double crossover events on some genetic intervals in oogenesis, but not spermatogenesis (Meneely et al. 2012; Meneely, Farago, and Kauffman 2002; Jonathan Hodgkin, Horvitz, and Brenner 1979), thereby generating a debate on whether spermatogenesis indeed has a lower level of crossover interference in comparison to oogenesis.

The use of whole genome sequencing can illuminate the crossover landscape and the genomic features that regulate crossover positioning and distribution. Prior studies have mapped and examined finer-scale features of crossover recombination in oocytes and shown an association of the crossover rate near sites where homologous chromosomes pair (Rockman and Kruglyak 2009). Additionally, fine-scale crossover mapping on a subset of the *X* chromosome in oocytes also showed a negative association of crossover sites with euchromatic histone modifications (Bernstein and Rockman 2016). Although the genomic landscape of chromatin modifications differs between *C. elegans* oogenesis and spermatogenesis (Tabuchi et al. 2018), it is unclear whether these fundamental features of chromosomes promote sex-specific crossover distributions. The lack of a genome-wide crossover landscape for *C. elegans* spermatogenesis has inhibited elucidating the mechanisms behind the sexually dimorphic crossover rates within the genome. Overall, the lack of hotspots combined with the clear sexual dimorphisms in both

the crossover landscape and the chromatin landscape of *C. elegans* presents an excellent opportunity to define the critical factors behind how crossing over is regulated by the native chromatin architecture.

To illuminate the both the fine-scale and broader scale genomic features contributing to the sexually dimorphic recombination landscapes in *C. elegans,* we performed high-resolution crossover mapping in individual products of both sperm and egg meioses. Using hundreds of thousands of SNP markers between the N2 Bristol and CB4856 Hawaiian strains, we mapped crossovers in spermatocytes and oocytes with an average resolution of one SNP every 300bp. Our analysis demonstrated that Chromosomes *I, II,* and *III* display the most sex-specific differences in recombination rates at both the kilobase and megabase scales. The global rate of double crossover events is nearly five-fold higher than oogenesis. Finally, we demonstrate the sex differences in the crossover landscapes are highly associated with specific chromatin states that regulate gene expression in the germline. Crossover formation in spermatogenesis is highly associated with the euchromatic histone modification H3K36me3, while oocytes crossovers display high association with the heterochromatic histone modification H3K27me3. Taken together, these results reveal that the mechanism(s) of homologous recombination in oogenesis and spermatogenesis differentially utilize multiple chromatin states to shape the final crossover landscape.

Results

High-resolution mapping of crossovers in single *C. elegans* genomes

To detect crossovers with high resolution in each sex, we performed whole genome sequencing of individual F2 progeny harboring recombinant chromosomes from single meioses in the F1 generation. Briefly, F1 hybrid progeny were generated by crossing N2 Bristol hermaphrodites and CB4856 Hawaiian males. F1 progeny were then backcrossed to individuals with a Bristol genetic background so that F2 progeny inherit singular recombinant chromosomes from sperm or egg meioses along with another N2 Bristol homolog. From the F2 generation, 300 oocyte-derived samples and 310 spermatocyte-derived samples were sequenced individually (see Methods; Figure 4.1A). The average read depth at SNP markers in F2 individuals was approximately 10 reads, and many samples were sequenced at 1-5X coverage (Supplemental Figure S4.1). Using 213,591 of the SNPs between the N2 Bristol and the CB4856 strains (see methods), we were able to map crossovers for each chromosome using our adaptation of the TIGER pipeline (Supplemental Figure S4.2) for HMM inference of crossover breakpoints which enables robust crossover detection against samples even sequenced at low coverage (Rowan et al. 2015). In our dataset of 610 genomes, our crossover mapping pipeline was able to successfully call crossovers on 297/300 oocyte samples and 300/310 spermatocyte samples. In the oocyte data, we detected 837 crossovers across 830 chromosomes, and 738 crossovers across 710 autosomes in the spermatocyte data (Figure 4.1B). The average size of the SNP intervals containing crossover breakpoints is 1568bp with a median resolution under 1kb for each chromosome (Figure 4.1C), indicating great resolution of detection in both sexes. The increased number of detected crossovers relative to the number of unique chromosomes represented is due

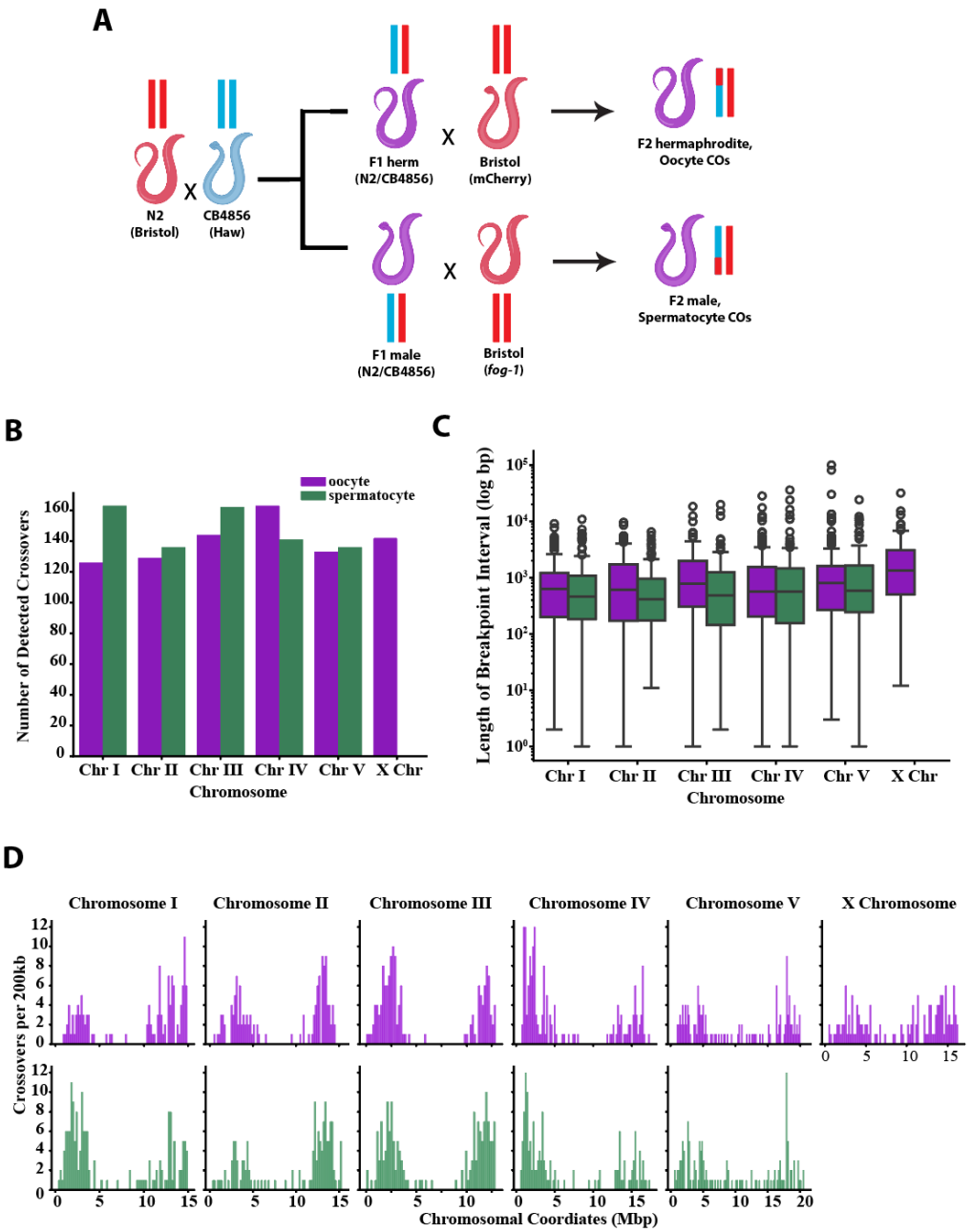


Figure 4.1. Sexually dimorphic crossover distributions in *C. elegans* meiosis. (A) Crossing scheme for the generation of F2 hybrids carrying recombinant chromosomes for sequencing and crossover detection. (B) A bar chart indicating the number of crossovers detected per chromosome in each sex. (C) Box and whisker plot describing the length of SNP intervals where crossover breakpoints were detected. (D) Histograms showing the global distribution of crossovers in oocytes versus oocytes on each chromosome. Crossovers were counted in nonoverlapping 200kb bins on each chromosome. Oocyte data is shown in purple and spermatocyte data is shown in green.

to the presence of double crossover events. On chromosome *I*, we detected approximately 23% fewer crossovers in oocyte samples, which could be due to some samples having insufficient read coverage or known genetic incompatibilities between the Bristol and Hawaiian backgrounds (Seidel, Rockman, and Kruglyak 2008; Seidel et al. 2011; Ben-David, Burga, and Kruglyak 2017). In total, the number of chromosomes with crossovers is within our expectation such that roughly 50% of the chromosomes inherited via random segregation from F1 meioses are recombinant.

The fine-scale crossover landscapes are sexually dimorphic in *C. elegans* meiosis

To assess whether then global distribution of crossovers was sexually dimorphic on each chromosome, we compared the distribution of crossovers across each chromosome for both spermatocyte and the oocyte recombination landscape data. Broadly, both oocyte and spermatocyte crossovers exhibit a bias towards the terminal thirds of all chromosomes. Notably, these results match previous studies describing the distribution pattern for crossovers (Rockman and Kruglyak 2009; Barnes et al. 1995). For chromosomes *I* and *III*, we determined that the crossover distributions are significantly different in each sex ($p < 0.01$ by Kolmogorov-Smirnov test). In contrast to spermatocytes, oocytes favor crossover formation on the right arm of chromosome *I* and the left arm of chromosome *III* (Figure 4.1D). In contrast to chromosomes *I, II, III,* and *IV*, we detected a greater proportion of crossovers in the central regions of chromosome *V* and the *X* chromosome. Overall, crossovers prefer to form in the terminal third of all chromosomes regardless of sex. Using our high-resolution crossover maps, we determined

whether regions within any of the chromosomes' arms had elevated crossover formation. Similar to previous studies (Rockman and Kruglyak 2009; Kaur and Rockman 2014), our data did not detect any distinct 1-2 kb regions of extremely elevated crossover rates that resemble recombination hot spots like those seen in mammals and budding yeast (Parvanov, Petkov, and Paigen 2010; F. Baudat et al. 2010; Gerton et al. 2000). Despite the lack of recombination hotspots, our data did identify 200 kb chromosomal regions with a greater density of crossovers (Figure 4.1D). We observed multiple 200 kb regions on chromosomes *I* and *IV*, in both oocytes and spermatocytes, that have 3-4 times as many crossovers as adjacent regions on the same chromosome arm (Figure 4.1D). Thus, while crossover formation is biased towards the chromosome "arms", we can see fine-scale variability in the crossover landscape suggesting that crossover formation is not random or uniform in these large domains. These high-resolution distributions for each sex, particularly on chromosomes *I, II*, and *III*, reflect earlier studies' broader conclusions of sexual dimorphisms using methods at multi-megabase resolution (Lim, Stine, and Yanowitz 2008; Meneely, Farago, and Kauffman 2002; Meneely et al. 2012; Wagner et al. 2010).

Sexually dimorphic rates of crossing over

Given that the spatial distribution of crossovers is sexually dimorphic on multiple autosomes, we first assessed if the genetic map lengths on each chromosome were also sexually dimorphic. A higher map length in one sex suggests a higher frequency of crossing over on a given chromosome. Our data shows that spermatocytes have the highest the map lengths on every chromosome except chromosome *IV* (Table 4.1). Given that the theoretical map length for

chromosomes that experience a single crossover is 50cM, this suggests that double crossovers were likely detected in both sexes and that the rate of crossing over is highest in spermatocytes.

Table 4.1. Calculated map lengths (cM) of chromosomes in each sex.

	I	***II***	***III***	***IV***	***V***	***X***
Oocyte	42.4	43.4	48.5	54.9	44.7	47.8
Spermatocyte	54.33	45.3	54	47	45.3	n/a

Elevated double crossovers in spermatocyte genomes

To better assess sexual dimorphisms in the crossover rate, we calculated the incidence of double crossovers (DCOs) at the global and chromosomal scales. The occurrence of DCOs in *C. elegans* is rare and often undetected (Barnes et al. 1995; Hammarlund et al. 2005; Hillers and Villeneuve 2003; Saito et al. 2009; 2012; 2013; Wagner et al. 2010). Our data set of 830 oocyte recombinant chromosomes and 710 spermatocyte recombinant chromosomes revealed the global rate of DCOs is 4.7-fold higher in spermatocytes than oocytes (3.94% spermatocytes vs 0.84% in oocytes) (Figure 4.2A). Notably, we found that this global rate was not equally shared across all chromosomes or limited to those with sexually dimorphic crossover distributions. In oocytes, we only detected DCOs on chromosomes *IV* (5/168, 3.1%) and the *X* chromosome (2/140, 1.42%), whereas in spermatocytes DCOs were detected on all autosomes. The spermatocyte DCO rate ranges from as low as 1.49% (2/134) on chromosome *II* to as high as 7.63% (10/131) on chromosome *IV*. Remarkably for spermatocytes, the rate of DCOs on chromosome *IV* is nearly

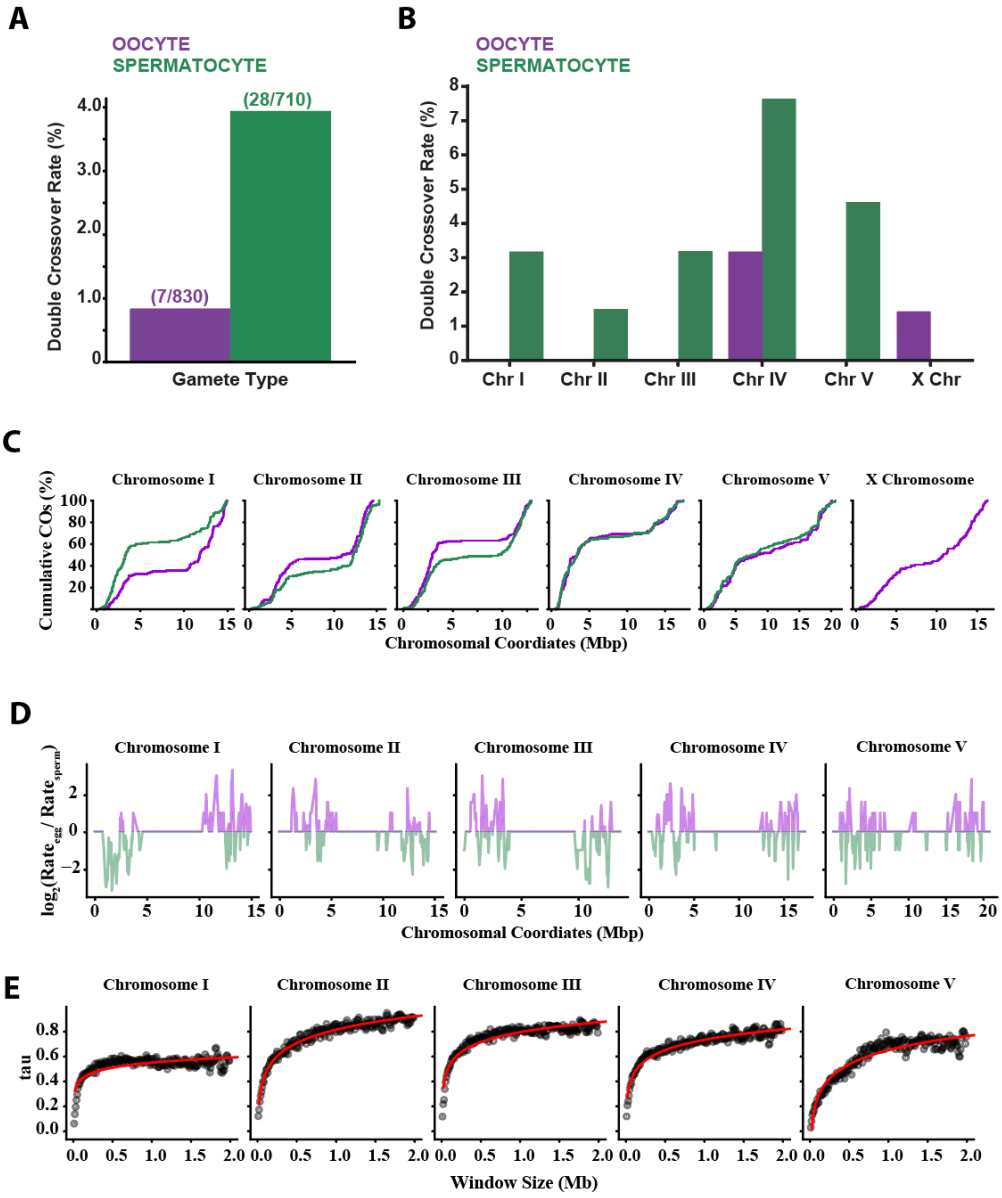


Figure 4.2. Broad and fine-scale variation in crossover rates in each sex. (A) Bar chart showing the global average frequency of double crossovers among all recombinant chromosomes. (B) Bar chart showing the frequency of double crossover events on each chromosome in sperm versus eggs. (C) Cumulative distribution plots showing the frequency of crossing over across each chromosome. (D) Line plot showing the Log2 values of the ratio of the recombination rate (cM/bp) in eggs versus sperm. Crossover rates were calculated in 200kb sliding windows with a 50% step size. (E) Scatterplot showing the correlation of oocyte and spermatocyte recombination rates in sliding windows of varying sizes. Kendall's tau was calculated as the correlation coefficient, all p-values < 0.05 except 10kb window size on chromosome V. Red line indicates the curve of best fit for each chromosome.

double the global average DCOs frequency in spermatocytes (Figure 4.2B). Thus, the overall crossover rate is not only different between the sexes but also between each chromosome.

Local differences in sex-biased crossover frequencies

To assess whether sexual dimorphisms in the frequency of crossing over persist at the megabase or kilobase scales, we began by examining cumulative frequency distributions of crossovers in each sex. The cumulative distributions of crossovers detected on each chromosome illustrate the expected frequency of crossing over between any two loci. We found marked differences between the sexes in crossover frequencies on the arms of chromosomes *I, II*, and *III* (Figure 4.2C). Spermatocytes display elevated frequency of crossover formation in the first five megabases of chromosome *I*. In comparison, oocytes have higher crossover frequencies between 2.5-5Mb on the left of chromosomes *II* and *III* (Figure 4.2C). In conclusion, these data show sex biases in the crossover landscape in multi-megabase domains on chromosomes *I*, *II*, and *III*.

Sexual dimorphisms in the crossover frequency at sub-megabase scales could indicate specific regions or features of chromosomes contributing to a sexually dimorphic crossover landscape. To test for sexual dimorphisms in the crossover rate at a finer scale, we performed a sliding window analysis of the crossover rate in 200 kb windows on each chromosome for each sex. Our approach detected many 200 kb regions where each sex has a higher local crossover rate in a non-favored chromosome arm (Figure 4.2D). Consistent with the spatial and frequency distributions of crossovers (Figures 4.1D and 4.2C, respectively), I found that chromosomes *I, II,* and *III* have large clusters of 200 kb regions with sexually dimorphic crossover rates (Figure

4.2D). Despite these large-scale patterns, we found 99 total instances on chromosomes *I* (n=31), *II* (n=40), and *III* (n=28) where neighboring 200 kb windows display opposite sex biases in the local crossover rate (Figure 4.2D). Further, while the chromosomal crossover distributions for chromosomes *IV* and *V* are not significantly different between the sexes at the megabase scale, we do see many instances (41 and 64 for chromosomes *IV* and *V*, respectively) where adjacent 200kb windows have higher crossover rates in the opposite sex (Figure 4.2D). Taken together, sexual dimorphisms in the crossover landscape persist at sub-megabase scales on each chromosome.

To determine how well the local crossover rates in each sex correlate across many scales, we repeated our sliding window analysis of crossover rates in a range of window sizes from 10 kb up to 2 Mb in 10 kb increments and calculated Kendall's *tau* as a correlation coefficient (Figure 4.2.3E). For autosomes *I, II, III*, and *IV*, crossover rates show little to no correlation (tau < 0.3) up to window sizes of 30-50 kb. Crossover rates on chromosome *V*, however, remain very weakly correlated up to a window size of approximately 200 kb. At larger window sizes, chromosome *I* is unique in that even in windows of 1 Mb or greater, crossover rates remain only moderately correlated (tau 0.4-0.6). Chromosomes *II, III, IV*, and *V*, in contrast, achieve much higher levels of correlation (tau 0.75-0.9) in megabase-scale windows. In conclusion, the rate of crossing over in spermatocytes versus oocytes is different at the finer scale of tens of kilobases with some differences persisting up to megabase scale domains. Further, we find that sexual dimorphisms in the crossover landscape are not uniformly shared in their distribution or magnitude across all chromosomes.

Proximity of crossovers to homolog pairing sites

To facilitate efficient access to the homologous chromosome as a repair template, homologous chromosomes pair along their lengths prior to crossover formation (Page and Hawley 2003). Prior studies suggest that crossovers are sometimes bias towards sequence-defined regions on each chromosome called pairing centers (PCs) (Rockman and Kruglyak 2009; Barnes et al. 1995; Lim, Stine, and Yanowitz 2008; Meneely et al. 2012; Saito et al. 2013; 2012; 2009; Phillips et al. 2009; MacQueen et al. 2005; Phillips and Dernburg 2006). Analyzing the location of crossovers relative to the PC for each chromosome revealed a sexual dimorphism for crossover placement relative to the PC. Specifically, the fraction of oocyte crossovers on the PC arm of chromosomes *I, II*, and *III* is much higher at 56%, 43.41%, and 61.11% respectively (Figure 4.3A-B). In spermatocytes, however, a minority of crossovers on chromosomes *I, II*, and *III* are formed in arm domains harboring each PC (30.06%, 30.88%, and 43.82%, respectively) (Figure 4.3A-B). These results indicate a sexual dimorphism for placement of crossovers relative to the PC "arm" and/or early-paired regions of the chromosome. Interestingly, when we measure the direct overlap of crossovers with clusters of sequence motifs that define the PCs (Phillips and Dernburg 2006; Phillips et al. 2009; MacQueen et al. 2005), we see that on average less than 3% of crossovers are formed within these PC sequence motif clusters except for spermatocyte crossovers on chromosome *IV* (Figure 4.3C). For chromosomes *I* and *II*, we find that the majority of both oocyte and spermatocyte crossovers are approximately 100 kb away from PC motif clusters (Figure 4.3D). This distance is variable across the other autosomes, with the approximate spacing in each sex being 60 kb on chromosome *III* and 200 kb on chromosomes

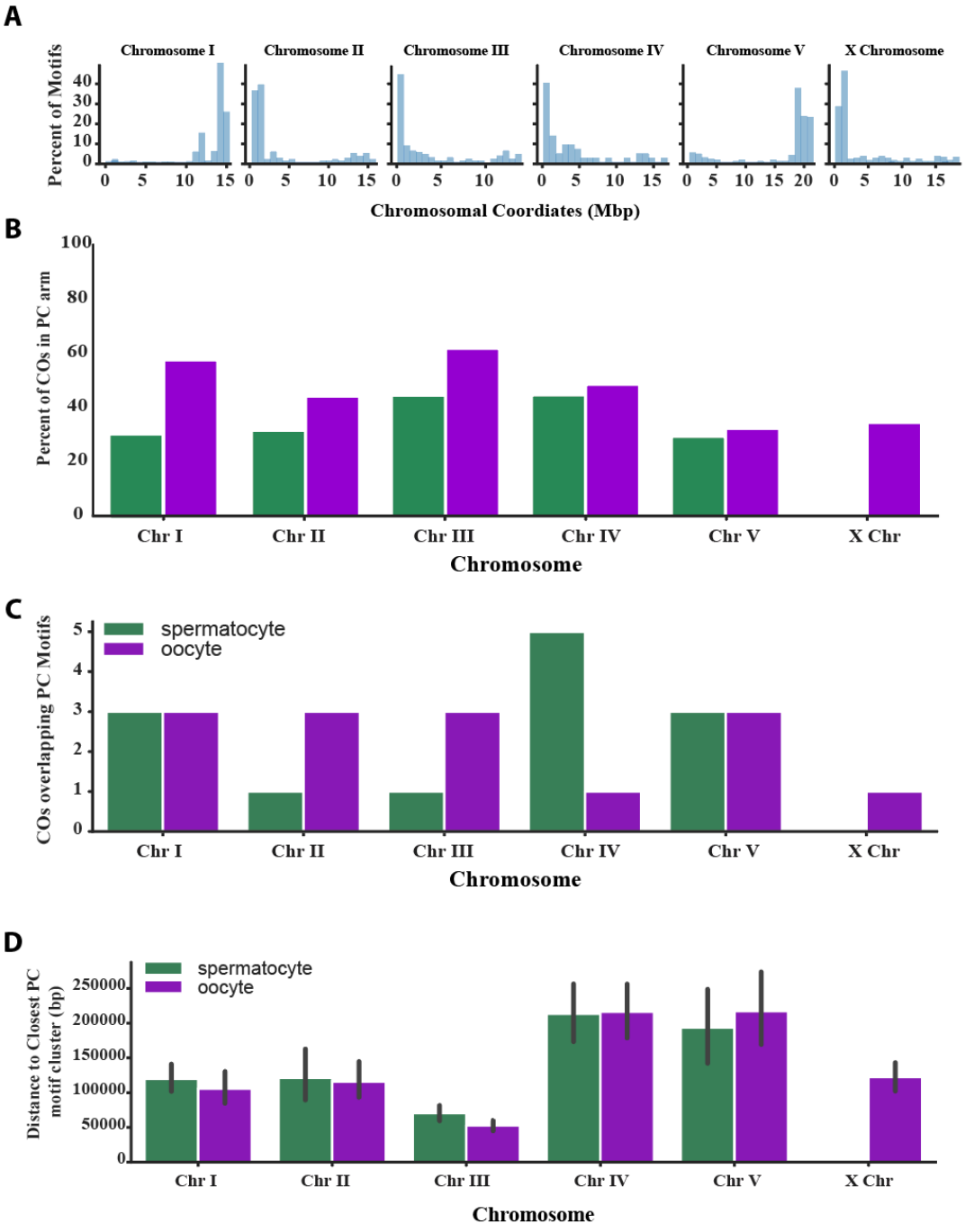


Figure 4.3. PC Motif cluster identification and association with crossovers. (A) Histogram showing all PC motif clusters identified by MCAST on each chromosome. (B) The percent of crossovers in each sex that are on the pairing center arm domains. (C) Bar chart showing the number of crossovers in each sex that directly overlap with PC motif clusters on each chromosome. (D) Bar chart showing the average distance of crossovers in each sex to the nearest PC motif cluster. Error bars indicate 95% confidence intervals.

IV and *V* (Figure 4.3D). Overall, we find that while crossovers are largely inhibited from forming directly within or adjacent to PC motif clusters, crossovers usually preferentially form on the PC "arm" of most chromosomes of oocytes.

Sex-specific crossover associations with germline gene expression

To determine whether any of the sexually dimorphic features of the crossover landscape arise from specific chromosomal features, we first tested the association of the crossover distribution with sequence level genome annotations. We tested for enrichment or depletion in intergenic regions, genes (including sub-gene annotations such as each UTR, exons, introns, and CDS), gene regulatory sequences (transcription factor binding sites, promoters, and enhancers) as well as other sequences such as non-coding RNAs (such as meiotically expressed piRNAs) and transposons (Supplemental Figure S4.3). For nearly all features tested, there was no significant enrichment or depletion determined by hypergeometric tests. In spermatocytes, we did find that crossovers were enriched in "ncRNAs" on chromosomes *I* and *III* (Log_2(fold) values of 1.58 and 2.77, respectively), which mirrored the sexually dimorphic crossover distributions on these chromosomes. Notably, in the genome annotations from Ensembl, "ncRNA" describes a subset of coding genes with non-coding splice variants of their mRNAs. These results suggest that a particular subset of the genes in *C. elegans* may drive sex differences in crossover formation.

To determine if gene expression in the germline shapes crossover distribution, we performed enrichment/depletion analyses only on genes expressed during meiosis using a published RNA-seq dataset that examined transcription levels specifically in the germline of

each sex (Tzur et al. 2018). We first labeled our complete set of annotated genes with specificity for oogenesis, spermatogenesis, genes commonly expressed in both germlines, or neither. Labels for shared or sex-specific expression were determined based on the presence of normalized read counts of ≥ 2 in the germline (Tzur et al. 2018). The distribution of meiotically expressed genes was greatest in the center of each chromosome, and most chromosomes had comparable counts of meiotic genes except for the *X* chromosome (Supplemental Figure S4.4). When we tested for fold enrichment or depletion of crossovers overlapping with these subsets of genes, we found several sex-specific associations. Spermatocyte crossovers had significant enrichments in genes expressed in germlines on chromosomes *III* and *V* (Log_2(fold) values of 0.52 and 0.60, respectively). In contrast, for oocyte crossovers, there were no statistically significant associations with meiotic gene expression states (Figure 4.4A). These trends indicate that chromosomal regions undergoing transcription may be shaping the crossover landscape differently in each sex.

Sex-specific associations of crossovers with chromatin states

In spermatocytes, an enrichment of crossovers in germline expressed genes indicates that the local chromatin structure may be influencing sexual dimorphisms in the crossover landscape. Gene expression states and the DSB landscape are influenced by chromatin states in multiple organisms (Pan et al. 2011; Powers et al. 2016; F. Baudat et al. 2010; C. L. Liu et al. 2005; Pokholok et al. 2005; Mikkelsen et al. 2007; Rando and Winston 2012; Ho et al. 2014). To determine whether sex-specific chromatin states contribute towards the sexually dimorphic recombination landscape, we tested the association of crossovers with sex-specific germline

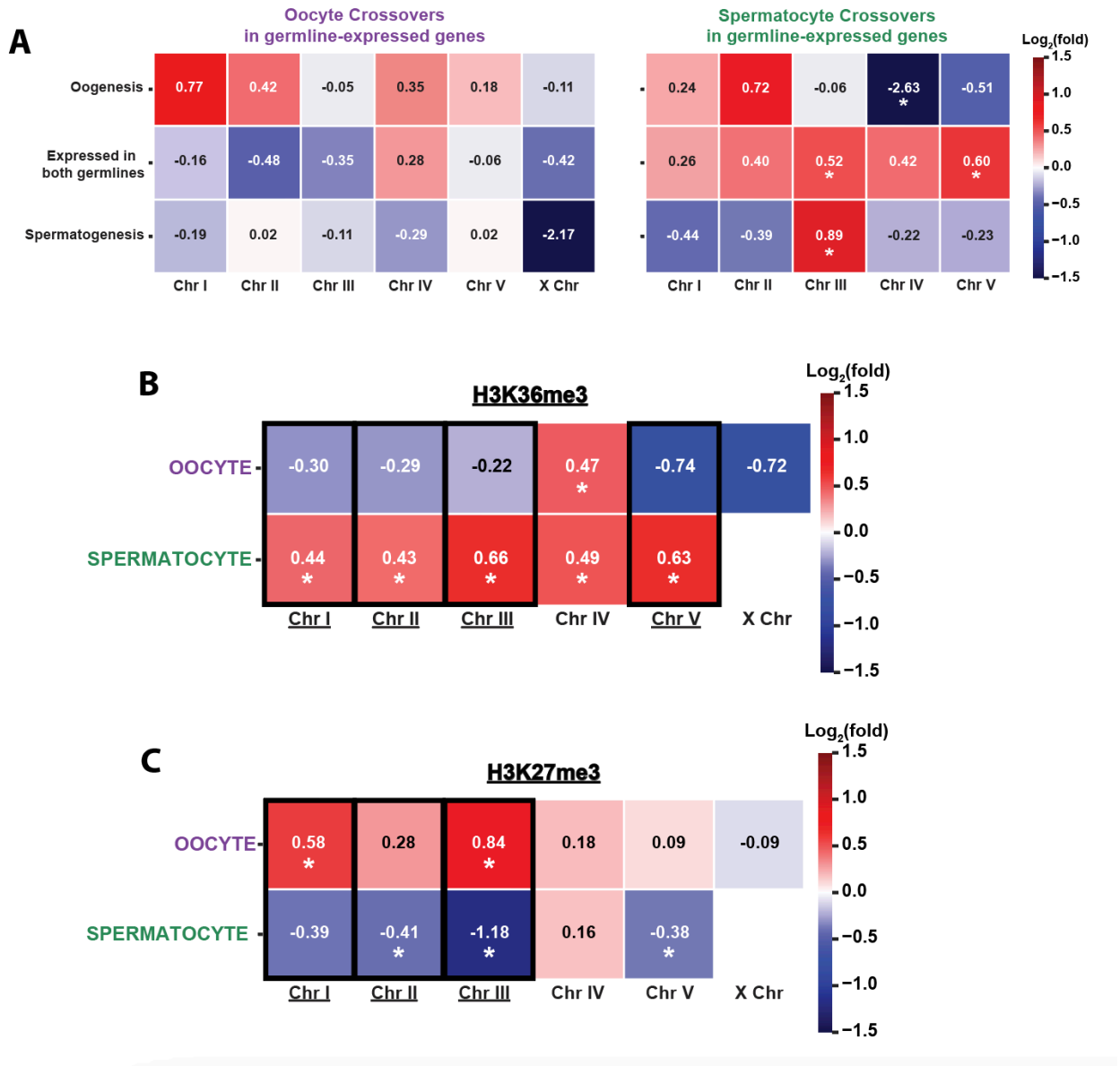


Figure 4.4. Differential associations of crossovers with chromatin states in each sex. (A) Heatmap showing the log2(fold) association of crossovers in germline expressed genes in oocytes (left) and spermatocytes (right). (B) Heatmap showing the log2(fold) association of crossovers with H3K36me3 ChIP-seq peaks. (C) Heatmap showing the log2(fold) association of crossovers with H3K27me3 ChIP-seq peaks. Asterisks indicate p-values < 0.05 by hypergeometric test when compared to simulated null distributions. Black boxes indicate p-values < 0.05 by hypergeometric test when comparing fold values in each sex.

landscapes of the euchromatic histone modification H3K36me3 and the heterochromatic histone modification H3K27me3 (Tabuchi et al. 2018). For the euchromatic modification H3K36me3, spermatocytes displayed significant enrichment of crossovers in euchromatic regions on all five autosomes (Log_2(fold) values 0.43-0.66, p-values < 0.05; Figure 4.4B). In contrast, oocytes showed a trend of crossovers depleted in H3K36me3 regions on chromosomes *I, II, III, V,* and *X* (Log_2(fold) values -0.22 to -0.74, Figure 4.4B). When we compared the difference of fold changes in spermatocytes versus oocytes, chromosomes *I, II, III*, and *V* were all significantly different (Figure 4.4B). Only for the euchromatic H3K36me3 landscape on chromosome *IV* did we find that both oocytes and spermatocytes have a significant association with the crossover landscape on chromosome *IV*, (Log_2(fold) values of 0.47 and 0.49, respectively; p-values < 0.05Figure 4.4B). In summary, we find that crossover formation in spermatocytes, but not oocytes, is strongly associated with chromatin marked by H3K36me3.

We then tested for sexually dimorphic association of crossovers with heterochromatin. For the heterochromatic mark H3K27me3, oocytes displayed an enrichment of crossovers in these regions on chromosomes *I* and *III* (Log_2(fold) values of 0.58 and 0.84, respectively; Figure 4.4C). In contrast, spermatocyte crossovers were depleted from heterochromatic regions on chromosomes *II, III*, and *V* (Log_2(fold) values of -0.38 to -1.18; Figure 4.4D). For chromosomes *I, II*, and *III*, the chromosomes with the most sexually dimorphic crossover landscapes, we found the difference in fold enrichments between the sexes to be statistically significant. Overall, our results suggest that the sex-specific crossover landscapes in *C. elegans* may be differentially

influenced by local differences in chromatin structure and accessibility near sites of active transcription in the germline.

Discussion

The importance of meiotic recombination for genome integrity and chromosome segregation is well understood, but how or why the sexes display differences in the recombination remains unclear. The studies we present here demonstrate that the global distribution and rate of crossing over is sexually dimorphic in *C. elegans* meiosis. Chromosomes *I* and *III* display the most drastic sex differences in the spatial distribution of crossovers, and all autosomes in spermatogenesis undergo double crossover events at rates much higher than oogenesis. Although most crossovers in either sex occur within 50-300kb of sites that facilitate homologous chromosome pairing, we do observe a greater proportion of oocyte crossovers on the pairing center arms of chromosomes *I, II*, and *III*. We also show that the distinct chromatin states along meiotic chromosomes are differentially associated with the crossover landscape in each sex. Crossover formation in spermatogenesis is enriched at sites of active gene expression associated with H3K36me3 marked euchromatin, whereas crossover formation in oocytes is enriched in H3K27me3 heterochromatin. Taken together, our high resolution maps and analyses of crossovers in spermatogenesis and oogenesis provide a platform to further investigate both mechanistic and evolutionary hypothesis about sexually dimorphic meiotic recombination.

Potential sources of sex-differences in crossover distribution

The DSB landscape. Crossovers frequencies in each sex are elevated in the terminal arm domains of each chromosome, yet oocytes preferentially form crossovers on the right and left arms of chromosomes *I* and *III*, respectively (Figure 4.2). Are crossovers in *C. elegans* spermatogenesis and oogenesis simply following the distribution of DSBs? The sex-differences in the crossover distribution on these chromosomes could be reflective of underlying sex differences in the distribution of DSBs that are the substrate for crossover recombination. In *C. elegans*, there are approximately four times as many DSBs present on the chromosome arms relative to the central domains throughout meiotic prophase I (Lascarez-Lagunas et al. 2023). While distinction has not yet been given to separately characterize the distribution of DSBs on the right versus left arms of each chromosome, it remains possible that there could be sexual dimorphisms as early in the recombination program as the induction of DSBs. Notably, other model systems display sexual dimorphisms at the initiation of recombination. Studies in mice demonstrate that in model systems with recombination hotspots induced by PRDM9 , there are multiple hotspot locations that display sex-specific activation (Brick et al. 2018). In models organisms that lack canonical hotspots (*e.g. C. elegans)*, it remains unclear how the underlying distribution of DSBs affects the resulting distribution of crossovers. Currently, there are no published or publicly available datasets from *C. elegans* that characterizes the global distribution of DSBs in either sex. This precludes the much-needed analysis of whether there are sex-specific DSB distributions and how this may lead to the sexually dimorphic crossover landscapes we observed.

Homologous Chromosome Pairing. Both sexes place most of their crossovers within 50-150 kb of pairing center motif clusters on chromosomes *I, II*, and *III* (Figure 4.3). In our dataset, we demonstrate a greater proportion of crossovers on the pairing centers in oocytes, with chromosomes *I, II*, and *III* displaying the greatest magnitude of differences when comparing the two sexes (Figure 4.2). These same chromosomes also have the greatest number of pairing center motif clusters relative to their total length (Phillips et al. 2009). Based on our data and published data from others, we hypothesize that the pairing of homologous chromosomes could play some role in shaping the crossover landscape in each sex. One proposed model for the procession of recombination describes how all DSB sites are competing for the accumulation of pro-crossover factors to determine the crossover versus non-crossover outcome (Liangyu Zhang et al. 2021; Morgan et al. 2021; Fozard, Morgan, and Howard 2023; C. Girard, Zwicker, and Mercier 2023). If the pairing of homologs enables faithful crossover formation, it is possible that DSB sites in early-paired regions of chromosomes have a temporal advantage in processing recombination intermediates and are more likely to form crossovers compared to DSB sites with less time in alignment. The total time of Prophase I, as well as time spent in the window for chromosome pairing, is nearly twice as long in *C. elegans* oogenesis (Jaramillo-Lambert et al. 2007). We therefore speculate that the prolonged duration of pairing in oogenesis likely contributes to the broad-scale differences in the crossover distribution on the PC "arms" of these chromosomes.

Crossover interference. We have demonstrated that the rate of crossing over on each autosome is higher in spermatogenesis as we see double crossovers at much higher frequencies than in oocytes (Figure 4.2). This supports the notion that the overall distribution and rate of crossing over between the sexes could also be a product of differences in the strength of

crossover interference. While multiple models have been proposed and effectors of interference have been identified (Libuda et al. 2013; Nabeshima, Villeneuve, and Hillers 2004; Liangran Zhang, Liang, et al. 2014; C. Girard, Zwicker, and Mercier 2023), the precise molecular mechanisms that lead to this phenomenon remains unclear. The sex with the lowest strength of interference, however, should have higher rates of crossover formation on individual chromosomes. Notably, sexual dimorphisms in the strength of crossover interference have been demonstrated in multiple model systems. In humans and some plants, interference is stronger, and the crossover rate is lower in males (Barth et al. 2000; Drouaud et al. 2007; Vizir and Korol 1990; Doniskeller 1987; Broman et al. 1998). We and others (Gabdank and Fire 2014; Henzel et al. 2011; Lim, Stine, and Yanowitz 2008) provide supporting evidence for a lower degree of interference in *C. elegans* spermatogenesis. Our data demonstrates that the frequency of double crossover events is higher on all autosomes during spermatogenesis. Our findings and others warrant further research analyzing a much greater number double crossover events on all chromosomes in *C. elegans* to more precisely characterize sex- and/or chromosome-specific variations in the strength of interference.

Crossovers in euchromatin versus heterochromatin

We demonstrate the crossing over in oogenesis is enriched in H3K27me3 marked heterochromatic regions on chromosomes *I* and *III* (Figure 4.4). In contrast, we note a negative association of crossovers with H3K27me3 in spermatogenesis and an enrichment of crossovers in H3K36me3 euchromatic regions. Therefore, we hypothesize that not only does sex-specific chromatin distributions influence the crossover landscape in each sex, but that the recombination

program in each sex is differentially affected by the presence of different chromatin modifications. Given that recombination intermediate processing is reliant on the recruitment and accumulation of pro-crossover factors at DSB sites, the processing of recombination intermediates in heterochromatin may be kinetically unfavored (Kelly et al. 2000; Jantsch et al. 2004; Bhalla et al. 2008; Nguyen et al. 2018; Liangyu Zhang et al. 2018; Yokoo et al. 2012). The prolonged duration of Prophase I in oocytes, however, could be amenable to the slower kinetics of DSB repair in heterochromatin. Despite this potential kinetic delay, there are known mechanisms for DSB processing in heterochromatin that support a model for the initiation and maturation of oocyte crossovers in these dense chromatin states. Evidence from studies in fruit flies demonstrate that DSB sites are relocated outside of heterochromatic compartments and processed for repair near the nuclear envelope in a more physically accessible environment (Chiolo et al. 2011; Caridi et al. 2017). Further, DSBs made in euchromatin versus heterochromatin are processed with similar kinetics, indicating that heterochromatic DSBs are not wholly refractory to crossover formation (Janssen et al. 2016).

We must note that our analysis of the overlap of oocyte crossovers in heterochromatin does not preclude the notion that these chromatin modifications are maintained and/or present during DSB processing and crossover formation. It remains to be shown whether marks such as H3K27me3 are remodeled around sites of active recombination, or whether these same mechanisms of relocating recombination intermediates to more repair-permissive environments is active in *C. elegans* oogenesis and promoting sex differences in recombination. For spermatocyte crossovers, it could be that these events would take place in regions of accessible chromatin that happen to coincide with transcription. Oocytes, however, display some preference

for silenced, heterochromatic regions on some chromosomes, and the reason behind this higher than expected frequency is worthy of further investigation.

Sexually dimorphic crossover distributions and gene evolution

Our results indicate that crossover recombination in *C. elegans* spermatogenesis is largely associated with genes that are actively express during in meiosis. Recombination in genes can be mutagenic (Arbel-Eden and Simchen 2019) and at the very least introduces genetic diversity in progeny through reciprocal exchange genetic content on homologs. Further, studies have found that transposable element activity, large scale differences in DNA methylation, and spermatocyte-specific euchromatic chromatin modifications create a "promiscuous" state that enables the differentiation of current and novel genes (Kurhanewicz et al. 2020; Kaessmann 2010). Thus, our detection of elevated crossover formation in genes during spermatogenesis support hypotheses that suggest meiotic genes are rapidly evolving (Van Oss and Carvunis 2019; Swanson and Vacquier 2002). Notably, the 'out of testis' hypothesis poses that male meioses may be a unique source of selection on reproductive genes (Kaessmann 2010). Our data aligns with these hypotheses and evidence from studies in fruit flies that demonstrated genes expressed in spermatogenesis are under positive selection, which promotes adaptation and fixation of beneficial alleles in populations (Betrán and Long 2003; Yang and Bielawski 2000). The exact molecular mechanism that is driving this preference for crossover in euchromatin in spermatocytes but not oocytes remains elusive, and further investigation is need to understand the rate at which recombination in spermatocytes may be promoting the evolution of meiotic genes in *C. elegans* populations.

Materials and Methods

Caenorhabditis elegans strains and maintenance

All strains were incubated at 20°C and maintained nematode on growth medium (NGM) plates seeded with the OP50 strain of *Escherichia coli.* Strains used in this experiment include the following: N2 (wildtype from Bristol, England), CB4856 (wildtype from Hawaii, United States), EG7841 (oxTi302 [eft-3p::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] I), and CB4108 (*fog-2*(q71) V). All genetic crosses were done by mating L4 stage males and hermaphrodites on NGM plates and screening for cross progeny after 3-4 days.

Crossing schemes for sex-specific crossover mapping

Parent (P0) N2 hermaphrodites were mated to CB4856 males to generate Bristol/Hawaiian F1 hybrids. Individual F1 progeny were placed onto their own plates and separated into two separate cross schemes. To assess oocyte recombination, F1 hermaphrodites at the L4 stage were then mated to EG7841 males, and F2 cross progeny marked with red fluorescent bodies were collected and briefly transferred to a new plate. To assess spermatocyte recombination, F1 males at the L4 stage were mated to CB4108 hermaphrodites, and male F2 cross progeny were collected and briefly transferred to a new plate before sucrose floatation.

Sucrose floatation and isolation of individual F2 progeny

To minimize bacterial contamination in downstream gDNA sample preps, we performed sucrose floatation on pooled worms from each cross scheme, respectively. Previously collected worms were immediately washed from plates with 8mL cold M9 buffer and transferred to 15mL

glass centrifuge tubes using a glass Pasteur pipette. Collected worms were centrifuged at 3000rpm at 4°C and washed in 4mL of fresh M9 twice. To separate worms from bacteria and other debris, 4mL of 60% sucrose solution was added to 4mL of M9 buffer and worms and vortexed briefly. The mixture was then spun at 5000 rpm at 4°C for 5 minutes. Using a glass pipette, the floating layer of worms were transferred to a new glass centrifuge tube on ice and brought up to 4mL in fresh M9. Worms were then incubated at room temp for 30 minutes and gently vortexed every 5 minutes. Worms were washed three times in equal volume of fresh M9 were performed before storing collected worms at -80°C in 1.5mL microcentrifuge tubes containing M9 buffer. To isolate individual F2 progeny before sequencing, 10-20 worms were transferred via glass Pasteur pipette into M9 buffer on a glass well slide. Individual worms were then transferred into 10uL of M9 in a single well of a 96-well PCR plate. Plates containing F2 progeny were then briefly stored at -80°C before whole-genome sequencing.

Illumina Whole Genome Sequencing and data processing

I developed a method for the high throughput analysis of meiotic recombination in individual genomes of *C. elegans* with as little as 0.75-1.0ng of DNA per sample. We sequenced 610 individual genomes (300 oocyte-derived samples and 310 spermatocyte-derived samples) at nearly half of the cost compared to available commercial methods. For Illumina short-read sequencing, library preparation was performed on individual worms for by the University of Oregon's Genomics and Cell Characterization Core Facility. The short-read libraries were then sequenced on an Illumina Novaseq (2 x 150bp). Illumina short reads from each individual F2 genome were trimmed using Trimmomatic (Bolger, Lohse, and Usadel 2014) to remove adapter

and barcode sequences. The trimmed reads were then aligned to the N2 Bristol reference genome (NCBI accession number PRJNA907379). All resulting variant positions comparing our N2 Bristol and CB4856 Hawaiian genomes are in relation to the N2 Bristol assembly. Aligned reads in SAM format were then sorted using SAMtools (Li et al. 2009) and converted to BAM files. Using Picard read groups were added via AddOrReplaceReadGroups, and duplicate reads were filtered using MarkDuplicates as described above. BAM files with filtered duplicate reads (Barnett et al. 2011) were used to call variant and homozygous reference sites using GATK HaplotypeCaller (McKenna et al. 2010) to generate per-sample VCF files. Each sample's VCF file was then filtered to only include homozygous SNP sites as determined in chapter I. SNP markers were then further filtered to exclude sites within repeats, low-complexity sequences, and transposons identified by RepeatMasker. We used 213,591 high quality SNP markers out of the 246,298 homozygous SNPs identified from comparing our lab's genome assemblies for N2 Bristol and CB4856 Hawaiian (Chapter 1, Table 1). In our N2 Bristol genome assembly, we improved upon previous SNP maps to cover more than 99.9% of each chromosome's length with 99.97% of the total genome covered. The average SNP spacing on each autosome (*I, II, III, IV*, and V) is 645bp, 354bp, 580bp, 674bp, and 262bp, and the average SNP spacing on the X chromosome is 1020bp (Supplemental Figure S4.1A). Many regions, such as the terminal "arm" domains of each chromosome have a much greater density of SNPs, and thus resolution of crossover detection is high. The centers of each chromosome have a much lower density of markers, with the largest gap between SNPs being 106kb on chromosome V.

Reconstruction of F1 chromosomal products of meiosis

To aid crossover detection, a Hidden Markov Model approach was used to reconstruct the chromosomes present after crossing over between the Bristol and Hawaiian homologous chromosomes in eggs and sperm of the F1 generation. A previously developed pipeline, TIGER (Rowan et al. 2015), was adapted for use on our samples and control for running each TIGER script was combined into a single shell script. Read counts and alleles for each SNP marker were generated from the output of GATK's HaplotypeCaller, and input files were prepared as described in (Rowan et al. 2015; https://github.com/betharowan/TIGER_Scripts-for-distribution). To review their framework, the ratio of read counts between the N2 Bristol and CB4856 Hawaiian backgrounds are transformed into a discrete alphabet and 6 states represented as AA, AU, AB, BU, BB, and UU. AA represents a homozygous N2 Bristol state, AB is the heterozygous state, U represents uncertainty in the homozygous state, and UU represents no reads/information at a given marker. A marker site with a minimum of 5 reads where 100% of the reads are from a single parental background are required to assign genotypes AA or BB. If there are fewer than five reads aligned or both parental alleles were observed at the same marker, genotypes were inferred from a multinomial distribution. Observing reads in a homozygous parental background should follow a binomial distribution such that 99% of reads (accounting for 1% sequencing errors) are from either N2 or CB4856, and 50% for the heterozygous state. Genotypes were then assigned according to the maximum value between the homozygous parental and heterozygous probabilities. Genotypes AU or BU were assigned unless there was an equal probability of each genotypic background at these markers with low coverage and/or uncertain parental origin.

An HMM is then generated for each individual genome to infer Bristol or Hawaiian haplotype blocks on each chromosome. Transmission and emission probabilities are estimated per sample for each model, and hidden states (homozygous N2, homozygous CB4856, or heterozygous) were then determined via the Viterbi algorithm. Output files for each sample that contain the intervals of each inferred hidden state were then converted to Pandas data frames in Python for further processing.

Feature engineering for Random Forest Classification

Random Forest Classification is a supervised machine learning method that can be used to automate the classification of data using a model trained on numerical features of the data. Our random forest classifier will label transitions between each haplotype block on a single chromosome as Crossover or "not crossover". Each haplotype interval is given as a pair of SNP coordinates denoting the start and end of each N2 or CB4856 interval. Between each haplotype interval, a new "transition" interval was created that is defined as two SNP markers whereby the start of the transition interval is the last SNP in a haplotype interval prior to transition, and the end of the transition interval is the first SNP in the following haplotype interval. To label these transition intervals as potential crossovers, numerical features associated with these transitions must be made, and these numerical data were calculated in varying window sizes from 10kb up to 10Mb centered on each transition interval. First, for each window size, the percentage of SNPs with CB4856 alleles upstream and downstream were calculated, and then the difference between these values was recorded as a separate feature. For an ideal crossover breakpoint, a transition interval will have the highest possible difference in the percentage of CB4856 alleles upstream

vs downstream. Secondly, for each window size, the cumulative distribution of both CB4856 and N2 alleles, respectively, was stored as an array such that each marker's index in the array contains the cumulative percentage of that given genotype up to that marker's position in the window. A third array was then generated as the difference between the CB4856 cumulative distribution array and the N2 array. Then, the values for the difference in cumulative genotype distributions for the marker positions at the start of the transition interval and the marker at its end were averaged to assign an overall difference in cumulative genotype distributions to the transition interval. For an ideal crossover breakpoint, a transition interval will have the highest possible difference between each genotype's cumulative distribution in that window. All these calculations were performed in windows rather than across the whole chromosome to increase the sensitivity of detection for both single and double crossover events on a single chromosome. Finally, in each individual chromosome, all numerical features associated with each transition interval were converted to a zero to one scale by normalizing to the maximum value of each feature across all transitions.

Training and Random Forest Classification of Crossover Sites

The training data for this classifier includes 217 of the 3,660 chromosomes sequenced. The training data includes a nearly even mixture of spermatocyte and oocyte samples as well as a mixture of non-recombinant, single crossover and double crossover chromosomes determined by manual visualization and confirmation. The classifier was set up as a forest of 500 decision trees, and the model was trained on 70% of the training dataset and then validated on the remaining 30%. Validation showed that the model had 99.7% accuracy in appropriately labeling crossover

versus non-crossover transition intervals. The most features that provided the greatest amount of accuracy to the model were calculations performed in window sizes of 500kb, 1MB, and 2MB. Transition intervals labeled as crossover breakpoints were retained for further analyses, and chromosomes with no labeled crossover breakpoints were only retained to calculate the total map length in centimorgans for each chromosome. Chromosomes were visually inspected to further validate automatic classification of crossover breakpoints. Notably, the first 50kb of the X chromosome was responsible for 40 samples having improper crossover calls. In each of these samples' X chromosomes, there were two crossovers identified: one in the 50kb region and another crossover as close as 1Mb away. Because three SNP intervals were improperly identified as crossovers in all X chromosomes in these 40 samples, these markers/intervals were then removed from the final crossover dataset given the extremely unlikely probability of such a high rate of crossing over in the same narrow region. Finally, 16 samples that were sequenced between below 1X read coverage did not have sufficient read data to confidently call crossovers. 18 crossovers on 17 chromosomes were recovered by visualization and manual inspection of the highest probable transition intervals.

Statistical Analysis of Crossing Over

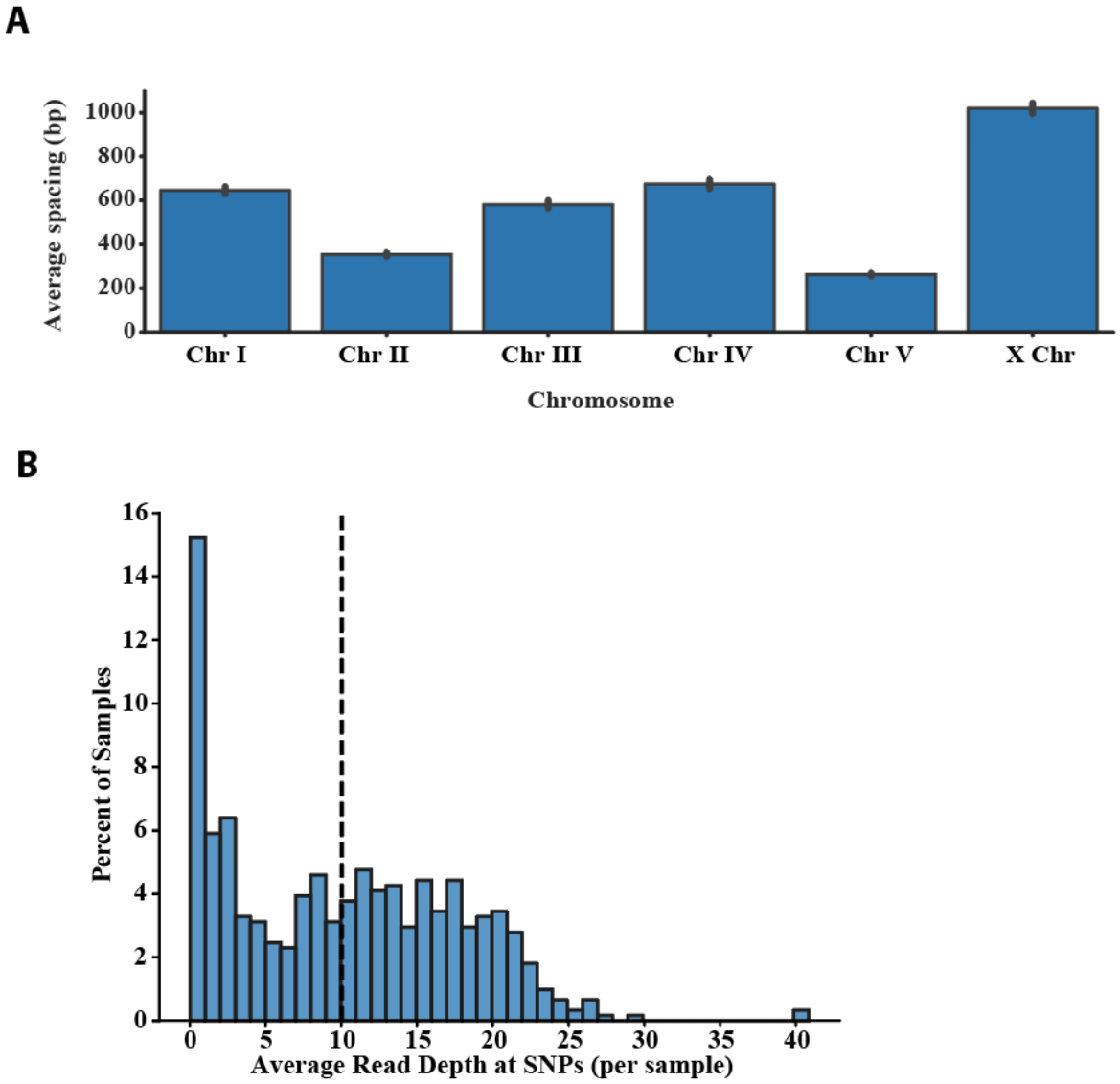
Whole chromosome map units (MU, centimorgans) were calculated using the formula MU = 100* (# of single crossovers (SCOs) + 2(# of DCOs)) / sample size. MU for specific intervals were calculated using the formula MU = Total MU for chromosome * (# of crossovers in interval / # of crossover in chromosome). To statistically define the "arm" versus central region of each chromosome, historically defined by a shared recombination rate across a multi-

megabase domain, piecewise linear regression was used. Briefly, the *pwlf* package in python (https://github.com/cjekel/piecewise_linear_fit_py) was used to perform segmented linear fits with the given parameters of two undefined breakpoints yielding three domains with a shared recombination rate. The "left tip" and "right tip" domains were added to each chromosome and defined as the non-recombining SNP intervals at the distal ends of each chromosome.

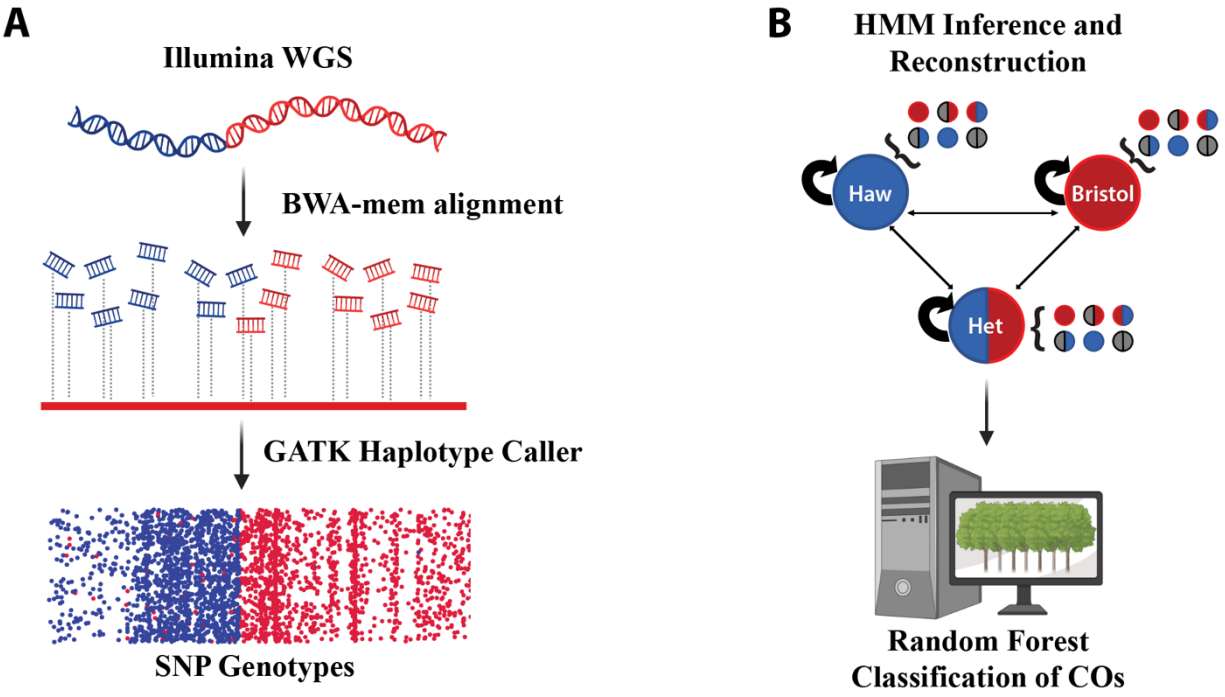
Genomic features correlated with crossing over

Gene annotations (gene, mRNA, exon, CDS, etc.) were downloaded from Ensembl for the ce11 genome assembly. Annotations for regulatory sequences, transcription factor binding sites, etc, were downloaded from Wormbase for the ce11 assembly as well. The coordinates for these sequence level annotations were then remapped onto our N2 Bristol genome assembly using Minimap2 alignments in the LiftOff program (Li 2018; Shumate and Salzberg 2021). Sex-specific germline RNA-seq annotations (Tzur et al. 2018) were taken and applied to our sequence level "gene" annotations in the remapped Ensembl dataset. Sex-specific germline ChIP-seq data (NCBI BioProject PRJNA475794) were aligned to the N2 genome and peaks were called using MACS3 (Feng et al. 2012) using parameters as previously described (Tabuchi et al. 2018). To identify pairing center regions, PC motifs from REF were identified and cluster analysis was performed in our genome assembly using MCAST (Bailey and Noble 2003) with maximum motif spacing of 100bp to be considered in the same cluster. Pairing center regions in the arms were defined as the interval from the occurrence of the first cluster to the last cluster detected in the arm region as defined by piecewise linear regression. To determine the fold enrichment (or depletion) of crossovers in each of these genomic features, the Genomic

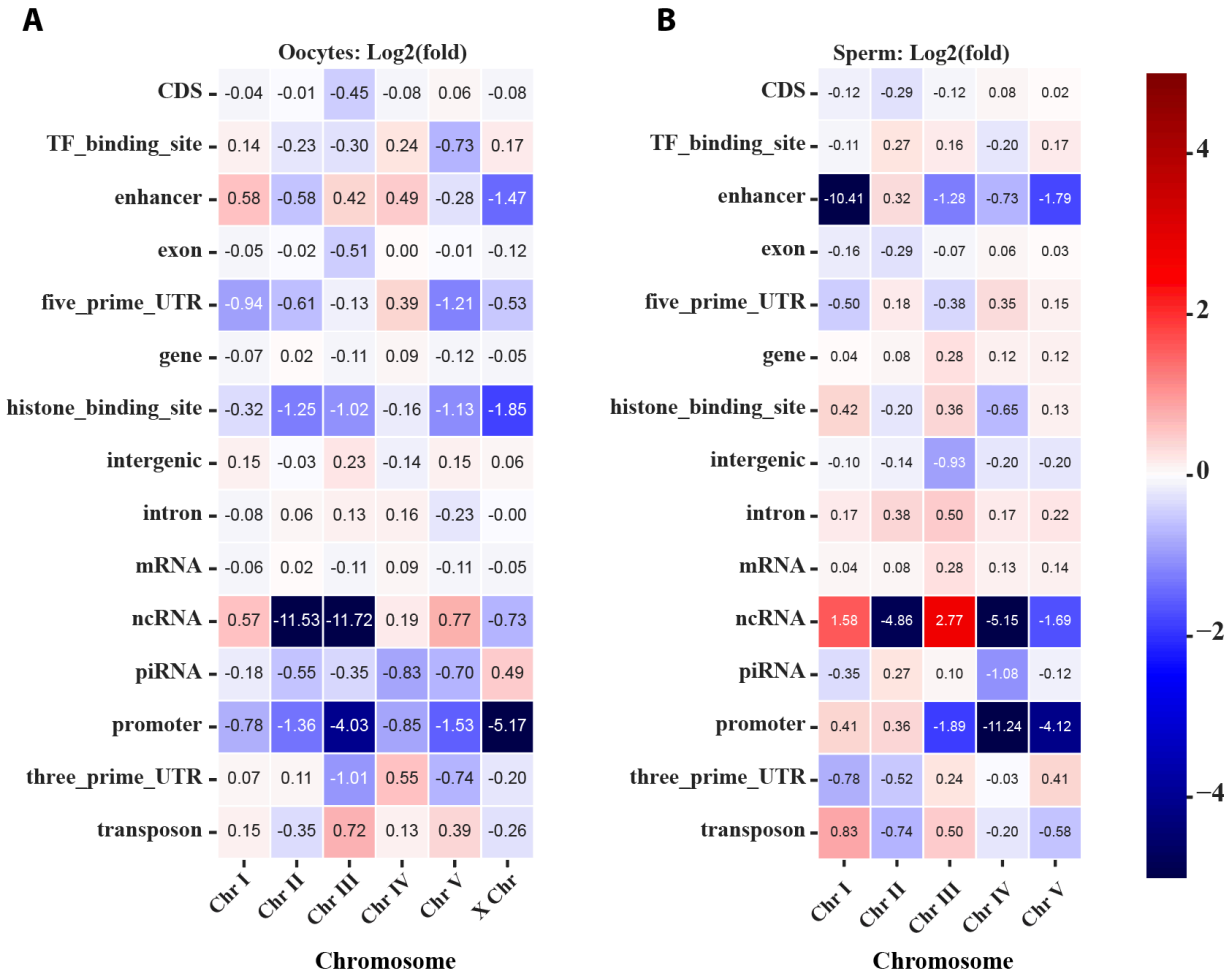
Association Tester (GAT, Heger et al. 2013) was used to simulate a null distribution of overlaps from 10,000 randomly shuffled intervals per annotation type. Significance of fold change (observed vs null/expected) was determined by hypergeometric test. Significance of the difference between fold changes in each sex was determined by examining the fold change ratio of egg vs sperm via hypergeometric test.



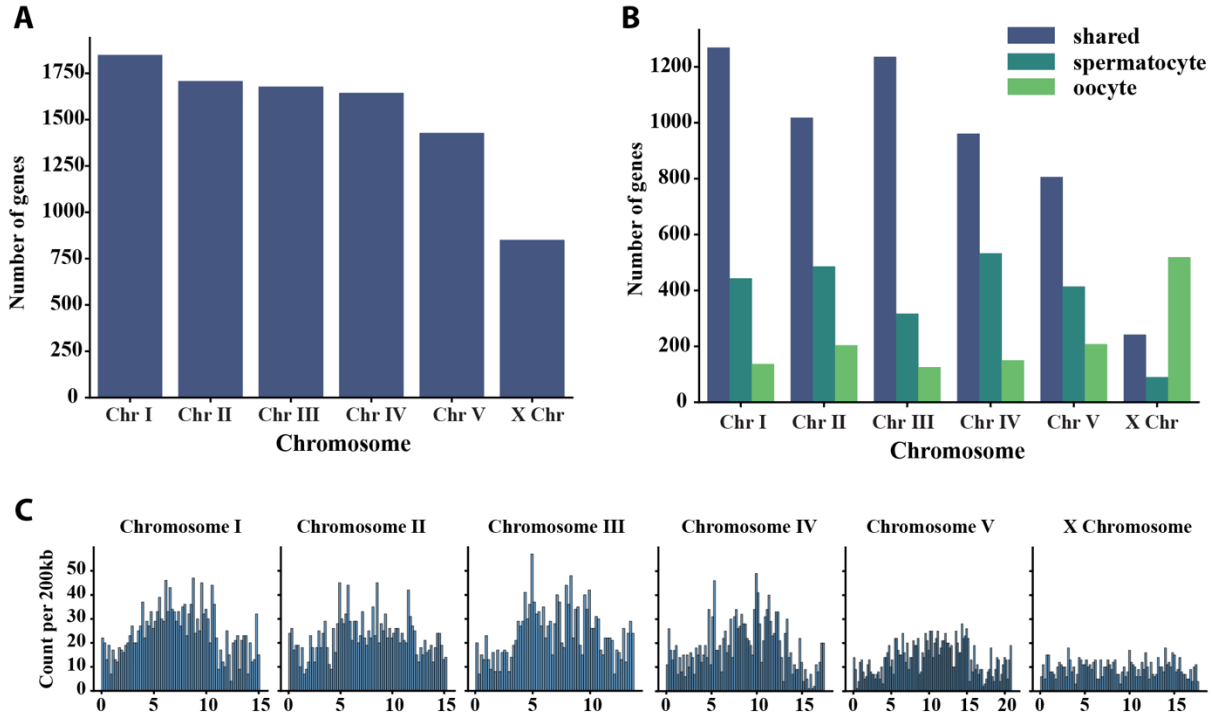
Supplemental Figure S4.1. Resolution and sequencing of F2 recombinant progeny. (A) A bar chart depicting the average distance between SNP markers on each chromosome in the genome. (B) A histogram depicting the distribution of genome coverage in individual samples. Genome coverage is depicted here as the global average read depth at SNP marker sites. Vertical dashed line indicates the average across all samples.



Supplemental Figure S4.2. Computational pipeline for processing of sequencing data. (A) Data collection and pre-processing including Illumina whole-genome sequencing, read alignment, and SNP calling. (B) Schematic of Hidden Markov Model based reconstruction of recombinant chromosomes and subsequent crossover detection via supervised machine learning and random forest classification.



Supplemental Figure S4.3. Association of crossovers with sequence-level annotations. (A) Heatmap showing the log2(fold) enrichment or depletion of oocyte crossovers with each sequence annotation. (B) Heatmap showing the log2(fold) enrichment or depletion of spermatocyte crossovers with each sequence annotation. All annotations were taken from the Ensembl ce11 genome annotation set and remapped to Libuda N2 Bristol genome assembly via LiftOff (See methods).



Supplemental Figure S4.4. Quantification of meiotically expressed genes in *C. elegans*. (A) Bar chart showing the total amount of all genes expressed in the germline on each chromosome. (B) Bar charts showing the number of genes expressed in both germlines or exclusively in oogenesis versus spermatogenesis on each chromosome. (C) Histogram showing the distribution of all genes expressed in the germline on each chromosome.

Bridge to Conclusion

In chapter 4, I presented a combination of methods that leverage our genome assemblies and SNPs identified in chapter 2 to map crossovers with high resolution in the chromosomal products of single meiotic events. By generating the first high resolution map of crossovers in *C. elegans* spermatogenesis, I was able to accurately compare the spermatocyte versus oocyte crossover landscape to show sex differences in the distribution and rate of crossing over. Further, the results of these studies reveal specific chromatin states as potential regulators of crossing over, with each chromatin state examined having opposing effects in the germline of either sex. I then speculated on other factors contributing to the multi-layered regulation of sex-specific crossover recombination and discussed the potential effects on sex-biased gene/genome evolution. In the conclusion, I will give one last summary of all the work presented in this dissertation with a final look at the directions future research can take to complete our understanding of how genomic integrity is maintained.

CHAPTER 5: CONCLUSION

In this dissertation, we illuminated and characterized the complex nature of large-scale genomic structural variants and the potential impacts that local differences in chromosome architecture play in shaping the landscape of genomic diversity. We then show how these same higher order chromatin structures influence distinct mechanisms of DNA repair in a sexually dimorphic manner to ensure faithful genome inheritance while also contributing to genetic diversification of progeny. This dissertation lays the foundation for future research that will illuminate the molecular mechanisms that both protect and diversify the information stored in our genomes.

In chapter 2, with my co-authors, we addressed the question of whether distinct chromatin states differentially affect the rate at which SNPs, indels, and SVs may appear across the genome. To do this, we generated reference genome sequences for the Bristol and Hawaiian isolates of *C. elegans*, and further demonstrated how the generation of reference genomes from the same assembly pipeline is the ideal approach for comparative genomic studies. By revisiting the genomic distribution of SNPs and indels, we revealed a greater density of genetic variation than previously uncovered by earlier studies using disparate sequencing technologies and methods to compare genome assemblies. Moreover, our methods enable the most accurate detection of SVs in the *C. elegans* genome accompanied by a nuanced dissection of the many different types and sizes of SVs affecting genomic organization. We also presented a novel method of tracking transposable DNA elements, which historically has been very computationally challenging due to the highly repetitive nature of their sequences. In those analyses, we demonstrate how a highly abundant and previously understudied family of

transposons are the most mobile genetic element between the Bristol and Hawaiian genomes. Lastly, we demonstrate a unique role for the heterochromatin state in contributing to the frequency of genomic SVs.

How does chromatin influence the rate at which different chromatin-associated variants arise? In these studies, we examined the genomic endpoints of the tens of thousands of generations of divergence between the Bristol and Hawaiian isolates of *C. elegans.* Our data strongly suggest that the heterochromatic versus euchromatic compartments of the genome likely have different influences on the rate of variant accumulation. We demonstrated that heterochromatic regions are relatively enriched for all types of variation, but it remains unclear as to what the specific rates of mutation accumulation are in these regions. Future studies leveraging experimental evolution methods could utilize genome-wide approaches as we employed here to assess the specific rates at which SNPs, indels, and SVs appear in distinct heterochromatin versus euchromatin states. Additionally, our results from chapter 4 suggest that sex-differences in germline chromatin may uniquely impact genomic diversity in different regions of the genome. Further research into whether there are sexually dimorphic rates of genomic variation in different chromatin structures/regions could further clarify how the spermatogenic versus oogenic genomes are especially ripe sources of genomic evolution.

In chapter 3, we discussed the ongoing genetic drift of laboratory lineages of the *C. elegans* model organism. There have been many reported cases of both genomic and phenotypic variance between laboratory lineages of the wild type strains most used for research. Our work in this chapter presents the first genome-wide analysis of the genomic variation between present-day lineages of both the Bristol and Hawaiian strains in multiple labs. We identified many SNPs

and indels predicted by prior studies of the mutation rate in *C. elegans,* and we build upon these findings by demonstrating the accumulation of many large and complex SVs over relatively short generational timescales. Our findings show that while most of the variation we identified is intergenic, it remains possible that sequence variants and chromosome rearrangements could be impacting hierarchical genome organization or gene regulatory sequences. Thus, present a comprehensive dataset of the genomic variation that likely underpins the phenotypic diversity observed across labs. Further, we discuss these data as a call to action for researchers to utilize strains with recently published genomes that reflect the organisms used in their research, as growing genomic divergence in laboratory wild types could impact the practice and interpretation of day-to-day research.

In research using model organisms, genetic drift between laboratory lineages of wild type strains is constantly ongoing and requires careful attention. We should aim to complement our understanding of genetic drift between natural populations in further research comparing the genomic distributions of SNPs, indels, and SVs between more laboratory strains and different model species. In this dissertation, we examined the quantity and distribution of genomic variation between different laboratory lineages of the Bristol and Hawaiian isolates, respectively, and found a very large fraction of the genome impacted by SVs. To more fully understand the problem genomic variants pose to model organism research, a careful examination of more laboratory lineages and more strains/isolates between labs is required. It remains possible that different strains and genomic backgrounds accumulate different types of variants at different rates. Laboratories passaging mutants that perturb certain chromatin structures, for example, may be accumulating variants that disrupt genome function at a higher rate than other strains. A more

comprehensive cross examination of many laboratory model organisms, then, would improve our understanding of how processes of genetic drift differ between natural and laboratory environments.

In chapter 4, we used both historical bioinformatics methods and novel applications of Random Forest Classification to develop a high-throughput method for whole genome sequencing and crossover mapping in single *C. elegans* genomes. We presented the highest resolution maps of crossover recombination in *C. elegans* to date, which includes the first genome wide map of crossovers in spermatogenesis. Using these incredibly precise recombination landscapes specific to each sex, we provide evidence that a fundamental mechanism of genome integrity is sexually dimorphic in developing sperm and eggs. Further, we leveraged simulations and enrichment analyses to show sex-specific crossover landscapes are differentially influenced by the landscape of different chemical modifications to chromatin.

What environmental and genetic factors could be differentially regulating crossover formation in each sex? Many biological processes are sexually dimorphic including aging and the response to different stressors. Future research should aim to resolve exactly how processes like aging affect the spatial distribution and rates of crossing over, and whether these effects on recombination are also sexually dimorphic. Additionally, we demonstrated that different chromatin states differentially influence the location of crossover formation in wild type sperm and eggs. To directly test how these distinct chromatin states regulate the crossover landscape, a genome-wide assessment of crossing over in mutant backgrounds that perturb specific chromatin modifications in the germline is required. Further, targeted and site-specific induction of the recombination program in different chromosome structures, not merely limited to histone

modifications, could further reveal how higher-order structures regulate recombination outcomes. Finally, this dissertation provides evidence for regulation of the crossover landscape considering the effects of sex and chromatin. To complete our understanding of how sexually dimorphic recombination is regulated by chromosome structures, future research is needed into how the landscape of noncrossover events fits into these regulatory schemes of recombination in the germline.

In conclusion, the research discussed in this dissertation highlight advancements in our fundamental understanding of genomic variation and integrity. The combination of next-generation DNA sequencing technology, simulations, and supervised machine learning presented here has provided novel insights into genome function. These data further support hypotheses about the interrelated nature of DNA structure, DNA repair, and genetic diversity while providing a platform for future research to further fill our gaps in knowledge of how the regulatory intersection of chromatin and/or biological sex influence the mechanisms that promote faithful genome inheritance.

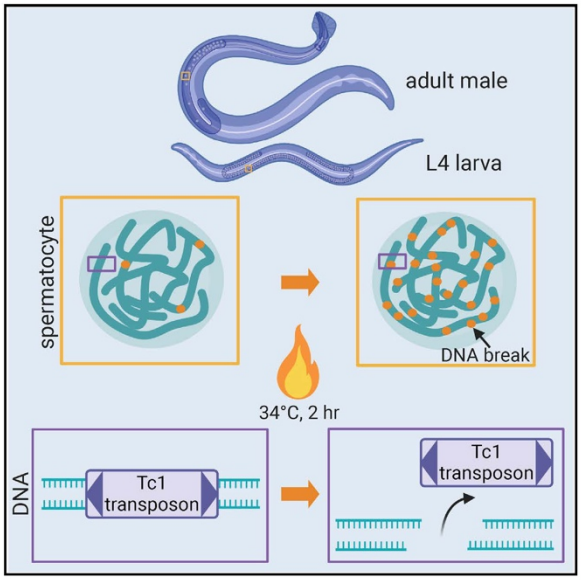
APPENDIX A

Report

Current Biology

Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes

Graphical Abstract



Authors

Nicole A. Kurhanewicz, Devin Dinwiddie, Zachary D. Bush, Diana E. Libuda

Correspondence

dlibuda@uoregon.edu

In Brief

Spermatogenesis in many organisms is sensitive to small temperature fluctuations, which can cause male infertility. Comparing *C. elegans* spermatocytes and oocytes, Kurhanewicz et al. find that heat-shocked spermatocytes exhibit large amounts of DNA breaks and Tc1/*mariner* transposon mobilization, which compromise genome integrity and male fertility.

Highlights

- Brief heat shock triggers DNA damage in spermatocytes
- Heat-induced DNA damage is SPO-11 independent and negatively impacts fertility
- Tc1/*mariner* transposons mobilize in the male germline after heat shock
- Female germline represses heat-induced DNA damage and Tc1 mobilization

Kurhanewicz et al., 2020, Current Biology *30*, 1–11
December 21, 2020
https://doi.org/10.1016/j.cub.2020.09.050

CellPress

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

Current Biology

CellPress
OPEN ACCESS

Report

Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes

Nicole A. Kurhanewicz,[1] Devin Dinwiddie,[1] Zachary D. Bush,[1] and Diana E. Libuda[1,2,*]
[1]Institute of Molecular Biology, Department of Biology, University of Oregon, 1229 Franklin Boulevard, Eugene, OR 97403, USA
[2]Lead Contact
*Correspondence: dlibuda@uoregon.edu
https://doi.org/10.1016/j.cub.2020.09.050

SUMMARY

Sexually reproducing organisms use meiosis to generate haploid gametes and faithfully transmit their genome to the next generation. In comparison to oogenesis in many organisms, spermatogenesis is particularly sensitive to small temperature fluctuations, and spermatocytes must develop within a very narrow isotherm [1–4]. Although failure to thermoregulate spermatogenetic tissue and prolonged exposure to elevated temperatures are linked to male infertility in several organisms, the mechanisms of temperature-induced male infertility have not been fully elucidated [5]. Here, we show that upon exposure to a brief 2°C temperature increase, *Caenorhabditis elegans* spermatocytes exhibit up to a 25-fold increase in double-strand DNA breaks (DSBs) throughout meiotic prophase I and a concurrent reduction in male fertility. We demonstrate that these heat-induced DSBs in spermatocytes are independent of the endonuclease SPO-11. Further, we find that the production of these heat-induced DSBs in spermatocytes correlate with heat-induced mobilization of Tc1/*mariner* transposable elements, which are known to cause DSBs and alter genome integrity [6, 7]. Moreover, we define the specific sequences and regions of the male genome that preferentially experience these heat-induced *de novo* Tc1 insertions. In contrast, oocytes do not exhibit changes in DSB formation or Tc1 transposon mobility upon temperature increases. Taken together, our data suggest spermatocytes are less tolerant of higher temperatures because of an inability to effectively repress the movement of specific mobile DNA elements that cause excessive DNA damage and genome alterations, which can impair fertility.

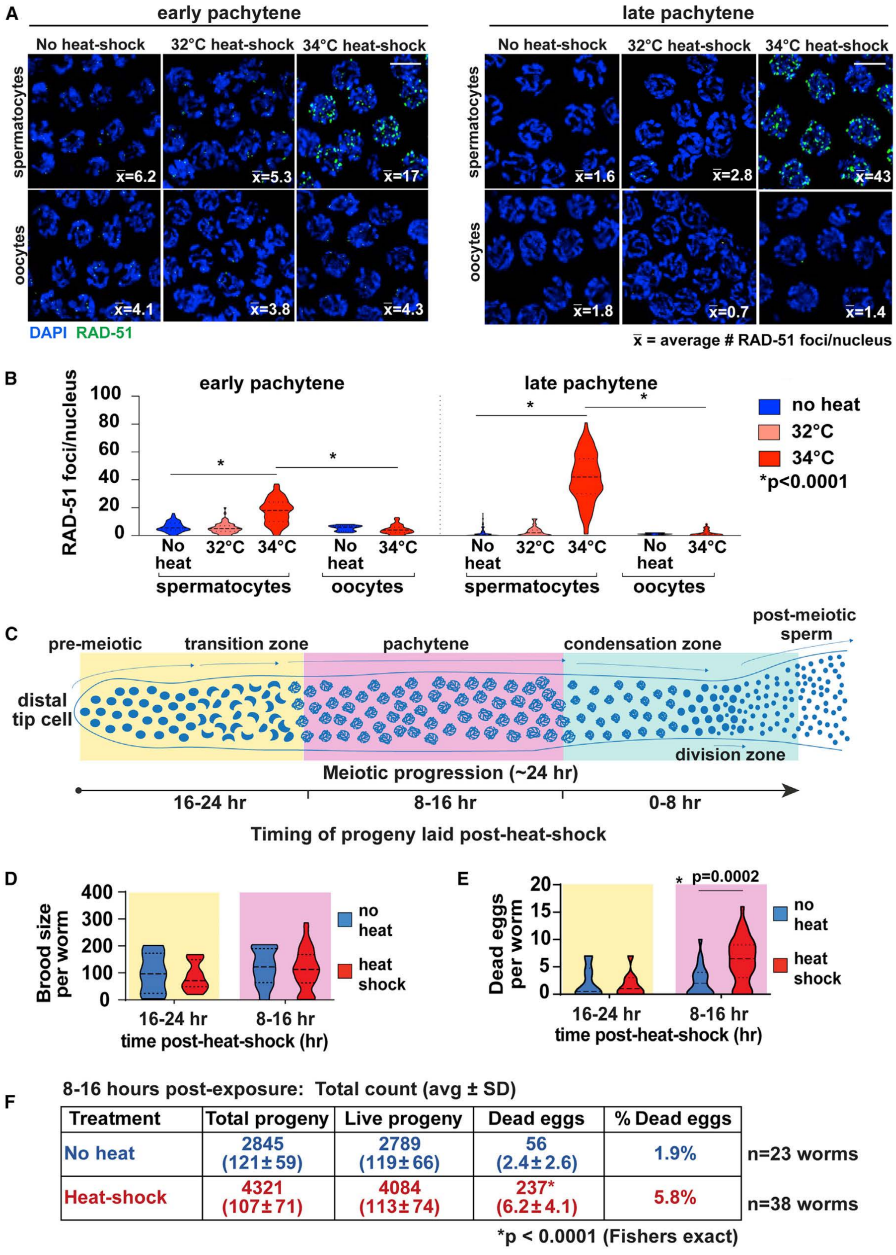
RESULTS AND DISCUSSION

Elevated Temperatures Produce DNA Damage in Spermatocytes during Meiotic Prophase I

Several studies indicate that spermatocytes in meiotic prophase I, as well as post-meiotic sperm, are particularly susceptible to thermal stress [1, 8, 9]. Further, the dysregulation of DNA damage and repair pathways during meiosis have been implicated as potential mechanisms linking heat stress and DNA damage in spermatocytes [10]. To investigate this link between heat stress and DNA damage in spermatocytes, we exploited the model system *Caenorhabditis elegans*, which allows for isolation, visualization, and comparison of both developing oocytes and spermatocytes.

To assess DNA damage during oogenesis and spermatogenesis, immunofluorescence for the recombinase RAD-51 [11] was used in the germlines of wild-type adult males (exclusively undergoing spermatogenesis) and adult hermaphrodites (exclusively undergoing oogenesis). Immediately following a 2-h heat shock of whole animals at 34°C (but not 32°C or extended exposure to 28°C), we found that, throughout meiotic prophase I, spermatocytes exhibit elevated RAD-51 foci (Figures 1, S1, and S2) that largely disappeared by 3 h post-heat shock (Figure S3). In contrast, heat-shocked oocytes did not display any increase in RAD-51 foci (Figure 1). In heat-shocked spermatocytes, RAD-51 foci increased nearly 3-fold in early pachytene nuclei (17 ± 8.7 foci/nucleus [n = 119] with heat shock versus 6.2 ± 3.6 foci/nucleus [n = 150] without heat shock; $p < 0.0001$; Figure 1) and late pachytene nuclei exhibited an approximately 25-fold increase in RAD-51 foci (41 ± 17 foci/nucleus [n = 110] versus 1.6 ± 2.7 foci/nucleus without heat shock [n = 146]; $p < 0.0001$; Figure 1). In contrast to meiotic prophase I, heat-induced DNA damage was not observed in pre-meiotic S phase or in mitotically dividing spermatocytes (Figure S1A). This restriction of heat-induced DNA damage to meiotic prophase I suggests that feature(s) unique to spermatocytes in prophase I permit the formation of heat-induced DNA damage. Further, during the early L4 larval stage of hermaphrodites, when their germline is exclusively undergoing spermatogenesis, a 34°C heat shock also generates DNA damage (Figure S1B). These results in the spermatocytes of the early L4 hermaphrodite indicates the heat-induced DNA damage is not sex specific but spermatocyte specific. Notably, heat-induced DNA damage in spermatocytes only occurs when the temperature threshold of 34°C is reached. At 36°C and 38°C, heat-induced DNA damage is no longer limited to prophase I and we observe RAD-51 foci in pre-meiotic S phase region of the germline as well as the mitotically dividing spermatocytes (Figure S2A). In addition, adult hermaphrodite germlines exposed to 36°C and 38°C also exhibit DNA damage continuously from the pre-meiotic region through

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050



F

8-16 hours post-exposure: Total count (avg ± SD)

Treatment	Total progeny	Live progeny	Dead eggs	% Dead eggs	
No heat	2845 (121±59)	2789 (119±66)	56 (2.4±2.6)	1.9%	n=23 worms
Heat-shock	4321 (107±71)	4084 (113±74)	237* (6.2±4.1)	5.8%	n=38 worms

*p < 0.0001 (Fishers exact)

(legend on next page)

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

meiotic prophase I (Figure S2B). Thus, once the threshold temperature has been exceeded, it is possible that additional mechanisms begin to contribute to the production of heat-induced DNA damage outside of meiotic prophase I.

Acute Exposure to Heat Stress Impairs Male Fertility

To determine whether heat-induced DNA damage affects male fertility, we assessed the fertility of wild-type adult males exposed to a 34°C heat shock. Both the spatial-temporal arrangement of the *C. elegans* germline and the established rate of spermatocyte progression through the male germline [12] enables comparing the fertility of heat-shocked spermatocytes that experienced temperature-induced DNA damage (8–16 h post-exposure) from spermatocytes that did not experience this damage (16–24 h post-exposure and no heat shock; Figure 1C). Brood size was comparable between heat-shocked and no heat shock cohorts of spermatids (Figure 1D), indicating the heat-damaged spermatocytes are competent to fertilize an oocyte. The production of inviable eggs increased 3-fold in the 8- to 16-h post-exposure cohort (average [avg] 6 ± 4 dead eggs/brood; n = 38 broods) compared with their no heat shock counterparts (avg 2 ± 2 dead eggs/brood; n = 23 broods; Figures 1E and 1F; Fisher's exact test; $p < 0.001$). Thus, the cohort of sperm that experienced heat-induced DNA damage also produced elevated numbers of dead eggs. These data suggest that, although heat-damaged spermatocytes can develop into functional spermatozoa capable of fertilization, their genetic content is likely compromised, therefore resulting in inviable progeny.

Temperature-Induced DNA Damage Is SPO-11 Independent

During early meiotic prophase I, double-strand DNA breaks (DSBs) are induced by the conserved endonuclease Spo11 [13]. To determine whether the heat-induced DNA damage was due to altered Spo11 activity, we examined RAD-51 foci in adults lacking the SPO-11 endonuclease [14]. As *spo-11*-null mutants lack programmed meiotic DSBs, few to no RAD-51 foci are present in the no heat shock male or hermaphrodite germlines (Figures 2A and 2B). Although heat-shocked *spo-11*-null mutant hermaphrodites looked similar to no heat shock *spo-11*-null hermaphrodites and males, heat-shocked, *spo-11*-null male germlines exhibited heat-induced RAD-51 foci at a similar level to that of heat-shocked wild-type males (Figures 2A and 2B). Using *spo-11*-null hermaphrodites at the molt for late L4 larval-young adult developmental transition stage, where both oogenesis and spermatogenesis occur simultaneously within the same germline, elevated RAD-51 foci were detected specifically in the spermatocytes and not the oocytes in the late L4-young adult transition heat-shocked germlines (Figure 2C). Together, these data demonstrate that the heat-induced RAD-51 foci in spermatocytes are a SPO-11-independent and spermatocyte-specific response to heat stress.

To determine whether heat-induced RAD-51 foci represent DSBs, we exploited the fact that crossover events require a DSB as an initiating event. Using the *C. elegans* pro-crossover factor COSA-1, which forms foci at nascent crossover sites *in vivo* [15], we assessed crossover formation in *spo-11*-null mutants following heat shock. Due to a lack of programmed meiotic DSBs, both male and hermaphrodite *spo-11*-null mutants fail to establish crossovers and therefore fail to form COSA-1 foci (0.16 ± 0.43 foci per nucleus; n = 49). Upon heat shock, COSA-1 foci were largely restored in the spermatocytes of *spo-11*-null mutant males (Figure 2D; 3.5 ± 1.8 foci per nucleus; n = 80). Thus, at least a subset of the heat-induced RAD-51 foci represent DSBs that are competent for recruitment of crossover repair machinery.

Heat Shock Activates Tc1 Transposons

Because movement of a transposable element (TE) can result in the formation of a DSB independent of SPO-11 [16–21], we tested whether heat shock activates the *mariner* TEs Tc1 and/or Tc3 first using quantitative PCR to measure Tc1 and Tc3 transposase expression. In males exposed to heat shock, Tc1 transposase expression increased (gene expression ratio = 2.4 ± 0.25) compared with their no heat shock counterparts

Figure 1. Acute Exposure to Heat Stress Produces DNA Damage in Male *C. elegans* Germlines and Impairs Male Fertility

(A) Representative immunofluorescence images of recombinase RAD-51 (green) from early and late pachytene regions of wild-type adult male germlines (spermatogenesis) and adult hermaphrodite germlines (oogenesis) with and without heat shock (32°C or 34°C for 2 h). Numbers on panels report average number of RAD-51 foci per nucleus for each group. Scale bar represents 5 μm.

(B) Quantification of RAD-51 foci per nucleus in early and late pachytene for wild-type spermatocytes and oocytes. Violin plots show frequency distribution of the data: center dashed line (median) and bottom and top dotted lines (first and third quartiles, respectively). Statistical significance between groups was determined using the Kruskal-Wallis non-parametric test, with Dunn's test to account for multiple comparisons. Statistically significant differences indicated by *, where $p < 0.0001$. Number of nuclei scored from male germlines: early pachytene no heat shock, n = 150; 32°C, n = 74; 34°C, n = 119; late pachytene no heat shock, n = 146; 32°C, n = 49; 34°C, n = 110. Hermaphrodite germlines: early pachytene no heat shock, n = 20; 32°C, n = 11; 34°C, n = 63; late pachytene no heat shock, n = 21; 32°C, n = 11; 34°C, n = 45.

(C–F) Male fertility was assessed using wild-type male worms and *fog-2*-null obligate females.

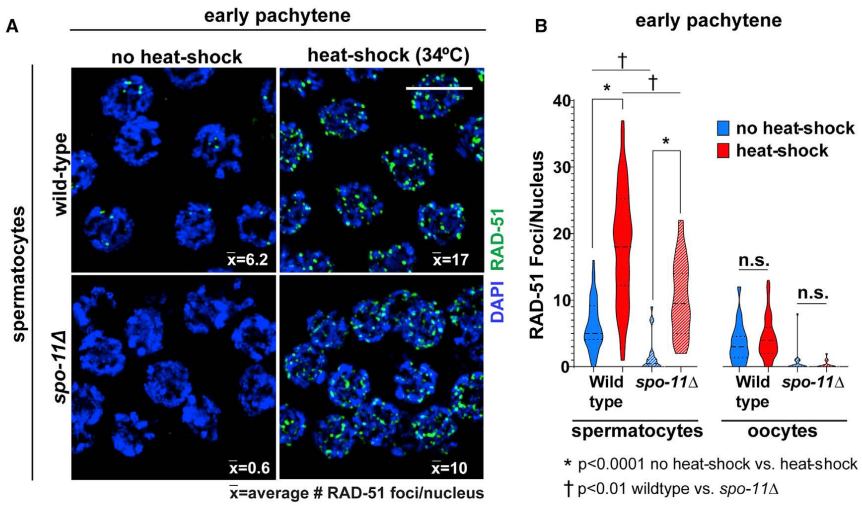
(C) Schematic of the *C. elegans* male germline with corresponding meiotic stages and 8-h intervals representing spermatocytes that were present at specific meiotic stages when heat shock occurred.

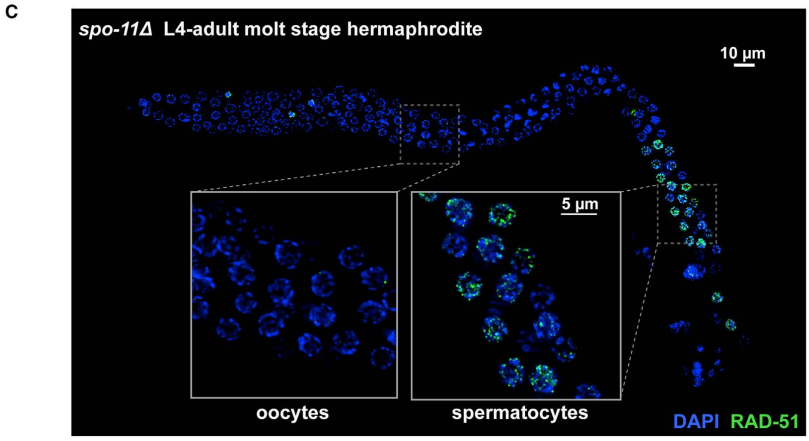
(D and E) Quantification of dead eggs and total brood size for each 8-h interval following heat shock. Violin plots show frequency distribution of the data: center dashed line (median) and bottom and top dotted lines (first and third quartiles, respectively). Statistical significance was determined using two-way ANOVA corrected for multiple comparisons. A statistically significant increase ($p = 0.0002$) in dead eggs per worm was found in the 8–16 h post-exposure window. Number of male worms scored for fertility: no heat shock 0–8 h, n = 8; 8–16 h, n = 23; 16–24 h, n = 6; heat shock 0–8 h, n = 10; 8–16 h, n = 38; 16–24 h, n = 11.

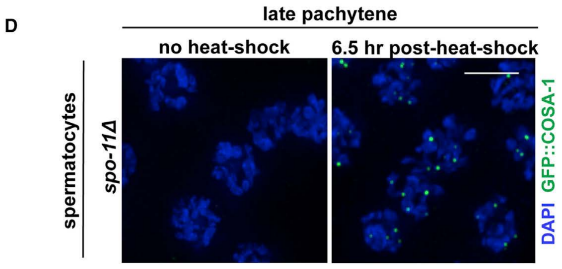
(F) Average ± standard deviation for total progeny, living progeny, and dead eggs per mated pair in the 8- to 16-h post-exposure window. Correlation between heat shock and production of dead eggs was assessed using a two-sided Fisher's exact test; total progeny no heat shock, n = 2,746; total progeny heat shock, n = 4,321; dead eggs no heat shock, n = 56; dead eggs heat shock, n = 237. A significant association was found between heat shock and the production of dead eggs $p < 0.0001$.

See also Figures S1–S3.

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050







(legend on next page)

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

(gene expression ratio = 1 ± 0.14; $p < 0.0001$). Notably, Tc1 transposase expression was significantly increased in males exposed to heat shock compared with hermaphrodites ($p = 0.0024$; Figure 3A, left). In contrast, Tc3 transposase expression remained unchanged upon heat shock in either males or hermaphrodites (Figure 3A, right). This result demonstrates that Tc1 and Tc3 transposase expression is differentially affected by heat stress, and males exhibit a significantly greater increase in Tc1 transposase expression following heat shock.

To determine whether the heat-induced Tc1 transposase expression leads to changes in transposon locations genome-wide, we used Illumina sequencing to map Tc1 and Tc3 locations throughout the genome (Figure 3B; see STAR Methods). All previously reported copies of Tc1 and Tc3 were represented in our dataset, including a second copy of Tc1 on chromosome X, which was not reported in Fischer et al. [22] but was later identified in Laricchia et al. [23]. Although ~90% of reads obtained mapped back to the known Tc1 and Tc3 locations previously identified in the N2 wild-type genome [22, 23], a number of *de novo* insertion sites (sites where Tc1 or Tc3 had not been previously reported; read threshold of >20 reads/site) were also identified for both Tc1 and Tc3 in males and hermaphrodites. Notably, we also identified Tc1 insertions at the site of the *unc-22(st136)* allele [24, 25], which contains a well-documented Tc1 insertion site identified during screens for worms with uncoordinated movement (neuron defects). The presence of *de novo* TE locations in the no heat shock control may represent Tc1 and Tc3 elements that have become recently mobilized and fixed in our N2 wild-type strain. However, given the relatively low read density at these *de novo* locations (avg = 405 ± 400 reads/location) compared with known Tc1 locations (avg = 4,114 ± 1,350 reads/location), this likely indicates a baseline level of Tc1 and Tc3 movement (Figure 3B).

Heat-shocked males exhibited a 65% increase in *de novo* Tc1 genomic locations (69 ± 14 *de novo* sites) compared with their no heat shock counterparts (42 ± 14 *de novo* sites; Figure 3C; Table S1), suggesting heat-induced mobilization of Tc1 in males. Although hermaphrodites exhibit a 1.6× increase in Tc1 expression following heat shock (Figure 3A), heat-shocked hermaphrodites exhibited a 16% reduction in *de novo* Tc1 locations (53 ± 15 *de novo* sites without heat shock versus 44 ± 11 *de novo* sites with heat shock; Figure 3C; Table S1). This result suggests that, unlike males, hermaphrodites can repress Tc1 transposase activity (or inhibit Tc1 translation) to prevent Tc1 mobilization. In concordance with the lack of sex-specific changes in Tc3 transposase expression changes upon heat shock, there was a much smaller incidence of novel Tc3 insertion in either sex (Figure 3C; Table S1). However, both males and hermaphrodites exhibited mild increases in novel Tc3 locations with heat shock (22% and 5%, respectively), which is likely due to the known temperature sensitivity of piwi-interacting RNA (piRNA) pathways that inhibit Tc3 transposition [26, 27]. Overall, these results indicate that heat shock increases the mobility of Tc1 specifically within the male genome.

The 2.4× increase in Tc1 transposase expression in males in comparison with the ~25× increase in DNA damage in spermatocytes upon exposure to heat shock suggests that other factors are likely contributing to the heat-induced DNA damage observed. For instance, Tc1 mobilization may not solely account for all heat-induced DNA damage. Although certain pathways regulate specific transposon families, including Tc3 regulation by the piRNA pathway [27, 28], there is also overlap between mechanisms that regulate various Tc families [29–31]. Notably, both the piRNA and mutator pathways have been reported to be heat sensitive [32, 33], and mutants for these pathways exhibit both elevated Tc transposase expression as well as transposon excision [7]. Thus, heat exposure may allow for germline activation of multiple transposons by disruption of individual or overlapping mechanisms, which may contribute to the production of heat-induced DNA damage. Alternatively, a single Tc1 transposase enzyme might work on additional DNA targets following mobilization of a Tc1 element. It is notable that nonlinear relationships between transposase expression and rates of transposition have been previously described for Tc1/mariner family elements [7, 27, 28]. Specifically, it was demonstrated that a 40× increase in Tc3 transposase expression led to a 100× increase in Tc3 transposition events [27]. Moreover, there may be repression or enhancement of Tc1 transposase activity, such as alterations to transposase binding to its 21-bp DNA target.

Intra- and Inter-chromosomal Mobilization of Tc1

To determine Tc1 mobility within individual chromosomes, Tc1 genomic location number was compared across each chromosome with known Tc1 copy number per chromosome. In males, *de novo* Tc1 genomic location number increased in frequency with the copy number of known Tc1 sites per chromosome

Figure 2. Temperature-Induced DNA Damage Occurs via a SPO-11-Independent Pathway

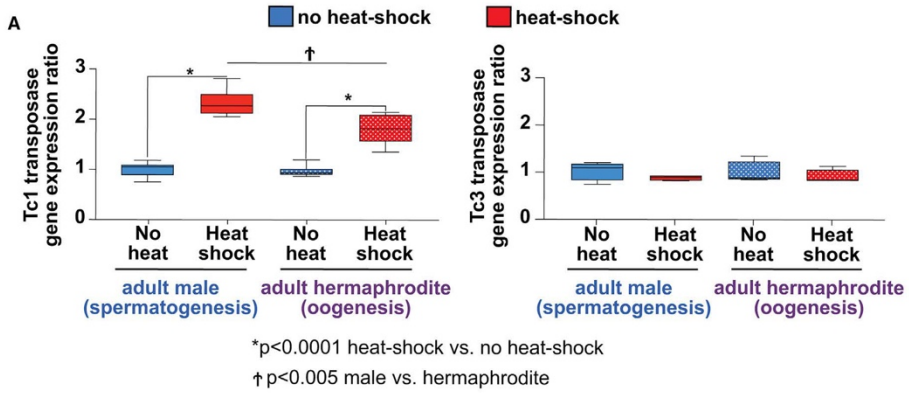
(A) Representative immunofluorescence images from early pachytene regions of male wild-type (top) and *spo-11*-null (bottom) germlines. Wild-type and *spo-11*-null mutant spermatocytes demonstrate similar relative increases in RAD-51 marked DNA damage with exposure to heat shock. Values reported within each panel report average number of RAD-51 foci per nucleus for each group. Scale bar represents 5 μm.

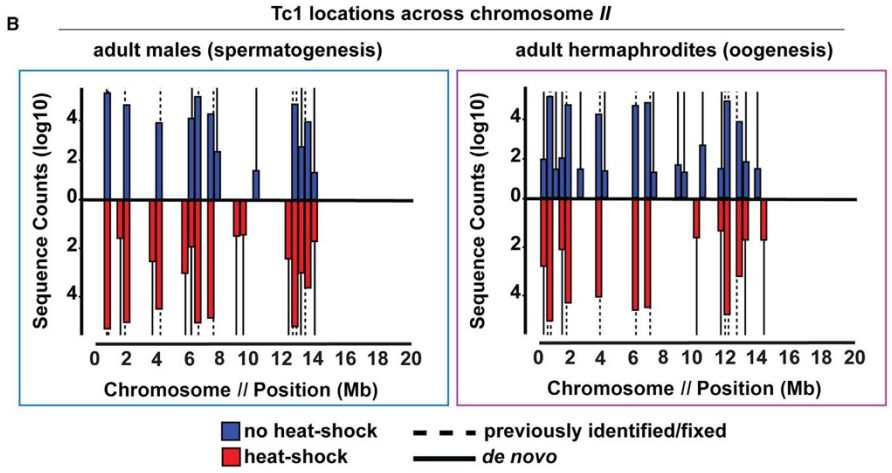
(B) Quantification of RAD-51 foci per nucleus during early pachytene for spermatocyes and oocytes of both wild-type and *spo-11*-null strains. Violin plots show frequency distribution of the data: center dashed line (median) and bottom and top dotted lines (first and third quartiles, respectively). Statistical significance between groups was determined using the Kruskal-Wallis non-parametric test, with Dunn's test to account for multiple comparisons. Statistially significant differences between heat- shock and no-heat-shock cohorts are indicated by * ($p < 0.0001$), while significant differences between wild-type and *spo-11* null strains are indicated by † ($p < 0.01$). Number of spermatocytes scored: wild-type no heat shock, n = 133; wild-type heat shock, n = 119; *spo-11* null no heat shock, n = 20; *spo-11* null heat shock, n = 19. Number of oocytes scored: wild-type no heat shock, n = 20; wild-type heat shock, n = 63; *spo-11* null no heat shock, n = 19; *spo-11* null heat shock, n = 15.

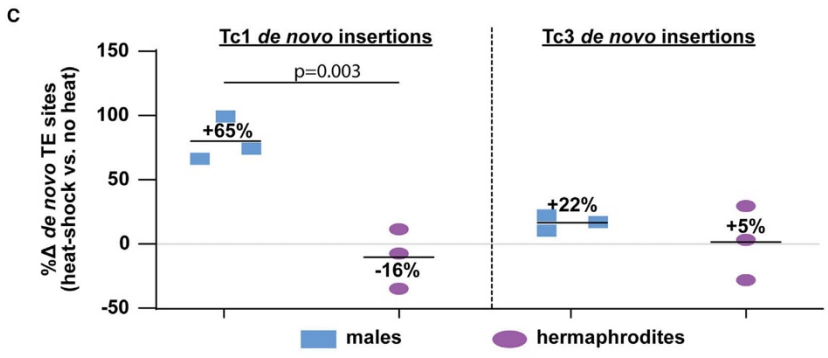
(C) Representative image of a *spo-11*-null hermaphrodite germline during the molt at the late L4-young adult developmental transition at 1-h post-heat shock. Although oocytes and spermatocytes are being produced in the same germline, only spermatocytes display a heat-induced increase in DNA damage. Enlarged panels display oogenic and spermatogenic regions of the gonad. Scale bars represent 5 μm in inset panels and 10 μm in whole gonad image.

(D) Formation of COSA-1 foci in late pachytene, indicating nascent crossover sites, was assessed in a *spo-11*-null strain with and without heat shock. Representative images from late pachytene of *spo-11*-null male germlines without heat shock and 6.5 h following heat shock are shown.

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050







(legend on next page)

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

($R^2 = 0.75$; $p = 0.026$; Figure 4A). In contrast, *de novo* Tc1 insertion sites in hermaphrodites do not strongly correlate with the copy number of known Tc1 sites per chromosome ($R^2 = 0.035$; $p = 0.72$). Similar to the anti-correlation of *de novo* Tc1 insertion sites in hermaphrodites, *de novo* Tc3 insertion sites in both males and hermaphrodites do not strongly correlate with the copy number of known Tc3 sites per chromosome ($R^2 = 0.034$, $p = 0.73$ and $R^2 = 0.47$, $p = 0.13$, respectively). Together, these results suggest that these chromosomes may contain preferential targets for Tc1 insertion. Alternatively, heat-induced Tc1 mobilization may preferentially undergo intrachromosomal movement. Further, because chromosomes with higher Tc1 copy number experience more Tc1 insertions, these chromosomes consequently may be more susceptible to amplification of Tc1 copy number due to preferential use of the homologous chromosome as a template for DSB repair during meiosis.

To determine where in the genome-specific Tc1 transposons move to upon heat shock, we analyzed sequencing reads that picked up known polymorphisms within the inverted repeat region of three Tc1 transposons in the genome [22]: (1) *TC1.6* (WBTransposon00000018 from Y48G9A), a pair of polymorphisms on chromosome *III*; (2) *TC1.9* (WBTransposon00000021 from Y46G5A), a pair of polymorphisms followed by a single-base-pair deletion on chromosome *II*; and (3) *TC1.24* (WBTransposon00000036 from T69A2AR), a single base deletion in the terminal inverted repeat followed by a polymorphism in the transposase-coding region on chromosome *IV* (Figure 4B). Reads were identified containing these polymorphisms and then parsed out by genomic location and experimental group, either heat shock or no heat (Figures 4C and 4D).

TC1.6 reads were found overwhelmingly at its established location (91% ± 4% of reads), regardless of heat shock or sex (Figures 4C and 4D). Because *TC1.6* on chromosome *III* contains a single-nucleotide polymorphism within the inverted repeat recognized by the Tc1 transposase [17], this result suggests that this polymorphism was sufficient to suppress mobility of *TC1.6*. Next, *TC1.9* reads were found widespread across the genome. Fewer than 0.5% of reads with the *TC1.9* polymorphism were found at its previously reported established location but were instead broadly distributed among the known Tc1 locations on all other chromosomes (Figure 4C). This result indicates that this polymorphism is no longer specific to any one copy of Tc1 in the genome of our wild-type N2 strain, suggesting it has been redistributed through the dynamic exchange of genetic information between TEs in the genome. Notably, reads containing this polymorphism were found to increase at *de novo* Tc1 insertion locations only in the male heat shock group ($p < 0.005$; Figure S4), indicating a general increase in transposon activity among the heat-shocked males. Lastly, the *TC1.24* reads were found more often at the *TC1.24* established location on chromosome *IV* than any other genomic position (35% ± 8% of total reads; Figure 4C). When identified at non-established sites, *TC1.24* was found primarily at other established Tc1 locations (Figure 4C). In heat-shocked males, *TC1.24* was located more often at non-established locations (79% ± 8% of reads from non-established location) compared with non-heat-shocked males (61% ± 15% of reads); whereas in hermaphrodites, *TC1.24* maintained a similar presence at its established location with and without heat shock (Figure 4C). This indicates *TC1.24* may either prefer to move into Tc1 sites that may have recently experienced an excision event or act as a preferred repair template for Tc1-induced DSB. Further, males with heat shock also exhibit an increase in *TC1.24* insertion into *de novo* locations (7% ± 6% of total reads) compared with their no-heat-shock counterparts (0.7% ± 0.3% of total reads). Overall, these data from three specific copies of Tc1 demonstrate the variability of transposon activity between individual copies of Tc1 and support our finding that Tc1 activity is globally increased upon heat shock specifically in males.

Tc1 Insertion Motifs

Previous studies reported a preferred Tc1 insertion motif, TATRTG (R = A or G) [16, 19, 34]. To determine whether Tc1 target sequences change following heat shock, sequencing reads were assessed for TATRTG [35]. Although TATRTG occurs at a relatively low rate across the genome (43,906 times across the 100,286,401 bp *C. elegans* genome), it was found within 11% ± 2% of our Tc1-adjacent reads. When reads are parsed out by sex and heat shock treatment, the TATRTG motif occurred consistently in reads from all experimental groups (Figure S4C). To assess the potential presence of novel insertion motifs in our dataset, we first aligned sequences and then quantified relative nucleotide frequencies at each of the first six positions 3′ adjacent to Tc1 (Table S2). Although the first five positions were consistent with the previously reported TATRTG motif, the sixth

Figure 3. Tc1 Transposase Expression and Increased Incidence of Tc1 Transposon Genomic Locations Are Associated with Temperature-Induced DNA Damage in Males

(A) Quantitative PCR of Tc1 and Tc3 transposase expression in both male and hermaphrodite populations with and without heat shock. Box and whisker plots show the distribution of the data: center line (median), box (first to the third quartile), and whisker lines (smallest and largest values recorded). An increase in Tc1 transposase expression (calculated as gene expression ratio using the Pfaffl method) is evident upon heat shock in both hermaphrodites and males (*$p < 0.0001$). Tc1 transposase expression is increased in males after heat shock compared with hermaphrodites after heat shock (†$p = 0.0019$; calculated from log-transformed gene expression ratio values). In contrast, Tc3 expression does not change in either sex or condition. Both *act-1* and *pmp-3* were used as reference genes. qPCR experiments included three biological replicates and were performed in triplicate.

(B) Inverse PCR paired with amplicon sequencing was used to identify Tc1 and Tc3 locations genome-wide in male and hermaphrodite populations (300 worms/group). Three biological replicates were produced for each experimental group. Illumina sequencing reads representing Tc1-adjacent sequences were aligned to the *C. elegans* genome (WS245). Histograms displaying read counts across each chromosome were constructed; previously reported Tc1 transposon locations were identified, as well as *de novo* Tc1 insertion sites. Representative plots for chromosome *II* from adult males and hermaphrodite populations from a single replicate are shown with known locations marked by dashed vertical lines and *de novo* locations marked by solid lines.

(C) Global *de novo* Tc1 and Tc3 insertion site number was quantified for males and hermaphrodites, and the total number of *de novo* insertion sites in heat-shocked worms was compared with no heat shock worms. Males demonstrated a 65% global increase in Tc1 *de novo* insertion sites (no heat shock: 42 ± 8; heat shock: 69 ± 8), although hermaphrodites demonstrated a 16% decrease (no heat shock: 53 ± 15; heat shock: 44 ± 11 *de novo* Tc1 locations).

See also Tables S1–S3.

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

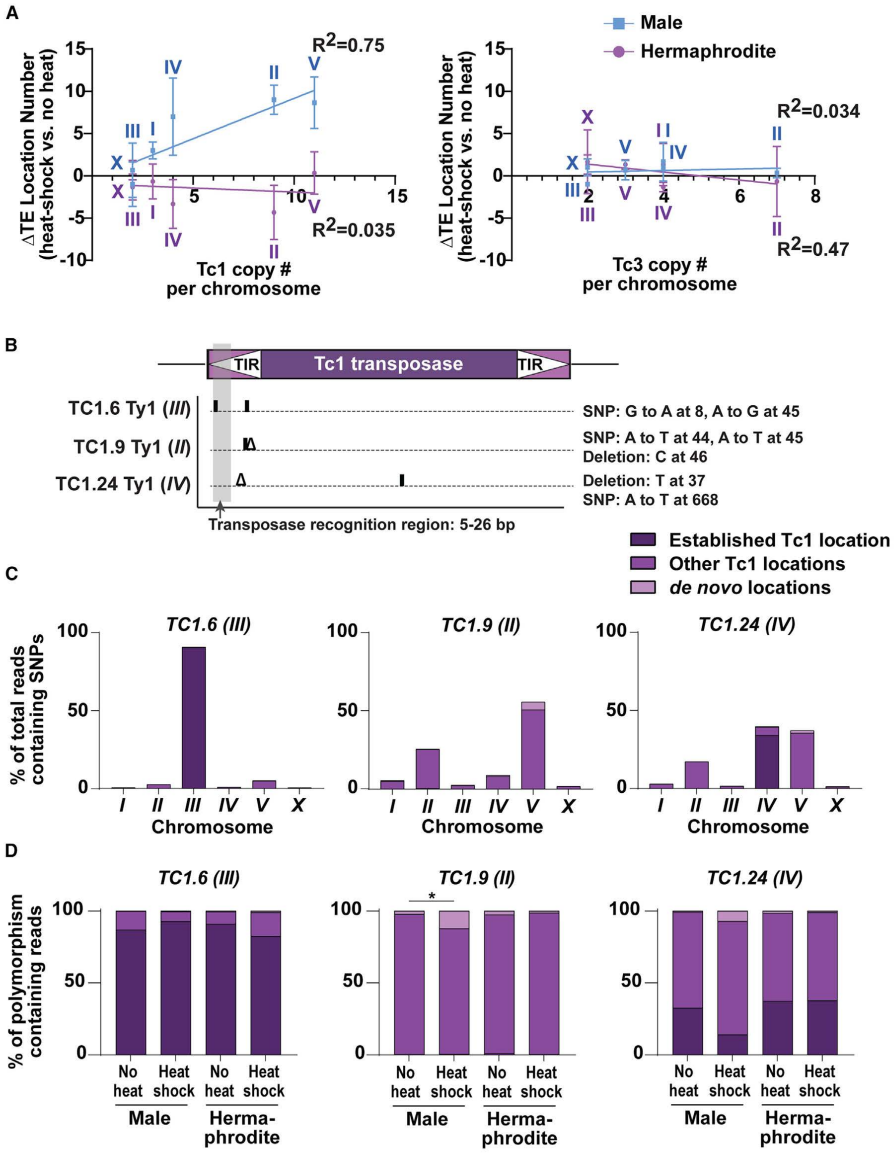


Figure 4. Tc1 Transposon Mobility Differentially Affected by Heat in Males and Hermaphrodites

(A) To evaluate transposon mobility across individual chromosomes, the global change in Tc1 and Tc3 transposon location number upon heat shock was plotted against Tc1 or Tc3 copy number per chromosome. Linear regression and correlation analysis were used to determine a coefficient of determination (R^2) value of 0.75 in males, representing a moderately strong correlation between Tc1 copy number per chromosome ($p = 0.026$) and indicating increased Tc1 occupancy at *de novo* insertion sites on chromosomes with more fixed copies of Tc1. Conversely, in hermaphrodites, Tc1 transposon *de novo* insertion number appears slightly reduced and does not correlate strongly with Tc1 copy number ($R^2 = 0.035$; $p = 0.72$). *De novo* Tc3 insertion sites in both males and hermaphrodites did not strongly correlate with the copy number of known Tc3 sites per chromosome ($R^2 = 0.034$, $p = 0.73$ and $R^2 = 0.47$, $p = 0.13$, respectively). Each point represents the

(legend continued on next page)

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

position was highly variable. To assess whether the G nucleotide in the sixth position might be dispensable for preferential Tc1 insertion, we considered a truncated version of the motif (TATRTN). The TATRTN motif, which occurs 233,427 times across the *C. elegans* genome, was present in 40% ± 4% of our Tc1-adjacent reads (Figure S4C). Tc1-adjacent reads containing the TATRTN motif were then aligned, and the relative nucleotide frequencies of the first six positions 3′ adjacent to Tc1 were quantified. Again, the first five positions were consistent with the previously reported insertion motif, but the sixth position was highly variable (Table S3). The usage of the TATRTN motif was increased in heat-shocked males related to other groups (Figure S4C, right graph). Overall, the previously reported TATRTG motif is preferred for Tc1 insertion, regardless of heat shock treatment, but the G in the sixth position may be dispensable for Tc1 transposase recognition and subsequent Tc1 insertion, especially in males following heat shock.

Conclusions

An increasing number of studies indicate that several meiotic prophase I processes differ between oogenesis and spermatogenesis [36]. Our data demonstrate that a single, transient exposure to heat stress produces intra- and inter-chromosomal transposon activity specifically in spermatocytes and not oocytes. Transposable elements constitute a large fraction of many eukaryotic and prokaryotic genomes, and their movement and accumulation are a major force shaping the structure and dynamics of the genomes they inhabit [37]. Movement of DNA transposons are known to drive the evolution of host genomes by facilitating the translocation of genomic sequences: from exon-shuffling to large-scale chromosomal rearrangements [37]. Additionally, the excision and subsequent repair of TE-associated DSBs provides the substrate for spontaneous mutation as well as the generation of allelic diversity [22, 37]. The germline is acutely at risk, as a DSB from a TE excision during meiotic prophase enables amplification of TE copy number due to recombination-based repair of DSBs [38].

Although oocytes are able to repress TE movement in response to heat shock, we show *C. elegans* spermatocytes lack the ability to effectively repress heat-induced TE movement within their genomes. The "out of testis" hypothesis suggests genes expressed in the male germline are more likely to generate new genes due to a combination of strong selective pressure and unique genomic features [39–41]. Given that TE movement can be an evolutionary driver [37], our data in *C. elegans* suggest that the inability of the male germline to repress heat-induced TE movement corroborates this hypothesis. Further, the difference in the relative numbers of spermatocytes versus oocytes produced throughout the lifetime of an organism may have contributed to the evolution of this differential response to heat stress. In most organisms, fewer oocytes are produced over the course of a lifetime compared with much larger number of spermatocytes produced by a male over the course of its lifetime. This difference may have enabled spermatocytes to tolerate a higher incidence of error while still maintaining enough sperm to promote fertility.

Our results in *C. elegans* may be an extreme manifestation of the inability of spermatocytes in most organisms to respond to TE mobility. A wild-type (Bristol) *C. elegans* population is naturally 99.9% hermaphrodite and therefore maintains are largely self population [42]. Our data show that heat-shocked spermatocytes of L4 larval hermaphrodites and adult males experience increased DNA damage and transposon mobility, which may increase genetic variation in sperm. Further, this increased DNA damage in L4 hermaphrodites may reduce self-fertility and thereby force outcrossing in response to stress.

Overall, our work suggests the combination of increased DNA damage and TE insertion into new, varied, and potentially harmful genomic positions in heat-stressed spermatocytes represents a dangerous mechanism for heritable transmission of an altered genome that generates male subfertility. Rising global temperatures and heatwaves are contributing to increased organism infertility (both human and insect) and risk of species extinction around the globe [43–45], often differentially impacting male or female fertility [46]. Understanding the mechanisms of heat-induced TE mobilization in spermatocytes may contribute to understanding how and why organism fertility and populations are shifting in response to climate change.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability

average (±standard deviation) change in TE location number between heat shock and no heat conditions from all three biological replicates. Data from each chromosome are labeled with the chromosome number.

(B) Schematic of Tc1 transposon structure with positions of polymorphisms (solid boxes) unique for three individual Tc1 copies are indicated.

(C) Assessment of individual Tc1 copies across individual chromosomes. Percent of polymorphism-containing reads across each chromosome; colors indicate the relative percent of reads representing Tc1 transposons present at their previously established location, at other known Tc1 locations, and at *de novo* locations. All SNP-containing reads from each exposure group are combined and parsed out by chromosome. Percentages represented by the stacked bars report the average of three replicates. *TC1.9* (*II*) is present across the genome (99.5% ± 0.3% of reads) but very rarely occurs at its established location (0.4% ± 0.3% of reads). In contrast, *TC1.6* (*III*) is usually present at its established location (91% ± 4% of reads). Finally, *TC1.24* (*IV*) is present the majority of the time at its established location (35% ± 8% of reads); however, it is also found elsewhere in the genome, at other known Tc1 locations and *de novo* insertion sites.

(D) Frequencies of SNP-containing reads identified in Figure 3C, now parsed out by sex and exposure condition. *TC1.9*-specific SNPs were found to increase at *de novo* insertion sites only in the male population exposed to heat shock (12% ± 4% of reads), compared with all other groups (male no heat: 2.2% ± 1%, hermaphrodite [herm] no heat: 2.7% ± 3%, and herm exposed to heat shock:1.3% ± 1%; $p < 0.005$ in all cases). Percentages represented by the stacked bars report the average of three replicates. Differences in percent occupancy at established, other Tc1, or *de novo* insertion sites were assessed via one-way ANOVA, with Sidak's test to account for multiple comparisons. Data for individual replicates are in Figure S4.

See also Figure S4 and Tables S2 and S3.

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.09.050.

ACKNOWLEDGMENTS

We thank A. Villeneuve for antibodies, both the CGC (funded by National Institutes of Health [NIH] P40 OD010440) and A. Villeneuve for strains, and A. Adamo for the pET30a plasmid. We thank B. Bowerman, C. Cahoon, and A. Villeneuve for comments on the manuscript. We also thank T. Schedl and P. Davis for involving us in officially establishing nomenclature for specific transposon positions in the *C. elegans* genome. Graphical abstract was made in part using BioRender. This work was supported by the National Institutes of Health F32GM130006 to N.A.K., National Institutes of Health T32HD007348 to Z.D.B., and National Institutes of Health R00HD076165 and Searle Scholar Award to D.E.L. D.E.L. is also a recipient of a March of Dimes Basil O'Connor Starter Scholar award and National Institutes of Health R35GM128890 award.

AUTHOR CONTRIBUTIONS

N.A.K., Z.D.B., and D.E.L. conducted experiments; D.D. performed bioinformatic analyses; and N.A.K. and D.E.L. designed the experiments and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 10, 2020
Revised: September 14, 2020
Accepted: September 15, 2020
Published: October 15, 2020

REFERENCES

1. Houston, B.J., Nixon, B., Martin, J.H., De Iuliis, G.N., Trigg, N.A., Bromfield, E.G., McEwan, K.E., and Aitken, R.J. (2018). Heat exposure induces oxidative stress and DNA damage in the male germ line. Biol. Reprod. *98*, 593–606.
2. Kim, B., Park, K., and Rhee, K. (2013). Heat stress response of male germ cells. Cell. Mol. Life Sci. *70*, 2623–2636.
3. Sage, T.L., Bagha, S., Lundsgaard-Nielsen, V., Branch, H.A., Sultmanis, S., and Sage, R.F. (2015). The effect of high temperature stress on male and female reproduction in plants. Field Crops Res. *182*, 30–42.
4. Rohmer, C., David, J.R., Moreteau, B., and Joly, D. (2004). Heat induced male sterility in Drosophila melanogaster: adaptive genetic variations among geographic populations and role of the Y chromosome. J. Exp. Biol. *207*, 2735–2743.
5. Durairajanayagam, D., Agarwal, A., and Ong, C. (2015). Causes, effects and molecular mechanisms of testicular heat stress. Reprod. Biomed. Online *30*, 14–27.
6. Hickman, A.B., and Dyda, F. (2016). DNA transposition at work. Chem. Rev. *116*, 12758–12784.
7. Wallis, D.C., Nguyen, D.A.H., Uebel, C.J., and Phillips, C.M. (2019). Visualization and quantification of transposon activity in *Caenorhabditis elegans* RNAi pathway mutants. G3 (Bethesda) *9*, 3825–3832.
8. Pérez-Crespo, M., Pintado, B., and Gutiérrez-Adán, A. (2008). Scrotal heat stress effects on sperm viability, sperm DNA integrity, and the offspring sex ratio in mice. Mol. Reprod. Dev. *75*, 40–47.
9. Chowdhury, A.K., and Steinberger, E. (1970). Early changes in the germinal epithelium of rat testes following exposure to heat. J. Reprod. Fertil. *22*, 205–212.
10. Paul, C., Melton, D.W., and Saunders, P.T. (2008). Do heat stress and deficits in DNA repair pathways have a negative impact on male fertility? Mol. Hum. Reprod. *14*, 1–8.
11. Alpi, A., Pasierbek, P., Gartner, A., and Loidl, J. (2003). Genetic and cytological characterization of the recombination protein RAD-51 in Caenorhabditis elegans. Chromosoma *112*, 6–16.
12. Jaramillo-Lambert, A., Ellefson, M., Villeneuve, A.M., and Engebrecht, J. (2007). Differential timing of S phases, X chromosome replication, and meiotic prophase in the C. elegans germ line. Dev. Biol. *308*, 206–221.
13. Keeney, S. (2008). Spo11 and the formation of DNA double-strand breaks in meiosis. Genome Dyn. Stab. *2*, 81–123.
14. Dernburg, A.F., McDonald, K., Moulder, G., Barstead, R., Dresser, M., and Villeneuve, A.M. (1998). Meiotic recombination in C. elegans initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. Cell *94*, 387–398.
15. Yokoo, R., Zawadzki, K.A., Nabeshima, K., Drake, M., Arur, S., and Villeneuve, A.M. (2012). COSA-1 reveals robust homeostasis and separable licensing and reinforcement steps governing meiotic crossovers. Cell *149*, 75–87.
16. Eide, D., and Anderson, P. (1988). Insertion and excision of Caenorhabditis elegans transposable element Tc1. Mol. Cell. Biol. *8*, 737–746.
17. Vos, J.C., van Luenen, H.G., and Plasterk, R.H. (1993). Characterization of the Caenorhabditis elegans Tc1 transposase in vivo and in vitro. Genes Dev. *7* (7A), 1244–1253.
18. Plasterk, R.H. (1996). The Tc1/mariner transposon family. Curr. Top. Microbiol. Immunol. *204*, 125–143.
19. Ketting, R.F., Fischer, S.E., and Plasterk, R.H. (1997). Target choice determinants of the Tc1 transposon of Caenorhabditis elegans. Nucleic Acids Res. *25*, 4041–4047.
20. Bessereau, J. (2006). Transposons in C. elegans. WormBook 1–13.
21. Riddle, D.L., Blumenthal, T., Meyer, B.J., and Priess, J.R. (1997). Mechanism of Tc1 transposition. C. elegans II, Second Edition (Cold Spring Harbor Laboratory).
22. Fischer, S.E.J., Wienholds, E., and Plasterk, R.H.A. (2003). Continuous exchange of sequence information between dispersed Tc1 transposons in the Caenorhabditis elegans genome. Genetics *164*, 127–134.
23. Laricchia, K.M., Zdraljevic, S., Cook, D.E., and Andersen, E.C. (2017). Natural variation in the distribution and abundance of transposable elements across the Caenorhabditis elegans species. Mol. Biol. Evol. *34*, 2187–2202.
24. Moerman, D.G., and Waterston, R.H. (1984). Spontaneous unstable unc-22 IV mutations in C. elegans var. Bergerac. Genetics *108*, 859–877.
25. Moerman, D.G., Kiff, J.E., and Waterston, R.H. (1991). Germline excision of the transposable element Tc1 in C. elegans. Nucleic Acids Res. *19*, 5669–5672.
26. Wang, G., and Reinke, V. (2008). A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. Curr. Biol. *18*, 861–867.
27. Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R.,

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol. Cell *31*, 79–90.

28. Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol. Cell *31*, 67–78.
29. Collins, J., Saari, B., and Anderson, P. (1987). Activation of a transposable element in the germ line but not the soma of Caenorhabditis elegans. Nature *328*, 726–728.
30. Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133–141.
31. Zhang, C., Montgomery, T.A., Gabel, H.W., Fischer, S.E., Phillips, C.M., Fahlgren, N., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2011). mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA *108*, 1201–1208.
32. Belicard, T., Jareosettasin, P., and Sarkies, P. (2018). The piRNA pathway responds to environmental signals to establish intergenerational adaptation to stress. BMC Biol. *16*, 103.
33. Uebel, C.J., Anderson, D.C., Mandarino, L.M., Manage, K.I., Aynaszyan, S., and Phillips, C.M. (2018). Distinct regions of the intrinsically disordered protein MUT-16 mediate assembly of a small RNA amplification complex and promote phase separation of Mutator foci. PLoS Genet. *14*, e1007542.
34. van Luenen, H.G., and Plasterk, R.H. (1994). Target site choice of the related transposable elements Tc1 and Tc3 of Caenorhabditis elegans. Nucleic Acids Res. *22*, 262–269.
35. Korswagen, H.C., Durbin, R.M., Smits, M.T., and Plasterk, R.H. (1996). Transposon Tc1-derived, sequence-tagged sites in Caenorhabditis elegans as markers for gene mapping. Proc. Natl. Acad. Sci. USA *93*, 14680–14685.
36. Cahoon, C.K., and Libuda, D.E. (2019). Leagues of their own: sexually dimorphic features of meiotic prophase I. Chromosoma *128*, 199–214.
37. Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. *41*, 331–368.
38. Engels, W.R., Johnson-Schlitz, D.M., Eggleston, W.B., and Sved, J. (1990). High-frequency P element loss in Drosophila is homolog dependent. Cell *62*, 515–525.
39. Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res. *20*, 1313–1326.
40. Kleene, K.C. (2001). A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. Mech. Dev. *106*, 3–23.
41. Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ravi Ram, K., Sirot, L.K., Levesque, L., Artieri, C.G., Wolfner, M.F., Civetta, A., and Singh, R.S. (2007). Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. Genetics *177*, 1321–1335.
42. Corsi, A.K., Wightman, B., and Chalfie, M. (2015). A transparent window into biology: a primer on Caenorhabditis elegans. Genetics *200*, 387–407.
43. Walsh, B.S., Parratt, S.R., Hoffmann, A.A., Atkinson, D., Snook, R.R., Bretman, A., and Price, T.A.R. (2019). The impact of climate change on fertility. Trends Ecol. Evol. *34*, 249–259.
44. Sales, K., Vasudeva, R., Dickinson, M.E., Godwin, J.L., Lumley, A.J., Michalczyk, Ł., Hebberecht, L., Thomas, P., Franco, A., and Gage, M.J.G. (2018). Experimental heatwaves compromise sperm function and cause transgenerational damage in a model insect. Nat. Commun. *9*, 4771.
45. Urban, M.C. (2015). Climate change. Accelerating extinction risk from climate change. Science *348*, 571–573.
46. Iossa, G. (2019). Sex-specific differences in thermal fertility limits. Trends Ecol. Evol. *34*, 490–492.
47. Libuda, D.E., Uzawa, S., Meyer, B.J., and Villeneuve, A.M. (2013). Meiotic chromosome structures constrain and respond to designation of crossover sites. Nature *502*, 703–706.
48. Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics *21*, 1859–1875.
49. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
50. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).
51. Colaiácovo, M.P., MacQueen, A.J., Martinez-Perez, E., McDonald, K., Adamo, A., La Volpe, A., and Villeneuve, A.M. (2003). Synaptonemal complex assembly in C. elegans is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. Dev. Cell *5*, 463–474.
52. Rinaldo, C., Ederle, S., Rocco, V., and La Volpe, A. (1998). The Caenorhabditis elegans RAD51 homolog is transcribed into two alternative mRNAs potentially encoding proteins of different sizes. Mol. Gen. Genet. *260*, 289–294.
53. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45.
54. R Development Core Team (2014). R: A language and environment for statistical computing. (R Foundation for Statistical Computing).

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-RAD-51	This study	N/A
Chicken polyclonal anti-RAD-51	This study	N/A
Rabbit anti-GFP	[15]	N/A
Bacterial and Virus Strains		
OP50 *Escherichia coli*	CGC	OP50
Deposited Data		
The sequencing data used in this report	NCBI BioProject Database: https://www.ncbi.nlm.nih.gov/bioproject/	Ascension #: PRJNA602000
Experimental Models: Organisms/Strains		
C. elegans N2 wild-isolate (Bristol)	CGC	N2
C. elegans DLW5 *spo-11(ok79) IV/nT1 [qIs51] (IV;V)*	This study	DLW5
C. elegans CB4108 *fog-2(q71) V*	CGC	CB4108
C. elegans AV761 *meIs8 [unc-119(+) Ppie-1::gfp::cosa-1] II; spo-11(me44) IV / nT1[qIs51] (IV;V)*	[47]	AV761
C. elegans DLW59 *meIs8 [unc-119(+) Ppie-1::gfp::cosa-1] II; spo-11(me44) IV / tmc9 [tm1s1221] IV*	This study	DLW59
Oligonucleotides		
See Table S2 for oligonucleotide information	N/A	N/A
Software and Algorithms		
softWoRx image acquisition software	Applied Precision	Provided by Deltavision
Imaris Image Analysis software	Bitplane	https://imaris.oxinst.com
Prism 8	GraphPad Software	http://www.graphpad.com
Gmap	[48]	https://bioconductor.org/packages/devel/bioc/vignettes/gmapR/inst/doc/gmapR.pdf
Trimmomatic	[49]	http://www.usadellab.org/cms/?page=trimmomatic
R	The R-project	https://www.r-project.org/
ggplot2	[50]	https://ggplot2.tidyverse.org/
Other		
RNAqueous-4PCR Total RNA Isolation kit	Ambion	AM1914

(*Continued on next page*)

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Superscript IV reverse transcriptase	Invitrogen	18090010
Random hexamers	Promego	C118A
PowerUp SYBR Green Master Mix	Applied Biosystems	A25776
DNeasy Blood and Tissue Kit	QIAGEN	69504
Mag-bind TotalPure NGS paramagnetic beads	Omega Bio-Tek	M1378-00
KAPA HyperPlus Library Preparation Kit	KAPA Biosystems	KK8503

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Diana E. Libuda (dlibuda@uoregon.edu)

Materials Availability

Antibodies generated in the course of this work are freely available to interested academic researchers through the Lead Contact

Data and Code Availability

The sequencing data used in this report have been deposited in NCBI's Sequencing Read Archive database (SRA). Original data have been deposited with links to BioProject accession number PRJNA602000 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

C. elegans strains, genetics, and culture conditions

All strains are from the Bristol N2 background, and were maintained at 20°C under standard conditions. All strains were maintained as mating stocks. Temperatures used for specific experiments are indicated in Figures and below.

The following strains were used in this study:
N2: Bristol wild-type strain
DLW5: *spo-11(ok79) IV/nT1* [*qIs51*] (*IV;V*)
CB4108: *fog-2(q71) V*
AV761: *meIs8[unc-119(+) Ppie-1::gfp::cosa-1] II*; *spo-11(me44) IV / nT1[qIs51] (IV;V)*
DLW59: *meIs8[unc-119(+) Ppie-1::gfp::cosa-1] II*; *spo-11(me44) IV / tmc9[tm1s1221] IV*

METHOD DETAILS

Heat shock

Heat-shock experiments were performed in a VWR refrigerated Peltier incubator (Model VR16P) at 34°C for 2 hours. Additional exposures were also done at 26°C, 28°C, 30°C, 32°C, 36°C and 38°C. Following heat-shock, an hour of recovery was given before sample collection.

Immunofluorescence

Immunofluorescence was performed as in Libuda et al., 2013 [47]. Briefly, gonads from adult male and hermaphrodite worms at 18–22 hr post-L4 stage were dissected together in 1x egg buffer with 0.1% Tween on the same VWR Superfrost Plus slides, fixed for 5 min in 1% paraformaldehyde, flash frozen with liquid nitrogen, and then fixed for 1 min in 100% methanol at −20°. Slides were washed 3 × 5 min in 1x PBS Tween (PBST) and blocked for 1 hr in 0.7% BSA in 1x PBST. Primary antibody dilutions were made in 1x PBST and added to slides. Slides were covered with a parafilm coverslip and incubated in a humid chamber overnight (14–18 hr). Slides were washed 3 × 10 min in 1x PBST. Secondary antibody dilutions were made at 1:200 in 1x PBST using Invitrogen goat or donkey Alexa Fluor-labeled antibodies and added to slides. Slides were covered with a parafilm coverslip and placed in a

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

humid chamber in the dark for 2 hr. Slides were washed 3 × 10 min in 1x PBST in the dark. All washes and incubations were performed at room temperature, unless otherwise noted. Next, 2 μg/ml DAPI was added to slides and slides were subsequently incubated in the dark with a parafilm coverslip in a humid chamber. Slides were washed once for 5 min in 1x PBST prior to mounting with Vectashield and a 20 × 40 mm coverslip. Slides were sealed with nail polish immediately following mounting and then stored at 4°C prior to imaging. All slides were imaged (as described below) within 2 weeks of preparation. The following primary antibody dilutions were used: rabbit anti-RAD-51 (1:5000), chicken anti-RAD-51 (1:1000) (RAD-51 antibodies were custom made, see "Antibody production" below), rabbit anti-GFP (1:1500) [15]. Immunofluorescence images from Figures 1, S1, and S2 utilized the rabbit anti-RAD-51 antibody. Immunofluorescence images from Figures 2 and S3 utilized the chicken-RAD-51 antibody.

Antibody production

Our existing RAD-51 antibodies were generated from a His-tagged fusion protein expressed by Genscript from plasmid pET30a containing the entire RAD-51S coding sequence (1385 bp, GenBank accession number AF061201) [51, 52]. Antibodies were produced in rabbit and chicken and affinity purified by Pocono Rabbit Farms.

Imaging and quantification of RAD-51 and COSA-1 foci

Immunofluorescence slides were imaged at 512 × 512 pixel dimensions on an Applied Precision DeltaVision microscope with a 63x lens and a 1.5x optivar. Images were acquired as Z stacks at 0.2 μm intervals and deconvolved with Applied Precision softWoRx deconvolution software. For quantification of RAD-51 and GFP::COSA-1 foci, early pachytene included the first four-to-five rows of nuclei immediately following the transition zone, while late pachytene included nuclei in the last four-to-five rows of nuclei preceding the condensation zone in males or the last four-to-five complete rows of nuclei in hermaphrodites. Foci per nucleus were quantified automatically from deconvolved three-dimensional (3D) stacks using Imaris software (Bitplane). Foci per nucleus counts were analyzed using one-way analysis of variance (ANOVA) to compare all groups in Prism 8 (Graphpad)

Male fertility assay

Wild-type (N2) males and *fog-2*Δ (CB4108) obligate females were isolated at the L4 developmental stage and aged 18-22 hours to adult stage. Individual obligate females were then singled and left at 20°C, while wild-type males were exposed to heat-shock (34°C, 2 hours). Immediately following heat-shock, individual males were paired with individual females and left at 20°C for 3 consecutive 8-hour intervals; during each successive interval single heat-shocked males were paired with virgin females. Living progeny, dead eggs, and unfertilized eggs were quantified for each female. Pairs with brood sizes of zero were excluded. Differences in the ability of heat-shocked males to produce viable progeny were analyzed using Fishers exact tests in Prism 8 (Graphpad). Due to poor mating performance of recently heat-shocked males, brood size from heat-shocked males was found to be reduced in the 0-8 hour post-exposure window (brood size per male with heat-shock = 39 ± 30 worms, versus without heat-shock = 87 ± 33). Experimental group sizes go as follows: 1) 0-8 hours post-heat-shock: no heat-shock, n = 8; heat-shock, n = 10; 2) 8-16 hours post-heat-shock: no heat-shock, n = 23; heat-shock, n = 38; and, 3) 16-24 hours post-heat-shock: no heat-shock, n = 6, heat-shock = 11. During the 0-8 hr post-exposure time point, heat-shocked males had difficulty mating immediately following exposure which significantly decreased brood size.

Quantitative PCR for Tc1 and Tc3 transposase

Sample preparation

200 wild-type (N2) L4 stage worms were collected for each experimental group (male without heat-shock, male with heat-shock, hermaphrodite without heat-shock, hermaphrodite with heat-shock). The L4 worms were aged 18-22 at 20°C hours to adult stage. Once at the adult stage, heat-shock groups were exposed to 34°C for 2 hours. Immediately following exposure, worms were washed off plates with M9 buffer, pelleted, flash frozen in liquid nitrogen and stored at −80°C.

RNA extraction, cDNA synthesis and quantitative PCR

RNA was isolated using the RNAqueous-4PCR Total RNA Isolation kit (Ambion). To quantify Tc1 and Tc3 transposase levels equal amounts of cDNA were synthesized using Superscript IV reverse transcriptase and random hexamers following the manufacturer-included protocol. Quantitative PCR was carried out using PowerUp SYBR Green Master Mix (Applied Biosystems) and 500nM of both forward and reverse primers (Table S3). *act-1* and *pmp-3* were amplified as internal controls. Gene expression ratios were calculated using the the Pfaffl method [53], and by ΔCt and ΔΔCt values to compare groups by one-way ANOVA analysis.

Evaluation of Tc1 and Tc3 transposon genomic locations

Sample preparation

300 wild-type (N2) L4 stage worms were collected for each experimental group (males without heat-shock, males with heat-shock, hermaphrodites without heat-shock, hermaphrodites with heat-shock). Three biological replicates were prepared for each experimental group. The L4 worms were aged 18-22 hours at 20°C to adult stage. Once at the adult stage, heat-shock groups were exposed to 34°C for 2 hours. Following a one-hour recovery period, genomic DNA was prepared using the QIAGEN DNeasy Blood and Tissue Kit using a modified protocol. Worms were transferred off plates and into 15 mL centrifuge tubes using M9 buffer, they were then pelleted by centrifugation at 2500 rpm at room temperature for 30 s. The pellets were washed twice with M9 and re-pelleted before resuspending pellet in 200 uL ATL buffer (from kit) and transferring solution to a 1.5 mL microcentrifuge tube. Samples were

Please cite this article in press as: Kurhanewicz et al., Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.09.050

incubated at −80°C overnight, then underwent 3 cycles of freeze (−20°C)/thaw (37°C). Proteinase K and RNase A was added according to the manufacturer-provided protocol, and the provided protocol was followed exactly thereafter.

Inverse PCR and amplicon sequencing

Genomic DNA was fragmented to ~800 bp using a Covaris M22 sonicator and checked for size by gel electrophoresis. Fragments were self-ligated using T4 ligase (NEB) and recovered using glycogen/isopropanol precipitation. Nested PCR was performed to amplify DNA sequence directly adjacent to Tc1 and Tc3 transposon ends. Two sets of primers were designed to capture sequence adjacent to both the 5′ and 3′ ends of the transposon. Primer sequences are provided in Table S3. Size selection was used to obtain products under 1 kb using Mag-bind TotalPure NGS paramagnetic beads (Omega Bio-Tek). End repair and adaptor ligation was carried out using reagents and protocols from the KAPA HyperPlus Library Preparation Kit (KAPA Biosystems) and Illumina TruSeq adapters. Samples were multiplexed and sequenced on an Illumina MiSeq v3 (25M reads) paired end 300.

Read alignment and quality filtering

Paired end reads were quality filtered for adaptor contamination and low-quality ends using trimmomatic [49]. After quality filtering an average of 1.5 million reads per paired end sample remained. Surviving reads were filtered to contain only sequence adjacent to Tc1 or Tc3 insertion sites by searching for and removing the inverted repeats and associated Tc1 and Tc3 sequences. Reads were then mapped to the *C. elegans* reference genome using Gmap [48] and uniquely aligned reads were imported into R [54], binned by mapping position, and plotted using ggplot2 [50].

Identification of specific Tc1 transposon polymorphisms and insertion motif analysis

Previously reported Tc1 polymorphisms were identified in our raw reads and sequences adjacent to insertion sites were mapped to the *C. elegans* reference genome using gmap [48], imported into R [54], and plotted using ggplot2 [50]. To investigate previously identified known Tc1 hotspot motifs around insertions sites, uniquely aligned reads were searched for the known motif sequence and reads containing the motif were trimmed to the 6 base pairs directly adjacent to insertion sites and plotted using ggplot2 [50].

Transposon nomenclature

In this report, we utilized newly established nomenclature to define the specific copies of Tc1 that were described in Fischer *et al.* [22]. A full description of these specific Tc1 copies goes as follows: *TC1.9* represents WBTransposon00000021 from Y46G5A on chromosome *II; TC1.6* represents WBTransposon00000018 from Y48G9A on chromosome *III*; and *TC1.24* represents WBTransposon00000036 from T69A2AR) on chromosome *IV.* This nomenclature was decided upon following a discussion with Paul Davis (WormBase, EMBL-EBI, Welcome Trust) and Tim Schedl (Washington University St. Louis).

QUANTIFICATION AND STATISTICAL ANALYSIS

All p values reported in Figures 1, 2, and S3 regarding the quantification of RAD-51 foci are calculated from Kruskal-Wallis one-way analysis of variance tests, which are robust non-parametric statistical tests appropriate for the relevant datasets. In all cases, Dunns test was used to account for multiple comparisons. For the time-course analysis in Figures 1D and 1E, we used a two-way mixed-effects ANOVA with a Sidak post hoc test to account for multiple comparisons. To assess dead egg production as shown in Figure 1F, we used Fisher's exact test (an appropriate test for 2 × 2 contingency tables) to evaluate whether dead eggs occurred at a higher frequency with exposure to heat stress independent of brood size. For Figure 3A, differences in gene expression were assessed using a one-way ANOVA on log-transformed gene expression ratios (calculated using the Pfaffl method, which calculates relative gene expression while accounting for differences in primer efficiencies), with a Sidak test used to account for multiple comparisons. Gene expression was also assessed by one-way ANOVA on ddCt values. The results agreed in both cases. The change in Tc1 and Tc3 de-novo site numbers, as shown in Figure 3C, was assessed via one-way ANOVA, with a Sidak test used to account for multiple comparisons. For Figure 4A the goodness-of-fit for the linear regression lines were reported as coefficient of determination (R^2; an appropriate goodness-of-fit statistic for fitted linear regression lines). In addition, a correlation analysis was run to test the null hypothesis that there was no correlation between the two variables, and the p value was reported. The changes in percent of Tc1 copies carrying specific polymorphisms residing at specific locations (established, other Tc1, or *de novo*) as shown in Figure 4D, as well as changes in the percentage of Tc1-adjacent reads containing TATRTG and TATRTN insertion motifs in Figure S4C, was assessed via one-way ANOVA, with a Sidak test used to account for multiple comparisons. All statistics were done using the Prism graphing program (GraphPad Software). For all tests described above, the assumptions of each test were met.

Current Biology, Volume *30*

Supplemental Information

Elevated Temperatures Cause Transposon-Associated DNA Damage in *C. elegans* Spermatocytes

Nicole A. Kurhanewicz, Devin Dinwiddie, Zachary D. Bush, and Diana E. Libuda

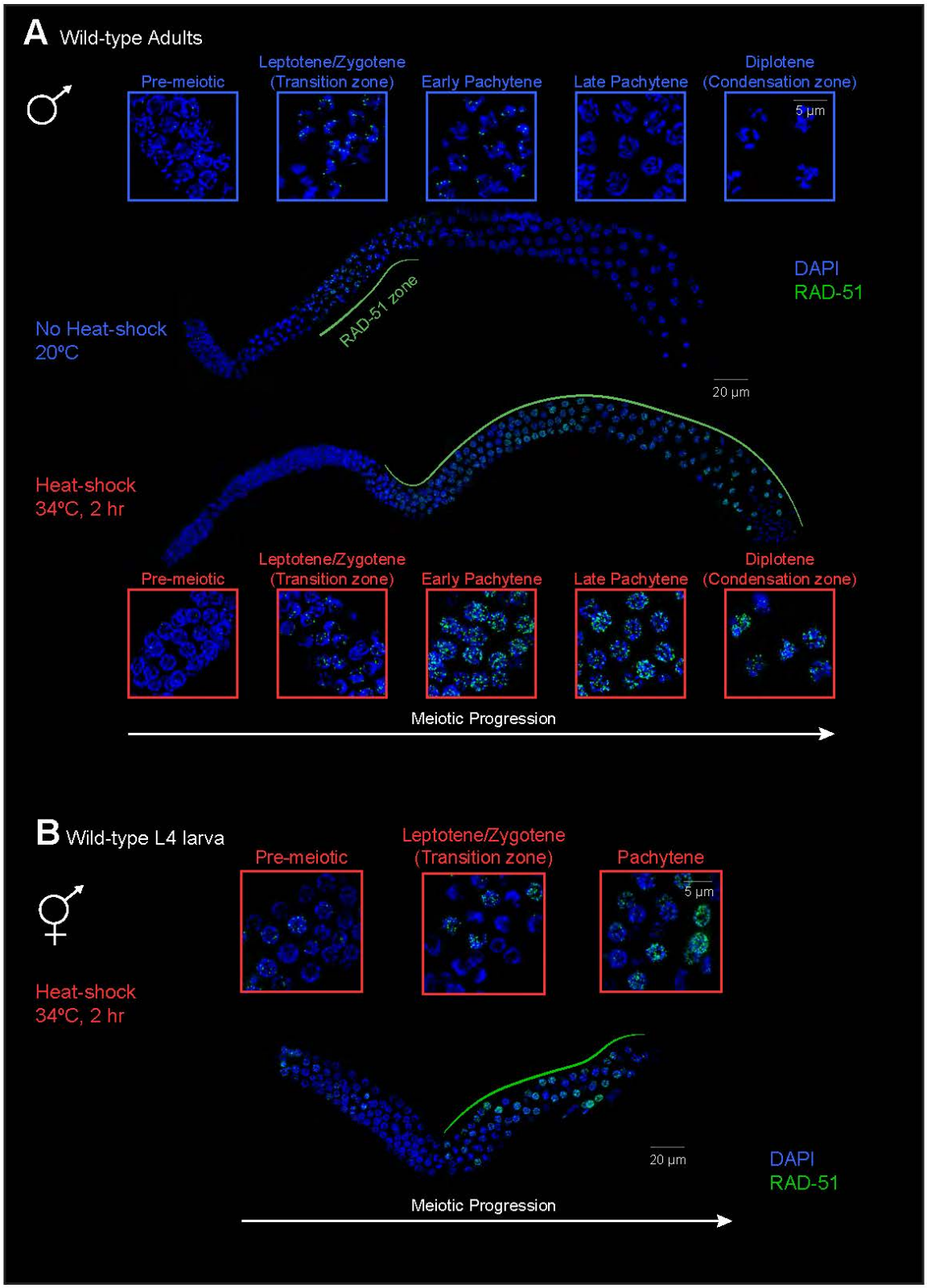

A Wild-type Adults
Pre-meiotic
Leptotene/Zygotene (Transition zone)
Early Pachytene
Late Pachytene
Diplotene (Condensation zone)
5 µm
DAPI
RAD-51
No Heat-shock 20°C
RAD-51 zone
20 µm
Heat-shock 34°C, 2 hr
Pre-meiotic
Leptotene/Zygotene (Transition zone)
Early Pachytene
Late Pachytene
Diplotene (Condensation zone)
Meiotic Progression
B Wild-type L4 larva
Pre-meiotic
Leptotene/Zygotene (Transition zone)
Pachytene
5 µm
Heat-shock 34°C, 2 hr
20 µm
DAPI
RAD-51
Meiotic Progression

Figure S1. Heat-exposure produces elevated RAD-51 foci throughout meiotic prophase I in the male germline. Related to Figure 1. Immunofluorescence for the recombinase RAD-51 was used to assess DNA damage throughout the germlines of wild-type *C. elegans* with and without heat-shock (34°C for 2 hours). (A) Extruded and fixed adult male gonads, which produce sperm, are shown without exposure (upper image) and with exposure to heat-shock (lower image). (B) An extruded and fixed L4 hermaphrodite gonad, exclusively producing sperm, is shown. Panels exhibit magnified images from each stage of meiosis. Scale bars represent 20 µm in whole gonad images and 5 µm in magnified panels. DAPI stain for DNA is shown in blue, RAD-51 foci marking sites of DNA damage are shown in green. The "RAD-51 zone" where RAD-51 foci are consistently present and associated with the nuclei of developing spermatocytes is indicated by a green line along the length of the gonad.

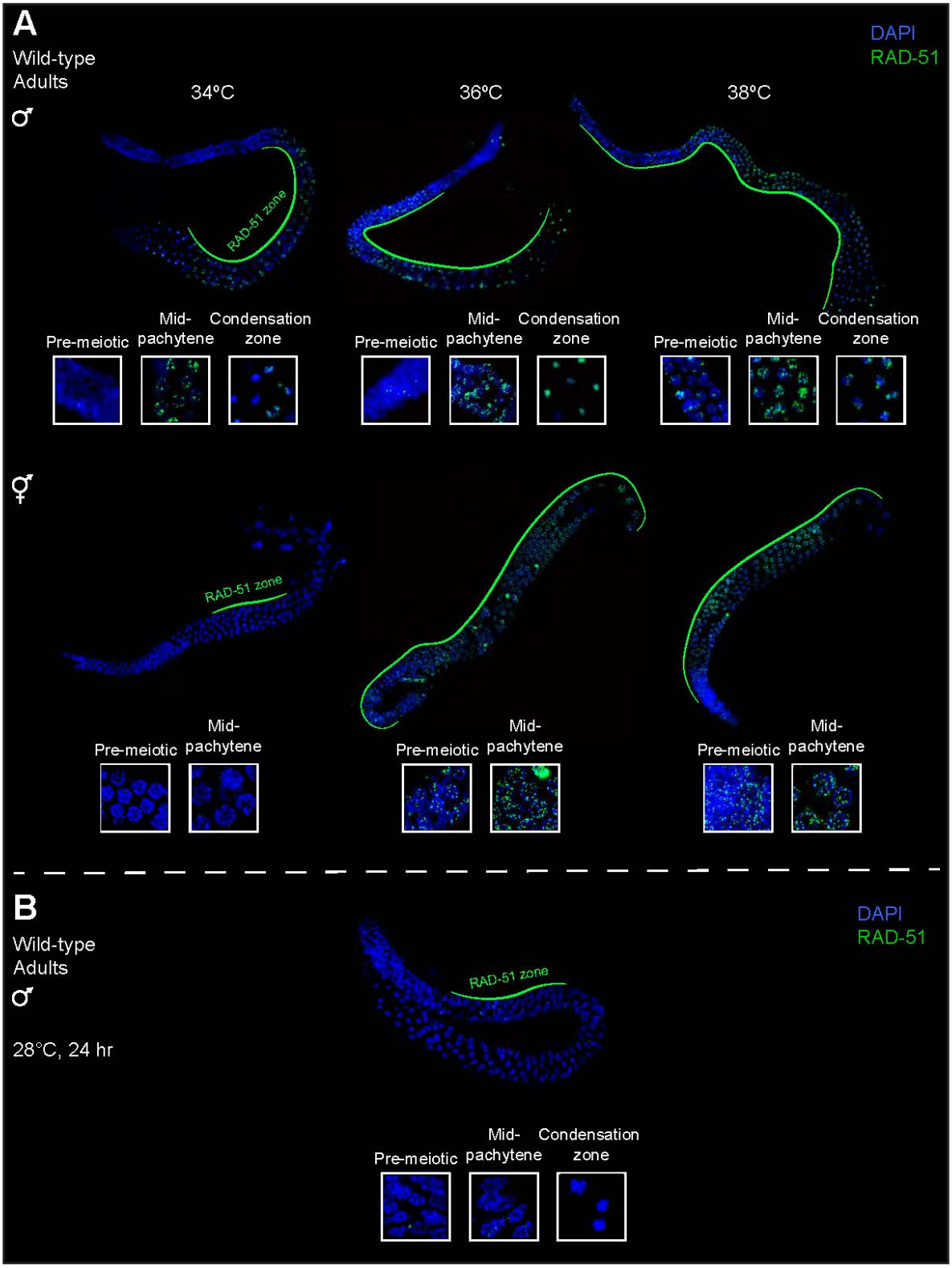

A
Wild-type
Adults
♂
DAPI
RAD-51
34°C
36°C
38°C
RAD-51 zone
Pre-meiotic
Mid-pachytene
Condensation zone
♀
B
28°C, 24 hr

Figure S2. Heat-induced DNA damage is exacerbated at temperatures above 34°C, but not by an extended exposure at 28°C. Related to Figures 1 and 2. (A) Immunofluorescence for the recombinase RAD-51 was used to assess DNA damage throughout the germlines of wild-type male and hermaphrodite *C. elegans* with exposure to heat-shock (34°C, 36°C, or 38°C for 2 hours). Extruded and fixed adult male gonads which produce sperm are shown on top, and hermaphrodite gonads which produce eggs are shown below. Magnified panels highlight the premeiotic and pachytene zones in both the male and hermaphrodite germline, as well as an additional panel highlighting the condensation zone in the male germline. Notably, the dissected hermaphrodite gonad ends at prophase I while the male germline includes stages encompassing meiotic divisions. Thus, this extra panel is only able to be shown in the male. DAPI stain for DNA is shown in blue, RAD-51 foci marking sites of DNA damage are shown in green. The "RAD-51 zone" where RAD-51 foci are consistently present and associated with the nuclei of developing spermatocytes is indicated by a green line along the length of the gonad. Temperatures above 34°C produce both an increase in heat-induced DNA damage, as well as an elongation of the RAD-51 zone in the male germline. Heat shock at 34°C does not illicit a DNA damage response in the hermaphrodite germline. However, exposure to 36°C and above produces DNA damage throughout the entire gonad. (B) The production of DNA damage was assessed throughout the germlines of wild-type male *C. elegans* with exposure to extended low-temperature heat stress (28°C for 24 hours). This temperature is permissive for the growth of wild-type *C. elegans* from egg to adult but results in sterility. An extruded and fixed adult male gonad is shown. Magnified panels highlight the premeiotic, pachytene zones and condensation zones. DAPI stain for DNA is shown in blue, RAD-51 foci marking sites of DNA damage are shown in green. Extended exposure at 28°C did not elevate the number of RAD-51 marked sites of DNA damage, nor was the RAD-51 zone extended.

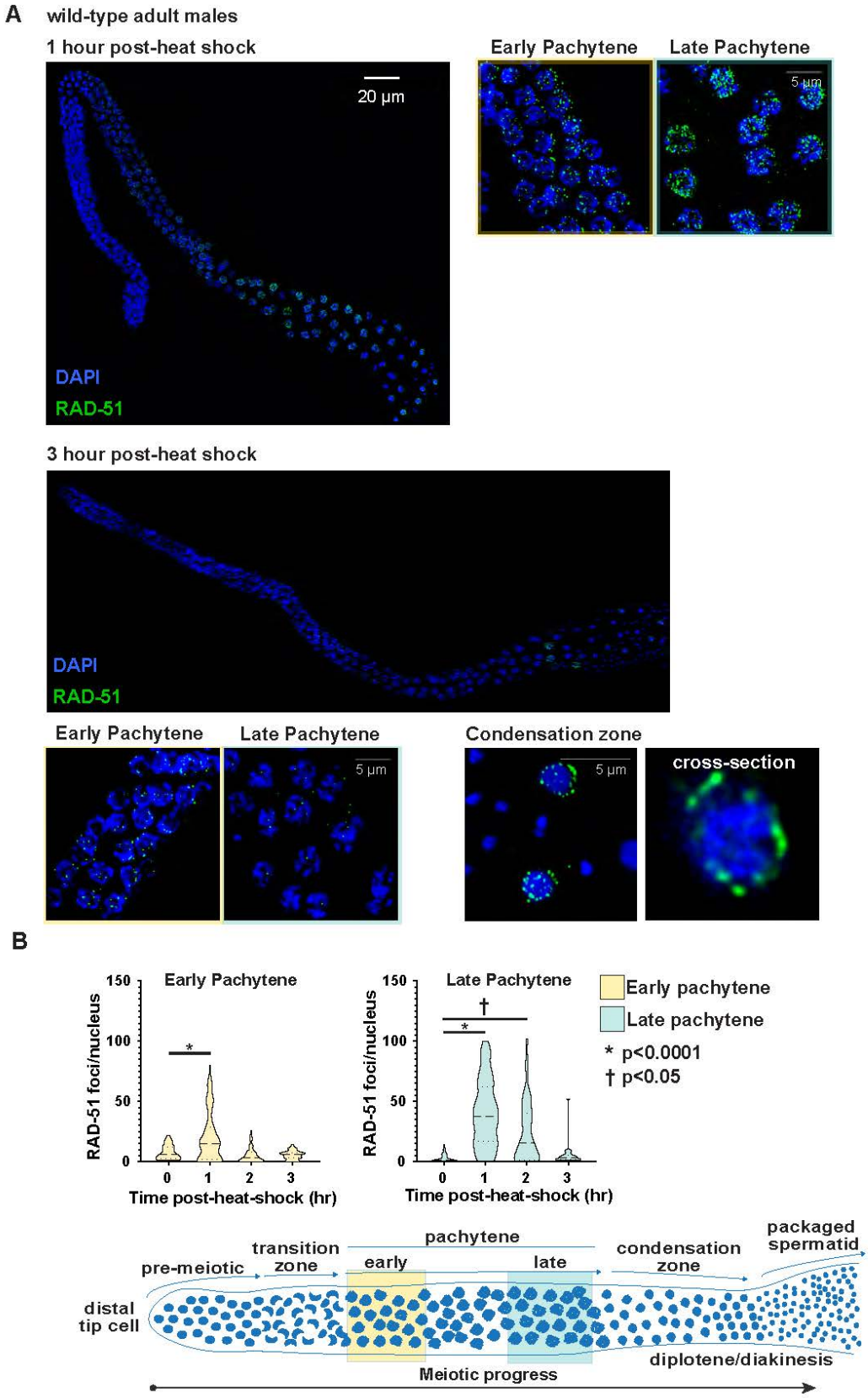

A
wild-type adult males
1 hour post-heat shock
20 µm
DAPI
RAD-51
Early Pachytene
Late Pachytene
5 µm
3 hour post-heat shock
DAPI
RAD-51
Early Pachytene
Late Pachytene
5 µm
Condensation zone
5 µm
cross-section
B
Early Pachytene
Late Pachytene
RAD-51 foci/nucleus
150
100
50
0
0 1 2 3
Time post-heat-shock (hr)
*
†
Early pachytene
Late pachytene
* p<0.0001
† p<0.05
pre-meiotic
transition zone
pachytene
early
late
condensation zone
packaged spermatid
distal tip cell
diplotene/diakinesis
Meiotic progress

Figure S3. Heat-induced DSB repair dynamics. Related to Figures 1, S1, and S2. To characterize the dynamics of resolution or repair of heat-induced DSBs, RAD-51 foci were visualized and quantified throughout the germlines of wild-type male *C. elegans* with exposure to heat-shock (34°C, 2 hours) at 1, 2, and 3-hours post-exposure. (A) Extruded and fixed adult male gonads, which produce sperm, are shown at 1-hour post-heat-shock (upper image) and 3-hour post-heat-shock (lower image). Magnified panels highlight the degree of heat-induced DNA damage in early and late pachytene. Scale bars represent 20 µm in whole gonad images and 5 µm in magnified panels. Notably, in the condensation region of heat-shocked male germ lines at the 3-hour time point, a subset of nuclei had massive amounts of unrepaired DSBs that were largely localized to the periphery of the nucleus heat-shock. DAPI stain for DNA is shown in blue, RAD-51 foci marking sites of DNA damage are shown in green. (B) Quantification of RAD-51 foci per nucleus in early and late pachytene for wild-type spermatocytes. Violin plots show frequency distribution of the data: the center dashed line represents the median, the bottom and top dotted lines represent the first and third quartiles respectively. The degree of damage at 1-hour post-heat-shock was more severe in late-pachytene (avg 40±3 foci/nucleus, n=76), than in early pachytene (avg 20±2 foci/nucleus, n=110). Similarly, RAD-51 marked DSBs levels returned to pre-heat-shock levels by 3 hours post-heat-shock in the late pachytene region, while a return to pre-heat-shock levels of RAD-51 marked DSBs only required 2 hours in the early pachytene region. Previous studies in *C. elegans* oogenesis indicate that distinct (or differential availability of) repair mechanisms are present in early versus late pachytene phases [S1, S2]. The difference in DSB repair dynamics in early vs late pachytene indicate that heat-shock induced DSBs may proceed down repair pathways that are either differentially available or differentially regulated during subsequent stages of meiotic prophase. Statistical significance between groups was determined using the Kruskal-Wallis non-parametric test, with Dunns test to account for multiple comparisons. Number of nuclei scored from male germlines: Early pachytene: no heat, n=16; 1 hour post-heat-shock, n=37; 2 hours post-heat-shock, n=33; 3 hours post-heat-shock, n=26; late pachytene: no heat, n=16; 1 hour post-heat-shock, n=27; 2 hours post-heat-shock, n=18; 3 hours post-heat-shock, n=13.

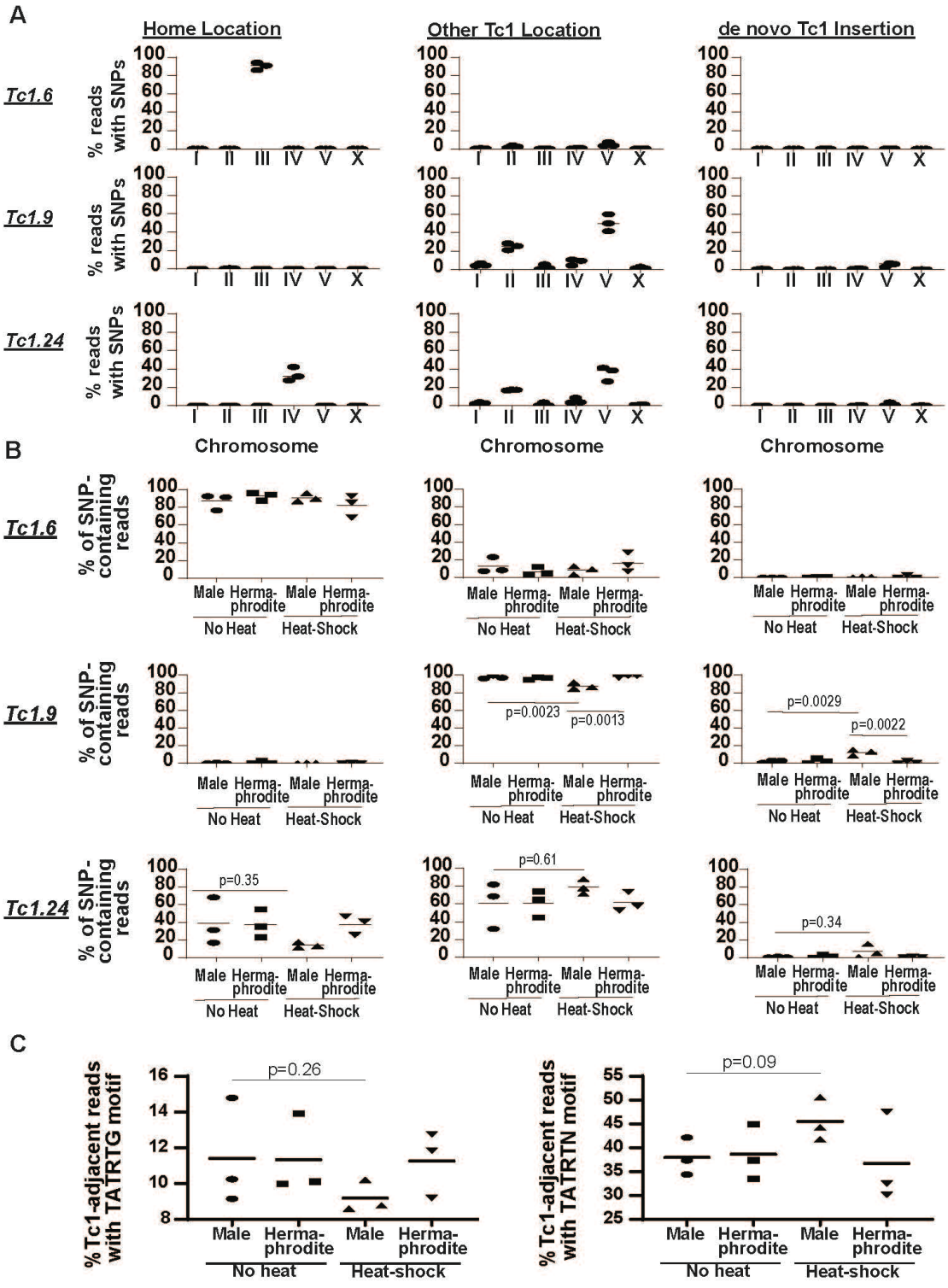

A
Home Location
Other Tc1 Location
de novo Tc1 Insertion
Tc1.6
Tc1.9
Tc1.24
% reads with SNPs
I II III IV V X
Chromosome
B
% of SNP-containing reads
Male Herma-phrodite Male Herma-phrodite
No Heat
Heat-Shock
p=0.0023
p=0.0013
p=0.0029
p=0.0022
p=0.35
p=0.61
p=0.34
C
% Tc1-adjacent reads with TATRTG motif
% Tc1-adjacent reads with TATRTN motif
p=0.26
p=0.09
No heat
Heat-shock

Figure S4. Characterization of three distinct copies of Tc1 and TATRTG consensus Tc1 insertion motif. Related to Figure 4 and Tables S2 and S3. Tc1 mobility was assessed at the level of individual copies of Tc1 using copy-specific polymorphisms. (A) Percent of polymorphism-containing reads across each chromosome for each copy of Tc1 t at their previously reported established location, at other known Tc1 locations, and at *de novo* locations. Values for each biological replicate are displayed. (B) Comparison of percentage of polymorphism-containing reads at each type of location between each exposure group. Values for each biological replicate are displayed. Differences in percent-occupancy of established, other Tc1, or *de novo* insertion sites were assessed vie one-way ANOVA, with Sidaks test to account for multiple comparisons. (C) Percent occurrence of TATRTG and TATRTN insertion motifs were quantified across each exposure group (3 biological replicates per group are represented on the graph). The TATRTG motif was found in ~11% of all Tc1-adjacent reads. There was a trend toward a decrease in the occurrence of the TATRTG motif in males exposed to heat-shock compared with unexposed males (heat-shock: 9.2 ± 0.8%, no heat: 11 ± 3%), however this was not statistically significant (p=0.26). The TATRTN motif was found in ~40% of all Tc1-adjacent reads. Males exposed to heat-shock demonstrated a slight increase in the occurrence of TATRTN compared with their no heat-shock counterparts (heat-shock: 46 ± 4%, no heat: 38 ± 4%), however, this was not statistically significant (p=0.09).

Tc1:	Total # sites occupied by Tc1		# known Tc1		# of *de novo* insertions		
	No heat shock	Heat shock	No heat shock	Heat shock	No heat shock	Heat shock	%Δ
Male	254 ± 20	283 ± 15	212 ± 9	214 ± 4	42 ± 8	69 ± 8	+64.3
Hermaphrodite	296 ± 22	247 ± 21	243 ± 9	202 ± 16	53 ± 15	44 ± 11	-16.9

Tc3:	Total # sites occupied by Tc3		# of known Tc3		# of *de novo* insertions		
	No heat shock	Heat shock	No heat shock	Heat shock	No heat shock	Heat shock	%Δ
Male	199 ± 27	217 ± 18	183 ± 20	197 ± 18	16 ± 7	20 ± 5	+25
Hermaphrodite	225 ± 6	211 ± 10	207 ± 5	192 ± 11	18 ± 3	19 ± 2	+6.9

Table S1. Increased incidence of *de novo* Tc1 insertions in males exposed to heat-stress. Related to Figure 3. Inverse PCR paired with amplicon sequencing was used to identify Tc1 and Tc3 locations genome wide in male and hermaphrodite populations (300 worms/group). Illumina sequencing reads (6851314 total reads) representing Tc1-adjacent DNA was aligned to the *C. elegans* genome (WS245). Genomic position was binned by 100 bp and a Tc1 or Tc3 location site was defined as a bin where a Tc1 or Tc3-adjacent read mapped at a frequency above 20. Tc1 and Tc3 locations were quantified for males and hermaphrodites. Total number of locations, number of known locations, and *de novo* insertion sites for each experimental group are reported.

Hermaphrodite No heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.06 ± 0.01	0.55 ± 0.03	0.15 ± 0	0.28 ± 0.04	0.12 ± 0.01	0.29 ± 0.03
C	0.38 ± 0.04	0.04 ± 0	0.26 ± 0.07	0.18 ± 0.02	0.08 ± 0.01	0.09 ± 0.01
G	0.05 ± 0.01	0.09 ± 0.01	0.36 ± 0.03	0.48 ± 0.03	0.05 ± 0	0.24 ± 0.02
T	0.51 ± 0.03	0.32 ± 0.03	0.24 ± 0.05	0.06 ± 0	0.74 ± 0.03	0.39 ± 0.03

Male No heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.07 ± 0.02	0.51 ± 0.03	0.16 ± 0.02	0.3 ± 0.06	0.15 ± 0.05	0.26 ± 0.05
C	0.4 ± 0.05	0.05 ± 0.02	0.21 ± 0.06	0.15 ± 0.02	0.12 ± 0.02	0.1 ± 0.04
G	0.06 ± 0.02	0.11 ± 0	0.4 ± 0.04	0.49 ± 0.03	0.04 ± 0	0.29 ± 0.01
T	0.47 ± 0.03	0.33 ± 0.04	0.22 ± 0.01	0.06 ± 0.01	0.68 ± 0.06	0.35 ± 0.07

Hermaphrodite Heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.07 ± 0.01	0.6 ± 0.12	0.18 ± 0.03	0.32 ± 0.12	0.11 ± 0.01	0.23 ± 0.03
C	0.34 ± 0.12	0.04 ± 0	0.26 ± 0.04	0.2 ± 0.03	0.1 ± 0.01	0.11 ± 0.05
G	0.05 ± 0	0.1 ± 0.02	0.32 ± 0.11	0.44 ± 0.14	0.05 ± 0.01	0.34 ± 0.03
T	0.54 ± 0.11	0.26 ± 0.11	0.25 ± 0.11	0.04 ± 0	0.75 ± 0.02	0.32 ± 0.07

Male Heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.07 ± 0.01	0.56 ± 0.02	0.19 ± 0.01	0.35 ± 0.05	0.17 ± 0.05	0.3 ± 0.02
C	0.37 ± 0.01	0.05 ± 0	0.23 ± 0.04	0.16 ± 0.02	0.11 ± 0.01	0.14 ± 0.01
G	0.05 ± 0.01	0.12 ± 0.01	0.33 ± 0.01	0.44 ± 0.05	0.05 ± 0	0.22 ± 0.03
T	0.51 ± 0	0.28 ± 0.01	0.25 ± 0.03	0.06 ± 0	0.67 ± 0.06	0.34 ± 0.02

Table S2. Analysis of all Tc1-adjacent reads for insertion motifs: Nucleotide frequency in first six positions immediately 3' of Tc1. Related to Figures 3 and S4, and Table S3. Sequences directly adjacent to inverted repeat of Tc1 were analyzed. The first six nucleotides directly adjacent to Tc1 were aligned and the frequency of each nucleotide at each position was quantified. An average and standard deviation are reported for all three replicates.

Hermaphrodite No heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.04 ± 0	0.66 ± 0.04	0.09 ± 0	0.38 ± 0.02	0.06 ± 0.01	0.36 ± 0.08
C	0.28 ± 0.05	0.03 ± 0	0.12 ± 0.06	0.1 ± 0.01	0.05 ± 0	0.14 ± 0.01
G	0.05 ± 0.01	0.06 ± 0	0.32 ± 0.01	0.48 ± 0.03	0.04 ± 0	0.15 ± 0.01
T	0.64 ± 0.05	0.25 ± 0.04	0.47 ± 0.05	0.04 ± 0	0.86 ± 0.01	0.35 ± 0.08

Male No heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.05 ± 0.02	0.58 ± 0.04	0.12 ± 0.04	0.27 ± 0.06	0.06 ± 0.02	0.27 ± 0.02
C	0.37 ± 0.03	0.03 ± 0	0.06 ± 0.01	0.06 ± 0.05	0.08 ± 0.04	0.18 ± 0.08
G	0.03 ± 0.01	0.09 ± 0.04	0.41 ± 0.03	0.63 ± 0.01	0.03 ± 0	0.22 ± 0.02
T	0.55 ± 0.02	0.3 ± 0.08	0.41 ± 0.02	0.03 ± 0	0.83 ± 0.06	0.34 ± 0.11

Hermaphrodite Heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.04 ± 0.01	0.71 ± 0.14	0.08 ± 0.01	0.39 ± 0.19	0.06 ± 0.01	0.19 ± 0.02
C	0.24 ± 0.13	0.03 ± 0	0.06 ± 0.01	0.06 ± 0	0.04 ± 0.01	0.22 ± 0.06
G	0.03 ± 0.01	0.05 ± 0.01	0.3 ± 0.16	0.52 ± 0.18	0.03 ± 0	0.24 ± 0.03
T	0.69 ± 0.13	0.22 ± 0.14	0.55 ± 0.16	0.03 ± 0.01	0.87 ± 0.02	0.35 ± 0.08

Male Heat shock	Nucleotide position immediately 3' of Tc1					
Frequency (Avg ± SD)	1	2	3	4	5	6
A	0.05 ± 0.02	0.6 ± 0.06	0.12 ± 0.05	0.36 ± 0.01	0.07 ± 0.01	0.26 ± 0.05
C	0.34 ± 0.05	0.03 ± 0	0.07 ± 0.01	0.07 ± 0.02	0.08 ± 0.02	0.23 ± 0.01
G	0.04 ± 0.01	0.09 ± 0.02	0.36 ± 0.07	0.54 ± 0.03	0.04 ± 0.01	0.14 ± 0.05
T	0.57 ± 0.04	0.27 ± 0.07	0.44 ± 0.01	0.04 ± 0	0.82 ± 0.03	0.36 ± 0.1

Table S3. Assessment of Tc1-adjacent reads containing TATRTN: Nucleotide frequency in first six positions immediately 3' of Tc1. Related to Figures 3 and S4, and Table S2. Reads containing the TATRTN motif were parsed out and the first six nucleotides directly adjacent to Tc1 were aligned. Similar to Table S2, the frequency of each nucleotide at each position was quantified. An average and standard deviation are reported for all three replicates.

Primer target	Primer sequence	Forward or Reverse
*Tc3	AATAGTCGCGGGTTGAGTTG	RV
*Tc3	GAGCGTTCACGGAGAAGAAG	FW
*Tc1	CACACGACGACGTTGAAACC	RV
*Tc1	AACCGTTAAGCATGGAGGTG	FW
**act-1*	GTACGTCCGGAAGCGTAGAG	RV
**act-1*	CACGGTATCGTCACCAACTG	FW
*(universal)	CTCAAGTGTCGTGGAGTCGGCAA	RV
†*pmp-3*	TGGCCGGATGATGGTGTCGC	FW
†*pmp-3*	ACGAACAATGCCAAAGGCCAGC	RV
Tc1 3' end, PCR #1	CAAACGGATACGCGACAAA	FW
Tc1 3' end, PCR #1	TGACAACGACGTGGCATC	RV
Tc1 5' end, PCR #1	GAAAGCCATTGTAGCTGG	FW
Tc1 5' end, PCR #1	ATCCAGTGCAACAAACAGAG	RV
Tc1 3' end, PCR #2 (nested)	CTGGTTGAATTTGGCATCTG	RV
Tc1 3' end, PCR #2 (nested)	TATCTTTTTGGCCAGCACTG	FW
Tc1 5' end, PCR #2 (nested)	CGAACAAGGAATACCCACGA	FW
Tc1 5' end, PCR #2 (nested)	CATTTCGCTTTATGCACACG	RV
Tc3 3' end, PCR #1	GAAATGTTCTCGTGGGATC	RV
Tc3 3' end, PCR #1	AGCGCGACCTCAGTTTTT	FW
Tc3 5' end, PCR #1	GAACCAGTCTGGAGATCCCAC	RV
Tc3 5' end, PCR #1	TCAAGCTTCCTGGTGCTCTC	FW
Tc3 3' end, PCR #2 (nested)	TCAAGCTTCCTGGTGCTCTC	RV
Tc3 3' end, PCR #2 (nested)	TCGGAAGTTCCTCAAACCTTCA	FW
Tc3 5' end, PCR #2 (nested)	AGCGCGACCTCAGTTTTTGA	RV
Tc3 5' end, PCR #2 (nested)	TACGCCATCCCCAAGAAACC	FW

*from [S3]
†from [S4]

Table S4. Primers used in this study. Related to STAR Methods.

Supplemental References

S1. Rosu, S., Libuda, D.E., and Villeneuve, A.M. (2011). Robust crossover assurance and regulated interhomolog access maintain meiotic crossover number. Science (New York, N.Y.) *334*, 1286-1289.

S2. Hayashi, M., Chin, G.M., and Villeneuve, A.M. (2007). C. elegans germ cells switch between distinct modes of double-strand break repair during meiotic prophase progression. PLoS Genet *3*, e191.

S3. Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., et al. (2008). Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to suppress Tc3 Transposon Mobility in the Caenorhabditis elegans germline. Molecular cell *31*, 79-90.

S4. Zhang, Y., Chen, D., Smith, M.A., Zhang, B., and Pan, X. (2012). Selection of Reliable Reference Genes in Caenorhabditis elegans for Analysis of Nanotoxicity. In PLoS One, Volume 7.

References Cited

Alonso-Blanco, Carlos, Leónie Bentsink, Corrie J. Hanhart, Hetty Blankestijn-de Vries, and Maarten Koornneef. 2003. "Analysis of Natural Allelic Variation at Seed Dormancy Loci of Arabidopsis Thaliana." *Genetics* 164 (2): 711–29. https://doi.org/10.1093/genetics/164.2.711.

Andersen, Erik C., Joshua S. Bloom, Justin P. Gerke, and Leonid Kruglyak. 2014. "A Variant in the Neuropeptide Receptor Npr-1 Is a Major Determinant of Caenorhabditis Elegans Growth and Physiology." Edited by Gregory P. Copenhaver. *PLoS Genetics* 10 (2): e1004156. https://doi.org/10.1371/journal.pgen.1004156.

Andersen, Erik C, Justin P Gerke, Joshua A Shapiro, Jonathan R Crissman, Rajarshi Ghosh, Joshua S Bloom, Marie-Anne Félix, and Leonid Kruglyak. 2012. "Chromosome-Scale Selective Sweeps Shape Caenorhabditis Elegans Genomic Diversity." *Nature Genetics* 44 (3): 285–90. https://doi.org/10.1038/ng.1050.

Aquilina, Gabriele, and Margherita Bignami. 2001. "Mismatch Repair in Correction of Replication Errors and Processing of DNA Damage." *Journal of Cellular Physiology* 187 (2): 145–54. https://doi.org/10.1002/jcp.1067.

Arbel-Eden, Ayelet, and Giora Simchen. 2019. "Elevated Mutagenicity in Meiosis and Its Mechanism." *BioEssays* 41 (4): 1800235. https://doi.org/10.1002/bies.201800235.

Arndt, Peter F., Terence Hwa, and Dmitri A. Petrov. 2005. "Substantial Regional Variation in Substitution Rates in the Human Genome: Importance of GC Content, Gene Density, and

Telomere-Specific Effects." *Journal of Molecular Evolution* 60 (6): 748–63. https://doi.org/10.1007/s00239-004-0222-5.

Avery, Oswald T., Colin M. MacLeod, and Maclyn McCarty. 1944. "STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES." *Journal of Experimental Medicine* 79 (2): 137–58. https://doi.org/10.1084/jem.79.2.137.

Bailey, Timothy L., and William Stafford Noble. 2003. "Searching for Statistically Significant Regulatory Modules." *Bioinformatics* 19 (suppl_2): ii16–25. https://doi.org/10.1093/bioinformatics/btg1054.

Bao, Weidong, Matthew G. Jurka, Vladimir V. Kapitonov, and Jerzy Jurka. 2009. "New Superfamilies of Eukaryotic DNA Transposons and Their Internal Divisions." *Molecular Biology and Evolution* 26 (5): 983–93. https://doi.org/10.1093/molbev/msp013.

Barnes, T M, Y Kohara, A Coulson, and S Hekimi. 1995. "Meiotic Recombination, Noncoding DNA and Genomic Organization in Caenorhabditis Elegans." *Genetics* 141 (1): 159–79. https://doi.org/10.1093/genetics/141.1.159.

Barnett, Derek W, Erik K Garrison, Aaron R Quinlan, Michael P Strömberg, and Gabor T Marth. 2011. "BamTools: A C++ API and Toolkit for Analyzing and Managing BAM Files." *Bioinformatics* 27 (12): 1691–92. https://doi.org/10.1093/bioinformatics/btr174.

Barth, Susanne, Albrecht E. Melchinger, Beate Devezi-Savula, and Thomas Lübberstedt. 2000. "A High-Throughput System for Genome-Wide Measurement of Genetic Recombination in Arabidopsis Thaliana Based on Transgenic Markers." *Functional & Integrative Genomics* 1 (3): 200–206. https://doi.org/10.1007/s101420000030.

Basenko, Evelina Y., Takahiko Sasaki, Lexiang Ji, Cameron J. Prybol, Rachel M. Burckhardt, Robert J. Schmitz, and Zachary A. Lewis. 2015. "Genome-Wide Redistribution of H3K27me3 Is Linked to Genotoxic Stress and Defective Growth." *Proceedings of the National Academy of Sciences* 112 (46). https://doi.org/10.1073/pnas.1511377112.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop, and B. De Massy. 2010. "PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice." *Science* 327 (5967): 836–40. https://doi.org/10.1126/science.1183439.

Baudat, Frédéric, Katia Manova, Julie Pui Yuen, Maria Jasin, and Scott Keeney. 2000. "Chromosome Synapsis Defects and Sexually Dimorphic Meiotic Progression in Mice Lacking Spo11." *Molecular Cell* 6 (5): 989–98. https://doi.org/10.1016/S1097-2765(00)00098-8.

Ben-David, Eyal, Alejandro Burga, and Leonid Kruglyak. 2017. "A Maternal-Effect Selfish Genetic Element in *Caenorhabditis Elegans*." *Science* 356 (6342): 1051–55. https://doi.org/10.1126/science.aan0621.

Bentsink, Leónie, Jemma Jowett, Corrie J. Hanhart, and Maarten Koornneef. 2006. "Cloning of *DOG1* , a Quantitative Trait Locus Controlling Seed Dormancy in *Arabidopsis*." *Proceedings of the National Academy of Sciences* 103 (45): 17042–47. https://doi.org/10.1073/pnas.0607877103.

Berchowitz, Luke, and Gregory Copenhaver. 2010. "Genetic Interference: Dont Stand So Close to Me." *Current Genomics* 11 (2): 91–102. https://doi.org/10.2174/138920210790886835.

Bernardi, Giorgio. 2000. “Isochores and the Evolutionary Genomics of Vertebrates.” *Gene* 241 (1): 3–17. https://doi.org/10.1016/S0378-1119(99)00485-0.

Bernstein, Max R, and Matthew V Rockman. 2016. “Fine-Scale Crossover Rate Variation on the *Caenorhabditis Elegans* X Chromosome.” *G3 Genes|Genomes|Genetics* 6 (6): 1767–76. https://doi.org/10.1534/g3.116.028001.

Bessereau, Jean-Louis. 2006. “Transposons in C. Elegans.” *WormBook : The Online Review of C. Elegans Biology*, January, 1–13. https://doi.org/10.1895/wormbook.1.70.1.

Betrán, Esther, and Manyuan Long. 2003. “*Dntf-2r* , a Young Drosophila Retroposed Gene With Specific Male Expression Under Positive Darwinian Selection.” *Genetics* 164 (3): 977–88. https://doi.org/10.1093/genetics/164.3.977.

Bhalla, Needhi, David J. Wynne, Verena Jantsch, and Abby F. Dernburg. 2008. “ZHP-3 Acts at Crossovers to Couple Meiotic Recombination with Synaptonemal Complex Disassembly and Bivalent Formation in C. Elegans.” Edited by R. Scott Hawley. *PLoS Genetics* 4 (10): e1000235. https://doi.org/10.1371/journal.pgen.1000235.

Bird, Adrian P. 1986. “CpG-Rich Islands and the Function of DNA Methylation.” *Nature* 321 (6067): 209–13. https://doi.org/10.1038/321209a0.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bradley, Michael D., Devin Neu, Fatmagul Bahar, and Roy D. Welch. 2016. “Inter-Laboratory Evolution of a Model Organism and Its Epistatic Effects on Mutagenesis Screens.” *Scientific Reports* 6 (1): 38001. https://doi.org/10.1038/srep38001.

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

Brick, Kevin, Sarah Thibault-Sennett, Fatima Smagulova, Kwan-Wood G. Lam, Yongmei Pu, Florencia Pratto, R. Daniel Camerini-Otero, and Galina V. Petukhova. 2018. “Extensive Sex Differences at the Initiation of Genetic Recombination.” *Nature* 561 (7723): 338–42. https://doi.org/10.1038/s41586-018-0492-5.

Broman, Karl W., Jeffrey C. Murray, Val C. Sheffield, Raymond L. White, and James L. Weber. 1998. “Comprehensive Human Genetic Maps: Individual and Sex-Specific Variation in Recombination.” *The American Journal of Human Genetics* 63 (3): 861–69. https://doi.org/10.1086/302011.

Busby, S. 1994. “Promoter Structure, Promoter Recognition, and Transcription Activation in Prokaryotes.” *Cell* 79 (5): 743–46. https://doi.org/10.1016/0092-8674(94)90063-9.

C. elegans Sequencing Consortium. 1998. “Genome Sequence of the Nematode C. Elegans: A Platform for Investigating Biology.” *Science* 282 (5396): 2012–18. https://doi.org/10.1126/science.282.5396.2012.

Caridi, P. Christopher, Laetitia Delabaere, Grzegorz Zapotoczny, and Irene Chiolo. 2017. “And yet, It Moves: Nuclear and Chromatin Dynamics of a Heterochromatic Double-Strand Break.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1731): 20160291. https://doi.org/10.1098/rstb.2016.0291.

Carreto, Laura, Maria F Eiriz, Ana C Gomes, Patrícia M Pereira, Dorit Schuller, and Manuel As Santos. 2008. “Comparative Genomics of Wild Type Yeast Strains Unveils Important

Genome Diversity." *BMC Genomics* 9 (1): 524. https://doi.org/10.1186/1471-2164-9-524.

Carvalho, Claudia M B, and James R Lupski. 2016. "Mechanisms Underlying Structural Variant Formation in Genomic Disorders." *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg.2015.25.

Chalopin, Domitille, Magali Naville, Floriane Plard, Delphine Galiana, and Jean-Nicolas Volff. 2015. "Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates." *Genome Biology and Evolution* 7 (2): 567–80. https://doi.org/10.1093/gbe/evv005.

Chebib, Jobran, Benjamin C. Jackson, Eugenio López-Cortegano, Diethard Tautz, and Peter D. Keightley. 2021. "Inbred Lab Mice Are Not Isogenic: Genetic Variation within Inbred Strains Used to Infer the Mutation Rate per Nucleotide Site." *Heredity* 126 (1): 107–16. https://doi.org/10.1038/s41437-020-00361-1.

Chiolo, Irene, Aki Minoda, Serafin U. Colmenares, Aris Polyzos, Sylvain V. Costes, and Gary H. Karpen. 2011. "Double-Strand Breaks in Heterochromatin Move Outside of a Dynamic HP1a Domain to Complete Recombinational Repair." *Cell* 144 (5): 732–44. https://doi.org/10.1016/j.cell.2011.02.012.

Chiruvella, K. K., Z. Liang, and T. E. Wilson. 2013. "Repair of Double-Strand Breaks by End Joining." *Cold Spring Harbor Perspectives in Biology* 5 (5): a012757–a012757. https://doi.org/10.1101/cshperspect.a012757.

Cole, Francesca, Liisa Kauppi, Julian Lange, Ignasi Roig, Raymond Wang, Scott Keeney, and Maria Jasin. 2012. "Homeostatic Control of Recombination Is Implemented

Progressively in Mouse Meiosis." *Nature Cell Biology* 14 (4): 424–30. https://doi.org/10.1038/ncb2451.

Connolly, Lanelle R., Kristina M. Smith, and Michael Freitag. 2013. "The Fusarium Graminearum Histone H3 K27 Methyltransferase KMT6 Regulates Development and Expression of Secondary Metabolite Gene Clusters." Edited by Hiten D. Madhani. *PLoS Genetics* 9 (10): e1003916. https://doi.org/10.1371/journal.pgen.1003916.

Cortes-Bratti, Ximena, Teresa Frisan, and Monica Thelestam. 2001. "The Cytolethal Distending Toxins Induce DNA Damage and Cell Cycle Arrest." *Toxicon* 39 (11): 1729–36. https://doi.org/10.1016/S0041-0101(01)00159-3.

Crick, F. H. 1958. "On Protein Synthesis." *Symposia of the Society for Experimental Biology* 12: 138–63.

Crombie, Tim A, Stefan Zdraljevic, Daniel E Cook, Robyn E Tanny, Shannon C Brady, Ye Wang, Kathryn S Evans, et al. 2019. "Deep Sampling of Hawaiian Caenorhabditis Elegans Reveals High Genetic Diversity and Admixture with Global Populations." Edited by Graham Coop, Diethard Tautz, and Asher Cutter. *eLife* 8 (December): e50465. https://doi.org/10.7554/eLife.50465.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A Albers, Eric Banks, Mark A DePristo, Robert E Handsaker, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics (Oxford, England)* 27 (15): 2156–58. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, Petr, and Shane A McCarthy. 2017. “BCFtools/Csq: Haplotype-Aware Variant Consequences.” *Bioinformatics (Oxford, England)* 33 (13): 2037–39. https://doi.org/10.1093/bioinformatics/btx100.

Daranlapujade, P, J Daran, P Kotter, T Petit, M Piper, and J Pronk. 2003. “Comparative Genotyping of the Laboratory Strains S288C and CEN.PK113-7D Using Oligonucleotide Microarrays.” *FEMS Yeast Research* 4 (3): 259–69. https://doi.org/10.1016/S1567-1356(03)00156-9.

Delcher, A L, S Kasif, R D Fleischmann, J Peterson, O White, and S L Salzberg. 1999. “Alignment of Whole Genomes.” *Nucleic Acids Research* 27 (11): 2369–76. https://doi.org/10.1093/nar/27.11.2369.

Denver, Dee R, Peter C Dolan, Larry J Wilhelm, Way Sung, J Ignacio Lucas-Lledó, Dana K Howe, Samantha C Lewis, et al. 2009. “A Genome-Wide View of Caenorhabditis Elegans Base-Substitution Mutation Processes.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (38): 16310–14. https://doi.org/10.1073/pnas.0904895106.

Dernburg, Abby F, Kent McDonald, Gary Moulder, Robert Barstead, Michael Dresser, and Anne M Villeneuve. 1998. “Meiotic Recombination in C. Elegans Initiates by a Conserved Mechanism and Is Dispensable for Homologous Chromosome Synapsis.” *Cell* 94 (3): 387–98. https://doi.org/10.1016/S0092-8674(00)81481-6.

Doniskeller, H. 1987. “A Genetic Linkage Map of the Human Genome.” *Cell* 51 (2): 319–37. https://doi.org/10.1016/0092-8674(87)90158-9.

Drouaud, Jan, Raphaël Mercier, Liudmila Chelysheva, Aurélie Bérard, Matthieu Falque, Olivier Martin, Vanessa Zanni, Dominique Brunel, and Christine Mézard. 2007. "Sex-Specific Crossover Distributions and Variations in Interference Level along Arabidopsis Thaliana Chromosome 4." Edited by Michael Lichten. *PLoS Genetics* 3 (6): e106. https://doi.org/10.1371/journal.pgen.0030106.

Duveau, Fabien, and Marie-Anne Félix. 2012. "Role of Pleiotropy in the Evolution of a Cryptic Developmental Variation in Caenorhabditis Elegans." Edited by Mohamed A. F. Noor. *PLoS Biology* 10 (1): e1001230. https://doi.org/10.1371/journal.pbio.1001230.

Eide, D, and P Anderson. 1985. "Transposition of Tc1 in the Nematode Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences* 82 (6): 1756–60. https://doi.org/10.1073/pnas.82.6.1756.

Ellegren, Hans, Nick Gc Smith, and Matthew T Webster. 2003. "Mutation Rate Variation in the Mammalian Genome." *Current Opinion in Genetics & Development* 13 (6): 562–68. https://doi.org/10.1016/j.gde.2003.10.008.

Emmons, Scott W., Lewis Yesner, Ke-san Ruan, and Daniel Katzenberg. 1983. "Evidence for a Transposon in Caenorhabditis Elegans." *Cell* 32 (1): 55–65. https://doi.org/10.1016/0092-8674(83)90496-8.

Fan, Hao, and Jia-You Chu. 2007. "A Brief Review of Short Tandem Repeat Mutation." *Genomics, Proteomics & Bioinformatics* 5 (1): 7–14. https://doi.org/10.1016/S1672-0229(07)60009-6.

Faustino, Nuno André, and Thomas A. Cooper. 2003. "Pre-mRNA Splicing and Human Disease." *Genes & Development* 17 (4): 419–37. https://doi.org/10.1101/gad.1048803.

Feng, Jianxing, Tao Liu, Bo Qin, Yong Zhang, and Xiaole Shirley Liu. 2012. “Identifying ChIP-Seq Enrichment Using MACS.” *Nature Protocols* 7 (9): 1728–40. https://doi.org/10.1038/nprot.2012.101.

Feschotte, Cédric, and Ellen J. Pritham. 2007. “DNA Transposons and the Evolution of Eukaryotic Genomes.” *Annual Review of Genetics* 41: 331–68. https://doi.org/10.1146/annurev.genet.40.110405.090448.

Fickett, James W., and Artemis G. Hatzigeorgiou. 1997. “Eukaryotic Promoter Recognition.” *Genome Research* 7 (9): 861–78. https://doi.org/10.1101/gr.7.9.861.

Fischer, Sylvia E J, Erno Wienholds, and Ronald H A Plasterk. 2003. “Continuous Exchange of Sequence Information Between Dispersed Tc1 Transposons in the *Caenorhabditis Elegans* Genome.” *Genetics* 164 (1): 127–34. https://doi.org/10.1093/genetics/164.1.127.

Fozard, John A, Chris Morgan, and Martin Howard. 2023. “Coarsening Dynamics Can Explain Meiotic Crossover Patterning in Both the Presence and Absence of the Synaptonemal Complex.” *eLife* 12 (February): e79408. https://doi.org/10.7554/eLife.79408.

Gabdank, Idan, and Andrew Z Fire. 2014. “Gamete-Type Dependent Crossover Interference Levels in a Defined Region of *Caenorhabditis Elegans* Chromosome V.” *G3 Genes|Genomes|Genetics* 4 (1): 117–20. https://doi.org/10.1534/g3.113.008672.

Gagliano, Sarah A., Sebanti Sengupta, Carlo Sidore, Andrea Maschio, Francesco Cucca, David Schlessinger, and Gonçalo R. Abecasis. 2019. “Relative Impact of Indels versus SNPs on Complex Disease.” *Genetic Epidemiology* 43 (1): 112–17. https://doi.org/10.1002/gepi.22175.

Garrison, Erik, and Gabor Marth. 2012. “Haplotype-Based Variant Detection from Short-Read Sequencing.” arXiv. https://doi.org/10.48550/ARXIV.1207.3907.

Gems, D, and D L Riddle. 2000. “Defining Wild-Type Life Span in Caenorhabditis Elegans.” *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 55 (5): B215-9. https://doi.org/10.1093/gerona/55.5.b215.

Gerton, Jennifer L., Joseph DeRisi, Robert Shroff, Michael Lichten, Patrick O. Brown, and Thomas D. Petes. 2000. “Global Mapping of Meiotic Recombination Hotspots and Coldspots in the Yeast *Saccharomyces Cerevisiae*.” *Proceedings of the National Academy of Sciences* 97 (21): 11383–90. https://doi.org/10.1073/pnas.97.21.11383.

Geurts, Aron M, Lara S Collier, Jennifer L Geurts, Leann L Oseth, Matthew L Bell, David Mu, Robert Lucito, et al. 2006. “Gene Mutations and Genomic Rearrangements in the Mouse as a Result of Transposon Mobilization from Chromosomal Concatemers.” Edited by Gregory S Barsh. *PLoS Genetics* 2 (9): e156. https://doi.org/10.1371/journal.pgen.0020156.

Gilbert, Clément, Jean Peccoud, and Richard Cordaux. 2021. “Transposable Elements and the Evolution of Insects.” *Annual Review of Entomology* 66 (1): 355–72. https://doi.org/10.1146/annurev-ento-070720-074650.

Girard, Chloe, David Zwicker, and Raphael Mercier. 2023. “The Regulation of Meiotic Crossover Distribution: A Coarse Solution to a Century-Old Mystery?” *Biochemical Society Transactions* 51 (3): 1179–90. https://doi.org/10.1042/BST20221329.

Girard, Lisa, and Michael Freeling. 1999. “Regulatory Changes as a Consequence of Transposon Insertion.” *Developmental Genetics* 25 (4): 291–96. https://doi.org/10.1002/(SICI)1520-6408(1999)25:4<291::AID-DVG2>3.0.CO;2-5.

Globus, Samuel T., and Scott Keeney. 2012. “The Joy of Six: How to Control Your Crossovers.” *Cell* 149 (1): 11–12. https://doi.org/10.1016/j.cell.2012.03.011.

Goel, Manish, Hequan Sun, Wen-Biao Jiao, and Korbinian Schneeberger. 2019. “SyRI: Finding Genomic Rearrangements and Local Sequence Differences from Whole-Genome Assemblies.” *Genome Biology* 20 (1): 277. https://doi.org/10.1186/s13059-019-1911-0.

Goerner-Potvin, Patricia, and Guillaume Bourque. 2018. “Computational Tools to Unmask Transposable Elements.” *Nature Reviews Genetics* 19 (11): 688–704. https://doi.org/10.1038/s41576-018-0050-x.

Gondo, Yoichi, Takeya Okada, Noriko Matsuyama, Yasushi Saitoh, Yoshiko Yanagisawa, and Joh-E Ikeda. 1998. “Human Megasatellite DNA RS447: Copy-Number Polymorphisms and Interspecies Conservation.” *Genomics* 54 (1): 39–49. https://doi.org/10.1006/geno.1998.5545.

Grelon, M. 2001. “AtSPO11-1 Is Necessary for Efficient Meiotic Recombination in Plants.” *The EMBO Journal* 20 (3): 589–600. https://doi.org/10.1093/emboj/20.3.589.

Guryev, Victor, Marco J. Koudijs, Eugene Berezikov, Stephen L. Johnson, Ronald H.A. Plasterk, Fredericus J.M. Van Eeden, and Edwin Cuppen. 2006. “Genetic Variation in the Zebrafish.” *Genome Research* 16 (4): 491–97. https://doi.org/10.1101/gr.4791006.

Hammarlund, Marc, M Wayne Davis, Hung Nguyen, Dustin Dayton, and Erik M Jorgensen. 2005. “Heterozygous Insertions Alter Crossover Distribution but Allow Crossover

Interference in *Caenorhabditis Elegans*." *Genetics* 171 (3): 1047–56. https://doi.org/10.1534/genetics.105.044834.

Haraksingh, Rajini R, and Michael P Snyder. 2013. "Impacts of Variation in the Human Genome on Gene Regulation." *Journal of Molecular Biology* 425 (21): 3970–77. https://doi.org/10.1016/j.jmb.2013.07.015.

Hardison, Ross C., Krishna M. Roskin, Shan Yang, Mark Diekhans, W. James Kent, Ryan Weber, Laura Elnitski, et al. 2003. "Covariation in Frequencies of Substitution, Deletion, Transposition, and Recombination During Eutherian Evolution." *Genome Research* 13 (1): 13–26. https://doi.org/10.1101/gr.844103.

Hartl, D. 1996. "Compensatory Nearly Neutral Mutations: Selection without Adaptation." *Journal of Theoretical Biology* 182 (3): 303–9. https://doi.org/10.1006/jtbi.1996.0168.

Hassold, Terry, and Patricia Hunt. 2001. "To Err (Meiotically) Is Human: The Genesis of Human Aneuploidy." *Nature Reviews Genetics* 2 (4): 280–91. https://doi.org/10.1038/35066065.

Heger, Andreas, Caleb Webber, Martin Goodson, Chris P Ponting, and Gerton Lunter. 2013. "GAT: A Simulation Framework for Testing the Association of Genomic Intervals." *Bioinformatics* 29 (16): 2046–48. https://doi.org/10.1093/bioinformatics/btt343.

Henzel, Jonathan V, Kentaro Nabeshima, Mara Schvarzstein, B Elizabeth Turner, Anne M Villeneuve, and Kenneth J Hillers. 2011. "An Asymmetric Chromosome Pair Undergoes Synaptic Adjustment and Crossover Redistribution During *Caenorhabditis Elegans* Meiosis: Implications for Sex Chromosome Evolution." *Genetics* 187 (3): 685–99. https://doi.org/10.1534/genetics.110.124958.

Hershey, A. D., and Martha Chase. 1952. “INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE.” *Journal of General Physiology* 36 (1): 39–56. https://doi.org/10.1085/jgp.36.1.39.

Hillers, Kenneth J. 2004. “Crossover Interference.” *Current Biology* 14 (24): R1036–37. https://doi.org/10.1016/j.cub.2004.11.038.

Hillers, Kenneth J., and Anne M. Villeneuve. 2003. “Chromosome-Wide Control of Meiotic Crossing over in C. Elegans.” *Current Biology* 13 (18): 1641–47. https://doi.org/10.1016/j.cub.2003.08.026.

Ho, Joshua W. K., Youngsook L. Jung, Tao Liu, Burak H. Alver, Soohyun Lee, Kohta Ikegami, Kyung-Ah Sohn, et al. 2014. “Comparative Analysis of Metazoan Chromatin Organization.” *Nature* 512 (7515): 449–52. https://doi.org/10.1038/nature13415.

Hodgkin, J, and T Doniach. 1997. “Natural Variation and Copulatory Plug Formation in Caenorhabditis Elegans.” *Genetics* 146 (1): 149–64. https://doi.org/10.1093/genetics/146.1.149.

Hodgkin, Jonathan, H Robert Horvitz, and Sydney Brenner. 1979. “NONDISJUNCTION MUTANTS OF THE NEMATODE *CAENORHABDITIS ELEGANS*.” *Genetics* 91 (1): 67–94. https://doi.org/10.1093/genetics/91.1.67.

Hodgkinson, Alan, and Adam Eyre-Walker. 2011. “Variation in the Mutation Rate across Mammalian Genomes.” *Nature Reviews Genetics* 12 (11): 756–66. https://doi.org/10.1038/nrg3098.

Houston, Brendan J, Brett Nixon, Jacinta H Martin, Geoffry N De Iuliis, Natalie A Trigg, Elizabeth G Bromfield, Kristen E McEwan, and R John Aitken. 2018. “Heat Exposure

Induces Oxidative Stress and DNA Damage in the Male Germ Line†." *Biology of Reproduction* 98 (4): 593–606. https://doi.org/10.1093/biolre/ioy009.

Hurles, Matthew E, Emmanouil T Dermitzakis, and Chris Tyler-Smith. 2008. "The Functional Impact of Structural Variation in Humans." *Trends in Genetics : TIG* 24 (5): 238–45. https://doi.org/10.1016/j.tig.2008.03.001.

Jain, Miten, Sergey Koren, Karen H Miga, Josh Quick, Arthur C Rand, Thomas A Sasani, John R Tyson, et al. 2018. "Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads." *Nature Biotechnology* 36 (4): 338–45. https://doi.org/10.1038/nbt.4060.

James, Kent W, Baertsch Robert, Hinrichs Angie, Miller Webb, and Haussler David. 2003. "Evolution's Cauldron: Duplication, Deletion, and Rearrangement in the Mouse and Human Genomes." *Proceedings of the National Academy of Sciences* 100 (20): 11484–89. https://doi.org/10.1073/pnas.1932072100.

Jamieson, Kirsty, Michael R. Rountree, Zachary A. Lewis, Jason E. Stajich, and Eric U. Selker. 2013. "Regional Control of Histone H3 Lysine 27 Methylation in *Neurospora*." *Proceedings of the National Academy of Sciences* 110 (15): 6027–32. https://doi.org/10.1073/pnas.1303750110.

Janssen, Aniek, Gregory A. Breuer, Eva K. Brinkman, Annelot I. Van Der Meulen, Sean V. Borden, Bas Van Steensel, Ranjit S. Bindra, Jeannine R. LaRocque, and Gary H. Karpen. 2016. "A Single Double-Strand Break System Reveals Repair Dynamics and Mechanisms in Heterochromatin and Euchromatin." *Genes & Development* 30 (14): 1645–57. https://doi.org/10.1101/gad.283028.116.

Jantsch, Verena, Pawel Pasierbek, Michael M. Mueller, Dieter Schweizer, Michael Jantsch, and Josef Loidl. 2004. "Targeted Gene Knockout Reveals a Role in Meiotic Recombination for ZHP-3, a Zip3-Related Protein in *Caenorhabditis Elegans*." *Molecular and Cellular Biology* 24 (18): 7998–8006. https://doi.org/10.1128/MCB.24.18.7998-8006.2004.

Jaramillo-Lambert, Aimee, Marina Ellefson, Anne M. Villeneuve, and JoAnne Engebrecht. 2007. "Differential Timing of S Phases, X Chromosome Replication, and Meiotic Prophase in the C. Elegans Germ Line." *Developmental Biology* 308 (1): 206–21. https://doi.org/10.1016/j.ydbio.2007.05.019.

Jones, G. H. 1984. "The Control of Chiasma Distribution." *Symposia of the Society for Experimental Biology* 38: 293–320.

Jones, Gareth H., and F. Chris H. Franklin. 2006. "Meiotic Crossing-over: Obligation and Interference." *Cell* 126 (2): 246–48. https://doi.org/10.1016/j.cell.2006.07.010.

Joyce, Eric F, Anshu Paul, Katherine E Chen, Nikhila Tanneti, and Kim S McKim. 2012. "Multiple Barriers to Nonhomologous DNA End Joining During Meiosis in *Drosophila*." *Genetics* 191 (3): 739–46. https://doi.org/10.1534/genetics.112.140996.

Kaessmann, Henrik. 2010. "Origins, Evolution, and Phenotypic Impact of New Genes." *Genome Research* 20 (10): 1313–26. https://doi.org/10.1101/gr.101386.109.

Kamath, Ravi S, Andrew G Fraser, Yan Dong, Gino Poulin, Richard Durbin, Monica Gotta, Alexander Kanapin, et al. 2003. "Systematic Functional Analysis of the Caenorhabditis Elegans Genome Using RNAi." *Nature* 421 (6920): 231–37. https://doi.org/10.1038/nature01278.

Kaur, Taniya, and Matthew V Rockman. 2014. “Crossover Heterogeneity in the Absence of Hotspots in *Caenorhabditis Elegans*.” *Genetics* 196 (1): 137–48. https://doi.org/10.1534/genetics.113.158857.

Keeney, Scott. 2008. “Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis.” In *Recombination and Meiosis*, edited by Richard Egel and Dirk-Henner Lankenau, 2:81–123. Genome Dynamics and Stability. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/7050_2007_026.

Keeney, Scott, Craig N Giroux, and Nancy Kleckner. 1997. “Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family.” *Cell* 88 (3): 375–84. https://doi.org/10.1016/S0092-8674(00)81876-0.

Kelly, Karen O, Abby F Dernburg, Gillian M Stanfield, and Anne M Villeneuve. 2000. “*Caenorhabditis Elegans Msh-5* Is Required for Both Normal and Radiation-Induced Meiotic Crossing Over but Not for Completion of Meiosis.” *Genetics* 156 (2): 617–30. https://doi.org/10.1093/genetics/156.2.617.

Kern, Andrew D, and Matthew W Hahn. 2018. “The Neutral Theory in Light of Natural Selection.” *Molecular Biology and Evolution* 35 (6): 1366–71. https://doi.org/10.1093/molbev/msy092.

Kianian, Penny M. A., Minghui Wang, Kristin Simons, Farhad Ghavami, Yan He, Stefanie Dukowic-Schulze, Anitha Sundararajan, et al. 2018. “High-Resolution Crossover Mapping Reveals Similarities and Differences of Male and Female Recombination in Maize.” *Nature Communications* 9 (1): 2370. https://doi.org/10.1038/s41467-018-04562-5.

Kim, Bernard Y, Jeremy R Wang, Danny E Miller, Olga Barmina, Emily Delaney, Ammon Thompson, Aaron A Comeault, et al. 2021. “Highly Contiguous Assemblies of 101 Drosophilid Genomes.” Edited by Graham Coop, Patricia J Wittkopp, and Timothy B Sackton. *eLife* 10: e66405. https://doi.org/10.7554/eLife.66405.

Kim, Chuna, Jun Kim, Sunghyun Kim, Daniel E Cook, Kathryn S Evans, Erik C Andersen, and Junho Lee. 2019. “Long-Read Sequencing Reveals Intra-Species Tolerance of Substantial Structural Variations and New Subtelomere Formation in C. Elegans.” *Genome Research* 29 (6): 1023–35. https://doi.org/10.1101/gr.246082.118.

Kim, Keun P., Beth M. Weiner, Liangran Zhang, Amy Jordan, Job Dekker, and Nancy Kleckner. 2010. “Sister Cohesion and Structural Axis Components Mediate Homolog Bias of Meiotic Recombination.” *Cell* 143 (6): 924–37. https://doi.org/10.1016/j.cell.2010.11.015.

Koch, R, H G van Luenen, M van der Horst, K L Thijssen, and R H Plasterk. 2000. “Single Nucleotide Polymorphisms in Wild Isolates of Caenorhabditis Elegans.” *Genome Research* 10 (11): 1690–96. https://doi.org/10.1101/gr.gr-1471r.

Koren, Sergey, Brian P Walenz, Konstantin Berlin, Jason R Miller, Nicholas H Bergman, and Adam M Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation.” *Genome Research* 27 (5): 722–36. https://doi.org/10.1101/gr.215087.116.

Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell* 128 (4): 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

Kurhanewicz, Nicole A., Devin Dinwiddie, Zachary D. Bush, and Diana E. Libuda. 2020. "Elevated Temperatures Cause Transposon-Associated DNA Damage in C. Elegans Spermatocytes." *Current Biology* 30 (24): 5007-5017.e4. https://doi.org/10.1016/j.cub.2020.09.050.

Lange, Julian, Shintaro Yamada, Sam E. Tischfield, Jing Pan, Seoyoung Kim, Xuan Zhu, Nicholas D. Socci, Maria Jasin, and Scott Keeney. 2016. "The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair." *Cell* 167 (3): 695-708.e16. https://doi.org/10.1016/j.cell.2016.09.035.

Lao, Jessica P., and Neil Hunter. 2010. "Trying to Avoid Your Sister." *PLoS Biology* 8 (10): e1000519. https://doi.org/10.1371/journal.pbio.1000519.

Lappalainen, Tuuli, Alexandra J. Scott, Margot Brandt, and Ira M. Hall. 2019. "Genomic Analysis in the Age of Human Genome Sequencing." *Cell* 177 (1): 70–84. https://doi.org/10.1016/j.cell.2019.02.032.

Laricchia, K.M., S. Zdraljevic, D.E. Cook, and E.C. Andersen. 2017. "Natural Variation in the Distribution and Abundance of Transposable Elements Across the Caenorhabditis Elegans Species." *Molecular Biology and Evolution* 34 (9): 2187–2202. https://doi.org/10.1093/molbev/msx155.

Lascarez-Lagunas, Laura I., Marina Martinez-Garcia, Saravanapriah Nadarajan, Brianna N. Diaz-Pacheco, Elizaveta Berson, and Mónica P. Colaiácovo. 2023. "Chromatin Landscape, DSB Levels, and cKU-70/80 Contribute to Patterning of Meiotic DSB Processing along Chromosomes in C. Elegans." Edited by Aimee Jaramillo-Lambert. *PLOS Genetics* 19 (1): e1010627. https://doi.org/10.1371/journal.pgen.1010627.

Lawson, Heather A., Yonghao Liang, and Ting Wang. 2023. “Transposable Elements in Mammalian Chromatin Organization.” *Nature Reviews Genetics* 24 (10): 712–23. https://doi.org/10.1038/s41576-023-00609-6.

Lee, Daehan, Stefan Zdraljevic, Lewis Stevens, Ye Wang, Robyn E. Tanny, Timothy A. Crombie, Daniel E. Cook, et al. 2021. “Balancing Selection Maintains Hyper-Divergent Haplotypes in Caenorhabditis Elegans.” *Nature Ecology & Evolution* 5 (6): 794–807. https://doi.org/10.1038/s41559-021-01435-x.

Lee, Heng-Chi, Weifeng Gu, Masaki Shirayama, Elaine Youngman, Darryl Conte, and Craig C. Mello. 2012. “C. Elegans piRNAs Mediate the Genome-Wide Surveillance of Germline Transcripts.” *Cell* 150 (1): 78–87. https://doi.org/10.1016/j.cell.2012.06.016.

Lee, Yeon, and Donald C. Rio. 2015. “Mechanisms and Regulation of Alternative Pre-mRNA Splicing.” *Annual Review of Biochemistry* 84 (1): 291–323. https://doi.org/10.1146/annurev-biochem-060614-034316.

Lesack, Kyle, Grace M. Mariene, Erik C. Andersen, and James D. Wasmuth. 2022a. “Different Structural Variant Prediction Tools Yield Considerably Different Results in Caenorhabditis Elegans.” *PLOS ONE* 17 (12): e0278424. https://doi.org/10.1371/journal.pone.0278424.

———. 2022b. “Different Structural Variant Prediction Tools Yield Considerably Different Results in Caenorhabditis Elegans.” *PLOS ONE* 17 (12): e0278424. https://doi.org/10.1371/journal.pone.0278424.

Lewis, Zachary A. 2017. “Polycomb Group Systems in Fungi: New Models for Understanding Polycomb Repressive Complex 2.” *Trends in Genetics* 33 (3): 220–31. https://doi.org/10.1016/j.tig.2017.01.006.

Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” *Bioinformatics* 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics (Oxford, England)* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Liao, L. W., B. Rosenzweig, and D. Hirsh. 1983. “Analysis of a Transposable Element in Caenorhabditis Elegans.” *Proceedings of the National Academy of Sciences of the United States of America* 80 (12): 3585–89. https://doi.org/10.1073/pnas.80.12.3585.

Libuda, Diana E., Satoru Uzawa, Barbara J. Meyer, and Anne M. Villeneuve. 2013. “Meiotic Chromosome Structures Constrain and Respond to Designation of Crossover Sites.” *Nature* 502 (7473): 703–6. https://doi.org/10.1038/nature12577.

Lim, Jaclyn G Y, Rachel R W Stine, and Judith L Yanowitz. 2008. “Domain-Specific Regulation of Recombination in *Caenorhabditis Elegans* in Response to Temperature, Age and Sex.” *Genetics* 180 (2): 715–26. https://doi.org/10.1534/genetics.108.090142.

Lister, Ryan, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, et al. 2009. “Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences.” *Nature* 462 (7271): 315–22. https://doi.org/10.1038/nature08514.

Liu, Chih Long, Tommy Kaplan, Minkyu Kim, Stephen Buratowski, Stuart L Schreiber, Nir Friedman, and Oliver J Rando. 2005. “Single-Nucleosome Mapping of Histone Modifications in S. Cerevisiae.” Edited by Peter Becker. *PLoS Biology* 3 (10): e328. https://doi.org/10.1371/journal.pbio.0030328.

Liu, Zhijie, Daria Merkurjev, Feng Yang, Wenbo Li, Soohwan Oh, Meyer J. Friedman, Xiaoyuan Song, et al. 2014. “Enhancer Activation Requires Trans-Recruitment of a Mega Transcription Factor Complex.” *Cell* 159 (2): 358–73. https://doi.org/10.1016/j.cell.2014.08.027.

Lloyd, Andrew. 2023. “Crossover Patterning in Plants.” *Plant Reproduction* 36 (1): 55–72. https://doi.org/10.1007/s00497-022-00445-4.

Lohe, Allan R., Daniel De Aguiar, and Daniel L. Hartl. 1997. “Mutations in the *Mariner* Transposase: The D,D(35)E Consensus Sequence Is Nonfunctional.” *Proceedings of the National Academy of Sciences* 94 (4): 1293–97. https://doi.org/10.1073/pnas.94.4.1293.

Machovina, Tyler S., Rana Mainpal, Anahita Daryabeigi, Olivia McGovern, Dimitra Paouneskou, Sara Labella, Monique Zetka, Verena Jantsch, and Judith L. Yanowitz. 2016. “A Surveillance System Ensures Crossover Formation in C. Elegans.” *Current Biology* 26 (21): 2873–84. https://doi.org/10.1016/j.cub.2016.09.007.

MacQueen, Amy J., Carolyn M. Phillips, Needhi Bhalla, Pinky Weiser, Anne M. Villeneuve, and Abby F. Dernburg. 2005. “Chromosome Sites Play Dual Roles to Establish Homologous Synapsis during Meiosis in C. Elegans.” *Cell* 123 (6): 1037–50. https://doi.org/10.1016/j.cell.2005.09.034.

Mahmoud, Medhat, Nastassia Gobet, Diana Ivette Cruz-Dávalos, Ninon Mounier, Christophe Dessimoz, and Fritz J Sedlazeck. 2019. “Structural Variant Calling: The Long and the Short of It.” *Genome Biology* 20 (1): 246. https://doi.org/10.1186/s13059-019-1828-7.

Makova, Kateryna D., and Ross C. Hardison. 2015. “The Effects of Chromatin Organization on Variation in Mutation Rates in the Genome.” *Nature Reviews Genetics* 16 (4): 213–23. https://doi.org/10.1038/nrg3890.

Malinin, Nikolay L, Li Zhang, Jeongsuk Choi, Alieta Ciocea, Olga Razorenova, Yan-Qing Ma, Eugene A Podrez, et al. 2009. “A Point Mutation in KINDLIN3 Ablates Activation of Three Integrin Subfamilies in Humans.” *Nature Medicine* 15 (3): 313–18. https://doi.org/10.1038/nm.1917.

Manni, Mosè, Matthew R Berkeley, Mathieu Seppey, Felipe A Simão, and Evgeny M Zdobnov. 2021. “BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.” *Molecular Biology and Evolution* 38 (10): 4647–54. https://doi.org/10.1093/molbev/msab199.

Martini, Emmanuelle, Robert L. Diaz, Neil Hunter, and Scott Keeney. 2006. “Crossover Homeostasis in Yeast Meiosis.” *Cell* 126 (2): 285–95. https://doi.org/10.1016/j.cell.2006.05.044.

Maydan, Jason S, Adam Lorch, Mark L Edgley, Stephane Flibotte, and Donald G Moerman. 2010. “Copy Number Variation in the Genomes of Twelve Natural Isolates of Caenorhabditis Elegans.” *BMC Genomics* 11 (January): 62. https://doi.org/10.1186/1471-2164-11-62.

McGrath, Patrick T., Matthew V. Rockman, Manuel Zimmer, Heeun Jang, Evan Z. Macosko, Leonid Kruglyak, and Cornelia I. Bargmann. 2009. “Quantitative Mapping of a Digenic Behavioral Trait Implicates Globin Variation in C. Elegans Sensory Behaviors.” *Neuron* 61 (5): 692–99. https://doi.org/10.1016/j.neuron.2009.02.012.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

Meneely, Philip M, Anna F Farago, and Tate M Kauffman. 2002. “Crossover Distribution and High Interference for Both the X Chromosome and an Autosome During Oogenesis and Spermatogenesis in *Caenorhabditis Elegans*.” *Genetics* 162 (3): 1169–77. https://doi.org/10.1093/genetics/162.3.1169.

Meneely, Philip M, Olivia L McGovern, Frazer I Heinis, and Judith L Yanowitz. 2012. “Crossover Distribution and Frequency Are Regulated by *Him-5* in *Caenorhabditis Elegans*.” *Genetics* 190 (4): 1251–66. https://doi.org/10.1534/genetics.111.137463.

Mikkelsen, Tarjei S., Manching Ku, David B. Jaffe, Biju Issac, Erez Lieberman, Georgia Giannoukos, Pablo Alvarez, et al. 2007. “Genome-Wide Maps of Chromatin State in

Pluripotent and Lineage-Committed Cells." *Nature* 448 (7153): 553–60. https://doi.org/10.1038/nature06008.

Miller, Danny E, Kevin R Cook, and R Scott Hawley. 2019. "The Joy of Balancers." *PLOS Genetics* 15 (11): e1008421.

Morgan, Chris, John A. Fozard, Matthew Hartley, Ian R. Henderson, Kirsten Bomblies, and Martin Howard. 2021. "Diffusion-Mediated HEI10 Coarsening Can Explain Meiotic Crossover Positioning in Arabidopsis." *Nature Communications* 12 (1): 4674. https://doi.org/10.1038/s41467-021-24827-w.

Muller, Hermann J. 1916. "The Mechanism of Crossing-Over." *The American Naturalist* 50 (592): 193–221. https://doi.org/10.1086/279534.

Munz, P. 1994. "An Analysis of Interference in the Fission Yeast Schizosaccharomyces Pombe." *Genetics* 137 (3): 701–7. https://doi.org/10.1093/genetics/137.3.701.

Murray, Andrew W., and Jack W. Szostak. 1985. "Chromosome Segregation in Mitosis and Meiosis." *Annual Review of Cell Biology* 1 (1): 289–315. https://doi.org/10.1146/annurev.cb.01.110185.001445.

Muzzey, Dale, Eric A Evans, and Caroline Lieber. 2015. "Understanding the Basics of NGS: From Mechanism to Variant Calling." *Current Genetic Medicine Reports* 3 (4): 158–65. https://doi.org/10.1007/s40142-015-0076-8.

Myers, Simon, Rory Bowden, Afidalina Tumian, Ronald E. Bontrop, Colin Freeman, Tammie S. MacFie, Gil McVean, and Peter Donnelly. 2010. "Drive Against Hotspot Motifs in Primates Implicates the *PRDM9* Gene in Meiotic Recombination." *Science* 327 (5967): 876–79. https://doi.org/10.1126/science.1182363.

Nabeshima, Kentaro, Anne M Villeneuve, and Kenneth J Hillers. 2004. “Chromosome-Wide Regulation of Meiotic Crossover Formation in Caenorhabditis Elegans Requires Properly Assembled Chromosome Axes.” *Genetics* 168 (3): 1275–92. https://doi.org/10.1534/genetics.104.030700.

Nattestad, Maria, and Michael C Schatz. 2016. “Assemblytics: A Web Analytics Tool for the Detection of Variants from an Assembly.” *Bioinformatics (Oxford, England)* 32 (19): 3021–23. https://doi.org/10.1093/bioinformatics/btw369.

Nesta, Alex V., Denisse Tafur, and Christine R. Beck. 2021. “Hotspots of Human Mutation.” *Trends in Genetics* 37 (8): 717–29. https://doi.org/10.1016/j.tig.2020.10.003.

Nguyen, Hanh, Sara Labella, Nicola Silva, Verena Jantsch, and Monique Zetka. 2018. “C. Elegans ZHP-4 Is Required at Multiple Distinct Steps in the Formation of Crossovers and Their Transition to Segregation Competent Chiasmata.” Edited by Sarit Smolikove. *PLOS Genetics* 14 (10): e1007776. https://doi.org/10.1371/journal.pgen.1007776.

Nicholas, W L, E C Dougherty, and E L Hansen. 1959. “Axenic Cultivation of Caenorhabditis Briggsae (Nematoda: Rhabditidae) with Chemically Undefined Supplements; Comparative Studies with Related Nematodes.” *Ann. N.Y. Acad. Sci* 77: 218–36.

Nurk, Sergey, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V. Bzikadze, Alla Mikheenko, Mitchell R. Vollger, et al. 2022. “The Complete Sequence of a Human Genome.” *Science* 376 (6588): 44–53. https://doi.org/10.1126/science.abj6987.

Otto, Sarah P., and Bret A. Payseur. 2019. “Crossover Interference: Shedding Light on the Evolution of Recombination.” *Annual Review of Genetics* 53 (1): 19–44. https://doi.org/10.1146/annurev-genet-040119-093957.

Page, Scott L., and R. Scott Hawley. 2003. “Chromosome Choreography: The Meiotic Ballet.” *Science* 301 (5634): 785–89. https://doi.org/10.1126/science.1086605.

Pan, Jing, Mariko Sasaki, Ryan Kniewel, Hajime Murakami, Hannah G. Blitzblau, Sam E. Tischfield, Xuan Zhu, et al. 2011. “A Hierarchical Combination of Factors Shapes the Genome-Wide Topography of Yeast Meiotic Recombination Initiation.” *Cell* 144 (5): 719–31. https://doi.org/10.1016/j.cell.2011.02.009.

Parvanov, Emil D., Petko M. Petkov, and Kenneth Paigen. 2010. “*Prdm9* Controls Activation of Mammalian Recombination Hotspots.” *Science* 327 (5967): 835–835. https://doi.org/10.1126/science.1181495.

Peterson, April L, and Bret A Payseur. 2021. “Sex-Specific Variation in the Genome-Wide Recombination Rate.” Edited by D Barbash. *Genetics* 217 (1): iyaa019. https://doi.org/10.1093/genetics/iyaa019.

Petronczki, Mark, Maria F Siomos, and Kim Nasmyth. 2003. “Un Ménage à Quatre.” *Cell* 112 (4): 423–40. https://doi.org/10.1016/S0092-8674(03)00083-7.

Phillips, Carolyn M., and Abby F. Dernburg. 2006. “A Family of Zinc-Finger Proteins Is Required for Chromosome-Specific Pairing and Synapsis during Meiosis in C. Elegans.” *Developmental Cell* 11 (6): 817–29. https://doi.org/10.1016/j.devcel.2006.09.020.

Phillips, Carolyn M., Xiangdong Meng, Lei Zhang, Jacqueline H. Chretien, Fyodor D. Urnov, and Abby F. Dernburg. 2009. “Identification of Chromosome Sequence Motifs That Mediate Meiotic Pairing and Synapsis in C. Elegans.” *Nature Cell Biology* 11 (8): 934–42. https://doi.org/10.1038/ncb1904.

“Picard Toolkit.” 2019. Java. https://broadinstitute.github.io/picard/.

Plasterk, Ronald H.A, Zsuzsanna Izsvák, and Zoltán Ivics. 1999. “Resident Aliens: The Tc1/ Mariner Superfamily of Transposable Elements.” *Trends in Genetics* 15 (8): 326–32. https://doi.org/10.1016/S0168-9525(99)01777-1.

Platt, Roy N., Laura Blanco-Berdugo, and David A. Ray. 2016. “Accurate Transposable Element Annotation Is Vital When Analyzing New Genome Assemblies.” *Genome Biology and Evolution* 8 (2): 403–10. https://doi.org/10.1093/gbe/evw009.

Pokholok, Dmitry K., Christopher T. Harbison, Stuart Levine, Megan Cole, Nancy M. Hannett, Tong Ihn Lee, George W. Bell, et al. 2005. “Genome-Wide Map of Nucleosome Acetylation and Methylation in Yeast.” *Cell* 122 (4): 517–27. https://doi.org/10.1016/j.cell.2005.06.026.

Powers, Natalie R., Emil D. Parvanov, Christopher L. Baker, Michael Walker, Petko M. Petkov, and Kenneth Paigen. 2016. “The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo.” Edited by Ian R Adams. *PLOS Genetics* 12 (6): e1006146. https://doi.org/10.1371/journal.pgen.1006146.

Prendergast, James Gd, Harry Campbell, Nick Gilbert, Malcolm G Dunlop, Wendy A Bickmore, and Colin Am Semple. 2007. “Chromatin Structure and Evolution in the Human Genome.” *BMC Evolutionary Biology* 7 (1): 72. https://doi.org/10.1186/1471-2148-7-72.

Quinlan, Aaron R, and Ira M Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Rando, Oliver J, and Fred Winston. 2012. “Chromatin and Transcription in Yeast.” *Genetics* 190 (2): 351–87. https://doi.org/10.1534/genetics.111.132266.

Ravanat, Jean-Luc, and Thierry Douki. 2016. “UV and Ionizing Radiations Induced DNA Damage, Differences and Similarities.” *Radiation Physics and Chemistry* 128 (November): 92–102. https://doi.org/10.1016/j.radphyschem.2016.07.007.

Richmond, Timothy J., and Curt A. Davey. 2003. “The Structure of DNA in the Nucleosome Core.” *Nature* 423 (6936): 145–50. https://doi.org/10.1038/nature01595.

Riehl, Kevin, Cristian Riccio, Eric A Miska, and Martin Hemberg. 2022. “TransposonUltimate: Software for Transposon Classification, Annotation and Detection.” *Nucleic Acids Research* 50 (11): e64–e64. https://doi.org/10.1093/nar/gkac136.

Rigau, Maria, David Juan, Alfonso Valencia, and Daniel Rico. 2019. “Intronic CNVs and Gene Expression Variation in Human Populations.” Edited by Colin Semple. *PLOS Genetics* 15 (1): e1007902. https://doi.org/10.1371/journal.pgen.1007902.

Robinson, James T, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. 2011. “Integrative Genomics Viewer.” *Nature Biotechnology* 29 (1): 24–26. https://doi.org/10.1038/nbt.1754.

Rockman, Matthew V, and Leonid Kruglyak. 2009. “Recombinational Landscape and Population Genomics of Caenorhabditis Elegans.” *PLOS Genetics* 5 (3): e1000419.

Rodgers, Kasey, and Mitch McVey. 2016. “Error-Prone Repair of DNA Double-Strand Breaks.” *Journal of Cellular Physiology* 231 (1): 15–24. https://doi.org/10.1002/jcp.25053.

Rowan, Beth A, Vipul Patel, Detlef Weigel, and Korbinian Schneeberger. 2015. “Rapid and Inexpensive Whole-Genome Genotyping-by-Sequencing for Crossover Localization and

Fine-Scale Genetic Mapping." *G3 Genes|Genomes|Genetics* 5 (3): 385–98. https://doi.org/10.1534/g3.114.016501.

Saito, Takamune T., Doris Y. Lui, Hyun-Min Kim, Katherine Meyer, and Monica P. Colaiácovo. 2013. "Interplay between Structure-Specific Endonucleases for Crossover Control during Caenorhabditis Elegans Meiosis." Edited by Michael Lichten. *PLoS Genetics* 9 (7): e1003586. https://doi.org/10.1371/journal.pgen.1003586.

Saito, Takamune T., Firaz Mohideen, Katherine Meyer, J. Wade Harper, and Monica P. Colaiácovo. 2012. "SLX-1 Is Required for Maintaining Genomic Integrity and Promoting Meiotic Noncrossovers in the Caenorhabditis Elegans Germline." Edited by Sharon E. Plon. *PLoS Genetics* 8 (8): e1002888. https://doi.org/10.1371/journal.pgen.1002888.

Saito, Takamune T., Jillian L. Youds, Simon J. Boulton, and Monica P. Colaiácovo. 2009. "Caenorhabditis Elegans HIM-18/SLX-4 Interacts with SLX-1 and XPF-1 and Maintains Genomic Integrity in the Germline by Processing Recombination Intermediates." Edited by Michael Lichten. *PLoS Genetics* 5 (11): e1000735. https://doi.org/10.1371/journal.pgen.1000735.

Sakamoto, Yoshitaka, Suzuko Zaha, Yutaka Suzuki, Masahide Seki, and Ayako Suzuki. 2021. "Application of Long-Read Sequencing to the Detection of Structural Variants in Human Cancer Genomes." *Computational and Structural Biotechnology Journal* 19: 4207–16. https://doi.org/10.1016/j.csbj.2021.07.030.

Sardell, Jason M., and Mark Kirkpatrick. 2020. "Sex Differences in the Recombination Landscape." *The American Naturalist* 195 (2): 361–79. https://doi.org/10.1086/704943.

Sarsani, Vishal Kumar, Narayanan Raghupathy, Ian T Fiddes, Joel Armstrong, Francoise Thibaud-Nissen, Oraya Zinder, Mohan Bolisetty, et al. 2019. “The Genome of C57BL/6J ‘Eve’, the Mother of the Laboratory Mouse Genome Reference Strain.” *G3 (Bethesda, Md.)* 9 (6): 1795–1805. https://doi.org/10.1534/g3.119.400071.

Schramm, Laura, and Nouria Hernandez. 2002. “Recruitment of RNA Polymerase III to Its Target Promoters.” *Genes & Development* 16 (20): 2593–2620. https://doi.org/10.1101/gad.1018902.

Schuster-Böckler, Benjamin, and Ben Lehner. 2012. “Chromatin Organization Is a Major Influence on Regional Mutation Rates in Human Cancer Cells.” *Nature* 488 (7412): 504–7. https://doi.org/10.1038/nature11273.

Schwacha, Anthony, and Nancy Kleckner. 1995. “Identification of Double Holliday Junctions as Intermediates in Meiotic Recombination.” *Cell* 83 (5): 783–91. https://doi.org/10.1016/0092-8674(95)90191-4.

———. 1997. “Interhomolog Bias during Meiotic Recombination: Meiotic Functions Promote a Highly Differentiated Interhomolog-Only Pathway.” *Cell* 90 (6): 1123–35. https://doi.org/10.1016/S0092-8674(00)80378-5.

Sedlazeck, Fritz J, Philipp Rescheneder, Moritz Smolka, Han Fang, Maria Nattestad, Arndt von Haeseler, and Michael C Schatz. 2018. “Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing.” *Nature Methods* 15 (6): 461–68. https://doi.org/10.1038/s41592-018-0001-7.

Seidel, Hannah S., Michael Ailion, Jialing Li, Alexander Van Oudenaarden, Matthew V. Rockman, and Leonid Kruglyak. 2011. “A Novel Sperm-Delivered Toxin Causes Late-

Stage Embryo Lethality and Transmission Ratio Distortion in C. Elegans." Edited by Laurence D. Hurst. *PLoS Biology* 9 (7): e1001115. https://doi.org/10.1371/journal.pbio.1001115.

Seidel, Hannah S., Matthew V. Rockman, and Leonid Kruglyak. 2008. "Widespread Genetic Incompatibility in *C. Elegans* Maintained by Balancing Selection." *Science* 319 (5863): 589–94. https://doi.org/10.1126/science.1151107.

Seisenberger, Stefanie, Simon Andrews, Felix Krueger, Julia Arand, Jörn Walter, Fátima Santos, Christian Popp, Bernard Thienpont, Wendy Dean, and Wolf Reik. 2012. "The Dynamics of Genome-Wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells." *Molecular Cell* 48 (6): 849–62. https://doi.org/10.1016/j.molcel.2012.11.001.

Serizay, Jacques, Yan Dong, Jürgen Jänes, Michael Chesney, Chiara Cerrato, and Julie Ahringer. 2020. "Distinctive Regulatory Architectures of Germline-Active and Somatic Genes in *C. Elegans*." *Genome Research* 30 (12): 1752–65. https://doi.org/10.1101/gr.265934.120.

Shirleen Roeder, G. 1990. "Chromsome Synapsis and Genetic Recombination: Their Roles in Meiotic Chromsome Segregation." *Trends in Genetics* 6: 385–89. https://doi.org/10.1016/0168-9525(90)90297-J.

Shumate, Alaina, and Steven L Salzberg. 2021. "Liftoff: Accurate Mapping of Gene Annotations." *Bioinformatics* 37 (12): 1639–43. https://doi.org/10.1093/bioinformatics/btaa1016.

Sijen, Titia, and Ronald H. A. Plasterk. 2003. "Transposon Silencing in the Caenorhabditis Elegans Germ Line by Natural RNAi." *Nature* 426 (6964): 310–14. https://doi.org/10.1038/nature02107.

Simão, Felipe A, Robert M Waterhouse, Panagiotis Ioannidis, Evgenia V Kriventseva, and Evgeny M Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” *Bioinformatics* 31 (19): 3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Slotkin, R. Keith, and Robert Martienssen. 2007. “Transposable Elements and the Epigenetic Regulation of the Genome.” *Nature Reviews Genetics* 8 (4): 272–85. https://doi.org/10.1038/nrg2072.

Smith, Zachary D., Michelle M. Chan, Tarjei S. Mikkelsen, Hongcang Gu, Andreas Gnirke, Aviv Regev, and Alexander Meissner. 2012. “A Unique Regulatory Phase of DNA Methylation in the Early Mammalian Embryo.” *Nature* 484 (7394): 339–44. https://doi.org/10.1038/nature10960.

Stankiewicz, Pawe\l, and James R Lupski. 2010. “Structural Variation in the Human Genome and Its Role in Disease.” *Annual Review of Medicine* 61 (1): 437–55. https://doi.org/10.1146/annurev-med-100708-204735.

Sterken, Mark G., L. Basten Snoek, Jan E. Kammenga, and Erik C. Andersen. 2015. “The Laboratory Domestication of Caenorhabditis Elegans.” *Trends in Genetics : TIG* 31 (5): 224–31. https://doi.org/10.1016/j.tig.2015.02.009.

Stewart, Mary K, Nathaniel L Clark, Gennifer Merrihew, Evan M Galloway, and James H Thomas. 2005. “High Genetic Diversity in the Chemoreceptor Superfamily of Caenorhabditis Elegans.” *Genetics* 169 (4): 1985–96. https://doi.org/10.1534/genetics.104.035329.

Stranger, Barbara E., Forrest Matthew S., Dunning Mark, Ingle Catherine E., Beazley Claude, Thorne Natalie, Redon Richard, et al. 2007. "Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes." *Science* 315 (5813): 848–53. https://doi.org/10.1126/science.1136678.

Strickland, Walter. 1958. "An Analysis of Interference in *Aspergillus Nidulans*." *Proceedings of the Royal Society of London. Series B - Biological Sciences* 149 (934): 82–101. https://doi.org/10.1098/rspb.1958.0053.

Sturtevant, A. H. 1913. "The Linear Arrangement of Six Sex-linked Factors in Drosophila, as Shown by Their Mode of Association." *Journal of Experimental Zoology* 14 (1): 43–59. https://doi.org/10.1002/jez.1400140104.

Sudmant, Peter H, Tobias Rausch, Eugene J Gardner, Robert E Handsaker, Alexej Abyzov, John Huddleston, Yan Zhang, et al. 2015. "An Integrated Map of Structural Variation in 2,504 Human Genomes." *Nature* 526 (7571): 75–81. https://doi.org/10.1038/nature15394.

Sulston, J E, and S Brenner. 1974. "The DNA of Caenorhabditis Elegans." *Genetics* 77 (1): 95–104. https://doi.org/10.1093/genetics/77.1.95.

Surzycki, S A, and W R Belknap. 2000. "Repetitive-DNA Elements Are Similarly Distributed on Caenorhabditis Elegans Autosomes." *Proceedings of the National Academy of Sciences of the United States of America* 97 (1): 245–49. https://doi.org/10.1073/pnas.97.1.245.

Swan, Kathryn A, Damian E Curtis, Kathleen B McKusick, Alexander V Voinov, Felipa A Mapa, and Michael R Cancilla. 2002. "High-Throughput Gene Mapping in Caenorhabditis Elegans." *Genome Research* 12 (7): 1100–1105. https://doi.org/10.1101/gr.208902.

Swanson, Willie J., and Victor D. Vacquier. 2002. “The Rapid Evolution of Reproductive Proteins.” *Nature Reviews Genetics* 3 (2): 137–44. https://doi.org/10.1038/nrg733.

Szostak, Jack W., Terry L. Orr-Weaver, Rodney J. Rothstein, and Franklin W. Stahl. 1983. “The Double-Strand-Break Repair Model for Recombination.” *Cell* 33 (1): 25–35. https://doi.org/10.1016/0092-8674(83)90331-8.

Szutorisz, Henrietta, Niall Dillon, and László Tora. 2005. “The Role of Enhancers as Centres for General Transcription Factor Recruitment.” *Trends in Biochemical Sciences* 30 (11): 593–99. https://doi.org/10.1016/j.tibs.2005.08.006.

Tabuchi, Tomoko M., Andreas Rechtsteiner, Tess E. Jeffers, Thea A. Egelhofer, Coleen T. Murphy, and Susan Strome. 2018. “Caenorhabditis Elegans Sperm Carry a Histone-Based Epigenetic Memory of Both Spermatogenesis and Oogenesis.” *Nature Communications* 9 (1): 4310. https://doi.org/10.1038/s41467-018-06236-8.

The 1000 Genomes Project Consortium. 2012. “An Integrated Map of Genetic Variation from 1,092 Human Genomes.” *Nature* 491 (7422): 56–65. https://doi.org/10.1038/nature11632.

The 1000 Genomes Project Consortium, Corresponding authors, Adam Auton, Gonçalo R. Abecasis, Steering committee, David M. Altshuler, Richard M. Durbin, et al. 2015. “A Global Reference for Human Genetic Variation.” *Nature* 526 (7571): 68–74. https://doi.org/10.1038/nature15393.

Thomas, Cristel G, Wei Wang, Richard Jovelin, Rajarshi Ghosh, Tatiana Lomasko, Quang Trinh, Leonid Kruglyak, Lincoln D Stein, and Asher D Cutter. 2015. “Full-Genome

Evolutionary Histories of Selfing, Splitting, and Selection in Caenorhabditis." *Genome Research* 25 (5): 667–78. https://doi.org/10.1101/gr.187237.114.

Thompson, Owen A, L Basten Snoek, Harm Nijveen, Mark G Sterken, Rita J M Volkers, Rachel Brenchley, Arjen van't Hof, et al. 2015. "Remarkably Divergent Regions Punctuate the Genome Assembly of the Caenorhabditis Elegans Hawaiian Strain CB4856." *Genetics* 200 (3): 975–89. https://doi.org/10.1534/genetics.115.175950.

Tørresen, Ole K, Bastiaan Star, Pablo Mier, Miguel A Andrade-Navarro, Alex Bateman, Patryk Jarnot, Aleksandra Gruca, et al. 2019. "Tandem Repeats Lead to Sequence Assembly Errors and Impose Multi-Level Challenges for Genome and Protein Databases." *Nucleic Acids Research* 47 (21): 10994–6. https://doi.org/10.1093/nar/gkz841.

Tzur, Yonatan B, Eitan Winter, Jinmin Gao, Tamar Hashimshony, Itai Yanai, and Monica P Colaiácovo. 2018. "Spatiotemporal Gene Expression Analysis of the *Caenorhabditis Elegans* Germline Uncovers a Syncytial Expression Switch." *Genetics* 210 (2): 587–605. https://doi.org/10.1534/genetics.118.301315.

Uchimura, Arikuni, Mayumi Higuchi, Yohei Minakuchi, Mizuki Ohno, Atsushi Toyoda, Asao Fujiyama, Ikuo Miura, Shigeharu Wakana, Jo Nishino, and Takeshi Yagi. 2015. "Germline Mutation Rates and the Long-Term Phenotypic Effects of Mutation Accumulation in Wild-Type Laboratory Mice and Mutator Mice." *Genome Research* 25 (8): 1125–34. https://doi.org/10.1101/gr.186148.114.

Van Oss, Stephen Branden, and Anne-Ruxandra Carvunis. 2019. "De Novo Gene Birth." *PLOS Genetics* 15 (5): e1008160. https://doi.org/10.1371/journal.pgen.1008160.

Van't Hof, Arjen E, Pascal Campagne, Daniel J Rigden, Carl J Yung, Jessica Lingley, Michael A Quail, Neil Hall, Alistair C Darby, and Ilik J Saccheri. 2016. "The Industrial Melanism Mutation in British Peppered Moths Is a Transposable Element." *Nature* 534 (7605): 102–5. https://doi.org/10.1038/nature17951.

Vergara, Ismael A, Allan K Mah, Jim C Huang, Maja Tarailo-Graovac, Robert C Johnsen, David L Baillie, and Nansheng Chen. 2009. "Polymorphic Segmental Duplication in the Nematode Caenorhabditis Elegans." *BMC Genomics* 10 (1): 329. https://doi.org/10.1186/1471-2164-10-329.

Vizir, I Yu, and A B Korol. 1990. "Sex Difference in Recombination Frequency in Arabidopsis." *Heredity* 65 (3): 379–83. https://doi.org/10.1038/hdy.1990.107.

Wagner, Cynthia R., Lynnette Kuervers, David L. Baillie, and Judith L. Yanowitz. 2010. "Xnd-1 Regulates the Global Recombination Landscape in Caenorhabditis Elegans." *Nature* 467 (7317): 839–43. https://doi.org/10.1038/nature09429.

Wang, Shunxin, Yanlei Liu, Yongliang Shang, Binyuan Zhai, Xiao Yang, Nancy Kleckner, and Liangran Zhang. 2019. "Crossover Interference, Crossover Maturation, and Human Aneuploidy." *BioEssays* 41 (10): 1800221. https://doi.org/10.1002/bies.201800221.

Wang, Yali, Yuan Gao, Chao Li, Hong Gao, Cheng-Cai Zhang, and Xudong Xu. 2018. "Three Substrains of the Cyanobacterium Anabaena Sp. Strain PCC 7120 Display Divergence in Genomic Sequences and *hetC* Function." Edited by Yves V. Brun. *Journal of Bacteriology* 200 (13). https://doi.org/10.1128/JB.00076-18.

Weber, Katherine P., Subhajyoti De, Iwanka Kozarewa, Daniel J. Turner, M. Madan Babu, and Mario De Bono. 2010. "Whole Genome Sequencing Highlights Genetic Changes

Associated with Laboratory Domestication of C. Elegans." Edited by Ben Lehner. *PLoS ONE* 5 (11): e13922. https://doi.org/10.1371/journal.pone.0013922.

Wegewitz, Viktoria, Hinrich Schulenburg, and Adrian Streit. 2008. "Experimental Insight into the Proximate Causes of Male Persistence Variation among Two Strains of the Androdioecious Caenorhabditis Elegans (Nematoda)." *BMC Ecology* 8 (1): 12. https://doi.org/10.1186/1472-6785-8-12.

Wicker, Thomas, Yeisoo Yu, Georg Haberer, Klaus F. X. Mayer, Pradeep Reddy Marri, Steve Rounsley, Mingsheng Chen, et al. 2016. "DNA Transposon Activity Is Associated with Increased Mutation Rates in Genes of Rice and Other Grasses." *Nature Communications* 7 (1): 12790. https://doi.org/10.1038/ncomms12790.

Wicks, Stephen R, Raymond T Yeh, Warren R Gish, Robert H Waterston, and Ronald H A Plasterk. 2001. "Rapid Gene Mapping in Caenorhabditis Elegans Using a High Density Polymorphism Map." *Nature Genetics* 28 (2): 160–64. https://doi.org/10.1038/88878.

Wolfe, Kenneth H., Paul M. Sharp, and Wen-Hsiung Li. 1989. "Mutation Rates Differ among Regions of the Mammalian Genome." *Nature* 337 (6204): 283–85. https://doi.org/10.1038/337283a0.

Xu, Guixia, Chunce Guo, Hongyan Shan, and Hongzhi Kong. 2012. "Divergence of Duplicate Genes in Exon–Intron Structure." *Proceedings of the National Academy of Sciences* 109 (4): 1187–92. https://doi.org/10.1073/pnas.1109047109.

Yalcin, B., J. Fullerton, S. Miller, D. A. Keays, S. Brady, A. Bhomra, A. Jefferson, et al. 2004. "Unexpected Complexity in the Haplotypes of Commonly Used Inbred Strains of

Laboratory Mice." *Proceedings of the National Academy of Sciences* 101 (26): 9734–39. https://doi.org/10.1073/pnas.0401189101.

Yang, Ziheng, and Joseph P. Bielawski. 2000. "Statistical Methods for Detecting Molecular Adaptation." *Trends in Ecology & Evolution* 15 (12): 496–503. https://doi.org/10.1016/S0169-5347(00)01994-7.

Yokoo, Rayka, Karl A. Zawadzki, Kentaro Nabeshima, Melanie Drake, Swathi Arur, and Anne M. Villeneuve. 2012. "COSA-1 Reveals Robust Homeostasis and Separable Licensing and Reinforcement Steps Governing Meiotic Crossovers." *Cell* 149 (1): 75–87. https://doi.org/10.1016/j.cell.2012.01.052.

Yoshimura, Jun, Kazuki Ichikawa, Massa J Shoura, Karen L Artiles, Idan Gabdank, Lamia Wahba, Cheryl L Smith, et al. 2019. "Recompleting the Caenorhabditis Elegans Genome." *Genome Research* 29 (6): 1009–22. https://doi.org/10.1101/gr.244830.118.

Yu, Zhi, Tim H. H. Coorens, Md Mesbah Uddin, Kristin G. Ardlie, Niall Lennon, and Pradeep Natarajan. 2024. "Genetic Variation across and within Individuals." *Nature Reviews Genetics*, March. https://doi.org/10.1038/s41576-024-00709-x.

Zetka, M C, and A M Rose. 1995. "Mutant Rec-1 Eliminates the Meiotic Pattern of Crossing over in Caenorhabditis Elegans." *Genetics* 141 (4): 1339–49. https://doi.org/10.1093/genetics/141.4.1339.

Zhang, Gaotian, Ye Wang, and Erik C. Andersen. 2022. "Natural Variation in *C. Elegans* Short Tandem Repeats." *Genome Research*, October, genome;gr.277067.122v2. https://doi.org/10.1101/gr.277067.122.

Zhang, Liangran, Zhangyi Liang, John Hutchinson, and Nancy Kleckner. 2014. "Crossover Patterning by the Beam-Film Model: Analysis and Implications." Edited by R. Scott Hawley. *PLoS Genetics* 10 (1): e1004042. https://doi.org/10.1371/journal.pgen.1004042.

Zhang, Liangran, Shunxin Wang, Shen Yin, Soogil Hong, Keun P. Kim, and Nancy Kleckner. 2014. "Topoisomerase II Mediates Meiotic Crossover Interference." *Nature* 511 (7511): 551–56. https://doi.org/10.1038/nature13442.

Zhang, Liangyu, Simone Köhler, Regina Rillo-Bohn, and Abby F Dernburg. 2018. "A Compartmentalized Signaling Network Mediates Crossover Control in Meiosis." *eLife* 7 (March): e30789. https://doi.org/10.7554/eLife.30789.

Zhang, Liangyu, Weston Stauffer, David Zwicker, and Abby F. Dernburg. 2021. "Crossover Patterning through Kinase-Regulated Condensation and Coarsening of Recombination Nodules." https://doi.org/10.1101/2021.08.26.457865.

Zhang, Xiaoyu, Oliver Clarenz, Shawn Cokus, Yana V Bernatavichute, Matteo Pellegrini, Justin Goodrich, and Steven E Jacobsen. 2007. "Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in Arabidopsis." Edited by James C Carrington. *PLoS Biology* 5 (5): e129. https://doi.org/10.1371/journal.pbio.0050129.

Zhang, Zhongge, and Milton H. Saier Jr. 2011. "Transposon-Mediated Adaptive and Directed Mutations and Their Potential Evolutionary Benefits." *Microbial Physiology* 21 (1–2): 59–70. https://doi.org/10.1159/000333108.

Zhao, Yuehui, Lijiang Long, Wen Xu, Richard F Campbell, Edward E Large, Joshua S Greene, and Patrick T McGrath. 2018. "Changes to Social Feeding Behaviors Are Not Sufficient

for Fitness Gains of the Caenorhabditis Elegans N2 Reference Strain." *eLife* 7 (October). https://doi.org/10.7554/eLife.38675.

Zierhut, Christian, Marc Berlinger, Christian Rupp, Akira Shinohara, and Franz Klein. 2004. "Mnd1 Is Required for Meiotic Interhomolog Repair." *Current Biology* 14 (9): 752–62. https://doi.org/10.1016/j.cub.2004.04.030.